ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/10613311

Complete Characterization of Posttranslational Modification Sites in the Bovine Milk Protein PP3 by Tandem Mass Spectrometry with Electron Capture Dissociation as the Last Stage

ARTICLE *in* ANALYTICAL CHEMISTRY · JUNE 2003

Impact Factor: 5.64 · DOI: 10.1021/ac026295b · Source: PubMed

CITATIONS
109

READS
30

5 AUTHORS, INCLUDING:

Frank Kjeldsen
University of Southern Denmark
71 PUBLICATIONS **2,238** CITATIONS
SEE PROFILE

Kim Haselmann
Novo Nordisk
45 PUBLICATIONS **2,429** CITATIONS
SEE PROFILE

Esben S Sørensen
Aarhus University
93 PUBLICATIONS **2,788** CITATIONS
SEE PROFILE

Roman Zubarev
Karolinska Institutet
251 PUBLICATIONS **10,935** CITATIONS
SEE PROFILE

Available from: Esben S Sørensen
Retrieved on: 29 January 2016

Anal. Chem. **2003,** *75,* 2355−2361

Complete Characterization of Posttranslational Modification Sites in the Bovine Milk Protein PP3 by Tandem Mass Spectrometry with Electron Capture Dissociation as the Last Stage

Frank Kjeldsen,[†,‡] Kim F. Haselmann,[†] Bogdan A. Budnik,[†,§] Esben S. Sørensen,[#] and Roman A. Zubarev*[,†,‡]

University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark, Department of Molecular and Structural Biology, Aarhus University, Aarhus, Denmark, and Division of Ion Physics, Department of Material Sciences, Uppsala University, Uppsala, Sweden

A comprehensive approach to protein identification and determination of sites of posttranslational modifications (PTMs) in heavily modified proteins was tested. In this approach, termed "reconstructed molecular mass analysis" (REMMA), the molecular mass distribution of the intact protein is measured first, which reveals the extent and heterogeneity of modifications. Then the protein is digested with one or several enzymes, with peptides separated by reversed-phase HPLC, and analyzed by Fourier transform mass spectrometry (FTMS). Vibrational excitation (collisional or infrared) or electron capture dissociation (ECD) of peptide ions provides protein identification. When a measured peptide molecular mass indicates the possibility of a PTM, vibrational excitation is applied to determine via characteristic losses the type and eventually the structure of the modification, while ECD determines the PTM site. Chromatographic peak analysis continues until full sequence coverage is obtained, after which the molecular mass is reconstructed and compared with the measured value. An agreement indicates that the PTM characterization was complete. This procedure applied to the bovine milk PP3 protein containing 25% modifications by weight yielded all known modifications (five phosphorylations, two O- and one N-glycosylation) as well as the previously unreported NeuNAc-Hex-[NeuNAc]HexNAc group O-linked to Ser60. With the FTMS performance improved, REMMA can serve as the basis for high-throughput, high-sensitivity PTM characterization of biological important proteins, which should speed up the proteomics research.

Posttranslational modifications (PTMs) in proteins are of immense importance for protein folding and function,[1] especially in higher organisms. Beside truncation of the nascent polypeptide chain at the N- or C-terminus, glycosylation and phosphorylation are by far the most common types of PTM found in proteins.[2] Glycan groups are usually covalently bound to either the hydroxyl functionality of serine or threonine residues (O-glycosylation) or to the amide functionality of asparagines (N-glycosylation). The predominant glycans found in glycoproteins are glucose (Glc), galactose (Gal), mannose (Man), fucose (Fuc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), and *N*-acetylneuraminic acid (NeuNAc). Phosphorylation mainly occurs on serine (∼90%) and threonine (∼10%) residues and to a lesser extent on tyrosine (∼0.05%) residues.[3] Reversible phosphorylation by protein kinases and protein phosphatases regulates principle fundamental cell functions, such as cell cycle control, DNA replication, transcription, translation, protein localization, energy metabolism, and signal transduction,[4,5] while the carbohydrate diversity (heterogeneity) modulates the solubility, folding, and recognition functions of glycoproteins.[6,7]

Since it is difficult in many cases to predict the type and position of modifications from autotranslated DNA sequences, experimentally obtained information about PTMs is highly relevant in the postgenomic era. The type and the position of modifications are organism-, cell-, time-, and process-dependent. Deciphering the global picture of protein-related processes in the cell (proteomics) is impossible without determination of the dynamic PTM map, which in turn requires the availability of a fast, reliable, and sensitive procedure for PTM characterization.

Currently, such a procedure is missing. The existing techniques for characterization of PTMs, such as Edman degradation[8] and chemical deglycosylation/phosphorylation (reductive β-elimination)[9,10] followed by mass spectrometry analysis, are unlikely to become sensitive and fast enough to satisfy the high-throughput, high-sensitivity requirements of the proteomics studies.

* Corresponding author: (address) Ångström Lab, Box 534, Uppsala S-75121 Sweden; (e-mail) roman.zubarev@angstrom.uu.se; (tel) +46 18 471 72 09; (fax) +46 18 55 57 36.

† University of Southern Denmark.

‡ Uppsala University.

§ Current address: Department of Biochemistry, Mass Spectrometry Resource, Boston University, Boston, MA 02118-2526.

Aarhus University.

(1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York, 2002; Chapter 2.
(2) Charlwood, J.; Bryant, D.; Skehel, J. M.; Camilleri, P. *Biomol. Eng.* **2001**, 229−240.
(3) Hunter, T. *Philos. Trans. R. Soc. London B* **1998**, *353*, 583−605.
(4) Krebs, E. G.; Beavo, J. A. *Annu. Rev. Biochem.* **1979**, *48*, 923−959.
(5) Hunter, T. *Cell* **1995**, *80*, 225−236.
(6) Varki, A. *Glycobiology* **1993**, *3*, 97−130.
(7) Parodi, A. J. *Annu. Rev. Biochem.* **2000**, *69*, 69−93.
(8) Bousfield, G. R.; Butnev, V. Y.; Butnev, V. Y. *Biol. Reprod.* **2001**, *64*, 136−147.

