Published in final edited form as:

J Proteome Res. 2008 September 5; 7(9): 4199–4208. doi:10.1021/pr800277y.

SeMoP: A New Computational Strategy for the Unrestricted Search for Modified Peptides Using LC–MS/MS Data

Christian Baumgartner^{*,†,‡}, Tomas Rejtar[†], Majlinda Kullolli[†], Lakshmi Manohar Akella[†], and Barry L. Karger^{*,†}

[†] *Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115*

[‡] *Research Group for Clinical Bioinformatics, Institute of Biomedical Engineering, University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria*

Abstract

A novel computational approach, termed Search for Modified Peptides (SeMoP), for the unrestricted discovery and verification of peptide modifications in shotgun proteomic experiments using low resolution ion trap MS/MS spectra is presented. Various peptide modifications, including post-translational modifications, sequence polymorphisms, as well as sample handling-induced changes, can be identified using this approach. SeMoP utilizes a three-step strategy: (1) a standard database search to identify proteins in a sample; (2) an unrestricted search for modifications using a newly developed algorithm; and (3) a second standard database search targeted to specific modifications found using the unrestricted search. This targeted approach provides verification of discovered modifications and, due to increased sensitivity, a general increase in the number of peptides with the specific modification. The feasibility of the overall strategy has been first demonstrated in the analysis of 65 plasma proteins. Various sample handling induced modifications, such as β -elimination of disulfide bridges and pyrocarbamidomethylation, as well as biologically induced modifications, such as phosphorylation and methylation, have been detected. A subsequent targeted Sequest search has been used to verify selected modifications, and a 4-fold increase in the number of modified peptides was obtained. In a second application, 1367 proteins of a cervical cancer cell line were processed, leading to detection of several novel amino acid substitutions. By conducting the search against a database of peptides derived from proteins with decoy sequences, a false discovery rate of less than 5% for the unrestricted search resulted. SeMoP is shown to be an effective and easily implemented approach for the discovery and verification of peptide modifications.

Keywords

mass spectrometry; post-translational modifications; unrestricted search; shotgun proteomics

Introduction

Most proteins undergo post-translational modifications (PTMs) that alter their physical and chemical properties, as well as their three-dimensional structure and stability, and such changes are often key regulators of protein function.^{1,2} Currently, several hundred PTMs are known,^{3,4} though only a few, such as phosphorylation, acetylation, methylation, and glycosylation,

* To whom correspondence should be addressed. E-mail: christian.baumgartner@umit.at (C.B.); b.karger@neu.edu (B.L.K.).

Supporting Information Available: Description of the algorithm for the unrestricted search, processed proteins, and verifications of modifications. This material is available free of charge via the Internet at <http://pubs.acs.org>.

are generally targeted in a proteomic analysis.⁵ In addition, various sequence polymorphisms attributed to alternative splicing or single nucleotide polymorphism are commonly observed.⁶ Moreover, a high number of peptide modifications can be introduced during sample preparation, and these modifications are often ignored or misinterpreted.

Shotgun proteomics is a widely used tool for global analysis of protein modifications;⁷ however, reliable and comprehensive identification of unknown modifications is one of the biggest challenges in experimental bioinformatics. In a typical LC-MS/MS experiment, hundreds of thousands of tandem mass spectra are typically collected; however, only 10–20% of these spectra are interpreted using current approaches,⁸ and it is expected that a number of unassigned MS/MS spectra may correspond to modified peptides. Regrettably, standard database search programs, such as Sequest or Mascot,^{9,10} are suitable only for identification of expected, user specified modifications.

Analysis of sequence tags¹¹ or full *de novo* sequencing has been utilized for identification of modified peptides.¹² Recently, a new set of algorithms to search in an unrestricted manner for modifications has been developed. However, to reduce the search complexity and thus the search time, most of the unrestricted search algorithms target modification only for proteins identified in the sample rather than for all potentially present proteins.¹³ ModifiComb, for example, allows identification of substoichiometrically modified peptides that elute from an RPLC gradient in close proximity to unmodified peptides.^{14,15} Another algorithm, termed TwinPeaks, uses spectral convolution, similar to the Sequest search algorithm, to compute the cross-correlation between predicted and experimental spectra in a search for a bimodal pattern in the cross-correlation function. Such a pattern is indicative of the presence of fragment ions with constant mass shift relative to the corresponding unmodified peptide.¹⁶ P-Mod calculates mass differences between search peptide sequences and a precursor and uses p-value statistics for estimating a sequence-to-MS/MS spectrum match.¹⁷ In another approach, PTM Explorer works in conjunction with a standard database search algorithm but includes all modifications from UniMod³ or other similar databases to facilitate identification of modified peptides.¹⁸

Recently, Tsur et al. proposed a Smith-Waterman-based spectrum alignment algorithm that is applicable to either *de novo* peptide sequencing, assembly of protein sequences from shotgun proteomic data or an unrestricted search for PTMs.^{19,20} An extended framework targeting peptide modifications based on this alignment approach, termed PTM Finder,²¹ as well as a strategy of selectively excluded mass screening analysis, SEMSA,²² have very recently been published.

This paper presents a new approach to the discovery of peptide modifications that is readily implemented in shotgun proteomics using low resolution MS/MS spectra. The approach is based on the coupling of an unrestricted search for peptide modifications with well established methods, allowing a broad search for modifications. SeMoP (search for modified peptides) is based on a three-step strategy: (1) a standard database search to identify proteins in a sample; (2) a comprehensive, unrestricted search to discover modified peptides within a specified mass range (± 200 Da) with respect to the corresponding unmodified peptides using only identified proteins; and (3) a targeted search using a standard database search for verifying selected modifications found in step 2.

A new algorithm is utilized for the unrestricted search for peptide modifications. The algorithm relies on detection of a constant shift between experimental and fragment ions predicted for unmodified peptides. The search generates a histogram, termed a ΔM plot, that displays this constant shift for matches of experimental fragment ions. It is important to emphasize that the strategy does *not* require the identification of unmodified peptides nor the specification of the modifications before the search. Furthermore, the algorithm can identify multiple

modifications per peptide, allow selection of the mass range for the search for modifications, and, most importantly, is simple to apply since the only parameter that must be specified is the mass range with respect to the unmodified peptide. It should be noted that the unrestricted algorithm is used for detection of potential modifications present in the sample rather than for detection and validation of modified peptides. The result of the unrestricted search (step 2) is a list of modifications that are then targeted in a standard database search to verify peptides with these targeted modifications (step 3).

