

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 July 2009 (23.07.2009)

PCT

(10) International Publication Number  
WO 2009/090188 A1

(51) International Patent Classification:  
G01N 33/50 (2006.01) G01N 33/68 (2006.01)

(74) Agent: HABETS, Winand Johannes Anto; P.O. Box 5096, NL-6130 PB Sittard (NL).

(21) International Application Number:  
PCT/EP2009/050365

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

(22) International Filing Date: 14 January 2009 (14.01.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
08150294.0 15 January 2008 (15.01.2008) EP

(71) Applicant (for all designated States except US): UNIVER-  
SITEIT UTRECHT HOLDING B.V. [NL/NL]; Yalelaan  
40, NL-3584 CM Utrecht (NL).

(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, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

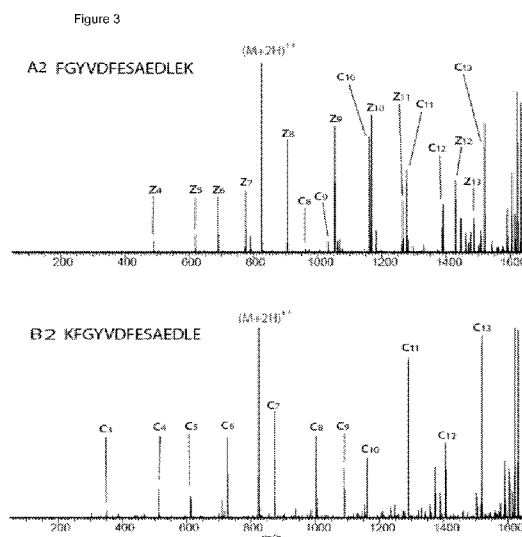
(72) Inventors; and

(75) Inventors/Applicants (for US only): HECK, Albert  
John Roeland [NL/NL]; Sorbonnelaan 16, NL-3584 CA  
Utrecht (NL). MOHAMMED, Shabaza [GB/NL]; Sor-  
bonnelaan 16, NL-3584 CA Utrecht (NL). BREUKELLEN,  
VAN, Bastiaan [NL/NL]; Sorbonnelaan 16, NL-3584 CA  
Utrecht (NL). TAOUATAS, Nadia [DK/NL]; Sorbon-  
nelaan 16, NL-3584 CA Utrecht (NL). DRUGAN,  
Madalina Mihaela [RO/NL]; Sorbonnelaan 16, NL-3584  
CA Utrecht (NL).

Published:  
— with international search report

[Continued on next page]

(54) Title: METHOD FOR DETERMINING THE AMINO ACID SEQUENCE OF PEPTIDES.



(57) Abstract: The invention is in the field of analytical methods suitable for biochemical applications and provides a method for determining the amino acid sequence of a peptide. The determination of amino acid sequences of proteins and peptides is useful in the study of biological systems. The invention relates to a method for determining at least part of the amino acid sequence of a protein comprising the steps of cleaving the protein into proteolytic peptides, ionizing the proteolytic peptides to generate peptide precursor ions, dissociating these peptide precursor ions using tandem mass spectrometry in order to obtain peptide fragment ions, followed by determining the amino acid sequence of a selected proteolytic peptide wherein the cleaving step generates at least one proteolytic peptide with an N-terminal lysine residue and wherein the dissociation of the peptide precursor ions is initiated by electron transfer.

WO 2009/090188 A1



— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

## METHOD FOR DETERMINING THE AMINO ACID SEQUENCE OF PEPTIDES.

### Field of the invention

5                   The invention is in the field of analytical methods suitable for biochemical applications and provides a method for determining the amino acid sequence of a peptide as well as a method for digesting a protein in a gel. The determination of amino acid sequences of proteins and peptides is useful in the study of biological systems.

10

### Background of the invention

                  The large scale study of structure and function of proteins and peptides is often termed proteomics. Proteins are vital parts of living organisms, as they are the main components of most physiological pathways in cells. The term "proteomics" was coined to make an analogy with genomics, the study of genes. The word "proteome" is a portmanteau of "protein" and "genome". The proteome of an organism is the set of proteins that are produced by the organism during its life, and its genome is its total set of genes.

15

                  Proteomics is often considered the next step in the study of biological systems, after genomics. The genome contains all the information required to construct an organism's protein complement. However, proteomics is much more complicated than genomics, mostly because while an organism's genome is rather constant, a proteome differs from cell to cell and constantly changes via biochemical interactions dictated by its immediate environment. One organism has the same genome in nearly every cell, nevertheless, it may have a radically different protein expression profile in different parts of its body, different stages of its life cycle and different environmental conditions. Another major difficulty is the complexity of proteins relative to genes. For example, the human genome consists of approximately 25 000 genes but it is estimated that more than 500 000 proteins can be derived from these genes. This increased complexity derives from mechanisms such as alternative splicing, post-translational protein modification (such as glycosylation or phosphorylation) and protein degradation.

20

25

30

35

                  Since proteins play a central role in the life of an organism, proteomics is instrumental in discovery of biomarkers, components that can indicate a particular disease. Current research in proteomics requires that the primary sequence

of proteins be resolved, sometimes on a massive scale.

Many techniques have been developed for protein sequencing, including deriving the amino acid sequence from a DNA or RNA sequence or directly from the protein itself, such as Edman degradation or analysis by Mass Spectrometry.

5 Nowadays, Mass spectrometry seems the method of choice for direct protein sequencing and a typical proteomics analysis may consist of the following five stages.

In stage 1, the proteins to be analysed are isolated from a biological source such as a cell lysate or tissue for instance by biochemical fractionation or affinity selection. This stage often includes a final step of two-dimensional gel  
10 electrophoresis, which usually separates proteins first by isoelectric point and then by molecular weight. Protein spots in a gel can be visualized using a variety of chemical stains or fluorescent markers. Proteins can often be quantified by the intensity of their stain. Once proteins are separated and quantified, they may be identified. Individual spots are cut out of the gel so that they contain a purified single protein species. The  
15 mass spectrum analysis of whole proteins is less sensitive than that of smaller peptides and the mass and charge of the intact protein by itself is insufficient for the identification of its primary amino acid sequence. Mass spectrometry can, in principle, sequence any size of protein, but the problem becomes computationally more difficult as the size increases. Proteolytic peptides are also easier to prepare for mass  
20 spectrometry than whole proteins, because they are more soluble. Therefore, proteins are preferably degraded into smaller proteolytic peptides in stage 2, for instance through enzymatic digestion. It should be noted that in certain cases, stage 1 may be omitted and the analyte of interest is directly subjected to stage 2.

In stage 2, degradation typically occurs enzymatically, for instance by  
25 trypsin digestion. Trypsin is a serine endoprotease found in the digestive system and catalyses the hydrolysis of peptide bonds, leading to proteolytic peptide fragments with C-terminally protonated amino acids. Trypsin predominantly cleaves proteins at the carboxyl side (or C-terminal side) of the amino acids lysine and arginine, except when either is followed by proline. In order to generate overlapping proteolytic peptide  
30 fragments, it is advantageous to use multiple enzymes with different specificities in this stage. Whereas trypsin is most commonly used, other enzymes employed for this purpose include pepsin, elastase, Lys-C, V8 (a Glu-C endoproteinase) and chymotrypsin.

In stage 3, the proteolytic peptide fragments are separated and  
35 delivered to the mass spectrometer. Separation may be achieved by one or more steps

of liquid chromatography (LC) such as high-pressure liquid chromatography (HPLC) using narrow-bore (often below 100 micron) columns. One method of delivering the peptides to the spectrometer is electrospray ionization (ESI). At the end of the HPLC column, the solution is sprayed out of a narrow nozzle charged to a high positive  
5 potential into the mass spectrometer. The charge on the droplets causes them to fragment until only intact protonated proteolytic peptides remain, often termed peptide precursor ions. Matrix-assisted laser desorption/ionization (MALDI) is another technique commonly used to volatilize and ionize the proteolytic peptides for mass spectrometric analysis. ESI ionizes the analytes out of a solution and is therefore  
10 readily coupled to liquid-based (for example, chromatographic and electrophoretic) separation tools, MALDI sublimates and ionizes the samples out of a dry, crystalline matrix via laser pulses. MALDI-MS is normally used to analyze relatively simple peptide mixtures, whereas integrated liquid-chromatography ESI-MS systems (LC-MS) are preferred for the analysis of complex samples.