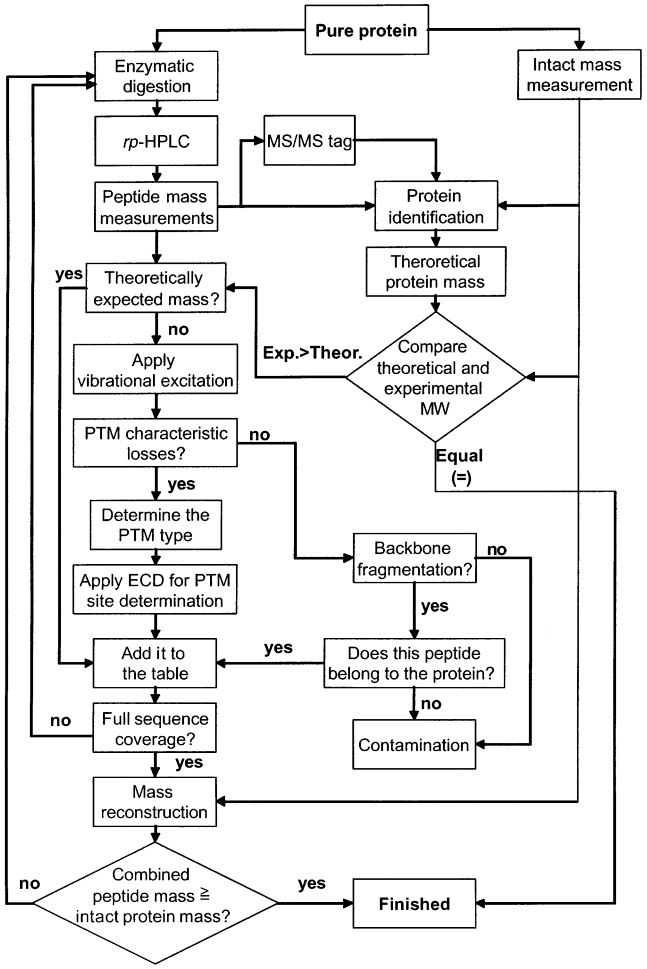
10.1021/ac026295b CCC: $25.00 © 2003 American Chemical Society
Published on Web 04/12/2003

To fulfill these requirements, the wet chemistry preparation steps should be limited to a minimum, and be replaced by rapid and sensitive techniques such as tandem mass spectrometry (MS^n, $n = 2$). While obtaining intact gas-phase protein ions with even a high degree of modification is possible with the modern soft ionization techniques, matrix-assisted laser desorption/ionization (MALDI)[11] and electrospray ionization (ESI)[12] mass spectrometric PTM analysis remains a challenge. This is mainly due to the lability in the gas phase of many modifications, which are easily lost under vibrational (collisional or infrared) excitation without revealing their positions.[13,14]

At the same time, the alternative to vibrational excitation techniques, electron capture dissociation (ECD),[15] has shown on several occasions its profound potential for PTM characterization. With ECD, the positions of phosphorylation,[16,17] N- and O-glycosylation,[14,18,19] sulfation,[19,20] and γ-carboxylation[20,21] can be easily established; these modifications are rapidly lost upon vibrational excitation. The cleavage of $N-C_\alpha$ backbone bonds in ECD is believed to happen within 10^{-12} s upon capture of a low-energy (<1 eV) electron by an ESI-generated polycation. This is much faster than intramolecular vibration energy redistribution in large molecules (up to 10^{-9} s); thus, ECD is nonergodic.[15] The high rate of bond cleavage and the moderate amount of excess energy prevent ECD losses of the labile PTM groups from the fragments.

Besides PTM determination in several mostly synthetic peptides, ECD has been used for determination of phosphorylation sites in the 24-kDa protein bovine β-casein.[16,17] While direct ECD of the intact protein ("top-down approach") identified the position of one phosphorylation and restricted three modifications to four possible sites and one more to four sites,[17] ECD of HPLC-separated enzymatic peptides of the same molecule ("bottom-up") gave the exact positions of all five phosphorylations.[16] Given that intact molecular weight information is extremely valuable in terms of the overall PTM content and heterogeneity, but does not provide information on PTM location, while the bottom-up approach does not guarantee finding all modification sites, we decided to combine both approaches into one, called "reconstructed molecular mass analysis" (REMMA). To the best of our knowledge, its originality is in obtaining in bottom-up analysis full sequence coverage and reconstruction of the molecular mass, which is then compared with the experimentally measured value. This ensures finding all statistically important PTM sites.

Chart 1. Flowchart of the REMMA Procedure



On a more detailed level, the REMMA approach is presented in Chart 1. In the beginning, the protein sequence is assumed unknown; therefore, the procedure includes protein identification. But first, molecular mass of the intact protein is measured to reveal the overall size and the heterogeneity of the modifications (top-down). Then the protein is digested with one or several different cleavage enzymes to ensure full sequence coverage. Since this is rarely obtained by applying only one protease, multiple digestions and enhanced chromatographic separation may be be needed in order to fulfill this important requirement. The enzymatic peptides are separated by reversed-phase HPLC and analyzed by Fourier transform mass spectrometry (FTMS), which besides its record resolving power and mass accuracy is currently the only MS technique compatible with ECD. HPLC fractions can be analyzed by HPLC/ECD-FTMS off-line as well as on-line in a manner suggested by several groups.[22,23] In the beginning, all peptide ions are fragmented using either vibrational excitation or ECD; the combination of the peptide mass and the partial sequence ("peptide sequence tag") is searched against the database to identify the protein, which is a part of the standard bottom-up