To demonstrate the SeMoP strategy, we initially focus on high abundant glycosylated proteins in human plasma samples. It was anticipated that such proteins will yield high sequence coverage and thus have a high potential for observing modified peptides. Tryptic digest samples were analyzed by LC-MS/MS, and a selected set of 65 glycoproteins, identified in all samples of the first discovery step, was identified for the unrestricted search for modified peptides. In this experiment, various modifications introduced during sample preparation such as desulfurization of cysteine, biologically induced modifications such as phosphorylation and methylation, and several amino acid substitutions have been identified. In a second experiment, 1367 proteins obtained from about 10 000 SiHa cells (cervical cancer cell line) were processed using the SeMoP approach leading to the detection of novel amino acid substitutions. The approach is shown to be promising as a tool to find modified peptides.

SeMoP Protocol

The SeMoP computational protocol consists of three steps schematically shown in Figure 1. First, a standard database search is conducted using Sequest to identify proteins in the sample (step 1). Then, an unrestricted search for modifications is performed using an in-house developed algorithm (step 2). A list of all peptides generated by *in silico* digestion of identified proteins is used as an input for the algorithm. As a result, modifications for peptides already identified in step 1 as well as additional modified peptides can be found. When substoichiometric modifications are targeted where both the modified and unmodified versions of the same peptide are expected in the sample, the unrestricted search is performed only for peptides identified in step 1, thus significantly speeding up the analysis. It should be noted that the unrestricted search is performed only on selected experimental MS/MS spectra, i.e. high quality spectra that were not assigned to peptides in step 1. The result of the unrestricted search is a ΔM plot that reveals candidate modifications. Specific modifications are selected for a targeted standard database search using again Sequest (step 3), and significant matches are verified using well established protocols. Importantly, the use of established tools to determine the significance of assignments of MS/MS spectra to modified peptides alleviates the need for the development of a new scoring algorithm for an unrestricted search. Furthermore, this final step typically leads to the detection of additional peptides with the targeted modification.

Unrestricted Search for Peptide Modifications

The algorithm for unrestricted searching relies on a comparison of the unmodified fragment list for the sequences with the fragment list created from experimental MS/MS spectra. Overlapped fragments are counted and interpreted as a measure of similarity via a ΔM plot. A workflow diagram of the algorithm is presented in Figure 2, and the basic principles of the algorithm are demonstrated in Figure 3.

Step A, Figure 2—For each unmodified peptide, the algorithm calculates *b*-ion $\{b_1, \dots, b_L\}$ and *y*-ion $\{y_1, \dots, y_L\}$ fragments with a charge state up to that of the precursor ion and then uses the *m/z* values of the fragment ions to create a list of mass differences of the form $b_{ij} = b_j - b_i$ and $y_{ij} = y_j - y_i$, where *i* is from 1 to *L*−1, *j* from 2 to *L*, with the condition *j* > *i*. *L* is the peptide length. The indices of spectral peaks *b_i* and *y_i* are retained to define the exact location of generated peptide fragments *b_{ij}* and *y_{ij}* in the predicted spectrum. Only experimental MS/

MS spectra which have the precursor ion within a specified window (± 200 Da) are selected for further comparison, reducing substantially the number of matching operations. Each experimental spectrum is first baseline corrected and denoised by a Hann window²³ to allow for the identification of peaks from background noise, and then, peaks are detected using a quadratic polynomial fit.

Step B, Figure 2—The algorithm uses a prescreening procedure to eliminate all peptide sequences that are unlikely to match by counting the number of exact matches between the m/z values of b - and y -ions from the predicted spectrum of an unmodified peptide to the m/z values found in the denoised spectrum within the mass tolerance of ± 0.7 Da. The unmodified fragment list for the sequences is matched to each MS/MS spectrum despite the mass difference between the peptide and precursor ion. Only spectral candidates with a specific predefined number N of matches of ions are selected as a measure of spectral similarity, while spectra not fulfilling this criterion are discarded. We set $N = 4$ for short peptides (8–10 residues),¹⁵ and linearly adjust N for longer peptides (see Supporting Information A, eq 1). The procedure strongly reduces the number of processed spectra in an extended search for modifications. This data reduction procedure is necessary to reduce the computer processing time since CPU-intensive operations are only conducted for spectra with a high likelihood of a match. Alternatively, when using a large number of nodes, this filtering procedure could be skipped leading to an improved sensitivity. On the other hand, it may also increase the false discovery rate. Next, an extended list of m/z values, defined by $pp_{ij} = p_j - p_i$, where p represent a spectral peak in the denoised MS/MS spectrum and $j > i$ is generated. Again, indices of spectral peaks p_i of selected peptide fragments pp_{ij} are retained. Analogous to spectral convolution, a comparison of the mass lists from the experimental and predicted spectra of the form $\langle b_{ij}, y_{ij}; b_i, y_i \rangle \leftrightarrow \langle p_{ij}; p_i \rangle$ is performed for all masses within a specified mass range $[-\Delta M$ to $+\Delta M]$. Fragments found to coincide in the experimental and predicted spectra within each bin, defined by the mass tolerance, are counted, and the accumulated number of fragment hits per bin is then plotted as the ΔM plot, see Figure 3. It should be noted that since low resolution ion trap MS/MS spectra are examined, the bin size of $\Delta M = 1$ Da is selected. For high resolution MS/MS spectra, the bin size could be adjusted to account for improved mass accuracy. The ΔM plot is then analyzed for the occurrence of major peaks using a Boolean acceptance criterion f_{hits} based on four ΔM plot specific parameters: (1) the number and (2) height of the major peaks, (3) their ΔM locations (presence of a peak at $\Delta M = 0$ is required) and (4) the median of the number of fragment hits within the specified search window $[-\Delta M, +\Delta M]$ to assign a match (for the definition of f_{hits} see Supporting Information A, eq 2).

Step C, Figure 2—Spectra passing the f_{hits} criterion are considered at this point as candidate matches. The mass difference between the experimental precursor and the predicted peptide mass is calculated and compared to the ΔM plot. Only in the case of the presence of a peak at the same mass difference in the ΔM plot will a peptide match be considered significant and reported in the final list of hits.

