

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
10 November 2005 (10.11.2005)

PCT

(10) International Publication Number
WO 2005/106923 A2

(51) International Patent Classification⁷: **H01J 49/40**,
49/00

(21) International Application Number:
PCT/US2005/012347

(22) International Filing Date: 13 April 2005 (13.04.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/561,938 14 April 2004 (14.04.2004) US

(71) Applicants (for all designated States except US): **APPLERA CORPORATION** [US/US]; 500 Old Connecticut Path, Framingham, MA 01701 (US). **MDS INC.** [CA/CA]; MDS SCIEX DIVISION, 71 Four Valley Drive, Concord, Ontario L4K 4V8 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ZHU, Xiangping** [CN/US]; 20 Valley View Drive, Grafton, MA 01536 (US). **HATTAN, Stephen, J.** [US/US]; 40 Haydn Street,

#2, Roslindale, MA 02131 (US). **CHEN, Xunming** [US/US]; 4 Reeves Road, Bedford, MA 01730 (US). **PAPAYANNOPOULOS, Ioannis, A.** [US/US]; 11 Independence Lane, Ashland, MA 01721 (US).

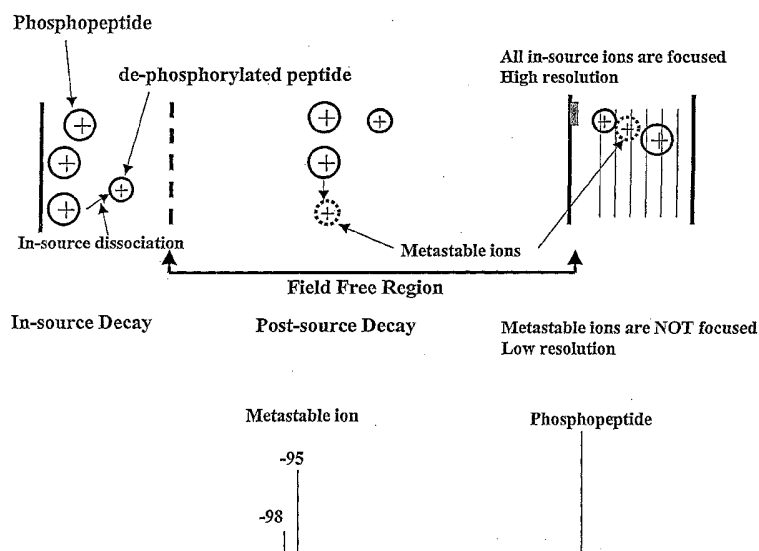
(74) Agent: **KARNAKIS, Andrew, T.**; Senior Patent Counsel, Applera Corporation, 500 Old Connecticut Path, Framingham, MA 01701 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,

[Continued on next page]

(54) Title: IDENTIFICATION OF MODIFIED PEPTIDES BY MASS SPECTROMETRY



(57) Abstract: In various aspects, provided are methods for identification of the presence of modified peptides using at least in part mass spectrometric signals from metastable ions that are formed in the field free region of the analyzer. For example, in various embodiments where the modified peptides are phosphopeptides, such metastable ions can appear approximately 95 u lower than the precursor ion m/z value from which they were formed when the peptide contains a phosphate group. In various embodiments, detection of such metastable ions in a MALDI-TOF mass spectrum can identify the modified peptide. In various aspects, provided are methods for identification of the presence of modified peptides using at least in part MS/MS data from two or more of modified peptides, demodified peptides and non-modified peptides by comparing the sequence and/or site of modification information for two or more of these three forms of a peptide.



FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO,
SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *without international search report and to be republished
upon receipt of that report*

IDENTIFICATION OF MODIFIED PEPTIDES BY MASS SPECTROMETRY

5 CROSS REFERENCES TO RELATED APPLICATIONS

The present application claims the benefit of U.S. Provisional Application No. 60/561,938 filed April 14, 2004. The entire contents of the foregoing provisional application are herein incorporated by reference.

10 INTRODUCTION

Mass spectrometry (MS), including MALDI-TOF MS, is a widely used analytical platform in numerous areas of biological research. Biomolecular samples (e.g., peptides, proteins, nucleic acids, carbohydrates) often contain moieties coupled to the biomolecule by labile chemical linkages that are susceptible to dissociation during analysis by the mass spectrometer. Such moieties are often of important significance in understanding the characteristics of the biomolecule and its relevance to the biological system under investigation. For example, proteins and or peptides containing phosphate linkages (-PO₄) are termed “phosphoproteins, phosphopeptides or phosphorylated peptides or proteins”. These peptides and proteins are prime examples of biological molecules that are susceptible to dissociation of their phosphate group during MS analysis. The addition or subtraction of phosphate groups on proteins is widely used by biological systems as chemical signals or triggers for regulation of cellular processes, and, as such, identification of phosphoproteins is one example of significant interest in numerous areas of proteomic research. Likewise, other post-translational modifications of peptides such as glycosylation have important biological ramifications, and peptides that contain such modifications may also undergo dissociation of their glycosylated moiety. In this regard, methods that rapidly identify the presence of such potential group modification (such as phosphorylation and glycosylation) in peptide fragments prepared by enzymatic or

chemical digestion of proteins, and selection of these peptides for further analysis by tandem mass spectrometry (MS/MS) techniques are highly desirable.

SUMMARY

5 The present teachings relate to a method for identification of modified peptides from mixtures of peptides when analyzed by mass spectrometry, e.g., axial matrix assisted laser desorption ionization (MALDI) time-of-flight (TOF) mass spectrometry.

 In accordance with various aspects of the present teachings, methods for using the metastable behavior of labile biomolecules, such as a peptide that has a phosphate or sugar
10 linkage, when analyzed by a mass spectrometer, such as a MALDI-TOF mass spectrometer, are provided that can be implemented as a means of detecting and identifying the labile biomolecule (e.g., phospho- and glyco-peptides). In various embodiments, the method includes identifying metastable ions from phosphopeptides (and/or glycopeptides and/or other modified peptides) that are susceptible to post source
15 decay in the field free region of a MALDI-TOF mass spectrometer.

 In various aspects, methods of detecting and identifying the labile biomolecule (e.g., phospho- and glyco-peptides) are provided by generating a MS/MS spectrum of the labile biomolecule. In various embodiments, the method includes using tandem mass spectrometry to determine all or part of the sequence of a modified peptide and the site of
20 modification based on searching a database using information from the sequence of the modified peptide, generating an MS/MS spectrum of a de-modified peptide that corresponds to the modified peptide and determining all or part of the sequence of the de-modified peptide and the site of modification based on searching a database using information from the sequence of the de-modified peptide. The sequence and the site of
25 modification determined for the modified peptide to that determined for the de-modified

peptide can be compared to identify that a modified peptide is present in a mixture of peptides if the sequence and the site of modification determined for the modified peptide substantially matches that determined for the de-modified peptide.

In various embodiments of the various aspects of the present teaching, a method of identifying the presence of a modified peptide, such as a post-translationally modified peptide, can be automated through the MS instrument software and facilitated by a computer algorithm that searches for and detects metastable ions in MALDI-TOF mass spectra that are collected in MS operation. Once identified, an experimentally derived mass unit value can be added to the detected metastable ion to identify the potential modified peptide, e.g., a potentially post-translationally modified peptide. In various embodiments, confirmation of the putative modified peptide can be carried out by tandem mass spectrometry techniques. Accordingly, in various aspects, the present teachings provide an article of manufacture where the functionality of one or more methods of the present teachings is embedded as computer-readable instructions on a computer-readable medium, such as, but not limited to, a floppy disk, a hard disk, an optical disk, a magnetic tape, a PROM, an EPROM, CD-ROM, or DVD-ROM.

These and other features of the present teaching are set forth herein.

BRIEF DESCRIPTION OF DRAWINGS

The skilled artisan will understand that the drawings, described below, are for illustration purposes only. The drawings are not intended to limit the scope of the present teachings in any way.

Fig. 1 depicts a schematic that highlights metastable ion formation in a MALDI-TOF mass spectrometer.

Fig. 2 depicts a MALDI-TOF-MS spectrum for a mixture of synthetically pure phosphopeptides LQRQPSpSPGPTPR (SEQ ID NO: 1), RYPRPVpSVPPSPSLSR (SEQ

ID NO: 2), GTSFMM_pTPYVVTRY_pYR (SEQ ID NO: 3) and CTNFMM_pTP_pYVVTRY_pYR (SEQ ID NO: 4) in reflector mode of operation (note: a “p” to the left of an amino acid residue identifies the site of phosphorylation as the residue immediately to the right of the “p”).

5 Fig. 3 depicts a MALDI-TOF-TOF spectrum for $m/z = 1500.6$, a putative phosphopeptide identified by the detection of a metastable ion at $m/z 1405.3$ in the MALDI-TOF-MS spectrum shown in Fig.2.

Fig. 4 depicts a MALDI-TOF-TOF spectrum for $m/z = 1874.8$, a putative phosphopeptide identified by the detection of a metastable ion at $m/z 1779.7$ in the
10 MALDI-TOF-MS spectrum shown in Fig.2.

Fig. 5 depicts a MALDI-TOF-MS spectrum of an in-solution trypsin digest of a protein, α -casein, collected in reflector mode of operation of the MALDI-TOF analyzer. The upper panel is an expansion of the MALDI-TOF MS spectrum that visualizes the metastable ion used to identify the phosphopeptide precursor ion.

15 Fig. 6 depicts a MALDI-TOF spectrum for a synthetically pure phosphopeptide DLDVPIPGRFDRRV_pSVAAE (SEQ ID NO: 5).

Fig. 7 depicts a flow chart of various embodiments of the current teachings that can be used for increasing confidence of detected peptides.

Fig. 8 schematically depicts phosphorylated serine (modified peptide) and its
20 dissociation into a de-phosphosphate serine (de-modified peptide) and serine (non-modified peptide).

Fig. 9 is a mass spectrum of the tryptic digest of a phosphoprotein, α -casein.