(9) Carlson, D. M. *J. Biol. Chem.* **1968**, *243*, 616−626.
(10) Resing, K. A.; Johnson, R. S.; Walsh, K. A. *Biochemistry* **1995**, *34*, 9477−9487.
(11) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299−2301.
(12) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.
(13) Mirgorodskaya, E.; Hassan, H.; Clausen, H.; Roepstorff, P. *Anal. Chem.* **2001**, *73*, 1263−1269.
(14) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Anal. Chem.* **1999**, *71*, 4431−4436.
(15) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265−3266.
(16) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1793−1800.
(17) Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal. Chem.* **2001**, *73*, 19−22.
(18) Håkansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Anal. Chem.* **2001**, *18*, 4530−4536.
(19) Haselmann, K. F.; Budnik, B. A.; Olsen, J. V.; Nielsen, M. L.; Reis, C. A.; Clausen, H.; Johnsen, A. H.; Zubarev, R. A. *Anal. Chem.* **2001**, *73*, 2998−3005.
(20) Kelleher, N. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T. *Anal. Chem.* **1999**, *71*, 4250−4253.
(21) Niiranen, H.; Budnik, B. A.; Zubarev, R. A.; Auriola, S.; Lapinjoki, S. *J. Chromatogr., A* **2002**, *962*, 95−103.

approach.[24] After the protein is identified, peptide ions are not fragmented if their accurately measured masses correspond to the expected enzymatic fragments; they are simply added to the table of identified masses. If a measured mass is not in the table of expected masses, which indicates the possibility of a posttranslational modification, vibrational activation is applied to the ion. At low excitation levels, two possibilities may occur. First, backbone fragmentation in the absence of labile modifications may reveal the primary sequence, which can identify the peptide as a nonspecific fragment, self-proteolytic fragment, contamination, or a peptide with a stable modification (acetylation, methylation, formylation, deamidation, disulfide bond formation, Met oxidation, etc.). Alternatively, characteristic losses from the parent ions (losses of 44 Da for γ-carboxylation, 80 Da for phosphorylation, 98 Da for phosphorylation and sulfation, 146, 161, 162, 203, or 291 Da for glycosylation, etc.) reveal the type and sometimes the structure of labile modifications. ECD then determines the PTM sites. The presence of peptides with the same primary sequence but different modifications indicates heterogeneity. After the full sequence coverage is obtained, the total molecular mass including all found PTMs is reconstructed and compared with the measured molecular mass distribution. A larger reconstructed mass than the measured intact protein mass indicates that not all modifications are present at once; then the assessment as to which modifications are more statistically significant is performed using the chromatographic peak intensities. The opposite case means that the PTM characterization has been incomplete and the procedure has to be repeated with higher chromatographic resolution, in the negative ion mode (an analogue of ECD (EDD)[25] for negative ions has been developed), or both.

REMMA has a similarity with the procedure proposed by VerBerkmoes et al. for high throughput protein identification.[26] They have integrated the "top-down" and the "bottom-up" approaches for identification of 868 proteins in *Shewanella oneidensis*. Unlike REMMA, this approach does not necessarily yield full sequence coverage or guarantee identification of all important PTM sites.

The goal of this study was to test the proposed ECD-based procedure. The need for such test arose from the fact that, despite a number of successful PTM site determinations in peptides, and the remarkable progress the McLafferty group achieved in ECD of large proteins without labile PTMs,[17] the ECD technique has not yet been 100% successful for naturally occurring heavily modified proteins. As such, the phosphorylated glycoprotein PP3 extracted from bovine milk was chosen. The choice was determined by the moderate size of the protein (15.3 kDa without modifications) and the extent of modifications (~25 wt %), as well as the fact that PTMs in this protein have previously been characterized by several groups, which provided an excellent reference point. Sørensen and Petersen performed one of the first studies on PP3 in 1993,[27] and the protein has since then been a subject of extensive research due to its involvement in a number of biological processes.[28] Sørensen and Petersen found in 1993 by Edman degradation that PP3 was modified at eight positions: by phosphorylation at Ser29, Ser34, Ser38, Ser40, and Ser46, O-glycosylation at Thr16 and Thr86, and N-glycosylation at Asn77.

Several groups have studied the structural diversity of the *N*- and *O*-glycans. In 1998, Coddeville et al.[29] investigated the glycan moiety of the O-glycosylation at the Thr86 residue by high-pH anion-exchange chromatography, ^{1}H NMR, and MALDI TOF mass spectrometry and found that it exits in three different isoforms: GalNAcα1-*O*-Thr (16%), Gal(β1−3)GalNAcα1-*O*-Thr (34%), and Gal(β1−4)GlcNAc(β1−6)[Gal(β1−3)]GalNAcα1-*O*-Thr (50%). Earlier studies by the same group discovered[30] that the moiety of the N-glycosylation at Asn77 includes eight different glycans, all biantennary of the "complex" type.

The prime objective was to identify the type and positions of PTMs in the PP3 sample. Since the PTM pattern in native PP3 was found to be very heterogeneous, as is the case for most naturally occurring proteins, the intention was to create a *portrait of the most likely member* of that heterogeneous family. This was performed by MS and MS^n analysis of the most abundant HPLC fractions of enzymatic peptide mixtures and continued to less abundant fractions until full sequence coverage was obtained.

EXPERIMENTAL SECTION

Materials and Sample Preparation. The source of the protein and the protein isolation procedure were the same as in ref 27. For intact protein mass measurements, PP3 from bovine milk (gi:741536) was dissolved in standard electrospray solvent (49:49:2 H_2O/MeOH/AcOH) to a concentration of ~10^{-6} M and sprayed through gold-coated nano-ESI capillary needles (MDS Proteomics, Odense, Denmark) at a flow of 100−1000 nL/h.