15 In stage 4, a mass spectrum of the peptides eluting at a particular time point is taken (MS1 spectrum, or 'normal mass spectrum'). Mass spectrometric measurements are carried out in the gas phase on ionized analytes (protonated proteolytic peptides, peptide precursor ions). By definition, a mass spectrometer consists of an ion source, a mass analyser that measures the mass-to-charge ratio  
20 ( $m/z$ ) of the ionized analytes, and a detector that registers the number of ions at each  $m/z$  value.

In stage 5, the computer generates a prioritized list of these peptide precursor ions for fragmentation and a series of tandem mass spectrometric or  
25 'MS/MS' analyses ensues. The first stage of tandem MS/MS isolates individual peptide precursor ions, and the second breaks the peptide precursor ions into peptide fragment ions and uses the fragmentation pattern to determine their amino acid sequences. The MS and MS/MS spectra are typically acquired for about one second each and stored for matching against protein sequence databases.

30 The outcome of the analysis is the amino acid sequence of the proteolytic peptide fragments and therefore the peptides and proteins making up the (purified) protein population.

In mass spectrometry, collision-induced dissociation (CID), referred to by some as collisionally activated dissociation (CAD), is currently the method of  
35 choice. CID is a mechanism by which to fragment peptide precursor ions in the gas phase. The peptide precursor ions are usually accelerated by some electrical potential

to high kinetic energy in the vacuum of a mass spectrometer and then allowed to collide with neutral gas molecules (often helium, nitrogen or argon). The collisions allow some of the kinetic energy to be converted into internal energy which results in bond breakage and the fragmentation of the peptide precursor ion into smaller fragments.

5 These fragment ions can then be analyzed by a mass spectrometer.

CID is frequently used as part of tandem mass spectrometry in proteomics analyses. While CID is currently the most popular method for standard tandem mass spectrometry, there are also other fragmentation methods, for example electron transfer dissociation (ETD) and electron capture dissociation (ECD).

10 These different fragmentation techniques lead to the appearance of different types of ion fragments. A nomenclature for various ion types was first suggested by P. Roepstorff and J. Fohlman (Proposal for A Common Nomenclature for Sequence Ions In Mass- Spectra of Peptides. Biomed. Mass Spectrom. 11, 601-601 (1984)) and subsequently modified as described by K. Biemann (Contributions of  
15 Mass-Spectrometry to Peptide and Protein- Structure. Biomed. and Env. Mass Spectrom. 16, 99-111 (1988)). Typically *y*, *b* (cleavage of the peptide bond) and *a* fragments (formally a loss of CO from a *b* ion) are observed in CID. This is schematically depicted in Figure 1.

ECD is usually considered a more direct fragmentation technique as  
20 compared to CID. In contrast to CID, ECD involves the introduction of low energy electrons to trapped gas phase ions. ECD produces significantly different types of fragment ions than CID. The unique (and complementary) fragments observed and the ability to fragment whole macromolecules effectively has been considered the most promising features of ECD. However, the low efficiencies and other technical difficulties  
25 have prevented wide spread use. ECD is primarily used in Fourier transform ion cyclotron resonance mass spectrometry.

ETD does not use free electrons but employs radical anions such as for example anthracene or azobenzene. When these anions react with positively charged peptide precursor ions an electron is transferred leading to the formation of *c*  
30 and *z* peptide fragment ions (figure 1). ETD cleaves peptide bonds randomly along the peptide backbone while side chains and modifications such as phosphorylation are usually left intact. The technique works well for higher charge state ions ( $z > 2$ ).

Analysis of some post-translational modifications (PTMs), such as phosphorylation, sulfonation, and glycosylation, is difficult with CID since the  
35 modification is often labile and preferentially lost over peptide backbone fragmentation,

resulting in little to no peptide sequence information. The presence of multiple basic residues also makes peptides exceptionally difficult to sequence by conventional CID mass spectrometry. In a recent review, the utility of ETD mass spectrometry for sequence analysis of post-translationally modified and/or highly basic peptides was investigated (Molina et al., Proceedings of the National Academy of Sciences of the United States of America 104, (2007) 2199-2204). Phosphorylated, sulfonated, glycosylated, nitrosylated, disulfide bonded, methylated, acetylated, and highly basic peptides were analyzed by CID and ETD mass spectrometry. It was concluded that ETD is an excellent method for localization of phosphorylation sites. This illustrates the utility of ETD as an advantageous tool in phosphoproteomics research.

Protein identifications using peptide CID spectra are more clear-cut than those achieved by mass mapping because, in addition to the peptide mass, the peak pattern in the CID spectrum also provides information about peptide sequence.

This information however, is not readily convertible into a full, unambiguous peptide sequence, therefore, CID is generally considered not very suitable for automated *de novo* sequencing. Instead, the CID spectra are scanned against comprehensive protein sequence databases using one of a number of different algorithms, each with its strengths and weaknesses.

The 'peptide sequence tag' approach extracts a short, unambiguous amino acid sequence from the peak pattern that, when combined with the mass information, is a specific probe to determine the origin of the peptide.

In the 'cross-correlation' method, peptide sequences in the database are used to construct theoretical mass spectra and the overlap or 'cross-correlation' of these predicted spectra with the measured mass spectra determines the best match.

In the third main approach, 'probability based matching', the calculated fragments from peptide sequences in the database are compared with observed peaks. From this comparison a score is calculated which reflects the statistical significance of the match between the spectrum and the sequences contained in a database.

In each of these methods the identified peptides are compiled into a protein 'hit list', which is the output of a typical proteomic analysis. Because protein identifications rely on matches with sequence databases, high-throughput proteomics is currently restricted largely to those species for which comprehensive sequence databases are available.

For those species where no genomic sequence information is

available, amino acid sequencing can only be done "*de novo*". i.e. without database matching. The spectra usually obtained in the above described prior art methods are hardly if at all suited for *de novo* sequencing since they do not provide unambiguous and easy to read sequence information.

5                   The present invention addresses this problem and provides a method for determining the amino acid sequence of a peptide wherein MS spectra are produced with very clear and unambiguous sequence information which is interpretable without the help of comprehensive databases. The method according to the invention provides MS/MS spectra containing predominantly c ions in sequential order. The  
10                   method according to the invention is therefore particularly suited for *de novo* sequencing. Moreover, it provides an improved method for the analysis of post-translational modifications of proteins.

#### Summary of the invention

15                   The invention relates to a method for determining at least part of the amino acid sequence of a protein comprising the steps of cleaving the protein into proteolytic peptides, ionizing the proteolytic peptides to generate peptide precursor ions, dissociating these peptide precursor ions using tandem mass spectrometry in order to obtain peptide fragment ions, followed by determining the amino acid  
20                   sequence of a selected proteolytic peptide, wherein the cleaving step generates at least one proteolytic peptide with an N-terminal lysine residue and wherein the dissociation of the peptide precursor ions is initiated by electron transfer.

                    In other words, the invention relates to the use of electron transfer dissociation in tandem mass spectrometry-based protein sequencing of proteolytic  
25                   peptides with an N-terminal Lysine residue.

                    The method and use according to the invention may advantageously be performed using an N-terminal Lysine-specific endoprotease.

                    It was also found that part of the method according to the invention may be performed on a slice from a gel. The invention therefore also relates to a  
30                   method for digesting a protein or peptide in a gel, comprising the steps of isolating a specific slice of the gel containing the protein or peptide and incubating the specific slice with an N-terminal Lysine-specific endoprotease in order to obtain proteolytic peptides.

35                   Detailed description of the invention

Peptide sequencing via tandem mass spectrometry (MS/MS) is one of the most powerful tools in proteomics for identifying proteins. Because complete genome sequences are accumulating rapidly, the recent trend in interpretation of MS/MS spectra has been through database search. However, *de novo* MS/MS spectral interpretation remains an open problem typically involving manual interpretation by expert mass spectrometrists. Several new algorithms have been developed for *de novo* interpretation that automatically learn fragment ion types and intensity thresholds from a collection of test spectra generated from any type of mass spectrometer (Dancik et al., J. Comp. Biol. 6, 327-342 (1999) and Performance Evaluation of Existing De Novo Sequencing Algorithms; Sergey Pevtsov, Irina Fedulova, Hamid Mirzaei, Charles Buck, and Xiang Zhang J. Proteome Res., 5 (11), 3018 -3028, (2006)).