Figs. 10 and 11 depict MS/MS spectra of various peaks in the mass spectrum of Fig. 9.

DESCRIPTION OF VARIOUS EMBODIMENTS

The present teachings provide a method of identifying the presence of various classes of biomolecules (e.g., peptides, proteins) that are modified in biological systems. For example, various constituents or moieties that modify biomolecules can have significant relevance to the state of the biomolecule and functional stability of the biological system to which the biomolecule pertains. In various embodiments for peptide identification, many of these moieties are linked to amino acid residues of peptides and proteins by labile chemical bonds that are susceptible to post source decay when analyzed by MS, such as in a MALDI-TOF mass spectrometer. The methods of the present teachings can be used with an axial MALDI-TOF MS such as the Voyager™ workstation (Applied Biosystems, Framingham, MA) or an axial MALDI-TOF-TOF MS instrument system such as the 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA).

In various embodiments involving MALDI-TOF MS analysis, biological samples of interest (e.g., peptides, proteins, nucleic acids, carbohydrates and the like) can be prepared by incorporation of the analyte sample into a solid phase matrix. The MALDI mass spectrometer instrument can use a specifically tuned laser to ionize the sample in a vacuum region called the MALDI source region. Liberated analyte ions, now in the gas phase, can be accelerated by application of an electrical potential to enter a field free region where they separate as a function of their mass to charge ratio and can thus be detected and categorized by this parameter. A phenomenon common to some ions with labile chemical linkages is dissociation into smaller ions in the field free region of the mass spectrometer in a process commonly known as “post source decay”. As used herein, ions that dissociate in the field free region are termed metastable ions. These metastable ions in the field free region have the same velocity as the precursor ions from which they were formed, but possess lower kinetic energy due to their lower mass. Typically MS instrument acquisition parameters are specifically tuned for the maximum sensitivity and

resolution for ions formed in the ion source region of the instrument. As a result, these ions are not focused to the same extent as the ions generated in the ion source region. Such lack of focusing of metastable ions results in a lower mass resolution as compared to the resolution of intact ions (or ions that dissociate in the ion source region) when
5 detected, e.g., using a TOF analyzer. Another reason for the lower measured resolution of metastable ion fragments is that some of the chemical energy released during metastable ion formation is converted to kinetic energy in the fragment ions. Such increased kinetic energy typically does not change the average energy in the time-of-flight direction but can increase the energy spread of the fragment ions. Also since these fragments are formed
10 outside of the ion source region, the application of delayed extraction techniques (which can impart velocity focusing properties) as described, e.g., in U.S. Patent No. 5,627,369 do not operate on such fragment ions. Such metastable ions are usually detected in MALDI-TOF mass spectra that are collected in a reflector mode of operation of these analytical instruments.

15 In various embodiments, the methods of the present teachings can identify the presence of phosphopeptides. In various embodiments, these methods take advantage of the chemical lability of the phosphopeptides linkage, which results in dissociation of the phosphopeptide bond forming phosphoric acid and a lower mass peptide without its original phosphate group. This dissociation can take place, e.g., either in an ion source
20 region after laser desorption/ionization or in a field free region of, e.g., a MALDI-TOF mass spectrometer. As used herein, the term de-phosphorylated peptide ions refers to ions that dissociate in the MALDI ion source region. Metastable ions generated in the field free region are not focused by the ion optics to the same extent as the ions generated in the ion source region resulting in lower mass resolution of these ions in TOF separation and
25 ion detection. In various embodiments, we have observed that the metastable ions can be

detected by a peak that is approximately 95 mass units (u) lower than the m/z value for the precursor ion from which it was generated. Identification of this metastable response in the MALDI-TOF reflector MS spectrum of a protein digest can be used to identify the presence of the putative phosphorylated peptide, i.e., as the peak that is about 95 u higher than the metastable ion and that has a substantially higher mass resolution than the metastable ion. The intact phosphorylated peptide can have a resolution that is substantially the same as the average resolution of other peptide peaks contained in the MALDI-TOF MS spectrum. However, metastable ions can have a resolution that is typically about 50% or lower than that of the average mass resolution over a given mass range. This observed mass difference (i.e., about 95 u) is not intuitive since the mass of a peptide which has lost its phosphate group should be theoretically 98 u lower than its phosphorylated counterpart and, indeed, this is the mass difference (-98 u) observed for phosphopeptides that dissociate in the ion source region of the MALDI-TOF mass spectrometer.

The value of mass difference between metastable ion and putative phosphopeptide (or other modified peptide) can be derived experimentally and can be dependent on instrument geometry. For example, analysis from reflector MS spectra collected on a 4700 Proteomics Analyzer from an LC MALDI study (in which 40 phosphorylated peptides were identified by the method of the present teachings) found the mass difference to be on average $95 \text{ u} \pm 1 \text{ u}$. In contrast when a Voyager DETM STR (Applied Biosystems) instrument was used, the experimentally derived mass difference was found to be nominally 94 u (see Fig. 6).

We have discovered that the lower than expected m/z ratio difference of the metastable ion compared to its precursor phosphopeptide ion can be explained with reference to the design of a MALDI TOF instrument with an ion mirror (reflector)

configuration. For this explanation, a reflector MALDI-TOF instrument can be thought of as having three distinct regions, a source region, a field free region and a mirror region. If we use D_s for the length of the source region, D_f for the length of the field free region, and D_m for the length of the mirror region, then the time-of-flight in each region can be
 5 determined by the following:

$$t_s = \alpha \frac{2D_s}{\sqrt{V_s}} \sqrt{m} \quad (1)$$

$$t_f = \alpha \frac{D_f}{\sqrt{V_s}} \sqrt{m} \quad (2)$$

$$10 \quad t_m = \alpha \frac{4D_m}{\sqrt{V_s}} \sqrt{m} \quad (3)$$

where t_s represents the time-of flight within the source region, t_f represents the time-of-flight within the field free region, t_m represents the time-of-flight within the mirror region, α represents a constant, m represents the mass of the precursor peptide ion,
 15 and V_s represents the source voltage.

Consider for example, metastable ions that are generated from a precursor peptide with a loss of mass, Δm (98 u for phosphopeptide), in the field free region. The metastable ions fly with their precursor ions at substantially the same velocity until they
 20 reach the mirror region. Thus, the time-of-flight difference between metastable ions and precursor ions arises in the mirror region.

Up to the mirror, the metastable ion has the same initial velocity, but different mass from its precursor ion. The time it takes for the metastable ion to travel through the mirror region can be given by:

$$25 \quad t_m^m = t_m \frac{m - \Delta m}{m} \quad (4)$$

where t_m^m represents the time the metastable ion takes to travel through the mirror region.

The time t and mass m relationship is typically determined by that of the precursor peptide ion:

$$5 \quad t = t_s + t_f + t_m = \alpha \sqrt{\frac{m}{V_s}} (2 D_s + D_f + 4 D_m) \quad (5)$$

or

$$m = t^2 V_s / (\alpha^2 (2 D_s + D_f + 4 D_m)^2) \quad (6)$$

The time-of-flight of the metastable ion can be expressed as $t^m = t - t_m + t_m^m$.

10 According to equation (6), the observed mass, m^m , for the metastable ion can then be given by:

$$m^m = (t - t_m + t_m^m)^2 V_s / (\alpha^2 (2 D_s + D_f + 4 D_m)^2) \quad (7)$$

Rearranging and substituting yields the following relation:

$$15 \quad m^m = m (1 - \beta \frac{\Delta m}{m})^2 \quad (8)$$

$$\text{with } \beta \equiv \frac{t_m}{t} = \frac{4 D_m}{2 D_s + D_f + 4 D_m} \quad (9)$$

Thus, the observed mass difference between the precursor peptide ion and its
20 metastable counterpart ion can be given by:

$$m - m^m = 2 \beta \Delta m - \beta^2 \frac{\Delta m^2}{m} \quad (10)$$

As described by Vestal et al. in J. Am. Soc. Mass Spectrom. 1998,9,892-911
("Vestal"), which is hereby incorporated by reference herein, for a MALDI-TOF
25 instrument with a single-stage source and mirror, the focusing condition can be given by the following condition:

$$D_f = 2 D_s + 4 D_m \quad (11)$$

Thus,

$$30 \quad \beta = \frac{2 D_m}{D_f} \quad (12)$$

for such an instrument.

As taught by Vestal, high mass resolution can be achieved when $\frac{4D_m}{D_f}$ is in the range of 0.98 to 1, which corresponds to a β in the range of 0.49 to 0.5. However, if one writes D_m in terms of D_f and D_s , a value for $\beta \equiv 0.5$ could correspond to an infinitely small source region, and thus could be a physically impractical value.

In various embodiments, a reasonable dimension would be a source region 5 mm long, a field free region 1000 mm long and a mirror length calculated according to equation (10). This ion optics configuration would have a β that is 0.495.

For a phosphopeptide, $\Delta m = 98$ u. The experimentally observed mass between the metastable ion, m^m , and its precursor ion, m , can be derived using equation (10):

$$m - m^m = 95.3 \pm 0.7 \text{ u, e.g., in the mass range of 1000 to 2500 u.}$$

This derived value (previously referred to as the experimentally derived mass difference) represents the average observed mass difference over a given mass range and is substantially in accord with results observed experimentally. We refer to this derived value herein as the adjusted mass difference.