The algorithm is implemented in LabView 8.0 (National Instruments, Austin, TX) and compiled in an executable file. To facilitate rapid processing, the algorithm is written in a form suitable for distributed computing, which is highly scalable and does not require additional software for operation. Thus, the algorithm can be deployed on networking computers running, for example, an MS Windows operating system.

The searches in this paper were carried out on a 12-CPU computer cluster and the processing of 1670 peptides and roughly 56 000 MS/MS spectra collected from the human plasma sample required about 60 h. The 4393 peptides and 46 500 spectra of the cancer cell line sample were processed in 145 h on the cluster. Importantly, the algorithm was designed for distributed

computing, and thus the implementation on a larger computer cluster would lead to a significant reduction in processing time. The program is freely available upon request.

Experimental Section

Sample Preparation and LC-MS

Human Plasma Sample—Six normal plasma samples were obtained from Bioreclamation (Hicksville, NY) were initially used to assess the SeMoP strategy. Samples were analyzed using the work flow described elsewhere.²⁴ Briefly, twelve highly abundant plasma proteins were depleted using a commercial immunoaffinity column (Beckman, Fullerton, CA), with subsequent enrichment of glycosylated proteins by an in-house developed Multi-Lectin Affinity Column (M-LAC).²⁵ Following denaturation, reduction, and alkylation with iodoacetamide, samples were digested with trypsin. After desalting with a reversed phase column, the digested samples were analyzed by 75 μ m i.d. RPLC column with a one hour gradient at 200 nL/min flow rate using a hybrid linear ion trap FT mass spectrometer (Thermo Fisher, San Jose, CA). It should be noted that LTQ FT instrument was operated in the high resolution MS mode (full profile mode) while the MS/MS spectra were obtained using low resolution (centroid mode). To demonstrate compatibility with broadly available LTQ instruments, the high mass accuracy of the precursor ions was not utilized in SeMoP strategy.

SiHa Cell Line—Sample preparation and analysis were described in detail elsewhere.²⁶ Briefly, an amount of total protein corresponding to a lysate obtained from roughly 10,000 SiHa cells (cervical cancer cell line) was separated by SDS-PAGE. The SDS-PAGE lane was divided into three sections that were individually in-gel digested and analyzed by 75 μ m i.d. RPLC column using a one hour gradient at 200 nL/min flow rate with a hybrid linear ion trap FT mass spectrometer (Thermo Fisher).

Determination of the False Discovery and Identification Success Rate

To determine the rate of false discovery modifications from the unrestricted search, a series of searches was conducted against a database consisting of sequences of the 65 selected human plasma proteins and against a separate decoy database consisting of randomized sequences of these 65 human proteins. All hits in the randomized database were considered false identifications.^{27,28}

To determine the identification success rate (sensitivity) of our algorithm for an unrestricted search, selected peptide modifications were specified in a subsequent Sequest database search. The results provided by Sequest were filtered using XCorr values greater than 1.9, 2.2, and 3.8 for 1+, 2+, and 3+, respectively. Because only a very small database (65 proteins) was used in this calculation, it was expected that, by applying the above strict criteria, the level of random identifications would be very low. Significant hits from the Sequest search were then compared with results obtained using our in-house algorithm for an unrestricted search to calculate the identification success rate as [# of hits from the unrestricted search]/[# of hits from the Sequest search].²⁸

Results and Discussion

In this study, we applied a new computational protocol, SeMoP, for an unrestrictive investigation of peptide modifications from 64 human and 1 bovine high abundant plasma glycoproteins and 1367 proteins of a cervical cancer cell line. Roughly 56 000 MS/MS spectra of the plasma and 46 500 MS/MS spectra of the cell line sample derived from a shotgun proteomic experiment were searched for modifications using sequences of unmodified peptides. Both expected and unusual modifications were detected. In particular, the human

plasma data set served as a model for demonstrating the feasibility of our approach, estimating the identification success and false discovery rate of the method.

Analysis of Human Plasma Proteins

Following step one of our protocol to identify proteins in the samples that would be used for unrestricted search, see Figure 1, a Sequest Cluster search engine within a CPAS system²⁹ was used for the database search of approximately 185 000 MS/MS spectra collected for the LC/MS analysis of 6 plasma samples. The search was conducted with a precursor mass tolerance of 2.5 Da and a fragment mass tolerance of 1 Da against the Swiss-Prot human protein database (release 52 with 15 498 protein sequences, downloaded in April 2007), appended with reversed protein sequences to facilitate the estimation of the discovery rate. Trypsin was specified as the digestion enzyme with up to 2 missed cleavages. Carbamidomethylation of cysteines was specified as the only modification in this search. Matches were accepted with peptide probabilities, estimated using a PeptideProphet value greater than 0.9 and Xcorr values greater than 1.9, 2.2, and 3.8 for 1+, 2+, and 3+ charged ions, respectively. On average, roughly 100 proteins were identified per sample. Out of the identified proteins, 64 highly abundant plasma glycoproteins plus the spiked internal standard, bovine fetuin, identified in all 6 samples, were selected for the subsequent unrestricted search. The list of these selected proteins is provided in the Supporting Information B. An in-house written perl script was used to generate a total of 1670 fully tryptic peptides (no missed cleavages) within a mass range from 700 to 3500 Da, and these peptide sequences were selected as input for the unrestricted search (Figure 1, steps 1 and 2). Out of a total of 185,000 MS/MS spectra, 56,065 in the initial Sequest search yielded cross-correlation values greater than 1.0, 1.5, and 2.0 for 1+, 2+, and 3+ charged ions, respectively, and these spectra were selected for the unrestricted search. This procedure allowed preferential selection of MS/MS spectra corresponding to peptide fragmentation. The MS/MS spectra of peptides identified in the Sequest search included peptides with carbamidomethylated cysteines that were retained in the data set to facilitate estimation of sensitivity of the unrestricted search.