In a few seconds, a tandem mass spectrometer is capable of ionizing a mixture of proteolytic peptides with different sequences and measuring their respective parent mass/charge ratios, selectively fragmenting each peptide into pieces and measuring the mass/charge ratios of the fragment ions. The peptide sequencing problem is then to derive the sequence of the peptides given their MS/MS spectra.

For an ideal fragmentation process and an ideal mass spectrometer the sequence of a peptide could be simply determined by converting the mass differences of consecutive ions in a spectrum to the corresponding amino acids. This ideal situation would occur if the fragmentation process could be controlled so that each peptide was cleaved between every two consecutive amino acids and a single charge was retained on only the N-terminal piece. In practice, the fragmentation processes in mass spectrometers are far from ideal. As a result, *de novo* peptide sequencing remains an open problem and even a simple spectrum may require tens of minutes for a trained expert to interpret.

In contrast to database searching, *de novo* sequencing allows to derive the peptide sequence directly from the mass spectrum. *De novo* sequencing becomes a necessity, if a peptide is not present in any database. The main application of *de novo* sequencing is for the study of non-sequenced organisms, detecting mutated peptides and post-translational modifications.

The main problem with standard (low mass accuracy, one fragmentation technique) *de novo* sequencing is the inability to distinguish the direction of the fragments (N- or C-terminal), and the low mass accuracy which leads to an excessive number of potential solutions.

Many different *de novo* sequencing approaches have been

suggested over the years. The simplest and most exhaustive approach is permutation of all conceivable amino acid combinations possible for a given peptide mass, and generation of theoretical mass spectra for each combination, "global approach". This is followed by comparison of all candidates with the real mass spectrum and selection of the best match.

Often, especially for low mass accuracy data, several theoretical spectra corresponding to different sequences will give an equally good match with the experimental spectrum. In many cases this is due to the fact that the b and y ion and/or c and z ion of a given sequence coincide in mass within the given mass accuracy.

Several groups have explored alternative proteases for MS/MS analysis of proteolytic peptides (Molina et al., Proceedings of the National Academy of Sciences of the United States of America 104 (7), 2199 (2007) and Garcia et al., Journal of Proteome Research 3 (6), 1219 (2004)). They explored the utility of Lys-C, Trypsin and Glu-C using CID and ETD and concluded that, for ETD applications, Lys-C offered a useful alternative to trypsin for the analysis of post-translational modifications in phosphoproteins.

Rao et al. demonstrated that a peptidyl-Lys metalloendoprotease could be used for the  $^{18}\text{O}$  *in vitro* labeling of proteolytic peptides and argued its usefulness in comparative proteomic studies (J. Proteome Research 4, 507-514 (2005) and Molec. Cell. Proteomics 4, 1550-1557 (2005)). They showed two CID MS/MS spectra of peptides with an N-terminal lysine residue. These spectra consisted of a typical CID pattern of b and y ions with a complexity that was not particularly suitable for de novo sequencing.

The term "Lys-N" as well as "endoproteases with N-terminal lysine cleavage specificity" is used herein to indicate a protein from the EC class 3.4.24.20 that cleaves peptidyl-lysine bonds (-X-Lys- in proteins and peptides).

Surprisingly, we now found that endoproteases with N-terminal Lysine (Lys-N) cleavage specificity were very suitable for de novo sequencing when used in combination with electron transfer dissociation (ETD) because this combination yielded straightforward and easy-to-interpret MS/MS spectra.

The MS/MS spectra generated with a method according to the invention consisted predominantly, if not exclusively of c-ions and presented those c-ions in consecutive order, thereby greatly facilitating de novo sequencing of the proteolytic peptides. Statistical analysis of the properties for the spectra obtained with LysN in combination with ETD indicate, that on average more than 90% of identified

peaks are c-ions. The invention therefore also relates to a method as described above wherein the tandem mass spectrometry results in a spectrum wherein on average more than 90% of identified peaks are c-ions, preferably more than 95% or even more preferably 100%.

5 In comparison to the use of Lys-C and trypsin, comparable specificity, sensitivity and selectivity were found for Lys-N proteases. Surprisingly, cleavage of peptides at the N terminus of a Lysine residue provided a clear advantage over other proteases due to the fact that the peptide ETD MS/MS spectra obtained with Lys-N consisted almost exclusively of c-type fragment ions.

10 Such clear spectra, providing simple sequence ladders, may prove to be a valuable strategy for the *de novo* sequence analysis of proteolytic peptides, such as for those peptides carrying post-translational modifications. The product ion statistics, mentioned above, allow development of computational efficient search algorithms for identifying peptides from the spectra obtained with LysN in combination  
15 with ETD.

Hence, the invention relates to the use of electron transfer dissociation in tandem mass spectrometry-based protein sequencing of proteolytic peptides with an N-terminal lysine residue.

In more detail, the invention may be described as a method for  
20 determining at least part of the amino acid sequence of a protein comprising the steps of cleaving the protein into proteolytic peptides, ionizing the proteolytic peptides to generate peptide precursor ions, dissociating these peptide precursor ions using tandem mass spectrometry in order to obtain peptide fragment ions, followed by determining the amino acid sequence of a selected proteolytic peptide, wherein the  
25 cleaving step generates at least one proteolytic peptide with an N-terminal lysine residue and wherein the dissociation of the peptide precursor ions is initiated by electron transfer.

Advantageously, the cleavage step is performed by an enzyme.

In the Examples section, two different metalloendoproteases are  
30 exemplified that contain one atom of zinc per molecule and which both generate N-terminal lysine fragments. The invention therefore relates to a method as described above wherein the enzyme is a metalloendopeptidase.

Enzymes may favorably be specific for an acyl-lysine bond; hence the method according to the invention may be described as a method wherein the  
35 enzyme is specific for an acyl-lysine bond. Such enzymes are categorized as EC

3.4.24.20, official name Peptidyl-Lys metalloendopeptidase.

Such enzymes may also favorably be obtained from a fungus (Nonaka et al., J. Biol. Chem. 272 (1997) 30032-30039). The invention therefore also concerns a method as described above wherein the enzyme is obtainable from a  
5 fungus. Enzymes having the amino acid sequence according to SEQ ID NO: 1 or SEQ ID NO: 2 (figure 6) obtainable from the fungi *Grifola frondosa* and *Pleurotus ostreatus* respectively, are preferred since they performed well in the method according to the invention.

Now that it is disclosed herein that the enzymes as described above  
10 may favorably be used for ETD MS/MS-based peptide sequencing, it is evident that alternative methods may be found by searching databases and screening other organisms for enzymes that have Lys-N specificity.

Such enzymes may be found by searching for nucleic acids with substantial homology with the sequences as disclosed herein (SEQ ID NO: 1 and SEQ  
15 ID NO: 2).

It is to be expected that enzymes with at least 85% sequence homology to SEQ ID NO: 1 or SEQ ID NO: 2 will perform at least equally well as the enzymes disclosed herein. Preferably, the homology is more than 90%, such as 92, 95, 96, 97, 98 or even more than 99%.

A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and  
20 more preferably at least about 95-98% of the nucleotide bases.

Alternatively, substantial homology or (similarity) exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is  
30 substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain  
35 embodiments will often be over a stretch of at least about nine nucleotides, usually at

least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

5 Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30 degrees Celsius. typically in excess of 37 degrees Celsius., and preferably in  
10 excess of 45 degrees Celsius. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur & Davidson, 1968. Suitable protocols for performing hybridisation experiments can be found in Sambrook et al. 2000. Molecular Cloning: A  
15 Laboratory Manual (Third Edition) Cold Spring Harbor Laboratory Press.

Methods for determining sequence homology between different strands of nucleic acids or proteins are known in the art. Homology is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin  
20 Biotechnology Center, 910 University Avenue, Madison, Wis. 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions in amino acid sequences typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine;  
25 serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

The invention therefore relates to a method as described above wherein the enzyme has an amino acid sequence derived from a DNA sequence which is substantially homologous to a DNA sequence encoding the polypeptides as disclosed herein (SEQ ID NO: 1 and SEQ ID NO: 2).

30 The invention also relates to a method as described above wherein the enzyme has an amino acid sequence having at least 85% homology to the amino acid sequence of a *Grifola frondosa* metalloendopeptidase according to SEQ ID NO: 1 or of a *Pleurotus ostreatus* metalloendopeptidase according to SEQ ID NO: 2.

35 Improved enzymes may be obtained by mutating the DNA sequences encoding the amino acid sequences of SEQ ID NO: 1 or SEQ ID NO: 2 and selecting

for enzymes with improved stability, specificity, or other desirable characteristics.