The same solution (1 μL) was applied to a sample target for MALDI experiments using 2,5-dihydroxybenzoic acid (DHB) as matrix.

The PP3 sample was also digested with Lys-C and Arg-C (Boehringer) and separated by RP-HPLC (Vydac C_{18}). To minimize the uncertainty with sample preparation, the HPLC fractions were prepared in a manner similar to that in ref 27. In all separations, a reversed-phase gradient was obtained by solvents A and B consisting of 0.1% TFA (A) and 80% acetonitril/H_2O containing 0.1% TFA (B) in a the total HPLC run of 1 h. In total, 68 and 55 fractions were collected from the Lys-C and Arg-C digestions, respectively. Of those, the most abundant 41 fractions of Lys-C and 25 fractions of Arg-C digest were analyzed by MS and MS^n. Previously, the same fractions were used in the test study for the new technique hot electron capture dissociation (HECD),[31] which identified 20 out of 25 Xle (Leu or Ile) residues.[32] All peptide fractions were lyophilized and then dissolved in an electrospray solvent to a final concentration of ~10^{-6} M; a few microliters of the solution from each fraction was consumed in the course of this work.

(22) Palmblad, M.; Tsybin, Y. O.; Ramstrom, M.; Bergquist, J.; Hakansson, P. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 988−992.
(23) Davidson, W.; Frego, L. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 993−998.
(24) Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390−4399.
(25) Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A. *Chem. Phys. Lett.* **2001**, *342*. 299−302.
(26) VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L. *J. Proteome Res.* **2002**, *1*, 239−252.
(27) Sørensen, E. S.; Petersen, T. E. *J. Dairy Res.* **1993**, *60*, 535−542.
(28) Girardet, J. M.; Linden, G.; Loye, S.; Courthaudon, J. L.; Lorient, D. J. *Dairy Sci.* **1993**, *76*, 2156−2163.
(29) Coddeville, B.; Dirardet, J.; Plancke, Y.; Campagna, S.; Linden, G.; Spik, G. *Glycoconjugate J.* **1998**, *15*. 371−378.
(30) Girardet, J.; Coddeville, B.; Plancke, Y.; Strecker, G.; Campagna, S.; Spik, G.; Linden, G. *Eur. J. Biochem.* **1995**, *234*, 939−946.

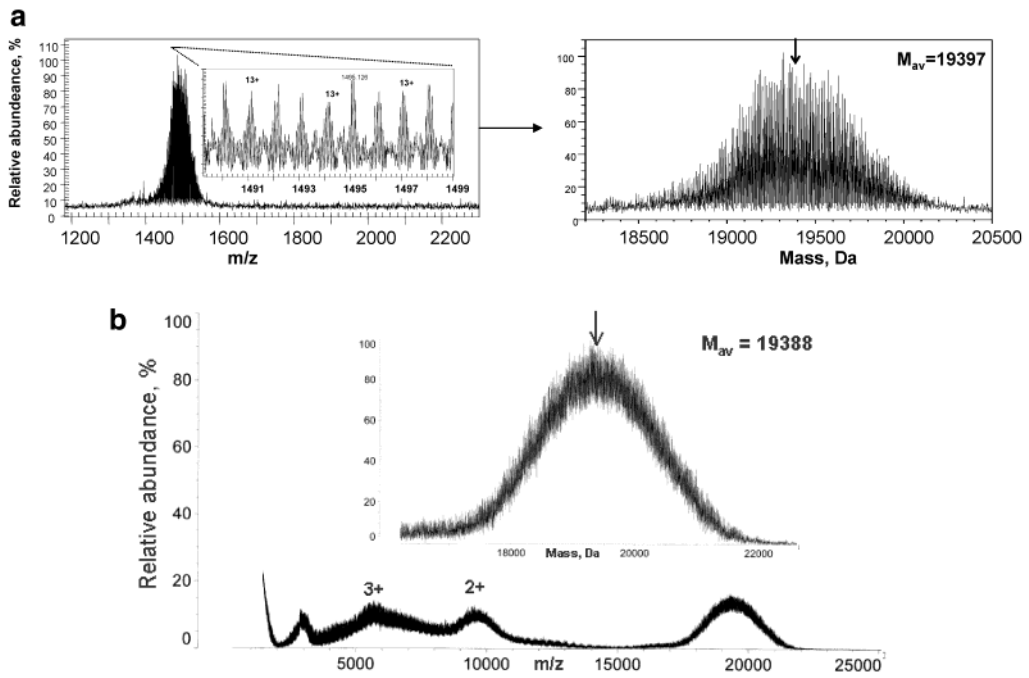


Figure 1. (a) ESI FT mass spectrum of intact PP3 (left) and its deconvolution (right). The average mass is determined to be 19 397 ± 18 Da. (b) MALDI TOF mass spectrum of intact PP3 and its deconvolution (insert). The average mass is 19 388 ± 15 Da.

Mass Spectrometry. A 4.7-T Ultima (Ionspec, Irvine, CA) was used in all FTMS experiments. Molecular ions were externally accumulated in a hexapole (Analytica of Bradford, MA) for 500−3000 ms and transferred to the cell via a 1-m-long rf-only quadrupole. The ions were captured in the cell using gated trapping.[33] For intact molecular mass measurements, the ions were cooled by a 2-ms short N_2 gas pulse. After the gas pulse, a 10-s pump-down period was applied, during which ramping of the trapping plate potential from 3 to 0 V and the inner ring potential from 2 to 0.25 V was performed to provide high resolution. For enzymatic peptides, gated trapping without gas pulse was followed by isolation of the precursor species with $Z \geq 2$ by a stored waveform. The isolated species, in case they could not be found in the list of expected masses, were first subjected to multipole storage-assisted dissociation (MSAD)[34] by increasing the accumulation period in the hexapole or collision activated dissociation (CAD) by pulsing the N_2 gas as for cooling, with simultaneous sustained off-resonance irradiation (SORI CAD)[35] of the ions of interest. If this revealed the presence of a posttranslational modification, ECD was immediately conducted by irradiation of the precursor ions by subthermal (∼0.2 eV) electrons for 50−500 ms. As a source of electrons, a 7-mm^2 indirectly heated dispenser cathode STD164 (HeatWave, CA) was used.[19,36] The voltage drop on the heater was 4.6 V, and the heating current was 1.1 A.