To illustrate the new search algorithm, the ΔM plots (1 Da resolution, mass range between ± 200 Da) for an MS/MS spectrum that corresponds to the unmodified and modified peptide, MATTTMIQSK, is shown in Figure 4. The single major peak at $\Delta M = 0$ in Figure 4A indicates correspondence between the predicted and observed fragments for the unmodified peptide as the majority of high intensity peaks in the experimental MS/MS spectrum could be directly explained by *b*- and *y*-ion fragments. Minor peaks in this plot could be attributed to noise (see the quality parameter for a peptide match in Supporting Information A, eq 3). (Random matches of peptide fragments could be suppressed using high mass accuracy MS/MS spectra.) On the other hand, the ΔM plot for an MS/MS spectrum corresponding to the modified peptide, shown in Figure 4B, exhibited two peaks, one at $\Delta M = 0$ and the other at $\Delta M = -48$ Da. The second peak clearly shows the presence of fragment ions in the experimental MS/MS spectrum that have a constant shift with respect to the predicted fragments of unmodified peptides, indicative of a modification. Further examination revealed that the mass shift can be interpreted as a dethiomethylation of oxidized methionine, a relatively uncommon modification reported in the UniMod database.³

Out of 56 065 MS/MS spectra processed, 2076 were reported as modified peptides (Step 2, Figure 1), as summarized in Figure 5 and, in more detail, Table 1. Despite the low mass accuracy of the LTQ instrument, a unique assignment was achieved for the majority of modifications including expected PTMs, missed peptide cleavages and sequence polymorphisms. The most common modifications were on cysteine residues that were introduced during sample preparation such as carbamidomethylation (+57 Da) from reduction and alkylation, the rarely reported N-terminal carboxyamidomethyl-Cys cyclization (+40 Da)

³⁰ and desulfurization (−34 Da),³¹ an unusual cysteine modification. It should be noted again that carbamidomethylated peptides were not excluded from the significant hits, since the high number of matches for this modification was found to be useful for estimating the identification success and false discovery rate after the targeted database search. Among biological modifications found in these samples, methylation and phosphorylation were the most common. Interestingly, several glutamic acid, isoleucine, and alanine substitutions were identified as well.

To evaluate the third step of the SeMoP strategy, see Figure 1, we next selected several frequent modifications out of the list of all identified modifications and conducted a targeted search using Sequest. Examples of modifications identified in the unrestricted search and comparison with the results of the subsequent Sequest search are presented below.

Evaluation of Modifications Identified Using Unrestricted Search for Plasma Sample

For evaluation of the performance of the unrestricted search algorithm, MS/MS spectra identified in step 1 were retained in the data set. As a result, the majority of modifications in the unrestricted search (1171 of 2076) were found to be carbamidomethylations (+57 Da) of cysteine. In total, 97.7% of these peptides were modified singly, 2.3% doubly and several triply. Importantly, roughly 1% of the cysteine containing peptides were found without any modification, indicating a low level of peptides with unreacted cysteines. Alkylation of N-terminal residues or histidines was not observed.

MS/MS spectra assigned to multiple carbamidomethylations can be used to illustrate the abilities of the unrestricted search algorithm to identify multiple modifications per peptide. As an example, Figure 6A presents a ΔM plot for a peptide with three cysteines modified with carbamidomethylation. As can be seen, the three major peaks reveal the presence of fragment ions for one, two, and three modifications (+57, +114, +171 Da). In addition, a peak at zero mass shift is observed, corresponding to the fragmentation of the unmodified residues. *In silico* MS/MS spectra of the peptide with three mass shifts of −57 Da introduced at various positions were generated and ΔM plots recalculated. The result is shown in Figure 6B in which the modifications are positioned on the cysteine residues, leading to a single high intensity peak with zero mass shift, indicating agreement between the experimental and the predicted MS/MS spectra. Similar examples of multiple modifications were also found for the combination of a variety of modifications with different masses, such as carbamidomethylation of cysteines and methionine oxidation.

Interestingly, Figure 5 reveals two other highly abundant cysteine modifications in this data set. One is N-terminal cyclization of carbamidomethylated cysteine (+40 Da), observed in 3.2% of the modified peptides, a known alteration but rarely reported.³⁰ The second is an unusual loss of 34 Da from cysteine found in 1.7% of all modifications. This latter change could be explained by the conversion of cysteine to dehydroalanine through β -elimination of disulfide bridges and is likely a result of heating during sample preparation.^{31–33}

Two of the three identified cysteine modifications, (1) carbamidomethylation (+57 Da) and (2) desulfurization (−34 Da), were subjected to a targeted database search using Sequest to verify identified peptides and also to assess the identification success rate (sensitivity) of our unrestricted search algorithm (step 3). The targeted search led to identification of 4703 peptides modified by carbamidomethylation, or roughly 3-fold increase compared to the number of peptides identified in the unrestricted search with the same modification (1171). Importantly, all carbamidomethylated peptides found in the unrestricted search were also identified in the targeted search demonstrating a high specificity of the unrestricted search. Similarly, all 36 peptides identified by the unrestricted search with desulfurized cysteine were among 141 peptides found by the targeted search demonstrating a similar gain in sensitivity for the targeted

search and again a high specificity of the unrestricted search. The relatively low sensitivity of the unrestricted algorithm for both modifications is likely a result of conservative filtering to minimize the false discovery rate, and thus the number of modifications to be validated, see next section.

Besides cysteine modifications, several peptides in Table 1 were detected with a +80 Da mass shift associated with serine residues, indicating phosphorylation. A subsequent targeted database search using Sequest with potential phosphorylation of S, T, and Y residues confirmed the presence of phosphorylation sites for 4 peptides of ITIH2_HUMAN and 3 peptides of FETUA_BOVIN using the unrestricted search. The targeted search led to identification of 5 additional phosphorylation sites, two in LAC_HUMAN, one in HEP2_HUMAN, one in KAC_HUMAN, and one in ITIH2_HUMAN. Importantly, well-known phosphorylation sites in ITIH2_HUMAN and FETUA_BOVIN were found, thus confirming correct matches. In addition, an increase in the number of modified peptides was achieved using the targeted database search.

Other modifications found using the unrestricted search include loss of ammonia, which was detected for 17 different peptides from 69 total MS/MS spectra. The loss of ammonia can be explained by formation of pyroglutamic acid from glutamine. A targeted Sequest search for this modification led to detection of 25 different peptides (47% increase) with a total of 106 matches (54% increase).