For different types of instruments, e.g., a different mirror length to drift tube length might be used, β can be different, resulting in different observed metastable ion mass. For example, if β is 0.49, the observed delta mass would be $m - m^m = 94.4 \pm 0.7 \text{ u, e.g., in the mass range of 1000 to 2500 u.}$

In various embodiments representing a typical workflow that can be carried out in accordance with the present teachings, a peptide MALDI sample can be prepared by mixing the sample with MALDI matrix such as α -cyano-4-hydroxycinnamic acid in a solution and then spotting the solution on a MALDI plate (MALDI target). After the solution is dried, the MALDI plate can be loaded into the ion source region of a MALDI

TOF mass spectrometer. The MALDI TOF MS instrument can be run in MS only mode to generate a mass spectrum of the mixture of peptides having a plurality of peaks corresponding to the peptide ions over a range of mass-to-charge values. Typically data collection can be carried out in a reflector mode of operation to identify component masses (or component mass-to-charge ratios) in the sample due to higher resolution and higher mass accuracy measurements when the MALDI TOF MS instrument is operated in this mode. Using techniques known to those of skill in the art, the resolution of each of the peaks within the range of mass-to-charge values can be calculated by determining the mass value at the peak divided by the width of the peak (measured at 50% of the peak intensity). From the individual resolutions, the average mass resolution for the peptide peaks over a given mass range can be determined. One or more of the peaks in the mass spectrum with a resolution substantially lower than the average resolution of the identified peaks (see, for example, Figs. 2, 5 and 6) can be selected as a putative metastable ion of an intact peptide precursor ion. The adjusted mass difference (e.g., 95 u in this example for phosphorylated peptides) can be added to the mass-to-charge value of the selected putative metastable peak to determine whether or not a peak exists in the spectrum at the newly calculated mass-to-charge value. If such a peak exists, it can be identified as a phosphopeptide precursor ion. The foregoing workflow can be repeated to identify the presence of other putative phosphopeptides present in the spectrum. According to the questions being addressed by the MS analysis, the data derived from collection of the MS spectra can be used for protein identification through protein database search of peptide mass fingerprinting (PMF) or the data can be used for the selection of precursor ions for further experiments such as, for example, using tandem mass spectrometry (MS/MS) to identify proteins or to confirm the phosphopeptide identified in accordance with the present teachings is indeed a phosphopeptide. In various embodiments, the instrument for

conducting MS/MS analysis is an Applied Biosystems 4700 Proteomics Analyzer, but other tandem MS instruments can be used also.

In various embodiments, the functionality of one or more of the methods of the present teachings described herein is implemented as computer-readable instructions on a general purpose computer. The computer can be separate from, detachable from, or
5 integrated into a mass spectrometry system. The computer-readable instructions can be written in any one of a number of high-level languages, such as, for example, FORTRAN, PASCAL, C, C++, or BASIC. The computer-readable instructions can be written in a script, macro, or functionality embedded in commercially available software, such as
10 EXCEL or VISUAL BASIC. The computer-readable instructions could be implemented in an assembly language directed to a microprocessor resident on a computer. For example, the computer-readable instructions could be implemented in Intel 80x86 assembly language if it were configured to run on an IBM PC or PC clone.

In various embodiments, the computer-readable instructions can be embedded on
15 an article of manufacture including, but not limited to, a computer-readable program medium such as, for example, a floppy disk, a hard disk, an optical disk, a magnetic tape, a PROM, an EPROM, CD-ROM or DVD-ROM. As used herein, the terms "algorithm" and "computer algorithm" refer to computer-readable instructions.

In various embodiments, the methods of the present teachings can be performed by
20 the MALDI-TOF MS instrument software using a computer algorithm that implements one or more methods of the present teachings to detect and identify peptide peaks that are modified with functional groups that undergo facile cleavage when analyzed by the MALDI-TOF mass spectrometer. For example, in various embodiments, an algorithm implements a method of the present teachings that uses the following strategy to identify
25 metastable ions that define the presence of post-translationally phosphate modified

peptides in complex mixtures such as are generated when mixtures of proteins are digested with, e.g., enzymes (such as trypsin, chymotrypsin, lysine endopeptidase, ARG-C, ASP-N, V8, or other enzymes known to those of ordinary skill in the art), chemical reagents (such as cyanogens bromide or the like), or combinations thereof, in proteomics experiments to facilitate protein identification and characterization. The mass resolution of the peaks in the mass spectrum (e.g., generated in linear mode, reflector mode, etc.) can be determined by MALDI TOF instrument software across a portion or the whole mass range. Peaks of lower resolution than values measured for typical peptide responses are identified as potential derivatives of phosphopeptides and stored in memory. For example, the algorithm can be set to identify peaks as putative metastable ions with resolutions ranging from 3,000 to 6,000 depending on instrument tuning conditions. Subsequently ions having m/z values 95 u greater than the detected ion of low resolution stored in the previous step and also having a resolution that is consistent with the mass resolution for peptides analyzed by MALDI-TOF MS (typically >10,000) are identified and stored as being predicative of potential phosphopeptides. Ion signals selected by this approach represent putative phosphopeptides in the analyzed mixture. Further confirmation of this prediction can be accomplished by further analysis of the putative phosphopeptide by tandem mass spectrometry (MS/MS) such as, but not limited to, MALDI-TOF-TOF-MS analysis.

In various embodiments, a method of the present teachings can be used with a wide variety of workflows as set forth in detail below. For example, various embodiments of the methods of the present teachings can also be used with either a protein-based or peptide-based workflow.

In various embodiments, protein mixtures that contain phosphoproteins can be separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE). Bands from gel electrophoresis separation of the protein mixture can be excised and proteins contained

therein digested by an in-gel (in-situ) enzymatic or chemical digestion strategy. Peptide mixtures that result from such digestion are analyzed by mixing the peptide mixture (sample) with matrix (such as alpha-cyano-4-hydroxycinnamic acid, 2,5-dihydroxybenzoic acid, sinnapinic acid or the like), dried and analyzed by MALDI-TOF-MS. One or more methods of the present teachings is then applied to identify putative phosphopeptides.

In various embodiments, protein mixtures that contain phosphoproteins can be separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE). Bands from gel electrophoresis separation of the protein mixture can be excised and proteins contained therein digested by an in-gel (in-situ) enzymatic or chemical digestion strategy. Liberated peptide mixtures can be further separated by high performance liquid chromatography (HPLC) methods such as, e.g., one dimension reversed phase, multiple dimensions in which two or more orthogonal dimensions of separation can be coupled to separate such complex mixture of peptides, and combinations thereof. Ultimately, separated peptide mixtures are co-deposited with a MALDI matrix onto a MALDI plate, or directly mixed with MALDI matrix, and deposited onto a MALDI plate. Resultant samples are analyzed by MALDI-TOF-MS. One or more methods of the present teachings is then applied to identify putative phosphopeptides.

In various embodiments, protein mixtures that contain phosphoproteins can be separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE). Resultant gels can be probed using a molecular scanner approach. By this approach proteins are electroblotted through a membrane that is impregnated with an enzyme (such as trypsin or the like). As the proteins transverse the enzyme membrane they are digested into peptides, which are subsequently captured on a peptide capture membrane. Subsequently the peptide capture membrane is sprayed with a MALDI matrix (such as alpha-cyano-4-hydroxycinnamic acid, 2,5-dihydroxybenzoic acid, sinnapinic acid or the like), dried and analyzed by

MALDI-TOF-MS. One or more methods of the present teachings is then applied to identify putative phosphopeptides.

In various embodiments, protein mixtures that contain phosphoproteins can be digested with an enzyme (such as trypsin, chymotrypsin, lysine endopeptidase, ARG-C, ASP-N, V8, or other enzyme known to artisans) or chemical reagents (such as cyanogens bromide or the like) as is performed during a global proteomics experiment. Liberated peptides can be separated using high performance liquid chromatography (HPLC) methods, such as, e.g., one dimension reversed phase, multiple dimensions in which two or more orthogonal dimensions of separation can be coupled to separate such complex mixture of peptides, and combinations thereof. Ultimately, separated peptide mixtures are co-deposited with a MALDI matrix onto a MALDI plate, or directly mixed with MALDI matrix, and deposited onto a MALDI plate as is practiced in the art of LC MALDI. Resultant samples are analyzed by MALDI-TOF-MS. One or more methods of the present teachings is then applied to identify putative phosphopeptides.

In various embodiments, phosphopeptide identification confidence can be increased by considering that the three forms of the peptide i.e., non-modified, phosphorylated (modified), and de-phosphate (de-modified), all bear the same amino acid sequence, but differ in regard to the degree of modification (e.g., phosphorylation). Therefore, confidence in phosphopeptide identification, can be increased by acquiring an MS/MS spectra for MS peaks corresponding to two or more of these three peptide forms to provide at least a partial sequence of each peptide form so analyzed, and performing a database search using these sequences, taking into account possible phosphorylation sites. A commercially available database that can be used for performing such searches is MASCOT® (Matrix Sciences Ltd.)

Sequencing peptides by tandem mass spectrometry (MS/MS) is typically achieved by selecting the molecular ion of a peptide (often referred to as “the precursor ion” or “parent ion”) with a first mass spectrometer (often referred to as the first dimension of mass spectrometry) and directing the precursor ion into an ion fragmentor (such as, e.g., a collision cell where it collides with an inert gas). The precursor ion is fragmented in the collision cell to a series of fragment ions (often referred to as “daughter ions”), among which are a ladder of ions with sequentially decreasing numbers of amino acids. Two prominent sets of ions are typically produced. One is a sequence ladder with amino acid deletions from the C-terminal end of the peptide. Another is a ladder with amino acid deletions from the N-terminal end. The fragment ions are then typically directed into a second mass spectrometer (often referred to as the second dimension of mass spectrometry) to resolve the fragmentation pattern of the precursor ion, i.e., the selected peptide.

For example, many database search approaches first identify database peptides with a mass matching that of the precursor under consideration. The next step is to match the precursor’s MS/MS spectrum to theoretical MS/MS spectra of the database peptides. Generally, the closest match is declared the winner. Modifications to peptides, e.g., phosphorylation, glycosylation, sulfation etc., can be taken into account in the database search, e.g., by addition of the expected mass of the modifying moiety to those peptides on which such modification generally occurs. Phosphorylation generally occurs on serine, threonine and tyrosine and can be taken into account by adding the expected mass of the phosphorylation (80 u) to the mass of the database peptides prior to identifying putative matches (hits). Then, a plurality of potential matches can be formed for each hit by phosphorylating each of the serine, threonine and tyrosine sites prior to forming the theoretical spectra.