Additional intact protein mass measurement was performed on a linear MALDI TOF instrument with pulsed delayed extraction (Kompact MALDI 2, Kratos Analytical, Manchester, U.K.). A total of 250 spectra each acquired with 1-ms pulse of a 337-nm N_2 laser were averaged.

RESULTS AND DISCUSSION

Molecular Mass Measurements. In the mass spectrum of the intact PP3 protein measured by ESI FTMS, only one charge state was observed, 13+ (Figure 1a). Upon deconvolution, peaks were found at every mass unit in the range from ∼18 500 to ∼20 100 Da. The average mass was determined as 19 397 ± 18 Da (13+, ∼1493 Th) with a rather large fwhm of 810 Da. The average mass determined by MALDI TOF (Figure 1b) from 1+ molecular ion mass distribution was 19 388 ± 15 Da with fwhm of 2350 Da, confirming the high degree of heterogeneity detected by ESI FTMS. The fact that the average mass values determined by two independent measurements coincided within the experimental error was encouraging. The sample heterogeneity and the large difference between the measured intact mass and the theoretical unmodified mass, 4102 ± 18 Da, was consistent with the presence of multiple modifications.

Protein Identification. Unique protein identity was revealed from peptide mass fingerprints and the long sequence tags provided by ECD. For instance, the tag (K/Q)E(K/Q)(I/L)V(I/L)R from the ECD spectrum of peptide 25−53 (Figure 2) gave by the nonredundant database search program PeptideSearch (EMBL Bioanalytical Research Group) 16 candidate proteins, only 1 of which was below 22 kDa. The found protein was the PP3 precursor some 18 amino acid residues longer than PP3. Detection in the Arg-C digest (fractions 33 and 36) of the PP3 N-terminal peptide that corresponds to the cleavage between the Ala18 and

(31) Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Jensen, F.; Zubarev, R. A. *Chem. Phys. Lett.* **2002**, *356*, 201−206.
(32) Kjeldsen, F.; Haselmann, K. F.; Sørensen, E. S.; Zubarev, R. A. *Anal. Chem.* **2003**, *75*, 1267−1274.
(33) Senko, M. W.; Hendrickson, C. L.; Emmett, M. R.; Shi, S. D.-H.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 970−976.
(34) Sannes-Lowery, K.; Griffey, R. H.; Kruppa, G. H.; Speir, J. P.; Hofstadler, S. A. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1957−1961.
(35) Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B. *Anal.Chim. Acta* **1991**, *246*, 199−210.
(36) Tsybin, Y. O.; Håkansson, P.; Budnik, B. A.; Haselmann, K. F.; Kjeldsen, F.; Gorshkov, M.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1849−1854.

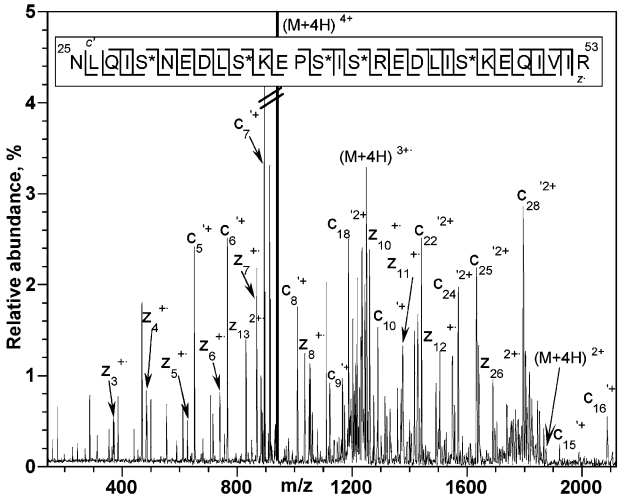


Figure 2. ECD FT mass spectrum of 4+ of peptide 25−53. Phosphorylation sites were determined at positions Ser29, Ser34, Ser38, Ser40, and Ser46.

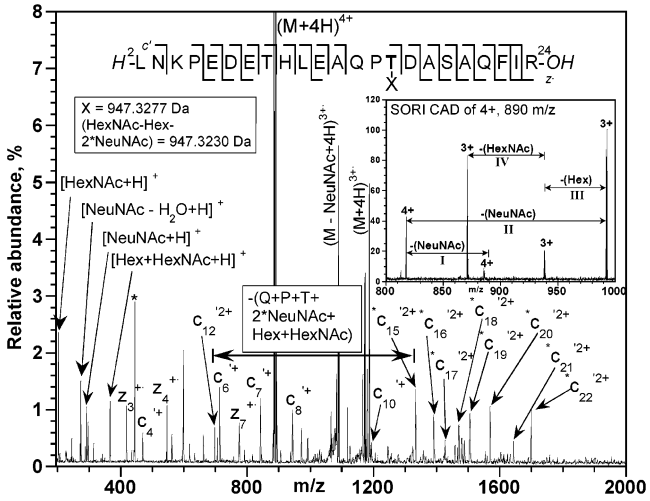


Figure 3. ECD FT mass spectrum of 4+ of peptide 2−24. ECD located O-glycosylation at Thr16. SORI CAD (inset) gave glycosidic losses in the following order: I, 291 Da (NeuNAc); II, 291 Da (NeuNAc); III, 162 Da (Hex); and IV, 203 Da (HexNAc). The accurately measured mass of the lost glycan is consistent with the composition $HexNAc_1Hex_1NeuNAc_2$.