Interestingly, several amino acid substitutions were detected as well, see Table 1. Peptide MYYSAVDPTK, derived from copper binding oxidoreductase ceruloplasmin, was detected with a D→E substitution, which is a known mutation that may present higher risk for Parkinson disease.³⁴ In another example, peptide FNKPFVFLMIEQNTK, belonging to α -1-antitrypsin, was found with an E → D substitution, which is a known variant of this protein.³⁵

In summary, a variety of peptide modifications including sample preparation-induced modifications as well as biologically relevant modifications, such as phosphorylation and sequence polymorphisms, were identified using an unrestricted search, followed by a targeted database search. The result illustrates the usefulness of the SeMoP approach to detect various classes of expected as well as unusual modifications.

Estimation of the False Discovery Rate (FDR) of the Unrestricted Search

Compared to the standard database search, the unrestricted search often leads to an increased rate of false identifications. Thus, care must be exercised when establishing the acceptance criteria for significant matches. We analyzed all significant matches reported by the unrestricted algorithm, and it was found that, out of 2076 detected modified peptides, 1890 could be assigned to known types of modifications using manual evaluation of MS/MS spectra. Next, we investigated the remaining pool of 186 matches (9%) that could not be confidently identified. It was found that roughly half of these spectra could be attributed to previously identified peptides but with an insufficient quality criterion for the MS/MS spectra to allow unambiguous identification, while the other half of the MS/MS spectra were attributed to random matches (90 out of 2076 detected modified peptides). Thus, an empirical false discovery rate was estimated to be roughly 5% of the total number of significant matches.

To determine the number of randomly assigned peptides more accurately, a decoy database search was employed.²⁷ Analogous to the standard database search, the unrestricted search was repeated against a database of peptides derived from proteins with random sequences. These peptides were generated by *in-silico* tryptic digestion of randomly reshuffled protein sequences; random rather than reversed protein sequences were used in this case to minimize the chances of creating homologous peptides. All positive matches in the decoy database were

considered to be incorrect and their quantity to be a measure of false discovery rate.²⁸ This search returned 98 significant hits, which is 4.7% of the total peptide matches, and these findings agree with our estimated empirical false discovery rate of 5%. For more information on the performance of our unrestricted search algorithm, see Supporting Information A, Tables 4A–B. In summary, the empirical as well as the statistically estimated false discovery rate were found to be in good agreement. Moreover, the false discovery rate below 5% allows for high confident identification of modified peptides without the need for further validation of matches.

Evaluation of Modifications Identified Using Unrestricted Search for SiHa Cells

To evaluate further the performance of SeMoP strategy, we investigated peptide modifications derived from a cervical cancer cell line (SiHa). Initially, all collected MS/MS spectra were analyzed by Sequest using the same filtering criteria as for the plasma protein samples, leading to the identification of a total of 1367 proteins (step 1). Only a selected set of MS/MS spectra with specific XCorr values was utilized for the unrestricted search (46 580 out of 68 000 MS/MS spectra), and MS/MS spectra already assigned to peptides using the Sequest search were retained in the data set. Compared to the plasma analysis where all potential peptides corresponding to proteins identified were considered, only substoichiometric modifications were targeted in this experiment and thus, only peptides identified by the Sequest algorithm (a total of 4393 peptides) with molecular weight ranging from 700 to 3500 Da were used in the unrestricted search.

Table 2A summarizes the modifications identified using the unrestricted search, most of the which are expected such as loss of ammonia. Nevertheless, several unusual modifications were found. For example, two peptides, VSFELFADKVPK from peptidyl-prolyl cis–trans isomerase A, a protein accelerating protein folding, and ISGLIYEETR from the histone H4 family, showed a +26 Da modification on a serine residue, which could be interpreted as a substitution of serine to leucine or isoleucine. Modification of +26 Da for these peptides has not yet been reported, but interestingly both serines were listed in the UniProt database to be potentially phosphorylated possibly suggesting adduct formation after beta elimination of the phosphate group. It is important to note that more modifications could be found with the analysis of a greater number of MS/MS spectra along with less stringent criteria, the latter of which would, however, increase the false discovery rate.

Following the SeMoP strategy, a targeted database search (step 3) of selected modifications, listed in Table 2B, was conducted to verify detected modifications, and to search for further peptide candidates with these modifications. All 10 selected peptide candidates, corresponding to 7 different modifications, e.g. tryptophan oxidation and formation of pyroglutamic acid, were confirmed by the targeted database search. As expected, an additional number of different peptides with the same modification, on average 5-times greater, was detected as well, demonstrating the usefulness of the targeted step. For example, cyclization of N-term carbamidomethylated cysteine was detected only once using the unrestricted database search, compared to the assignment of six MS/MS spectra to the same modified peptides accomplished in the targeted search. Annotated MS/MS spectra of all peptides listed in Table 2B are provided in Supporting Information C.

Several proteins found in this study were homologous, and thus several peptides identified in the initial Sequest search differed by only a single amino acid residue. Since the algorithm for unrestricted search considers amino acid substitution simply a modification, for homologous peptides a single MS/MS spectrum should be assigned to all homologous peptides with the modification corresponding to their amino acid substitutions. We tested the ability of the unrestricted search to find such homologous peptides. For example, an MS/MS spectrum that was matched directly to peptide IMNTFSVMPSPK from tubulin beta-2A chain with no modification was also matched to a modified peptide IMNTFSVM³²PSPK. However, such

a modification could be explained by M → V substitution (a difference of −32 Da) and indeed peptide IMNTFSVVPSPK can be found in a homologous protein, tubulin beta-2C chain, also identified in the sample. Table 3 summarizes all identified homologous peptides found in this data set using the unrestricted search, denoting modification sites and mass shifts of detected amino acid substitutions. The results show that all matches determined using the unrestricted database search were verified by homologous peptides. In addition, the exact modification sites of all amino acid substitutions could be determined, demonstrating the ability of the approach to identify single amino acid modifications. In summary, these results further confirm the usefulness of our unrestricted search algorithm.