This process can be repeated when searching for a match to the de-phosphate peptide precursor. Instead of adding the phosphorylation mass (80 u) to the database peptide, the mass of the lost water molecule (18 u) is subtracted from the unmodified peptide's mass. Similar to the phosphorylation-related search, each of the serine, threonine and tyrosine site can be de-phosphated prior to forming the theoretical spectra. Similarly, a database search can be performed on the precursor for the peak corresponding to the unmodified peptide. In various embodiments, if all three searches yield a peptide with the same amino acid sequence, the confidence in the search result can be significantly increased.

Fig. 7 illustrates a flowchart of various embodiments of the current teachings. At 705 a MS scan is performed. This scan can then be searched for peak relationships corresponding to the phosphorylation or other peptide modifications, e.g., glycosylation and sulfation. Phosphorylation can be present, for example, if there are peaks that are 18 u lower and 80 u higher than the current peak. In this way, groups of the peaks corresponding to a de-phosphate (de-modified) peptide, its corresponding non-modified peptide and its corresponding phosphorylated (modified) peptide can be formed (710). The process can proceed by acquiring an MS/MS spectra for the de-phosphate peptide precursor and searching the database taking into account the loss of the water molecule (720). A similar acquisition and search can be performed for the precursor corresponding to the phosphorylated peptide (730). And an acquisition and search can be performed for the non-modified peptide (740). Finally, the results of two or more of the searches can be compared to identify a common peptide (750). Additional control logic can be added (702, 715, 760) in order to process multiple groups in a list.

One skilled in the art will appreciate that any two of the three acquisition/search steps can be used and that the order of the acquisition/searches steps is generally not

significant. As well, the exact type of database search algorithm is generally not important so long as it permits taking into account the various modifications. Some embodiments will compute the confidence in the result by accounting for differences in the ranking of the search results in each of the steps. For example, if two of the three searches produce a common peptide with a high score and the same peptide in the third search has a low score, the final confidence in the result may be lower than if all three searches yielded a common peptide with a high score.

Fig. 8 schematically illustrates the dissociation of a phosphorylated serine peptide **802** (modified peptide) with a formal residue mass of about 167 u. When irradiated with light (hv), this can lead to fragmentation to produce: (i) de-phosphate serine **804** (the de-modified peptide corresponding to the modified peptide) with a formal residue mass of about 69 u and a H_3PO_4 moiety (expected mass of about 98 u), and/or (ii) serine **806** (the non-modified peptide corresponding to the modified peptide) with a formal residue mass of about 87 u and a HPO_3 moiety (expected mass of about 80 u).

Aspects of the present teachings may be further understood in light of the following examples, which should not be construed as limiting the scope of the present teachings in any way.

EXAMPLES

Example 1: Formation of metastable ions within a MALDI-TOF mass spectrometer.

The schematic shown in Fig. 1 depicts a representation of the formation of metastable ions in a MALDI-TOF mass spectrometer. Here the ion represented by the smaller dashed circle formed in the field free region of the spectrometer is the metastable ion. As such, this ion will not be focused to the same extent as ions formed in the ion source region (as represented by the solid circles). Additionally, the ions formed in the

field free region have increased kinetic energy spread from chemical bond dissociation, but the average of the gain in kinetic energy is zero. Both attributes can combine to result in a lower apparent resolution of this ion population, which helps to identify such metastable ions and ultimately, in various embodiments of the present teachings, the peptides from which they were generated.

Example 2: Detection of metastable ions in a mixture of phosphopeptides when analyzed by MALDI-TOF-MS in reflector mode operation.

A MALDI-TOF mass spectrum collected in reflector mode of operation of a sample containing phosphopeptides LQRQSpSPGTPR (SEQ ID NO: 1), RYPRPVpSVPPSPSLSR (SEQ ID NO: 2), GTSFMMpTPYVVTRYR (SEQ ID NO: 3) and CTNFMMpTPpYVVTRYR (SEQ ID NO: 4) on an Applied Biosystems 4700 Proteomics Analyzer is shown in Fig. 2. The spectrum was an average of 20 sub-spectra for a total 1000 laser shots per spectrum. As seen across the mass range (900-2500), the observed resolution for most peptides is over 10,000. However, the mass resolution of certain ions is much lower (e.g., the ion at a m/z value of 1405.3 (**201**) was measured with a resolution of 3458 and the ion at a m/z value of 1779.7 (**202**) was measured with a resolution 4213). These poorly resolved peaks represent metastable ions. Both of these metastable ion associate peaks have corresponding modified peptide peaks about 95 u higher than themselves, and as such, these metastable ion peaks were identified as potentially originating from phosphorylated peptides. In this example the peak detected at $m/z = 1500.6$ (**203**) was predicted to be a phosphopeptide corresponding the metastable ion detected at $m/z = 1405.3$, and the peak detected at $m/z = 1874.9$ (**204**) was predicted to be a phosphopeptide corresponding the metastable ion detected at $m/z = 1779.7$.

Example 3: Confirmation that peptide of $m/z = 1500.6$ was a phosphopeptide.

A MALDI-TOF/TOF spectrum for the ion detected at $m/z = 1500.6$ was collected with an average of 25 sub-spectrum for a total 2500 laser shots per spectrum on an Applied Biosystems 4700 Proteomics Analyzer with high energy (1Kev) collision induced dissociation and is presented in Fig. 3. This spectrum depicts the typical 98 u loss from the precursor ion and many of the y and b-series ions identified in this spectrum. This tandem MS (MS/MS) spectrum of $m/z = 1500.6$ led to confirmation that this peptide was a phosphopeptide.

10

Example 4: Confirmation that peptide of $m/z = 1874.8$ was a phosphopeptide.

A MALDI-TOF/TOF spectrum for the ion detected at $m/z = 1874.8$ was collected on an Applied Biosystems 4700 Proteomics Analyzer with high energy collision induced dissociation and was an averaged of 25 sub-spectrum for a total 2500 laser shots per spectrum and is presented in Fig. 4. This spectrum depicts the typical 98 u loss from the precursor ion and many of the y and b-series ions identified in this spectrum. This tandem MS (MS/MS) spectrum of $m/z = 1874.8$ led to confirmation that this peptide was a phosphopeptide.

15

Example 5: Detection of metastable ions in a trypsin digest of α -casein when analyzed by MALDI-TOF-MS in reflector mode operation.

20

The phosphorylated protein, α -casein, was cleaved into fragments by an enzyme, bovine trypsin. These fragments are tryptic peptides which mean that all the peptides have either lysine or arginine at their C-terminus. These fragments are either non-phosphorylated peptides or phosphorylated peptides. This sample was mixed with

25

MALDI matrix, α -cyano-4-hydroxycinnamic acid, and was spotted on MALDI plate using a dry-droplet method. A MALDI-TOF reflector MS spectrum of an in-solution trypsin digest of this protein was collected on an Applied Biosystems 4700 Proteomics Analyzer with an average of 20 sub-spectra for a total 1000 laser shots per spectrum and is presented Fig. 5. A method using the metastable ion signal to identify a modified peptide, here a metastable ion signal of a phosphorylated peptide where an adjusted mass difference of 95 u was added to the mass-to-charge value of the metastable ion was applied to analysis of the trypsin digest of α -casein. In this example the method implemented successfully determined 3 potential phosphopeptides VPQLEIVPNpSAEER (SEQ ID NO: 6), YKVPQLEIVPNpSAEER (SEQ ID NO: 7), KYKVPQLEIVPNpSAEER (SEQ ID NO. 8), whereas a peptide mass fingerprinting (PMF) experiment only identified 2 phosphopeptides, VPQLEIVPNpSAEER, YKVPQLEIVPNpSAEER, with one missed cleavage site in the MASCOT database search. An example is that the method identified a peak at 1952.0 (501) as a phosphopeptide whose sequence is YKVPQLEIVPNpSAEER. This peptide sequence was confirmed by a peptide mass fingerprinting using a MASCOT database search.

Example 6: Detection of metastable ions in a phosphopeptide when analyzed by Voyager STR MALDI-TOF-MS in reflector mode operation.

A method of the current teachings using the adjusted mass difference was applied to analysis of a phosphorylated peptide precursor ion DLDVPIPGRFDRRVpSVAAE (SEQ ID NO: 5) on another type of MALDI-TOF mass spectrometer of Applied Biosystems, the Voyager STR. In the method using the metastable ion signal to identify a modified peptide, an adjusted mass difference of 94 u was added to the mass-to-charge value of the metastable ion. The spectrum of this phosphopeptide was collected on an

Applied Biosystems Voyager DE™ STR with an average of 200 laser shots per spectrum and is presented in Fig. 6. In this example the method implemented by the algorithm successfully determined the presence of the phosphopeptide at 2191.7544 (**601**) and the ion at 2097.8196 (**602**) is a metastable ion.

5

Example 7: Continuation of detection of metastable ions in a phosphopeptide.

The phosphorylated protein, α -casein, was cleaved into fragments by an enzyme, bovine trypsin, and subject to mass spectrometry analysis using an Applied Biosystems 4700 Proteomics Analyzer to produce the spectrum shown in Fig. 9. A method using the

10 metastable ion signal to identify a modified peptide, here a metastable ion signal of a phosphorylated peptide where an adjusted mass difference of 95 u was added to the mass-to-charge value of the metastable ion, identified a putative phosphopeptide at 1952.0 u (**901**) and its de-phosphate peptide was at 1854.0 u (**902**) and the non-modified peptide was at 1872.0 u (**903**). To confirm these findings, an MS/MS spectrum (Fig.10) was taken

15 with the putative phosphopeptide (1952.0 u) as the precursor ion that was fragmented. The resultant MS/MS spectrum was subjected to a MASCOT database search using variable phosphorylation of serine, threonine and tyrosine. The peptide ion score from the MASCOT search was 50 and its confidence of identification was 99.7%. Another MASCOT database search was conducted of the MS/MS spectrum of the de-phosphate

20 peptide (1854.0 u), but this time using de-H₂O as the variable modification of serine, threonine and tyrosine. The results are shown in Fig. 11. In this instance, the peptide ion score was 85 and its confidence of identification was 100%. The peptide sequence found in both data base searches was the same, and thus, this example shows that the precursor phosphopeptide found at 1952.0 u was confirmed to be a phosphorylated peptide.