Table 1. List of Identified Peptides Providing Full Sequence Coverage

peptide	fraction	enzyme	type of modification	degree of modification +/partial/−
1−35	35	Lys-C	O-glycosylation, phosphorylation	partial
1−24	33	Arg-C	O-glycosylation	+
1−24	36	Arg-C	O-glycosylation	−
2−24	30	Arg-C	O-glycosylation	+
3−24	32	Arg-C	O-glycosylation	−
25−53	43	Arg-C	Phosphorylation	+
48−64	18	Lys-C	O-glycosylation	+
54−76	41	Arg-C	none	−
74−99	15	Lys-C	N-glycosylation, O-glycosylation	+
74−99	17	Lys-C	N-glycosylation, O-glycosylation	partial
100−106	21	Lys-C	none	−
107−120	22	Lys-C	none	−
110−123	50	Arg-C	none	−
123−135	39	Lys-C	none	−

Ile19 residues in the precursor sequence revealed that the protein under consideration was indeed PP3 and not its precursor. Full sequence coverage (with five overlaps) was provided by six Lys-C (fractions 15, 18, 21, 22, 35, and 39) and four Arg-C (fractions 30, 41, 43, and 50 (nonspecific cleavage)) peptides (Table 1), of which four peptides included all modifications (fractions Arg-C 30, Arg-C 43, Lys-C 15, and Lys-C 18).

In the 66 fractions analyzed, more than 60 unique peptide masses were determined. This diversity (specific-only Arg-C and Lys-C digestion would produce only 19 peptides) was due to PTM heterogeneity and nonspecific cleavages.

PTM Site Determination. All phosphorylated residues were found in the peptide from the Arg-C fraction 43, which molecular mass was determined as 3752 Da. In the SORI CAD spectrum of 4+ molecular ions (not shown), consecutive losses of 98 Da (H_3PO_4) revealed the presence of at least three phosphate groups. The measured mass was consistent with a pentaphosphorylated peptide 25−53 (NLQISNEDLSKEPSISREDLISKEQIVIR) that contains five potential phosphorylation sites. ECD of the 4+ parent ion at m/z 939 gave totally 25 c′ and 24 z• fragment ions (Figure 2) corresponding to all 27 possible interresidue cleavages. The number of peaks in the spectrum of this most highly modified phosphopeptide successfully analyzed by ECD exceeded 500 for S/N = 2. The ECD spectrum shown in Figure 2 confirmed phosphorylations at Ser29, Ser34, Ser38, Ser40, and Ser46, consistent with the literature data.[27]

Glycan moieties were found in three peptides, in fractions Lys-C 15, Lys-C 18, and Arg-C 30. SORI CAD of 4+ of the latter peptide (3556 Da) gave consecutive losses of glycoside units in the following order: I and II, 291 Da (NeuNAc); III, 162 Da (Hex); and IV, 203 Da (HexNAc) (Figure 3, inset), but no sequence information. ECD of the same ions (Figure 3) gave both the sequence identification (18 c′ and 11 z• fragment ions, corresponding to all but two possible cleavages) and the O-linked glycan position in this 2−24 product (Thr16). The absence of the first amino acid residue (Ile1) can be due to an unexpected cleavage by Arg-C or posttranslational deletion in the primary sequence. One of the N−C_α bonds remaining intact in ECD was located on the N-terminal side of the glycosylated Thr16 residue, which can be explained by the sterical hindrance by the glycan units to the charge solvation on the carbonyl oxygen.[15] Despite this, the glycan position could be unambiguously assigned since the missed sequence (Gln14−Pro15) could not contain a glycosylation site. The two remaining potential sites (Thr9 and Ser19) were excluded by the present ECD cleavages. The glycan structure at Thr16 was determined by the order of losses in the SORI CAD spectrum. Since NeuNAc was a terminal glycoside unit, the only possible structure for the glycan moiety was NeuNAc-Hex-[NeuNAc]-HexNAc-*O*-Thr, which is one of the typical conformations of O-linked glycosidic groups.[37] These four glycosides weigh together 947.3230 Da, while the accurate mass of the glycan moiety at Thr16 was determined as 947.328 Da. The 5 ppm deviation was well within the average experimental error (±10 ppm for this measurement).

The peptide weighing 5599 Da isolated from the Lys-C 15 fraction also showed abundant losses of glycosidic units in the

(37) Lisowska, E. *Cell. Mol. Life Sci.* **2002**, *59*, 445−455.

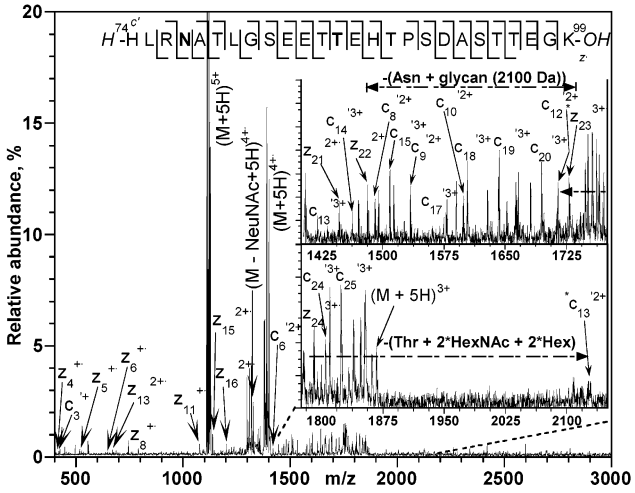


Figure 4. ECD FT mass spectrum of 5+ of peptide 74−99. Two out of 10 potential sites were glycosylated, by N-linkage at Asn77 (add-on mass 2100.7765 Da) and O-linkage at Thr86 (add-on mass 730.2535 Da).