Conclusions

In this paper, we have presented a straightforward strategy, SeMoP, for the discovery and verification of peptide modifications from LC–MS/MS data in shotgun proteomics data processing. SeMoP relies on coupling standard database searching to identify proteins present in the sample with a new algorithm for an unrestricted search of peptide modifications, followed by a second standard database search using modifications discovered by the unrestricted search. Importantly, since the standard database search is used for identification of modified peptides, well-established algorithms can be applied to determine significance of matches. In addition, due to the high sensitivity of standard database search, SeMoP leads to identification of a greater number of modified peptides than unrestricted search alone.

SeMoP was applied to data sets of human plasma proteins and a cervical cancer cell line, detecting a number of expected as well as some unusual peptide modifications. The SeMoP approach utilizes a user specified mass range, ±200 Da, to search for peptide modifications. This mass range can be modified based on the application and the type of MS instrumentation used.

We have demonstrated the feasibility of the approach using as MS widely employed low mass resolution ion trap. High mass accuracy in the MS and, even more importantly in the MS/MS mode, could provide a significant improvement and permit not only discrimination between modifications with a similar mass but also provide stricter criteria leading to a decrease of false matches.³⁶ A simple procedure employing a targeted database search was used to assess the sensitivity of the unrestricted algorithm. In addition, the false discovery rate of the unrestricted was estimated using a search against database with random protein sequences. In addition, SeMoP strategy could be readily adopted to novel fragmentation techniques such as ETD. In summary, our experimental results demonstrate that SeMoP is a useful and simple tool for global analysis of protein modifications in shotgun proteomics with the potential to be extended to high-mass accuracy MS/MS data.

Acknowledgements

We thank the NIH GM15847 for support of this work and Mr. D. Wang of the Barnett Institute for providing SiHa cell LC–MS/MS data. C.B. thanks the Max Kade Foundation and the Austrian Genome Program (GEN-AU, BIN II) for support of a fellowship. Contribution 920 from the Barnett Institute.

References

1. Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* 2007;4(10):798–806. [PubMed: 17901869]
2. Pang CN, Hayen A, Wilkins MR. Surface accessibility of protein post-translational modifications. *J Proteome Res* 2007;6(5):1833–1845. [PubMed: 17428077]
3. Creasy DM, Cottrell JS. Unimod: Protein modifications for mass spectrometry. *Proteomics* 2004;4(6):1534–1536. [PubMed: 15174123]

4. Farriol-Mathis N, Garavelli JS, Boeckmann B, Duvaud S, Gasteiger E, Gateau A, Veuthey AL, Bairoch A. Annotation of post-translational modifications in the Swiss-Prot knowledge base. *Proteomics* 2004;4(6):1537–1550. [PubMed: 15174124]
5. Larsen MR, Trelle MB, Thingholm TE, Jensen ON. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* 2006;40(6):790–798. [PubMed: 16774123]
6. Roth MJ, Forbes AJ 2nd, Kim YB, Robinson DE, Kelleher NL. Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry. *Mol Cell Proteomics* 2005;4(7):1002–1008. [PubMed: 15863400]
7. MacCoss MJ, McDonald WH, Saraf A, Sadygov S, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A, Clark JI, Yates JR. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci USA* 2002;99(12):7900–7905. [PubMed: 12060738]
8. MacCoss MJ. Computational analysis of shotgun proteomics data. *Curr Opin Chem Biol* 2005;9(1):88–94. [PubMed: 15701459]
9. Eng JK, McCormick AL, Yates JR III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spec* 1994;5:976–989.
10. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectroscopy data. *Electrophoresis* 1999;20(18):3551–3567. [PubMed: 10612281]
11. Tabb DL, Saraf A, Yates JR. GutenTag: High-Throughput Sequence Tagging via an Empirically Derived Fragmentation Model. *Anal Chem* 2003;75(23):6415–6421. [PubMed: 14640709]
12. Searle BC, Dasari S, Wilmarth PA, Turner M, Reddy AP, David LL, Nagalla SR. Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. *J Proteome Res* 2005;4(2):546–554. [PubMed: 15822933]
13. Craig R, Beavis RC. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun Mass Spectrom* 2003;17(20):2310–2316. [PubMed: 14558131]
14. Nielsen ML, Savitski MM, Zubarev RA. Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Mol Cell Proteomics* 2006;5(12):2384–2391. [PubMed: 17015437]
15. Savitski MM, Nielsen ML, Zubarev RA. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol Cell Proteomics* 2006;5(5):935–948. [PubMed: 16439352]
16. Havilio M, Wool A. Large-scale unrestricted identification of post-translation modifications using tandem mass spectrometry. *Anal Chem* 2007;79(4):1362–1368. [PubMed: 17297935]
17. Hansen BT, Davey SW, Ham AJ, Liebler DC. P-Mod: an algorithm and software to map modifications to peptide sequences using tandem MS data. *J Proteome Res* 2005;4(2):358–368. [PubMed: 15822911]
18. Chamrad DC, Korting G, Schafer H, Stephan C, Thiele H, Apweiler R, Meyer HE, Marcus K, Bluggel M. Gaining knowledge from previously unexplained spectra-application of the PTM-Explorer software to detect PTM in HUPO BPP MS/MS data. *Proteomics* 2006;6(18):5048–5058. [PubMed: 16912973]
19. Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA. Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol* 2005;23(12):1562–1567. [PubMed: 16311586]
20. Bandeira N, Tsur D, Frank A, Pevzner PA. Protein identification by spectral networks analysis. *Proc Natl Acad Sci USA* 2007;104(15):6140–6145. [PubMed: 17404225]
21. Tanner S, Payne SH, Dasari S, Shen Z, Wilmarth PA, David LL, Loomis WF, Briggs SP, Bafna V. Accurate Annotation of Peptide Modifications through Unrestrictive Database Search. *J Proteome Res* 2007;7(1):170–181. [PubMed: 18034453]
22. Seo J, Jeong J, Kim YM, Hwang N, Paek E, Lee KJ. Strategy for Comprehensive Identification of Post-translational Modifications in Cellular Proteins, Including Low Abundant Modifications: Application to Glyceraldehyde-3-phosphate Dehydrogenase. *J Proteome Res* 2008;7(2):587–602. [PubMed: 18183946]
23. Oppenheim, AV.; Schafer, RW. *Discrete-Time Signal Processing*. Prentice-Hall; Upper Saddle River, NJ: 1999. p. 468–471.