In a comparative experiment (example 4), Lys-C generated proteolytic peptides of BSA were subjected to ETD. In Figure 2, the resulting ETD induced fragmentation spectrum is given (spectrum A), which revealed the formation of multiple c and z peptide fragment ions with the z ions marginally being more dominant.

In parallel, the performance of a Lys-N endopeptidase was evaluated under the same experimental conditions, i.e. on BSA with analysis by ETD. When the Lys-N proteolytic peptides were subjected to ETD, the only peptide fragment ions observed were N-terminal c ions, as depicted in Figure 2, spectrum B. Most notably, these c-ions were displayed in consecutive order.

To test the general applicability of Lys-N and its feasibility for biological samples, a whole cell lysate of HEK 293 cells was digested in parallel with Lys- N and Lys-C. The generated complex peptide mixtures were analyzed by ETD. In Figure 3, spectrum A2 shows a typical peptide fragmentation obtained for a Lys-C generated proteolytic peptide, whereas spectrum B2 originated from a Lys-N peptide. These findings are quite similar to the above findings and resemble those observed for BSA (figure 2), namely a dominance of c ions in consecutive order in the spectrum obtained with Lys-N.

Yet another peptide obtained from a whole cell lysate of HEK 293 cells was digested in parallel with Lys- N and Lys-C and analyzed by ETD. In Figure 4, spectrum A3 shows a typical peptide fragmentation obtained for a Lys-C generated peptide, whereas spectrum B3 originated from a Lys-N peptide. Again, this confirmed the above findings and showed a spectrum with predominantly c-ions (one z-ion was identified) with the c-ions in consecutive order.

Summarizing, using Lys-N in combination with ETD we observed cleavage at almost every amide bond in the peptide backbone, providing easy to interpret sequence ladders.

Without wanting to be bound by theory, the following is provided as an ex post facto analysis of a possible mechanism underlying the observed advantages of Lys-N. Lys-C generated peptides have a basic entity at both termini. Therefore, the fragmentation of Lys-C generated peptides may result in the formation of both c and z type ions under ETD conditions. In contrast, in Lys-N generated peptides the protons or charges will be preferentially attached to the N-terminus due to the presence of two free amine groups, which may result in the exclusive generation of c-type fragment ions.

The above experiments showed that the Lys-C enzyme yielded peptide fragment that dissociated under ETD conditions into a complex mixture containing about equal amounts of both c- and z- ions. As a consequence, the MS/MS spectra were more difficult to interpret. In contrast, the spectra obtained with Lys-N  
5 generated peptides were solely dominated if not exclusively formed by c-ion sequence ladders and were therefore easier to interpret. This is considered a huge advantage for *de novo* sequencing of proteins and peptides.

We also compared the performance of Lys-N generated proteolytic peptides in ETD and CID. As expected, the Lys-N proteolytic peptides resulted in a  
10 spectrum consisting of a mixture of b and y ions when CID was used for the ionization of the proteolytic peptides. However, when ETD was used, this resulted in nice clean sequence ladders of c-ions without any z ions present. This is illustrated in the representative example shown in Figure 5.

In addition, protein analyses may also be performed on analytes that  
15 originate from one and two-dimensional sodium-dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE) separations as well as on iso-electrofocussing gel separations. We again tested the applicability of the Lys-N protease to proteins (i.e. BSA) after being migrated through a gel. As Lys-C is known to be less well adaptable to protein spots from gels, we compared the performance of Lys-N with trypsin.

20 Surprisingly, Lys-N proved to be at least as efficient as trypsin for in-gel proteolysis, as evidenced in table 1. Consequently, the invention relates to a method for digesting a protein or peptide in a gel, comprising the steps of isolating a specific slice of the gel containing the protein or peptide and incubating the specific slice with an N-terminal Lysine-specific endoprotease in order to obtain proteolytic  
25 peptide fragments. Preferred are such methods wherein the enzyme is a metalloendopeptidase, preferably with specificity for an acyl-lysine bond.

In such a method, gels may preferably be selected from the group consisting of one-dimensional gels, two-dimensional gels, polyacrylamide gels, iso-electric focusing gels

Table 1

Enzyme	Suitable for in-gel digestion	Obtained sequence coverage	Number of identified peptide fragments	Number of peptides dominated by c-ions	Required protease / protein ratio
Trypsin	YES				
Lys-C	NO	59%	28	0% (0/19)	1:50
Lys-N	YES	56%	27	100% (14/14)	1:85

Table 1: One-to-one comparison of Lys-C and Lys-N.

5 This comparison was made using BSA as model protein. The amount of proteases needed, the obtained sequence coverage, and the number of identified peptide fragments are similar. However, as described herein, the ETD MS/MS spectra were dominated by c-ions for the Lys-N peptides. Moreover, Lys-N is suitable for in-gel digestion.

10 Reagents for biochemical methods are often commercially obtainable in the form of a kit of parts. Such kits then contain the reagents in combination with a product insert that describes the envisaged use of the parts. The method according to the invention may therefore be provided in the form of a kit for degrading a protein into proteolytic peptides, containing at least an N-terminal Lysine-specific endoprotease and a manual for digesting a protein or peptide in order to make it suitable for tandem  
15 mass spectrometry-based protein sequencing in combination with electron transfer dissociation.

In conclusion, in this study we evaluated a Lys-N protease for ETD MS/MS proteomics analyses. In general, as far as sensitivity, selectivity and specificity are concerned; the Lys-N protease performs at least equally well as currently used  
20 proteases. Remarkably, we found that Lys-N provides a clear advantage over other proteases due to the fact that the peptide ETD MS/MS spectra almost exclusively consisted of c-type fragment ions, resulting in simple-to-read sequence ladders for the peptides of interest. It was found that Lys-N, in combination with ETD, therefore provides a valuable strategy for the analysis of post-translational modifications and *de*  
25 *nov*o sequencing. Moreover, Lys-N was found to be suitable for in-gel digestion.

The following literature describes useful background information and is herewith incorporated by reference.

References

1. H. Kang, L. Pasa-Tolic, and R. D. Smith, *Journal of the American Society for Mass Spectrometry* 18 (7), 1332 (2007);
- 5 2. P. Mallick, M. Schirle, S. S. Chen et al., *Nature Biotechnology* 25 (1), 125 (2007).
3. A. I. Nesvizhskii, O. Vitek, and R. Aebersold, *Nature methods* 4 (10), 787 (2007);
4. Shevchenko, M. Wilm, O. Vorm et al., *Analytical Chemistry* 68 (5), 850 (1996).
- 10 5. K. G. Standing, *Current opinion in structural biology* 13 (5), 595 (2003).
6. T. Keough, R. S. Youngquist, and M. P. Lacey, *Proceedings of the National Academy of Sciences of the United States of America* 96 (13), 7131 (1999).
7. E. C. Peters, D. M. Horn, D. C. Tully et al., *Rapid Communications in Mass Spectrometry* 15 (24), 2387 (2001).
- 15 8. J. E. P. Syka, J. J. Coon, M. J. Schroeder et al., *Proceedings of the National Academy of Sciences of the United States of America* 101 (26), 9528 (2004);
9. Y. Xia, H. P. Gunawardena, D. E. Erickson et al., *Journal of the American Chemical Society* (2007);
10. R. A. Zubarev, N. L. Kelleher, and F. W. McLafferty, *Journal of the American Chemical Society* 120 (13), 3265 (1998);
- 20 11. A. Chi, D. L. Bai, L. Y. Geer et al., *International Journal of Mass Spectrometry* 259 (1-3), 197 (2007);
12. L. M. Mikesh, B. Ueberheide, A. Chi et al., *Biochimica et biophysica acta* 1764 (12), 1811 (2006).
- 25 13. H. Molina, D. M. Horn, N. Tang et al., *Proceedings of the National Academy of Sciences of the United States of America* 104 (7), 2199 (2007).
14. B. A. Garcia, S. A. Busby, C. M. Barber et al., *Journal of Proteome Research* 3 (6), 1219 (2004).
15. E. S. Witze, W. M. Old, K. A. Resing et al., *Nature methods* 4 (10), 798 (2007).
- 30 16. V. H. Wysocki, G. Tsaprailis, L. L. Smith et al., *Journal of Mass Spectrometry* 35 (12), 1399 (2000).
17. D. L. Swaney, G. C. McAlister, M. Wirtala et al., *Analytical Chemistry* 79 (2), 477 (2007);
- 35 18. N. Leymarie, C. E. Costello, and P. B. O'Connor, *Journal of the American Chemical Society* 125 (29), 8949 (2003).