All literature and similar material cited in this application, including, but not limited to, patents, patent applications, articles, books, treatises, and web pages, regardless of the format of such literature and similar materials, are expressly incorporated by reference in their entirety. In the event that one or more of the incorporated literature and similar materials differs from or contradicts this application, including but not limited to defined terms, term usage, described techniques, or the like, this application controls.

The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described in any way.

While the present teachings are described in conjunction with various embodiments, it is not intended that the present teachings be limited to such embodiments or examples. On the contrary, the present teachings encompass various alternatives, modifications, and equivalents, as will be appreciated by those of skill in the art.

The claims should not be read as limited to the described order or elements unless stated to that effect. It should be understood that various changes in form and detail may be made without departing from the scope of the appended claims. Therefore, all embodiments that come within the scope and spirit of the following claims and equivalents thereto are claimed.

CLAIMS

1. A method of identifying the presence of a modified peptide in a mixture of peptides by a mass spectrometry instrument comprising the steps of:
 - 5 generating a mass spectrum of a mixture of peptides, the mass spectrum comprising a plurality of peaks over a range of mass-to-charge values;
 - determining a resolution value for at least a portion of the peaks within the range of mass-to-charge values;
 - setting a peptide resolution value based at least on the determined resolution value;
 - 10 selecting one or more peaks with a resolution lower than the peptide resolution value as a putative metastable ion of a modified peptide;
 - determining an adjusted mass difference between a modified peptide and a metastable ion derived from the dissociation of a moiety from an ion of the modified peptide;
 - 15 adding about the adjusted mass difference to the mass-to-charge value of the selected peak corresponding to the putative metastable ion to derive an empirical mass-to-charge value for a precursor ion of the modified peptide ;
 - and
 - 20 identifying that a modified peptide is present in the mixture of peptides if a peak exists in the mass spectrum at the empirical mass-to-charge value.
2. The method of claim 1, wherein the modified peptide is a phosphorylated peptide.
- 25 3. The method of claim 1, wherein the modified peptide is one or more of a glycosylated peptide and a sulfonated peptide.
4. The method of claim 2, wherein the moiety is H_3PO_4 .
- 30 5. An article of manufacture comprising a computer-readable media with computer readable instructions embodied thereon for performing the method of claim 1.
6. A method of identifying the presence of a modified peptide in a mixture of peptides by mass spectrometry comprising the steps of:

generating a MS/MS mass spectrum of a modified peptide in a mixture of peptides;
determining at least a portion of the sequence of the modified peptide and the site of modification based at least on the search of a database using
5 information from the MS/MS mass spectrum of the modified peptide;
generating a MS/MS mass spectrum of a de-modified peptide corresponding to the modified peptide in the mixture of peptides;
determining at least a portion of the sequence of the de-modified peptide and the site of modification based at least on the search of a database using
10 information from the MS/MS mass spectrum of the de-modified peptide;
comparing the at least portion of the sequence and the site of modification determined for the modified peptide to that determined for the de-modified peptide; and
identifying that a modified peptide is present in the mixture of peptides if the
15 sequence and the site of modification determined for the modified peptide substantially matches that determined for the de-modified peptide.

7. The method of claim 6, further comprising before the step of generating a MS/MS mass spectrum of a modified peptide, comprising the steps of:
20 generating a mass spectrum of a mixture of peptides, the mass spectrum comprising a plurality of peaks over a range of mass-to-charge values; and identifying the mass-to-charge values in the mass spectrum associated with one or more modified peptides.
- 25 8. The method of claim 7, the step of identifying the mass-to-charge values in the mass spectrum associated with one or more modified peptides, comprising the steps of:
determining a resolution value for at least a portion of the peaks within the range of mass-to-charge values;
30 setting a peptide resolution value based at least on the determined resolution value;
selecting one or more peaks with a resolution lower than the peptide resolution value as a putative metastable ion of a modified peptide;

determining an adjusted mass difference between a modified peptide and a metastable ion derived from the dissociation of a moiety from an ion of the modified peptide;

adding about the adjusted mass difference to the mass-to-charge value of the selected peak corresponding to the putative metastable ion to derive an empirical mass-to-charge value for an ion of the modified peptide ; and identifying the mass-to-charge values in the mass spectrum associated with one or more modified peptides as the empirical mass-to charge-value if a peak exists in the mass spectrum at the empirical mass-to charge-value.

9. The method of claim 6, wherein the modified peptide is a phosphorylated peptide.

10. The method of claim 6, wherein the modified peptide is one or more of a glycosylated peptide and a sulfonated peptide.

11. The method of claim 8, wherein the moiety is H_3PO_4 .

12. The method of claim 6, further comprising the steps of:

generating a MS/MS mass spectrum of a non-modified peptide

corresponding to the modified peptide in the mixture of peptides;

determining at least a portion of the sequence of the non-modified peptide based at least on the search of a database using information from the MS/MS mass spectrum of the non-modified peptide;

comparing the at least portion of the sequence determined for the modified peptide, the de-modified peptide and the non-modified peptide to each other; and

where the step of identifying that a modified peptide is present in the mixture of peptides comprises:

identifying that a modified peptide is present in the mixture of peptides if

the sequences determined for the modified peptide, de-modified peptide and non-modified peptide substantially match each other, and the site of modification determined for the modified peptide substantially matches that determined for the de-modified peptide.

13. An article of manufacture comprising a computer-readable media with computer readable instructions embodied thereon for performing the method of claim 6.
14. An article of manufacture comprising a computer-readable media with computer readable instructions embodied thereon for performing the method of claim 12.
15. A method of identifying the presence of a modified peptide in a mixture of peptides by a mass spectrometry instrument comprising the steps of:
- generating a mass spectrum of a mixture of peptides, the mass spectrum comprising a plurality of peaks over a range of mass-to-charge values; determining a resolution value for at least a portion of the peaks within the range of mass-to-charge values; setting a peptide resolution value based at least on the determined resolution value;
- selecting one or more peaks with a resolution lower than the peptide resolution value as a putative metastable ion of a modified peptide; determining an adjusted mass difference between a modified peptide and a metastable ion derived from the dissociation of a moiety from an ion of the modified peptide where the adjusted mass difference is given by
- $$m - m^m = 2 \beta \Delta m - \beta^2 \frac{\Delta m^2}{m}$$
- where, m represents the mass of the precursor peptide ion, m^m represents the observed mass for the metastable ion and β is a constant that depends on geometry of the instrument;
- adding about the adjusted mass difference to the mass-to-charge value of the selected peak corresponding to the putative metastable ion to derive an empirical mass-to-charge value for a precursor ion of the modified peptide ; and
- identifying that a modified peptide is present in the mixture of peptides if a peak exists in the mass spectrum at the empirical mass-to-charge value.
16. The method of claim 1, wherein the adjusted mass difference is given by

$$m - m^m = 2 \beta \Delta m - \beta^2 \frac{\Delta m^2}{m}$$

5 where, m represents the mass of the precursor peptide ion, m^m represents the observed mass for the metastable ion and β is a constant that depends on geometry of the instrument.

- 10 17. The method of claim 15, wherein the modified peptide is a phosphorylated peptide.
18. The method of claim 15, wherein the modified peptide is one or more of a glycosylated peptide and a sulfonated peptide.
- 15 19. The method of claim 17, wherein the moiety is H_3PO_4 .
20. A method of identifying the presence of a modified peptide in a mixture of peptides by mass spectrometry comprising the steps of:
- generating a MS/MS mass spectrum of a modified peptide in a mixture of peptides;
 - 20 determining at least a portion of the sequence of the modified peptide and the site of modification based at least on the search of a database using information from the MS/MS mass spectrum of the modified peptide;
 - generating a MS/MS mass spectrum of a non-modified peptide
 - 25 corresponding to the modified peptide in the mixture of peptides;
 - determining at least a portion of the sequence of the non-modified peptide based at least on the search of a database using information from the MS/MS mass spectrum of the non-modified peptide;
 - comparing the at least portion of the sequence and the site of
 - 30 modification determined for the modified peptide to the at least portion of the sequence for the non-modified peptide; and
 - identifying that a modified peptide is present in the mixture of peptides if the sequence and the site of modification determined for the modified

peptide substantially matches the sequence determined for the non-modified peptide.

21. An article of manufacture comprising a computer-readable media with
5 computer readable instructions embodied thereon for performing the method of
claim 20.