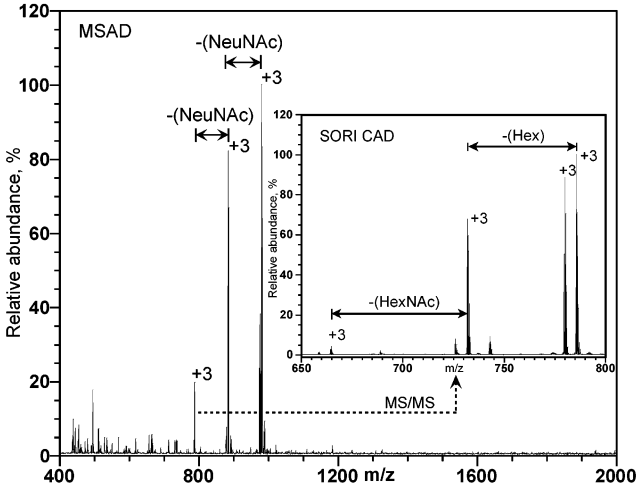


Figure 5. MSAD and SORI CAD (inset) FT mass spectra of 3+ of the peptide from fraction Lys-C 18 showing consecutive losses of two units of NeuNAc (291 Da), one unit of Hex (162 Da), and one unit of HexNAc (203 Da).

CAD spectrum of 5+. Again, the peptide sequence and the sites of modifications were established by ECD (Figure 4). The most abundant fragment peaks in the ECD spectrum were due to the small losses and the loss of a NeuNAc residue, consistent with protonation of the glycosidic bond. Closer inspection of the spectrum revealed extensive backbone fragmentation (20 c′ and 15 z• ions), with all but one interresidue bond broken. This fragmentation was consistent with the peptide being 74−99 product containing PTMs in 2 out of 10 potential sites: *N*-glycan at Asn77 (consensus tripeptide N−X-T/S(C), X ≠ P)[38] and *O*-glycan at Thr86. The weights of the glycans were determined as 2100.777 and 730.254 Da, respectively. To the best of our knowledge, this was the most heavily modified glycopeptide (50% by mass) characterized to date by ECD. The stoichiometry of each modification could be determined unambiguously from their masses. The *O*-glycan corresponded to $HexNAc_2Hex_2$ (730.2644 Da, 14 ppm deviation from the experimental mass). This configuration has been described earlier, and it is consistent with the largest *O*-glycan structure found at that position.[29] The *N*-glycan corresponded to $HexNAc_5Hex_4NeuNAc_1Fuc_1$ (2100.7615 Da, deviation of 7 ppm from the measured mass), which was also in good agreement with the earlier finding for this glycan moiety.[30] This *N*-glycan was by far the heaviest single modification determined so far by ECD.

A counterpart of this peptide containing only the N-glycosylation was found in the fraction Lys-C 17 and subjected to SORI CAD (not shown). As expected, the spectrum was dominated by abundant glycoside losses from the parent 5+ ions, and the complex fragmentation pattern revealed internal heterogeneity even within this species. All N-linked glycans are known to consist of a common pentasaccharide core Man(α1−3)-[Man(α1−6)]-Man(β1−4)-GlcNAc(β1−4)-GlcNAc(β1-N).[39] SORI CAD confirmed the presence of this oligosaccharide core, plus an additional Fuc-(α1−6) (146 Da) unit known to be attached as a common variant to the first GlcNAc(β1-N) glycoside.[37] The existence of several isoforms were revealed by the losses corresponding to different glycosides (mostly Hex, 162 Da, and HexNAc, 203 Da) originating from parent ions with identical masses. Complete characterization of the heterogeneous glycosylation in this peptide required much better chromatographic resolution and was outside the scope of the current study.

Although all modifications previously described in the literature were found, full sequence coverage was not yet obtained. Further analysis of less abundant fractions revealed an unknown glycopeptide in the fraction Lys-C 18. An increase in the accumulation time in the hexapole from 500 to 1500 ms (MSAD) produced losses of two NeuNAc glycoside units (291 Da) from the parent ion (2752 Da). Figure 5 shows this mass spectrum. The 3+ fragment ion at *m*/*z* 788 was isolated and further subjected to SORI CAD, which gave a loss of one hexose (162 Da) followed by a loss of one *N*-acetylhexosamine (203 Da) (insert). Two ECD spectra were obtained for this peptide. Figure 6a shows the ECD spectrum of 3+ at *m*/*z* 981, and Figure 6b demonstrates ECD performed after MSAD on 3+ ions at *m*/*z* 788. The latter spectrum contains cleavages of all possible interresidue bonds while the former lacks just one cleavage. Both ECD spectra were consistent with the missing sequence 48-EQIVIRSSRQPQSQNPK-64. Three potential O-glycosylation sites exist in this peptide (Ser54, Ser55, and Ser60), but only Ser60 was found modified. Combined SORI CAD and ECD data revealed the glycan moiety, NeuNAc-Hex-[NeuNAc]HexNAc-*O*-Thr, which was the same as in the O-glycosylation found in peptide 2−24, with its mass fitted to the measured increment of 947.316 Da with an accuracy of 7 ppm. With this finding, full sequence coverage was obtained.

Sequence Reconstruction. To complete the analysis, the total mass of the fully modified protein was reconstructed from the masses of the found peptides and compared with the mass measured for the intact protein. The reconstructed mass was 20 420 Da, which is 1023 ± 18 Da heavier than the average measured protein mass, and is close to the heaviest isoform mass found by ESI FTMS (Figure 1a). This means that our approach to full PTM characterization has been successful and that the fully modified form of the PP3 protein looks as shown in Chart 2. At

(38) Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631−664.
(39) Takatani, M.; Nakama, T.; Kubo, K.; Manabe, S.; Nakahara, Y.; Ito, Y.; Nakahara Y. *Glycoconjugate J.* **2000**, *17*, 361−375.