19. H. L. Han, Y. Xia, and S. A. McLuckey, *Journal of Proteome Research* 6 (8), 3062 (2007).

20. R. E. March, *Journal of Mass Spectrometry* 32 (4), 351 (1997).

Legend to the figures.

5 Figure 1: Nomenclature for various ion types

Figure 2: Schematic representation of the ETD induced fragmentation in Lys-C and Lys-N generated doubly charged peptide ions and ETD MS/MS spectra of doubly charged ions originating from A1) Lys-C and B) Lys-N generated peptides from BSA (A) QTALVELLK and (B1) KQTALVELL. The spectrum in the B panel shows a  
10 nearly complete sequence ladder of c-ions.

Figure 3: Schematic representation of the ETD induced fragmentation in Lys-C and Lys-N generated doubly charged peptide ions and ETD MS/MS spectra of doubly charged ions originating from A2) Lys-C and B2) Lys-N generated peptides from HEK293 cells. (A2) FGYVDFESAEDLEK and (B2) KGYVDFESAEDLE. The spectrum  
15 in the B2 panel shows a nearly complete sequence ladder of c-ions.

Figure 4: Schematic representation of the ETD induced fragmentation in Lys-C and Lys-N generated doubly charged peptide ions and ETD MS/MS spectra of doubly charged ions originating from A3) Lys-C and B3) Lys-N generated peptides from HEK293 cells. (A3) GADFLVTEVENGGSLSK and (B3) KGADFLVTEVENGGSLSG.  
20 The spectrum in the B3 panel shows a nearly complete sequence ladder of c-ions.

Figure 5: Typical CID and ETD MS/MS spectra of a Lys-N generated doubly charged peptide ion. (A) CID mass spectrum of Lys-N generated KNQVALNPQNTVFDA. (B) ETD mass spectrum of Lys-N generated KNQVALNPQNTVFDA. Typically, the ETD spectra are much easier to interpret and  
25 provide straightforward sequence ladders.

Figure 6: Sequence listings of amino acid sequences of Lys-N metalloendoproteases from *Grifola fondosa* and *Pleurotus ostreatus*.

Examples

30

Example 1: Materials and methods.

Protease inhibitor cocktail and Lys-C were obtained from Roche Diagnostics (Mannheim, Germany). Metalloendopeptidase from *Grifola Frondosa* (Lys-N) was obtained from Seikagaku Corporation (Tokyo, Japan). Alternatively, a  
35 metalloendopeptidase from *Pleurotus ostreatus* may be used (Nonaka et al., *J. Biol.*

Chem. 272 30032-30039 (1997)). Bovine Serum Albumin (BSA) and Iodoacetamide were obtained from Sigma-Aldrich (Steinheim, Germany). DL-Dithiothreitol was obtained from Fluka biochemical (Steinheim, Germany). HEK293 cells were a gift from the ABC Protein Expression Center (Utrecht University, The Netherlands). HPLC-S  
5 gradient grade acetonitrile was purchased from Biosolve (Valkenwaard, The Netherlands). Acetic acid was obtained by MERCK KGaA (Damstadt, Germany) and high purity water obtained from Milli-Q system (Millipore, Bedford, MA).

Human Embryonic kidney (HEK) 293T cells (HEK293 cells) were harvested at a density of approx  $1.5 \times 10^6$  cells/mL and stored at  $-30^\circ\text{C}$ . Cells were  
10 thawed and resuspended in ice-cold lysisbuffer (15 mL PBS, 150  $\mu\text{L}$  Tween 20 and protease inhibitor cocktail). After Dounce homogenizing on ice, the lysate was stored at  $0^\circ\text{C}$  for 10 min. Subsequently centrifugation at  $20000 \times g$  in a tabletop centrifuge (Eppendorf, Hamburg, Germany) at  $4^\circ\text{C}$  yielded separation of soluble and insoluble protein fractions. The soluble fraction was collected and the concentration determined  
15 by a Bradford assay. The lysate was dissolved in 50 mM ammonium bicarbonate to a concentration of 4 mg/mL.

#### Example 2: HEK293 Lysate and BSA in-solution digestion.

Hundred mg digested protein/lysate was reduced with 45 mM  
20 dithiothreitol ( $50^\circ\text{C}$ , 15 min) followed by alkylation using 110 mM Iodoacetamide (dark, RT, 15min). Buffer exchange was performed with 50 mM ammonium bicarbonate using 5 kD spin columns. The resulting solutions were dried in a vacuum centrifuge and resuspended in 50 mM ammonium bicarbonate. The purified digests were aliquoted. One part was digested with Lys-C and an equal amount with Lys-N. Lys-C was added  
25 to the samples at a 1:50 (w/w) ratio and incubated at  $25^\circ\text{C}$  over night and Lys-N was added at a ratio of 1:85 (w/w) and also incubated over night at  $25^\circ\text{C}$ .

#### Example 3: BSA in gel-digestion.

Gel bands containing BSA were cut out of the gel and washed with  
30 water. After shrinking the gel pieces with acetonitrile they were reduced with 10 mM of DTT ( $60^\circ\text{C}$ , 1 hour) followed by alkylation using 55 mM Iodoacetamide (dark, RT, 30min). After shrinking the gel pieces with acetonitrile they were incubated with trypsin or Lys-N (10 ng /  $\mu\text{L}$ ) and left on ice for 30 min. Excess trypsin or Lys-N was then removed and enough AMBIC was added to cover the gel pieces. The gel pieces were  
35 incubated over night at  $37^\circ\text{C}$ . Supernatant was transferred to new eppendorf tubes.

Peptide were extracted by adding 5 % FA to the gel pieces following heating (65 °C, 2 min) and shaking (RT, 20 min). The supernatant was added to the previous supernatant.

5 Example 4: CID and ETD analysis.

Digested BSA and HEK293 lysate were subjected to nanoscale liquid chromatography tandem mass spectrometry (nano-LC-MS/MS) analysis, performed on an Agilent 1100 HPLC system (Agilent Technologies) connected to a LTQ XL Linear Ion Trap Mass Spectrometer with an ETD source at the back from Thermo Fisher Scientific, Inc. (Waltham, MA). The samples (0.5 µg digested HEK293 lysate and 50 fmol of BSA) were diluted in 5 % formic acid and injected on the trap column (Aqua C18 (phenomenex, Torrance, CA)), 20 mm x 100 µm I.D.) at a flow rate of 5 µL / min. The peptides were transferred with a split-reduced flow rate of 100 nL / min solvent A (0.1 M acetic acid) onto an analytic column (Reprosil C18 RP (Dr Maisch, Germany), 15 20 cm x 50 µm I.D.). Elution of peptides from digested lysate was achieved with a linear gradient from 0 to 35% B (acetonitrile/water (v/v) containing 0.1 M acetic acid) in 95 min and digested BSA was eluted with a linear gradient from 0 to 40% B in 45min. The column effluent was directly introduced into the ESI source of the MS.

The mass spectrometer was operated in positive ion mode, from 350 to 1500 m/z in MS mode and with an AGC value of 1.00e+05. Parent ions were isolated for a more accurate mass measurement by performing a SIM scan and fragmented by CID or ETD in data-dependent mode with an AGC value of 1.00e+04.

Ions were fragmented using CID with normalized collision energy of 35 and 30 ms activation time. ETD fragmentation was performed with supplemental activation, fluoranthene was used as reagent anion and ion/ion reaction in the ion trap was taking place for 100 ms.

Example 5: data analysis

After MS measurements data was analyzed with the MASCOT software version 2.2.0 (www.matrixscience.com). The database search was made with parameters set to consider a peptide tolerance of ± 0.5 Da, a fragment tolerance of ± 0.9 Da, a static modification of + 57 Da on cystein residues (Carbamidomethylation) and a differential modification of + 16 Da on methionine (oxidation). CID and ETD spectra of digested lysate were searched in NCBI nr 20070713 (5269953 sequences; 35 1825351362 residues) database and BSA in a BSA database.