10

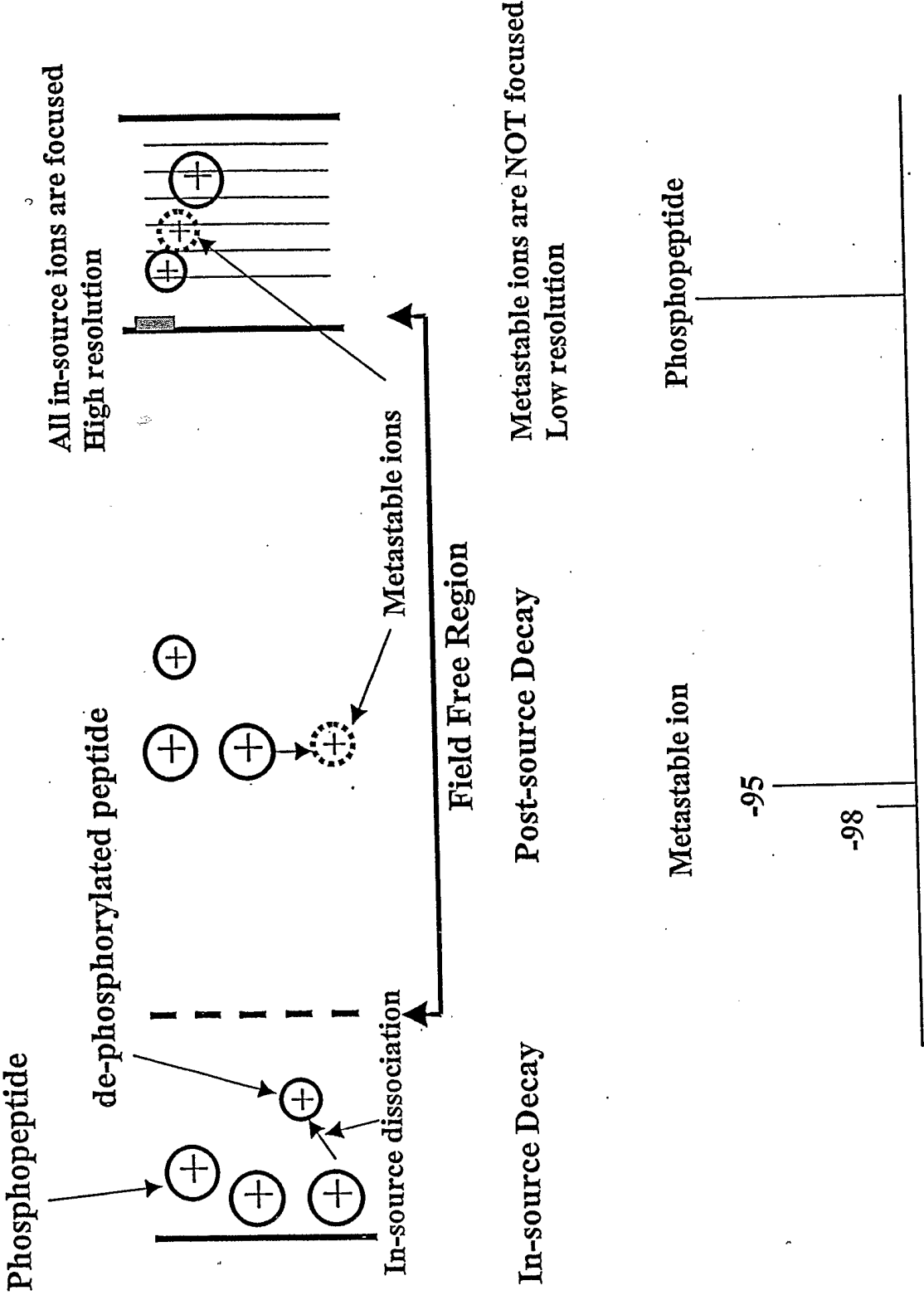


Fig. 1

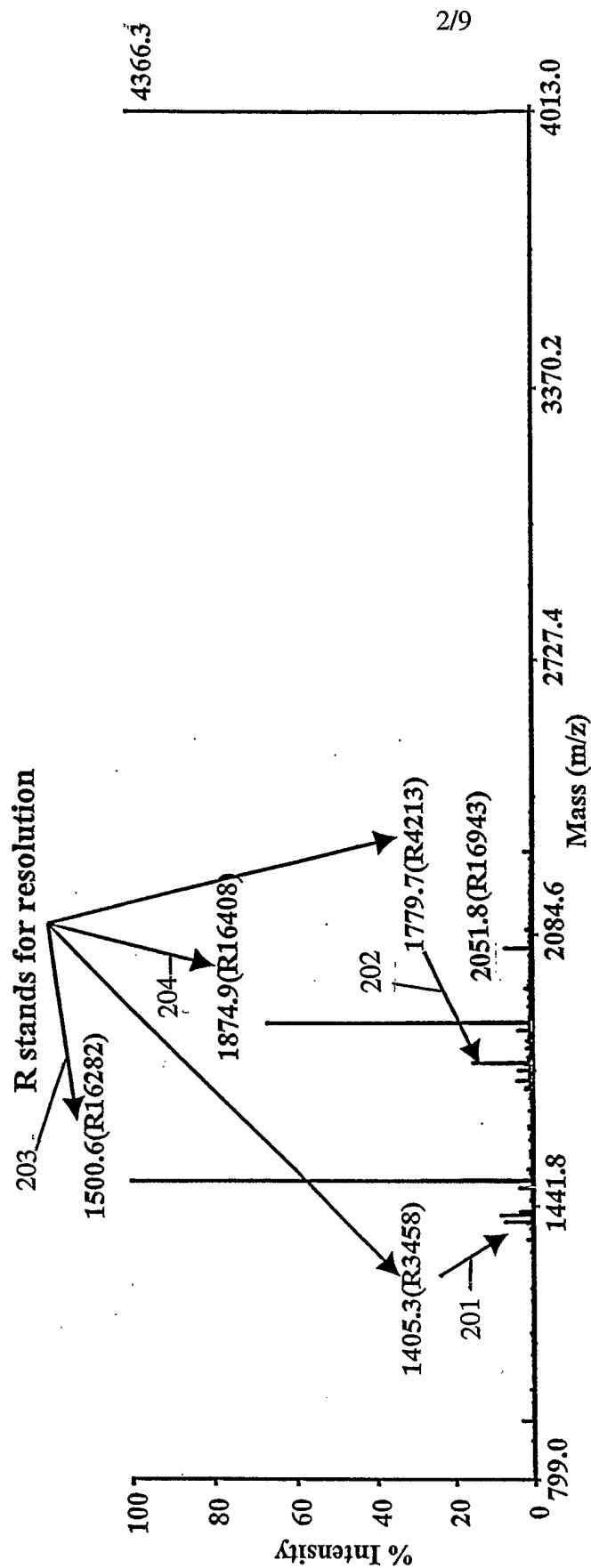


Fig. 2

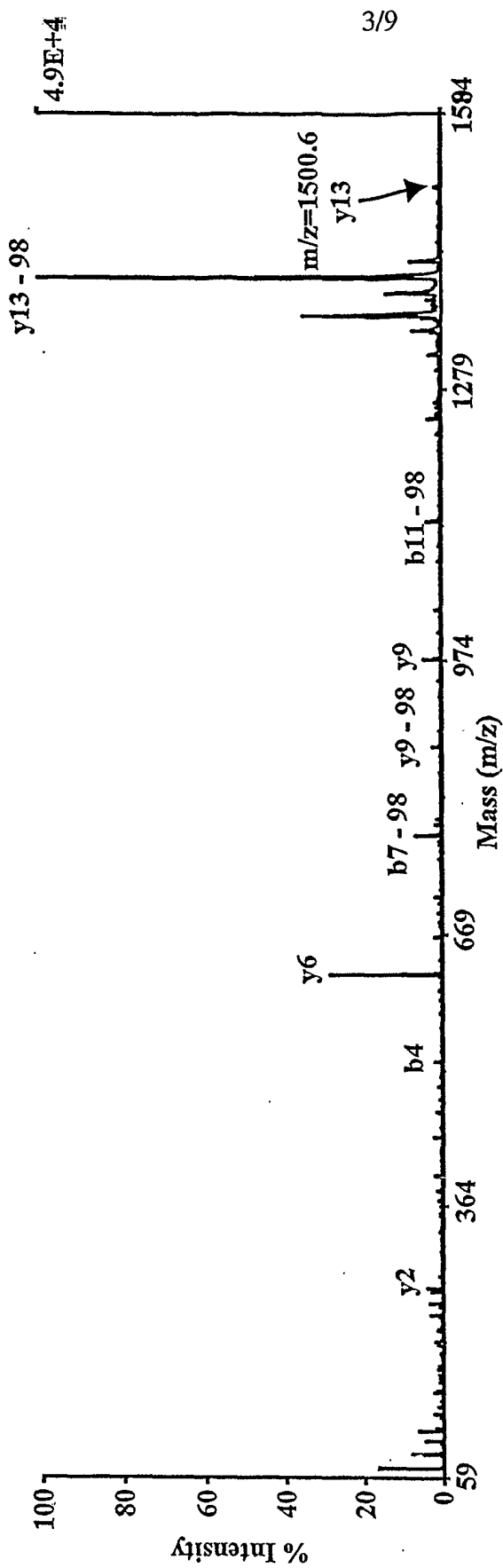


Fig. 3

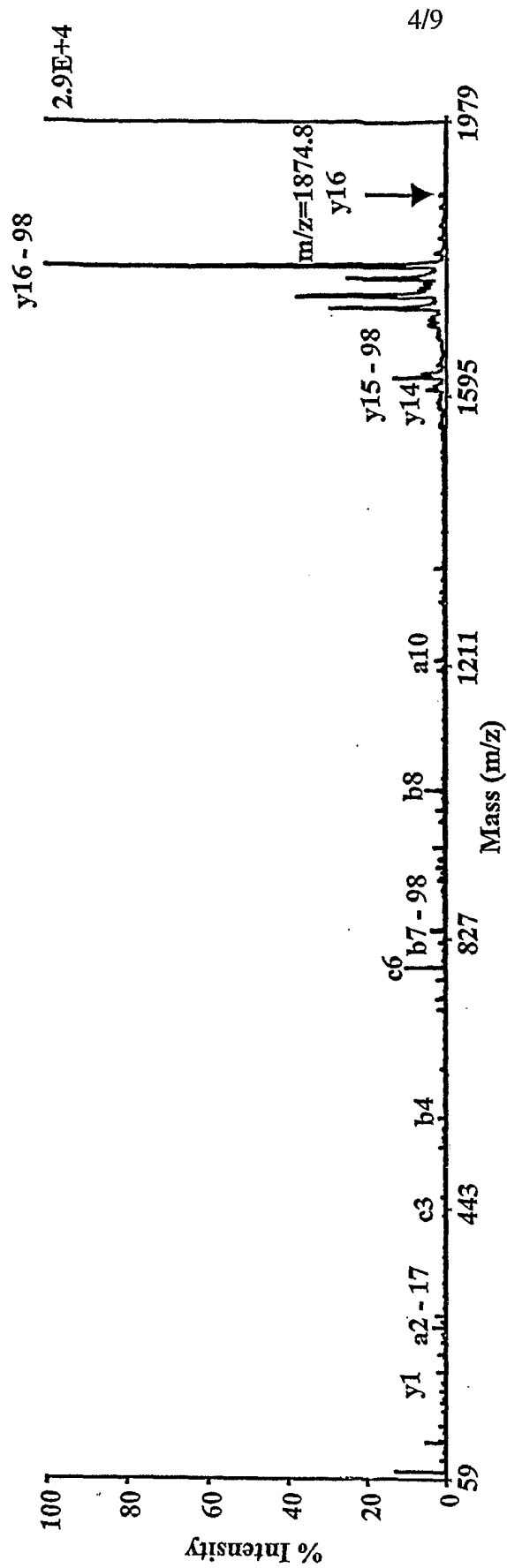