24. Plavina T, Wakeshull E, Hancock WS, Hincapie MJ. Combination of Abundant Protein Depletion and Multi-Lectin Affinity Chromatography (M-LAC) for Plasma Protein Biomarker Discovery. *J Proteome Res* 2007;6(2):662–671. [PubMed: 17269723]
25. Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J Chromatogr A* 2004;1053:79–88. [PubMed: 15543974]
26. Gu Y, Wu SL, Meyer JL, Hancock WS, Burg LJ, Linder J, Hanlon DW, Karger BL. Proteomic analysis of high-grade dysplastic cervical cells obtained from ThinPrep slides using laser capture microdissection and mass spectrometry. *J Proteome Res* 2007;6(11):4256–4268. [PubMed: 17902640]
27. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007;4(3):207–214. [PubMed: 17327847]
28. Kall L, Storey JD, MacCoss MJ, Noble WS. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *J Proteome Res* 2008;7(1):40–44. [PubMed: 18052118]
29. Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang R, Faca V, Gaffken P, Zhang H, Whitaker J, States D, Hanash S, Paulovich A, McIntosh MW. Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J Proteome Res* 2006;5(1):112–121. [PubMed: 16396501]
30. Glazer, AN.; Delange, RJ.; Sigman, DS. Chemical modification of proteins: Selected methods and analytical procedures. Work, TS.; Work, E., editors. North Holland; Amsterdam: 1975. p. 99-101.Ch. 3
31. Pentelute BL, Kent SB. Selective desulfurization of cysteine in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation. *Org Lett* 2007;9(4):687–690. [PubMed: 17286375]
32. Cohen SL, Price C, Vlasak J. Beta-elimination and peptide bond hydrolysis: two distinct mechanisms of human IgG1 hinge fragmentation upon storage. *J Am Chem Soc* 2007;129(22):6976–6977. [PubMed: 17500521]
33. Rejtar, T.; Baumgartner, C.; Kullolli, M.; Karger, BL. Beta-elimination of disulfide bridges: a common sample preparation induced protein modification. Proceeding of the 56th ASMS Conference; Denver, CO. 2008.
34. Hochstrasser H, Tomiuk J, Walter U, Behnke S, Spiegel J, Krüger R, Becker G, Riess O, Berg D. Functional relevance of ceruloplasmin mutations in Parkinson's disease. *FASEB J* 2005;19(13):1851–1853. [PubMed: 16150804]
35. Graham A, Hayes K, Weidinger S, Newton CR, Markham AF, Kalsheker NA. Characterisation of the alpha-1-antitrypsin M3 gene, a normal variant. *Hum Genet* 1990;85(3):381–382. [PubMed: 2394452]
36. Shen Y, Tolic N, Hixson KK, Purvine SO, Pasa-Tolic L, Qian WJ, Adkins JN, Moore RJ, Smith RD. Proteome-Wide Identification of Proteins and Their Modifications with Decreased Ambiguities and Improved False Discovery Rates Using Unique Sequence Tags. *Anal Chem* 2008;80(6):1871–1882. [PubMed: 18271604]

Figure 1.
Workflow diagram of the SeMoP protocol.

Figure 2. Flow diagram of the algorithm for unrestricted search for peptide modifications. The terms “shifted” and “non-shifted” refer to “constant mass shifts” and “zero mass shifts” between experimental and predicted fragment ions.

Figure 3. Simplified scheme to illustrate the principle of the algorithm for the unrestricted search. Based on the charge state of the precursor ion, a predicted MS/MS spectrum is generated. The predicted and denoised experimental spectra are transformed to a normalized ΔM plot by comparing matching fragment ions for shifted and nonshifted fragment peaks over ΔM .

Figure 4.

Example of a normalized ΔM plot for the unmodified and modified peptide MATTTMIQSK from the inter- α -trypsin inhibitor heavy chain H2 protein. (A) ΔM plot for the unmodified peptide with one major peak corresponding to nonshifted fragments, that is, $\Delta M = 0$. (B) ΔM plot of the modified peptide with a mass shift of -48 Da, corresponding to a dethiomethylation of oxidized methionine. Background noise, which is evenly distributed over the ΔM plots, is caused by matches of noise peaks in the experimental MS/MS spectrum within the given mass tolerance of ± 0.7 Da.

Figure 5.

Distribution of most frequent peptide modifications in the region ± 200 Da identified in the 65 high abundant plasma glycoproteins. SeMoP found PTMs, amino acid substitutions, missed peptide cleavages, C- and N-terminal sequence polymorphisms, as well as random matches evenly distributed on the ΔM scale. The calculated false discovery rate was less than 5%.

Figure 6.

Identification of multiple modifications per peptide. (A) ΔM plot of the triply modified peptide LCHCPVGYTGPFCDVDTK. The peaks at $\Delta M = +57$, $+114$, $+171$ Da arise from a carbamidomethylation of cysteine residues. The central peak at $\Delta M = 0$ represents nonshifted matches. (B) Verification of sites of modification: The figure depicts the single peak ΔM plot based on the predicted spectrum confirming all three cysteine modification sites. The plot is similar to the ΔM plot of an unmodified peptide (see Figure 4A).

Table 1Summary of Peptide Modifications Found in 65 High Abundant Glycoproteins in Plasma^a

ΔM (Da)	modification	number of observed peptides	protein	examples of modified peptides
-58	Glu → Ala substitution	19	ITIH3_HUMAN	LWAYLTIEQLLE ⁻⁵⁸ K
-48	Dethiomethylation	4	ITIH2_HUMAN	M ⁻⁴⁸ ATTMIQSK
-34	Desulfurization of Cys	36	APOH_HUMAN	18 peptides VC ⁻³⁴ PFAGILENGAVR
			CFAH_HUMAN	DTSC ⁻³⁴ VNPPTVQNAYIVSR
			CO3_HUMAN	VYAYYNLEESC ⁻³⁴ TR
-17	Pyro-Glu from Gln	69		17 peptides
-14	Glu → Asp	1	A1AT_HUMAN	FNKPFVFLM I E ⁻¹⁴ QNTK
	Ile → Val	4	FETUA_BOVIN	TPIVGQPSI ⁻¹⁴ PGGPVR
+14	Asp → Glu	1	CERU_HUMAN	MYYSAV D ⁺¹⁴ PTK
+16	Oxidation	60		15 peptides
+28	Ala → Val	2	A1BG_HUMAN	AEAQAQYSAVA ⁺²⁸ K
		2	ITIH4_HUMAN	A ⁺²⁸ FQPFVELTMPYSVIR
	Unknown modification	3	A2MG_HUMAN	SGLSTGWTQLS ⁺²⁸ K
+40	Carbamidomethyl-Cys cyclization (N-terminus)	65	CFAH_HUMAN	12 peptides C ⁺⁴⁰ FEGFGIDGPAIAK
+57	Carbamidomethylation	1171		133 peptides
+80	Phosphorylation	4	ITIH2_HUMAN	SLPGES ⁺⁸⁰ EEMMEEVDQVTLYSYK
		3	FETUA_BOVIN	HTFS ⁺⁸⁰ GVASVESSSGEAFHV GK