## CLAIMS

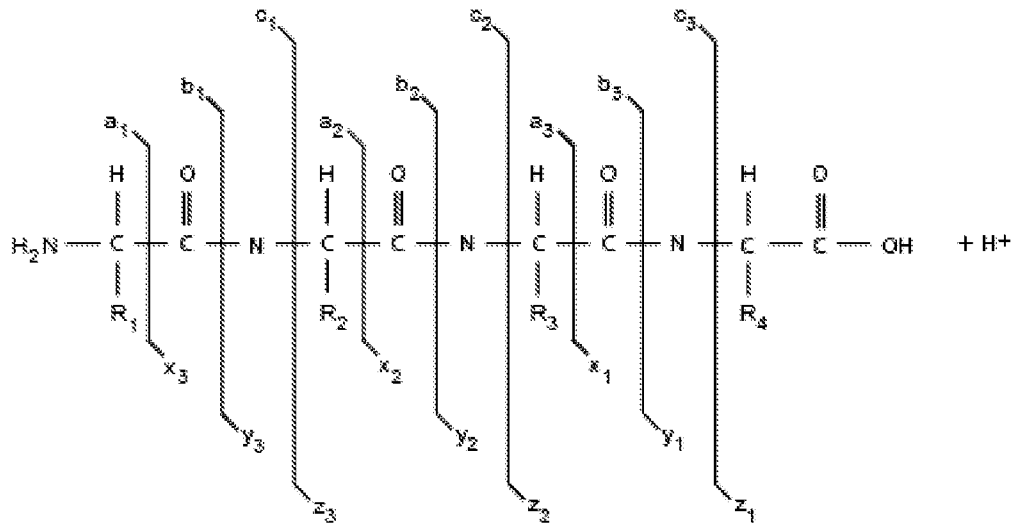
1. Method for determining at least part of the amino acid sequence of a protein comprising the steps of
  - 5 - cleaving the protein into proteolytic peptides,
  - ionizing the proteolytic peptides to generate peptide precursor ions,
  - dissociating these peptide precursor ions using tandem mass spectrometry in order to obtain peptide fragment ions, followed by
  - determining the amino acid sequence of a selected proteolytic peptide,
- 10 wherein said cleaving results in at least one proteolytic peptide with an N-terminal lysine residue and wherein said dissociating of the peptide precursor ions is initiated by electron transfer.
2. Method according to claim 1 wherein said cleaving is performed using an enzyme.
3. Method according to claim 2 wherein said enzyme is a metalloendopeptidase.
- 15 4. Method according to claims 2 or 3 wherein said enzyme is specific for an acyl-lysine bond.
5. Method according to claims 2 – 4 wherein said enzyme is obtainable from a fungus.
- 20 6. Method according to claims 2 – 5 wherein said enzyme has an amino acid sequence having more than 85% homology to the amino acid sequence of a *Grifola frondosa* metalloendopeptidase or of a *Pleurotus ostreatus* metalloendopeptidase according to SEQ ID NO: 1 or SEQ ID NO: 2.
7. Method according to claims 1 – 6 wherein said step of determining the amino acid sequence of a selected proteolytic peptide is done *de novo*.
- 25 8. Use of electron transfer dissociation in tandem mass spectrometry-based protein sequencing of proteolytic peptides with an N-terminal lysine residue.
9. Use according to claim 8 wherein an enzyme specific for an acyl-lysine bond is used to generate said peptides with an N-terminal lysine residue.
10. Use according to claim 9 wherein said enzyme is obtainable from a fungus.

11. Use according to claims 9 – 10 wherein the enzyme has an amino acid sequence having more than 85% homology with the amino acid sequence of a *Grifola frondosa* metalloendopeptidase or of a *Pleurotus ostreatus* metalloendopeptidase according to SEQ ID NO: 1 or SEQ ID NO: 2.
- 5 12. Method for digesting a protein or peptide in a gel, comprising the steps of isolating a specific slice of the gel containing the protein or peptide and incubating the specific slice with an enzyme in order to obtain peptides with an N-terminal lysine residue.
- 10 13. Method according to claim 12 wherein said enzyme is an N-terminal lysine-specific endoprotease
14. Method according to claims 12 or 13 wherein said enzyme is a metalloendopeptidase.
15. Method according to claims 12 – 14 wherein said enzyme is specific for an acyl-lysine bond.
- 15 16. Method according to claims 12 – 15 wherein said enzyme is obtainable from a fungus.
- 20 17. Method according to claims 12 – 16 wherein said enzyme has an amino acid sequence having more than 85% homology to the amino acid sequence of a *Grifola frondosa* metalloendopeptidase or of a *Pleurotus ostreatus* metalloendopeptidase according to SEQ ID NO: 1 or SEQ ID NO: 2.
18. Method according to claims 12 – 17 wherein said gel is selected from the group consisting of one-dimensional gels, two-dimensional gels, polyacrylamide gels and iso-electric focusing gels.