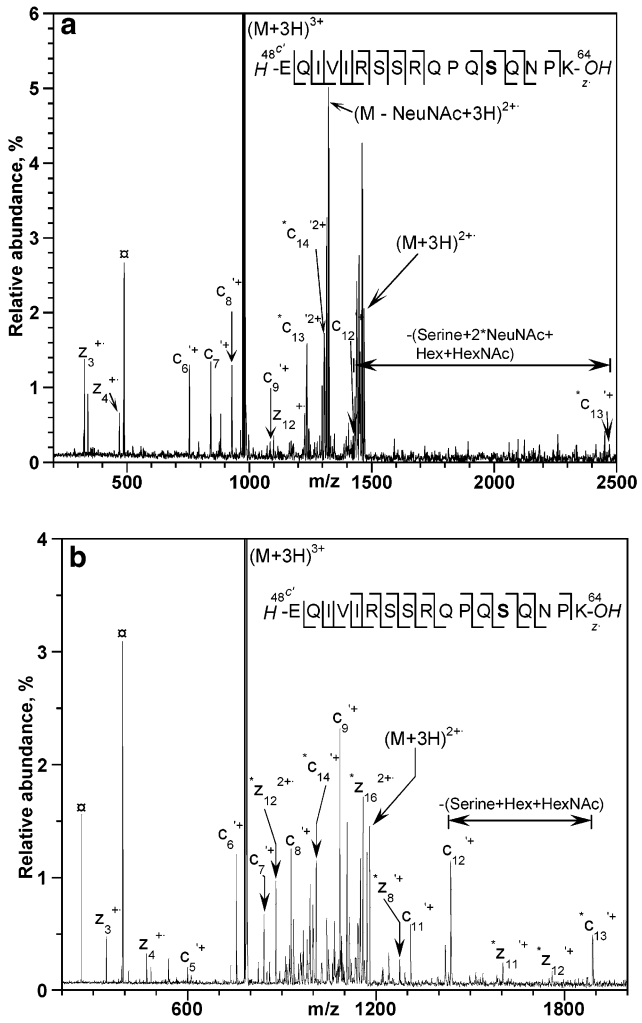


Figure 6. (a) ECD FT mass spectrum of 3+ of peptide 48−64 (*m*/*z* 981) revealing O-glycosylation of Ser60 by $HexNAc_1Hex_1NeuNAc_2$. (b) ECD FT mass spectrum of 3+ of the MSAD fragment at *m*/*z* 788 of the same peptide. The glycosylation moiety at Ser60 is $HexNAc_1Hex_1$.

Chart 2. Sequence and the Map of Post-Translational Modifications (G, Glycosylation; P, Phosphorylation) of the Bovine PP3 Protein[a]

ILNKPEDETH LEAQPTDASA QFIRNLQISN EDLSKEPSIS REDLISKEQI

VIRSSRQPQS QNPKLPLSIL KEKHLRNATL GSEETTEHTP SDASTTEGKL

MELGHKIMRN LENTVKETIK YLKSLFSHAF EVVKT

[a] O-Glycosylation at Ser60 has not been previously reported.

the same time, it appears that statistically average natural PP3 protein may not necessarily carry all modifications, which is consistent with the detection in the digests of peptides with different degrees of modification for all glycosidic modifications (Table 1). Alternatively, both MALDI and ESI produced losses of PTM from the intact protein, which we find unlikely given the similarity of both mass values.

CONCLUSION

The primary goal of this study, to test the REMMA procedure, has been achieved by determination of all known PTM sites in PP3; in addition, a previously unreported modification, O-glycosylation at Ser60, was found. To achieve the same result with alternative MS approaches would be difficult. For example, direct ECD of the intact protein at the observed charge state 13+ would be unlikely to yield the same level of detail as did ECD of the constituent peptides with the combined charge 24+. Charging the intact protein to a similar value, even if possible, would make it unstable against facile PTM losses. Higher peptide average charging resulted not only in nearly full backbone fragmentation in ECD bust also in higher FT ICR resolution that generally improves with lowering *m*/*z* and provides in turn higher mass accuracy. At the same time, strict a bottom-up approach not supported by overall molecular weight data would lead to MS^n analysis of many more low-abundance chromatographic peaks, which undoubtedly contain a great variety of PTMs occurring with low probabilities. The reconstructed molecular mass analysis avoids this pitfall by raising a "stop" flag as soon as full sequence coverage is obtained.

The overall labor consumption amounted to three weeks, which was much shorter than would have been with the strict bottom-up approach. Such a labor consumption does not exclude the use of the proposed REMMA procedure in high-throughput analyses because almost all steps involved are possible to automate. The high-sensitivity requirement is however more difficult to meet. Although only picomole amounts of each sample were consumed during the study (such amounts are readily available for many natural proteins, e.g., as those from milk, body fluids, and muscles), the standard protein separation on 2D gels in proteomics requires femtomole sensitivity. Despite improvements in the sensitivity,[40] we doubt the possibility of routine HPLC-ECD analysis at femtomole levels using the existing FT ICR instrumentation. This, however, should be available on the coming generation of FT mass spectrometers. Alternatively, ECD may be implemented in the future in other high-sensitivity instruments, e.g., quadrupole ion traps or qTOF mass spectrometers. But the current study shows that, even with the existing technology, comprehensive data can be obtained from highly modified proteins from natural sources.

ACKNOWLEDGMENT

The work was supported by the Danish National Research Councils (Grants SNF 9891448, 51-00-0358, and STFV 0001242 to R.A.Z., STFV 26-01-0058 to R.A.Z. and F.K.) and the Swedish Wallenberg Consortium North for Functional Genomics (grant to R.A.Z.).

Received for review November 19, 2002. Accepted January 30, 2003.

AC026295B

(40) Håkansson, K.; Emmett, M. R.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 3605−3610.