^aFETUA_BOVIN was spiked in all samples. Bold-marked amino acids indicate the modification site. When a high number of modified peptides were found, only the total number of modified peptides is listed.

Table 2

Summary of Modifications Found in SiHa (Cervical Cancer) Cell Line and Comparison of Selected Modifications Found in the SiHa Cell Line Using an Unrestricted and Targeted Search^a

A: Summary of modifications found in SiHa cell line				
ΔM (Da)	modification type	number of observed peptides	protein	examples of modified peptides
-30	Thr → Ala	1	COF1_HUMAN	YALYDATYET ⁻³⁰ K
	Glu → Val	1	K2C1_HUMAN	YEE ⁻³⁰ LQITAGR
-17	Pyro-glu from Gln	3	2AAA_HUMAN	Q ⁻¹⁷ LSQSLLPAIVELAEDAK
+16	Oxidation	42		28 peptides
+26	Ser → Ile or Ser → Leu	1	PPIA_HUMAN	VS ⁺²⁶ FELFADKVPK
		2	H4_HUMAN	IS ⁺²⁶ GLIYEETR
+32	Dioxidation	1	LDH6B_HUMAN	LIIVSNPVDILTYVAW ⁺³² K
+40	Carbamidomethyl-Cys cyclization (N-terminus)	1	MGN2_HUMAN	C ⁺⁴⁰ LVFSLIGLHFK
+42	Unknown modification	3	H2B1B_HUMAN	A ⁺⁴² MGIMNSFVNDIFER
B: Comparison of selected modifications found in the SiHa cell line using an unrestricted and targeted search				
ΔM (Da)	modified peptides	number of observed peptides in unrestricted search	peptides found in targeted search ^b	
-30	YALYDAT ⁻³⁰ YETK	1	9	
-27	N ⁻²⁷ LDLDSIIAEVK [#]	4	10	
-17	Q ⁻¹⁷ LSQSLLPAIVELAEDAK	1	28	
	Q ⁻¹⁷ KDYETATLSEIK	1		
	Q ⁻¹⁷ LLQANPILEAFGNAK	1		
+26	VS ⁺²⁶ FELFADKVPK	1	13	
	IS ⁺²⁶ GLIYEETR	2		
+32	LIIVSNPVDILTYVAW ⁺³² K	1	3	
+40	C ⁺⁴⁰ LVFSLIGLHFK	1	6	
+42	A ⁺⁴² MGIMNSFVNDIFER	3	20	

^a Bold-marked amino acids indicate the modification site.

^b Spectra with XCorr values greater than 1.9, 2.2, and 3.8 for 1+, 2+, and 3+, respectively, were considered to be correct hits. All peptides found in the unrestricted search could be verified by the targeted search. See text for details.

Table 3Validation of the Unrestricted Search Algorithm for Peptide Pairs in Homologous Proteins in the SiHa Cell Line^a

peptide pairs	proteins	modification sites (mass shift in Da)	found (Y/N) ^b
IMNTFSVMPSPK	TBB2A_HUMAN	M → V (-32)	Y
IMNTFSVVPSPK	TBB2C_HUMAN		N
LISQIVSSITASLR	TBA4A_HUMAN	S → G (-30)	Y
LIGQIVSSITASLR	TBA1A_HUMAN		N
LVSIGAEIIVDGNVK	ACTN1_HUMAN	V → A (-28)	Y
LVSIGAEIIVDGNAK	ACTN4_HUMAN		N
NDEELNKLLGR	H2A1B_HUMAN	R → K (-28)	Y
NDEELNKLLGK	H2A1D_HUMAN		N
NLDLDSIIAEVK	K22E_HUMAN	N → S (-27)	Y
SLDLDSIIAEVK	K2C1_HUMAN		Y
SLDLDSIIAEVK	K2C7_HUMAN	G → S (+30)	Y
SLDLDSIIAEVK	K2C1_HUMAN		Y
NLDLDSIIAEVKAQYEEIANR	K2C5_HUMAN	N → Q (+14)	Y
NLDLDSIIAEVKAQYEEIAQR	K22E_HUMAN		N
ITGEAFVQFASQELAEK	HNRPF_HUMAN	I → S (-26)	Y
STGEAFVQFASQELAEK	HNRH1_HUMAN		N
LFIGGLSFETTDESRLR	ROA2_HUMAN	E → D (-14)	Y
LFIGGLSFETTDESRLR	ROA1_HUMAN		N
LFIGGLSFETTDDSLR	ROA3_HUMAN	D → E (+14)	Y
LFIGGLSFETTDESRLR	ROA1_HUMAN		N
LQLWDIAGQER	RAB32_HUMAN	I → T (-12)	Y
LQLWDTAGQER	RAB34_HUMAN		N
LQVWDTAGQER	RAB13_HUMAN	V → L (+14)	Y
LQLWDTAGQER	RAB34_HUMAN		N
AVFVDLEPTVIDEVR	TBA3_HUMAN	V → I (+14)	Y
AVFVDLEPTVIDEIR	TBA4A_HUMAN		N
DAGTIAGLNVLR	HSP7C_HUMAN	L → M (+18)	Y
DAGTIAGLNVMR	GRP78_HUMAN		N
ISVYYNEATGGK	TBB5_HUMAN	S → N (+27)	Y
INVYYNEATGGK	TBB2C_HUMAN		N

^a Bold-marked amino acids indicate the modification site.^b Modification was detected (Y) or not detected (N) in unrestricted search. See text for more details.