039 WO

Figures 1 / 6

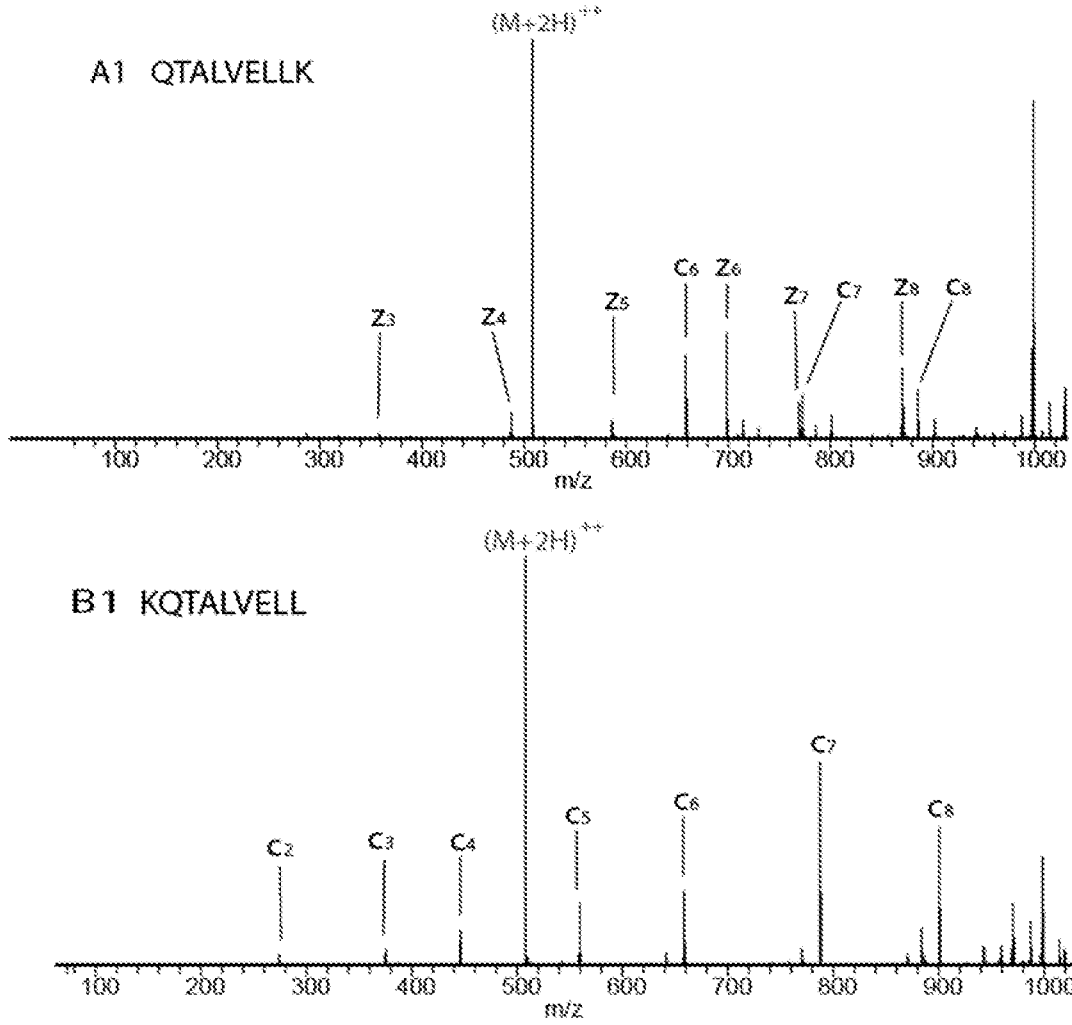
Figure 1



039 WO

Figures 2 / 6

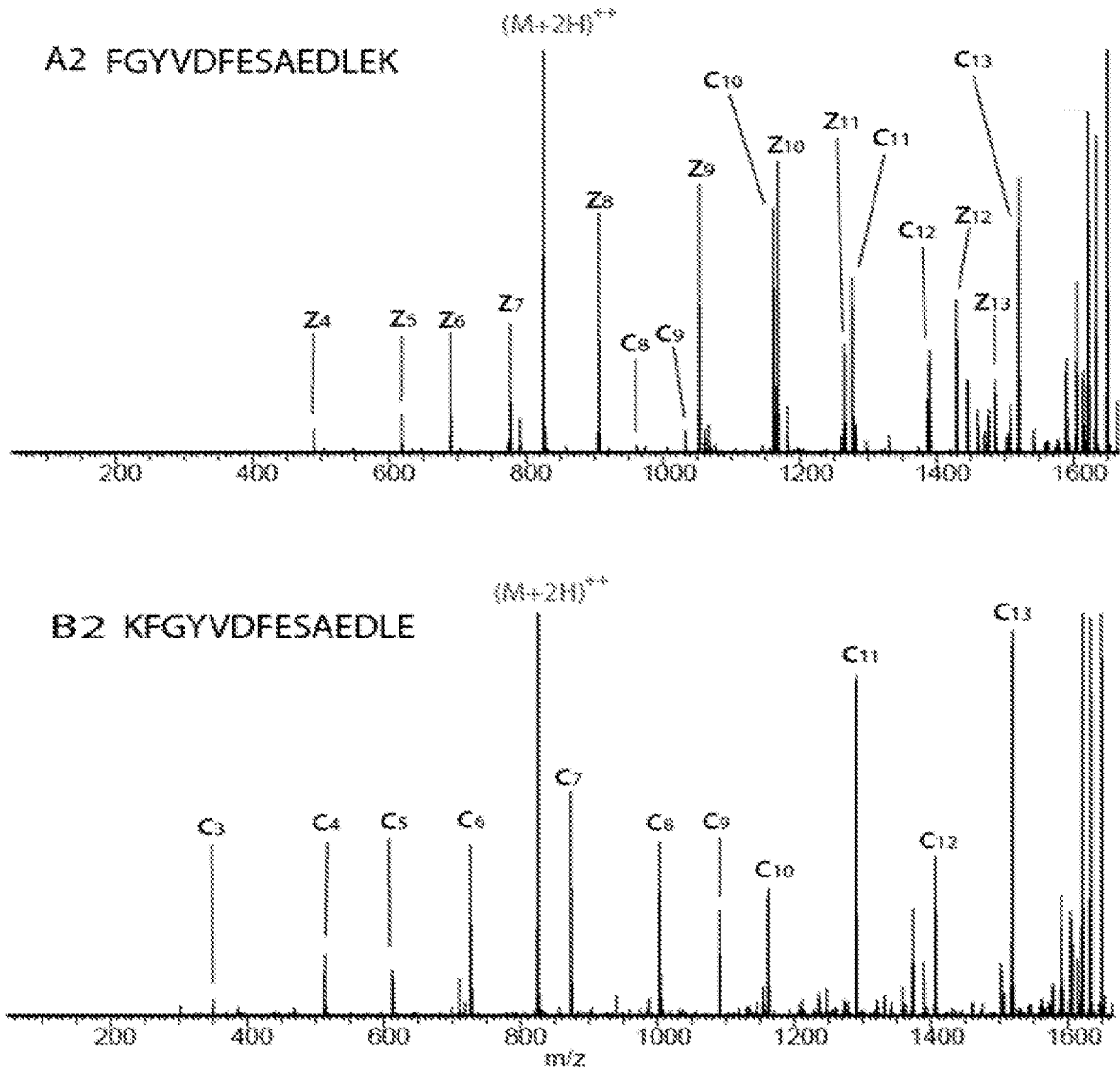
Figure 2



039 WO

Figures 3 / 6

Figure 3



039 WO

Figures 4 / 6

Figure 4

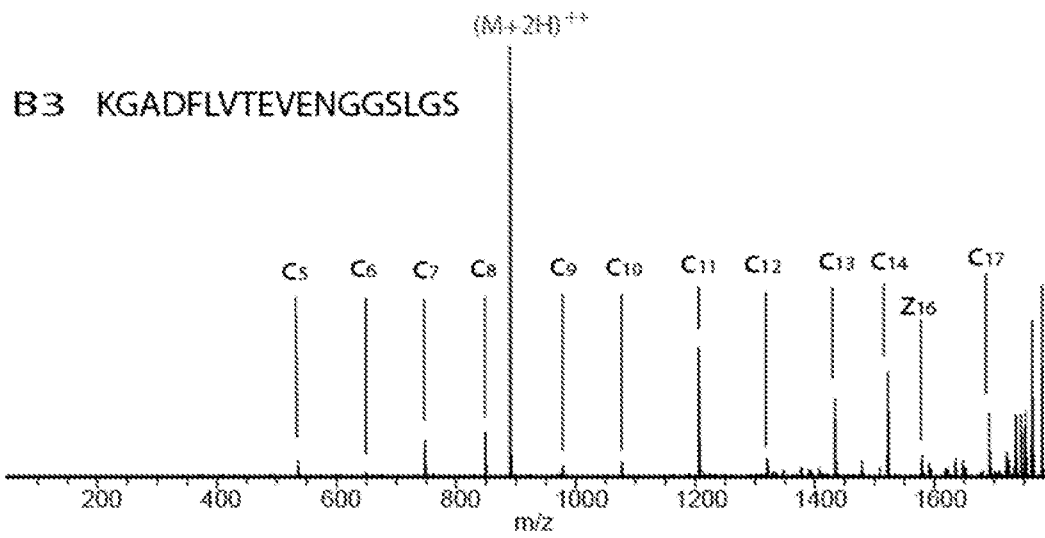
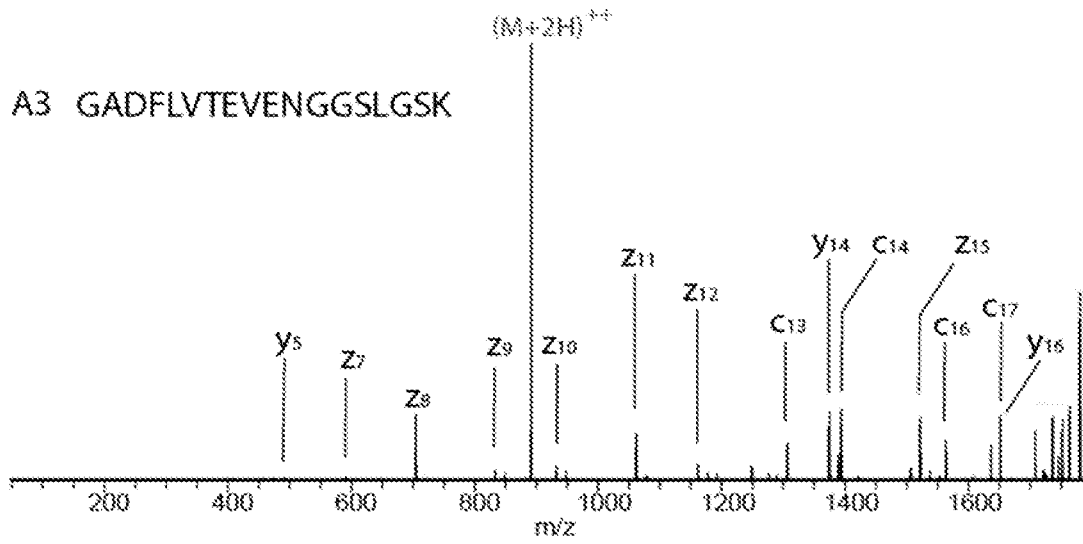
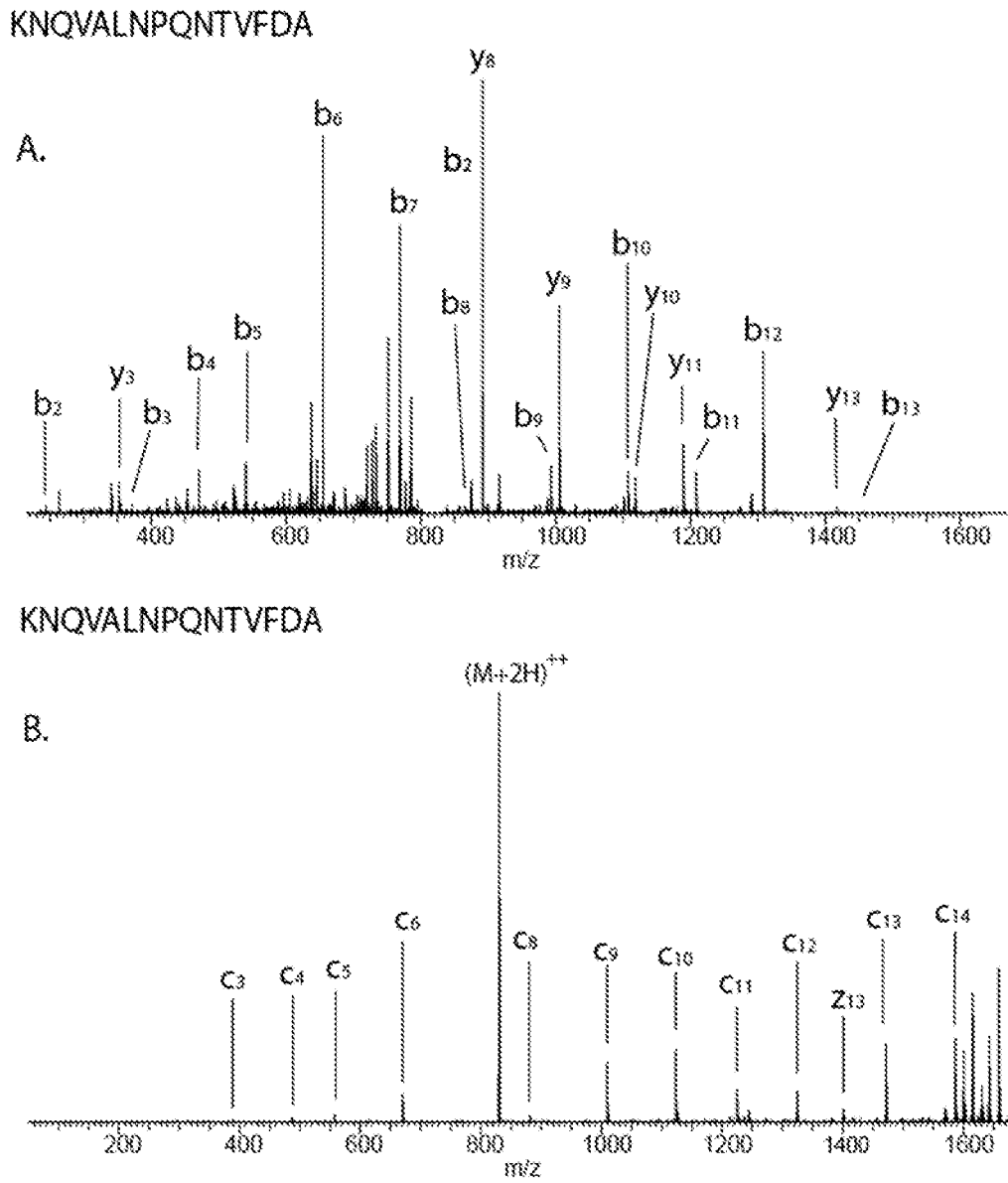