Fig. 4

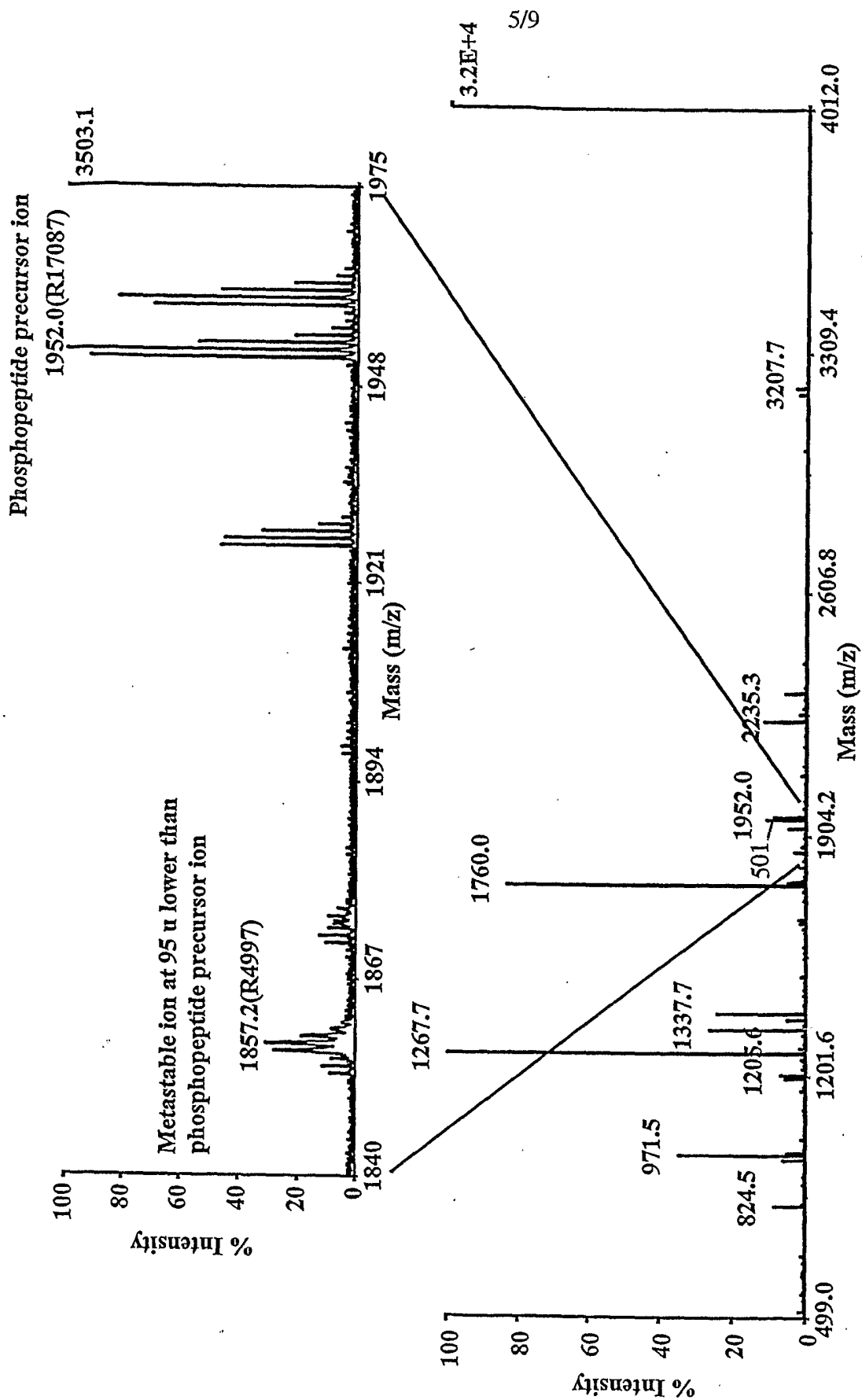


Fig. 5

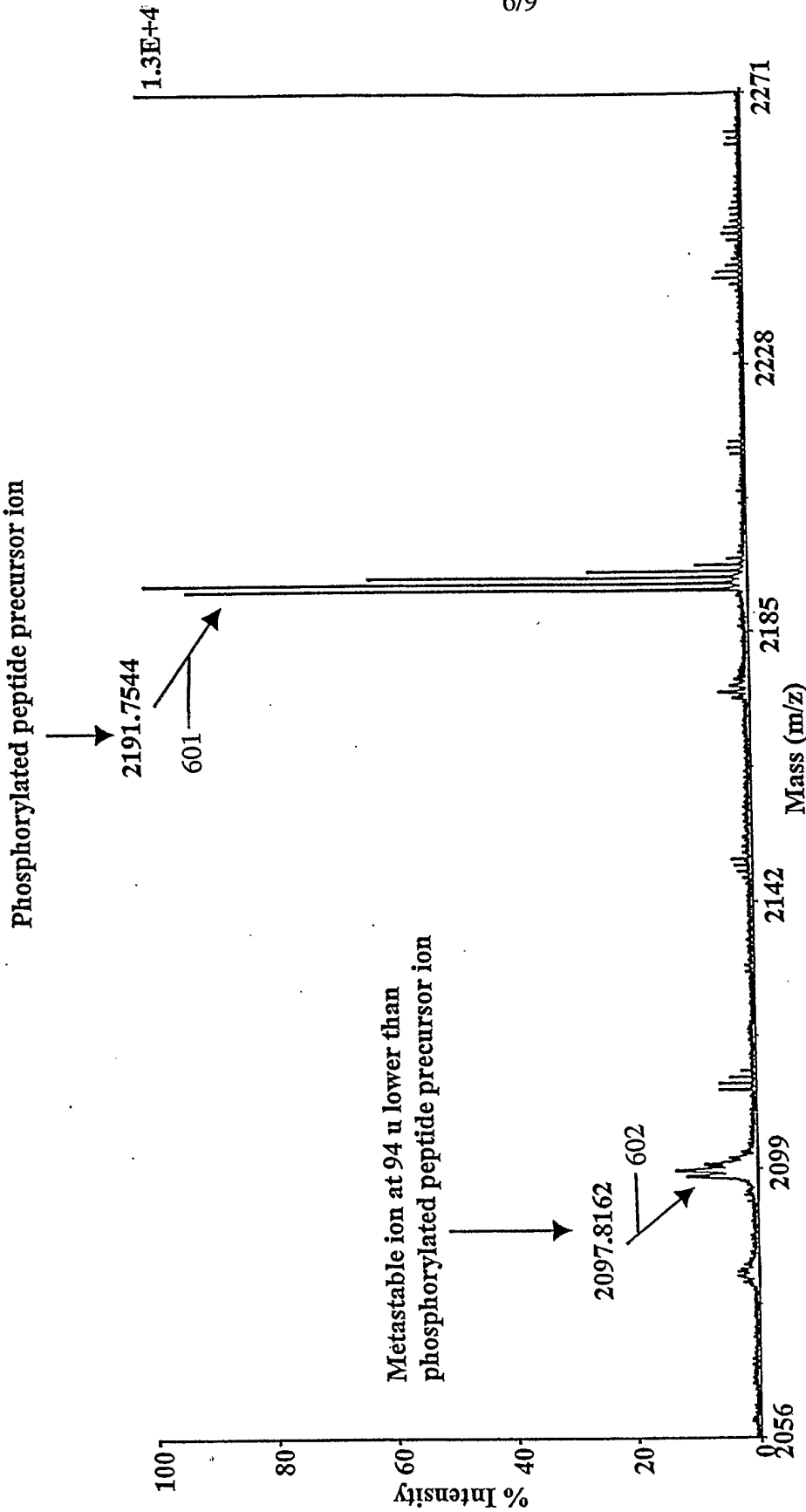


Fig. 6

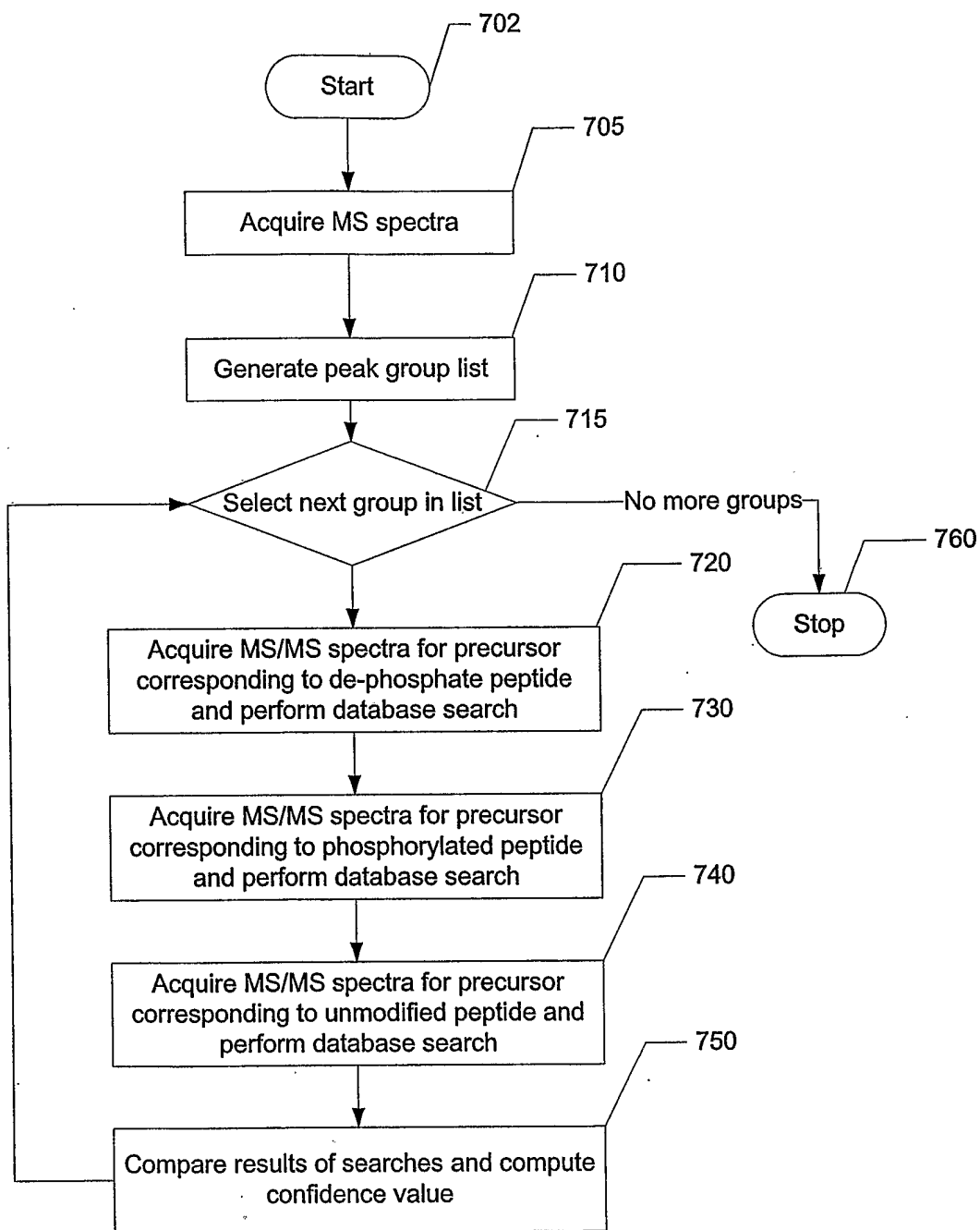


Fig. 7