Figure 5



039 WO

Figures 6 / 6

Figure 6

SEQ ID NO: 1:

TYNGCSSSEQSALAAAAASAAQSYVAESLSYLQHTAATPRYTTWFGSYISSRHSTVLQH  
YDMNSNDFSSYSFDCTCTAAGTFAYVYPNRFGTVYLCGAFWKAPTTGTDSQAGTLVHES  
SHFTRNGGTDYAYGQAAAKSLATMDPKAVMNADNHEYFSENNPAQS

SEQ ID NO: 2

ATFVGCSATRQTQLNAAASQAQTYAANALSYLNSHTSSTTRYTTWFGTFVTSRYNTVLSH  
FSSISSNTFSSYTFDCTCSDSGTYAFVNPSNFGYVTLCGAFWNAPVAGTDSRGGTLIHES  
SHFTRNGGTDHVVYQAGAQLARSNPAQAIDNADSHEYFAENNPALA

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2009/050365

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. G01N33/50      G01N33/68				
According to International Patent Classification (IPC) or to both national classification and IPC				
<b>B. FIELDS SEARCHED</b>				
Minimum documentation searched (classification system followed by classification symbols) G01N				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched				
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data				
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
Y	DORMEYER WILMA ET AL: "Targeted analysis of protein termini." JOURNAL OF PROTEOME RESEARCH DEC 2007, vol. 6, no. 12, 10 October 2007 (2007-10-10), pages 4634-4645, XP002474094 ISSN: 1535-3893 page 4635 - page 4643  -----  -/--	1-18		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.				
* Special categories of cited documents :				
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none; vertical-align: top;"> <ul style="list-style-type: none"> <li>*A* document defining the general state of the art which is not considered to be of particular relevance</li> <li>*E* earlier document but published on or after the international filing date</li> <li>*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</li> <li>*O* document referring to an oral disclosure, use, exhibition or other means</li> <li>*P* document published prior to the international filing date but later than the priority date claimed</li> </ul> </td> <td style="width: 50%; border: none; vertical-align: top;"> <ul style="list-style-type: none"> <li>*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</li> <li>*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</li> <li>*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</li> <li>*Z* document member of the same patent family</li> </ul> </td> </tr> </table>			<ul style="list-style-type: none"> <li>*A* document defining the general state of the art which is not considered to be of particular relevance</li> <li>*E* earlier document but published on or after the international filing date</li> <li>*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</li> <li>*O* document referring to an oral disclosure, use, exhibition or other means</li> <li>*P* document published prior to the international filing date but later than the priority date claimed</li> </ul>	<ul style="list-style-type: none"> <li>*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</li> <li>*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</li> <li>*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</li> <li>*Z* document member of the same patent family</li> </ul>
<ul style="list-style-type: none"> <li>*A* document defining the general state of the art which is not considered to be of particular relevance</li> <li>*E* earlier document but published on or after the international filing date</li> <li>*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</li> <li>*O* document referring to an oral disclosure, use, exhibition or other means</li> <li>*P* document published prior to the international filing date but later than the priority date claimed</li> </ul>	<ul style="list-style-type: none"> <li>*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</li> <li>*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</li> <li>*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</li> <li>*Z* document member of the same patent family</li> </ul>			
Date of the actual completion of the international search		Date of mailing of the international search report		
7 May 2009		13/05/2009		
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer  Gonçalves Mauger, M		

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2009/050365

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>PINKSE MARTIJN W H ET AL: "Highly robust, automated, and sensitive online TiO<sub>2</sub>-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster."            JOURNAL OF PROTEOME RESEARCH FEB 2008, vol. 7, no. 2, February 2008 (2008-02), pages 687-697, XP002474095            ISSN: 1535-3893            published online 23.11.2007            page 688 - page 690</p>	1-18
A	<p>VAN BREUKELEN BAS ET AL: "Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra."            RAPID COMMUNICATIONS IN MASS SPECTROMETRY : RCM 2006, vol. 20, no. 16, 2006, pages 2490-2496, XP002474096            ISSN: 0951-4198            abstract</p>	1-18
A	<p>EP 1 617 223 A (AGILENT TECHNOLOGIES INC [US]) 18 January 2006 (2006-01-18)            claims 13-18; example 1</p>	1-18
Y	<p>US 2004/121477 A1 (THOMPSON DEAN R [US] ET AL) 24 June 2004 (2004-06-24)            page 4 - page 6; claims; example 1</p>	1-18
Y	<p>WO 02/08767 A (PROCTER &amp; GAMBLE [US]) 31 January 2002 (2002-01-31)            page 8 - page 30</p>	1-18
Y	<p>SYKA J E P ET AL: "PEPTIDE AND PROTEIN SEQUENCE ANALYSIS BY ELECTRON TRANSFER DISSOCIATION MASS SPECTROMETRY"            PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, DC, US,            vol. 101, no. 26,            29 June 2004 (2004-06-29), pages 9528-9533, XP009051646            ISSN: 0027-8424            page 9529 - page 9532</p>	1-18

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2009/050365

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1617223	A	18-01-2006	CN 1749269 A 22-03-2006
			CN 1766636 A 03-05-2006
			US 2006014210 A1 19-01-2006
US 2004121477	A1	24-06-2004	EP 1447668 A2 18-08-2004
WO 0208767	A	31-01-2002	AU 7356801 A 05-02-2002
			AU 2001273568 B2 24-11-2005
			CA 2414215 A1 31-01-2002
			CN 1551984 A 01-12-2004
			EP 1356297 A2 29-10-2003
			JP 2004505248 T 19-02-2004