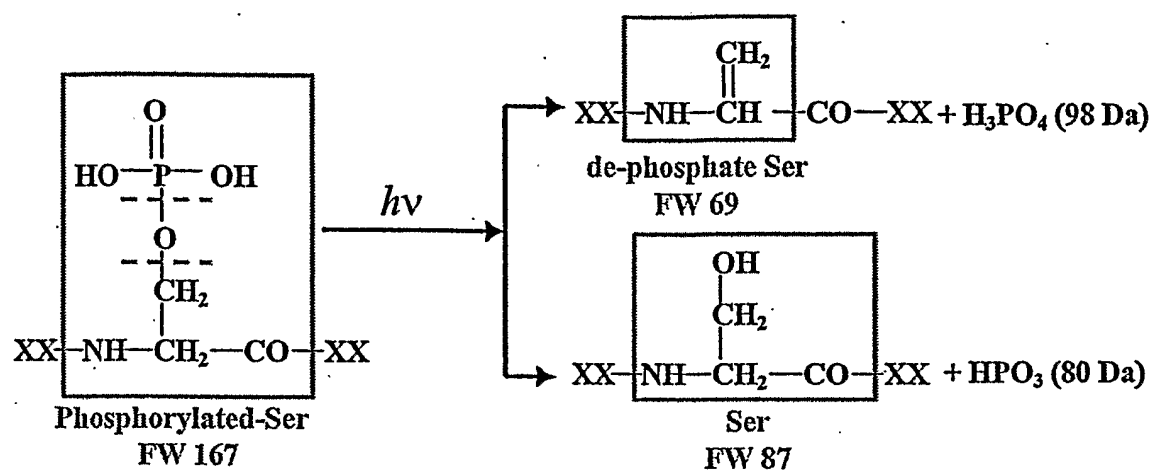


Fig. 8

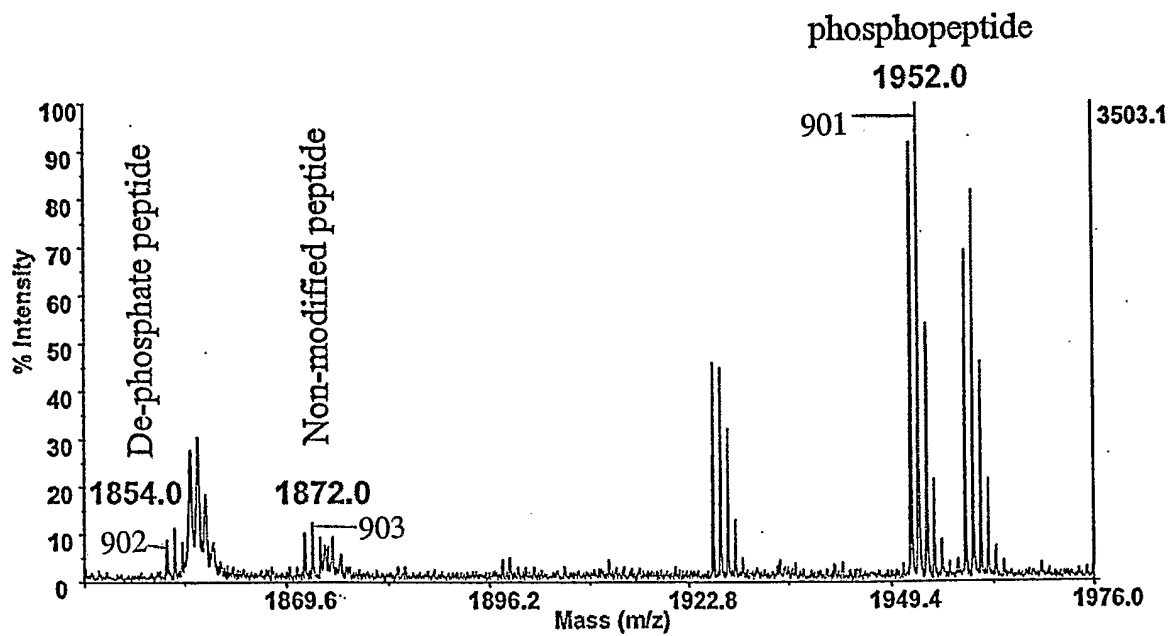


Fig. 9

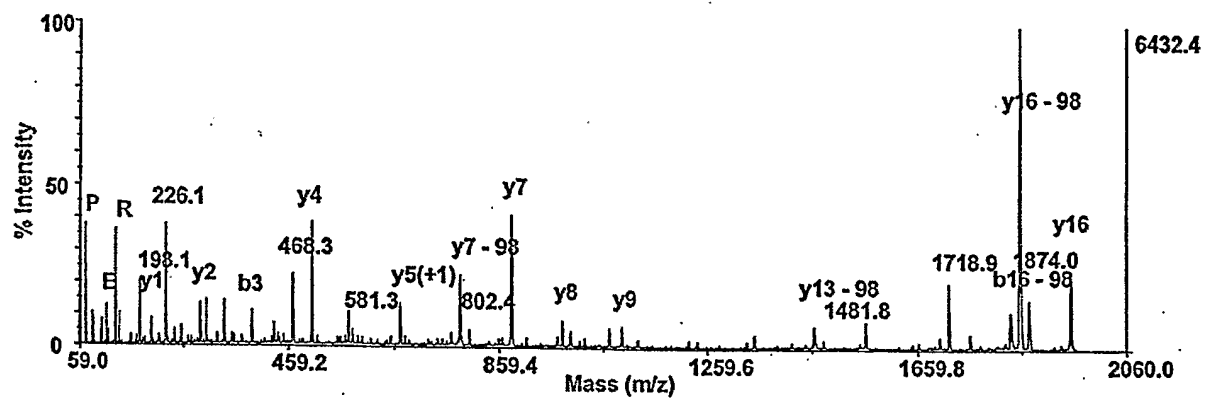


Fig. 10

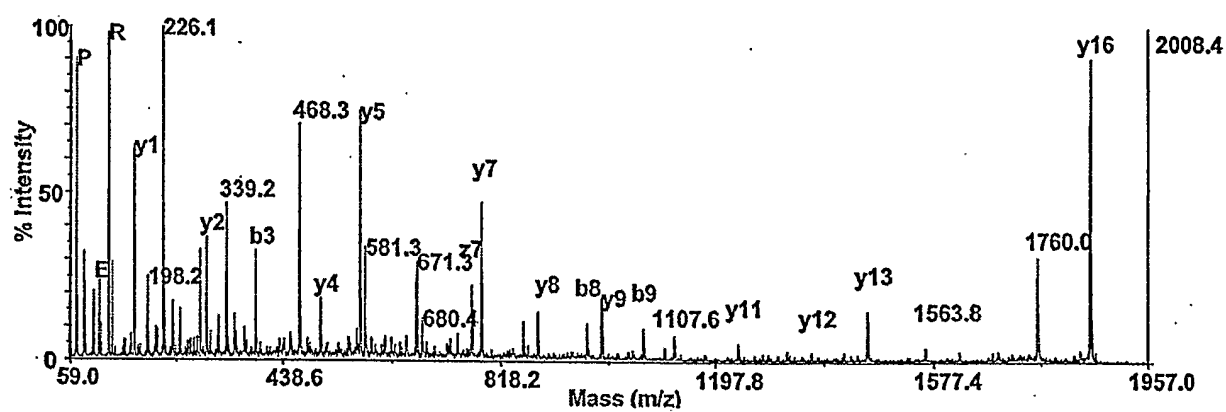


Fig. 11