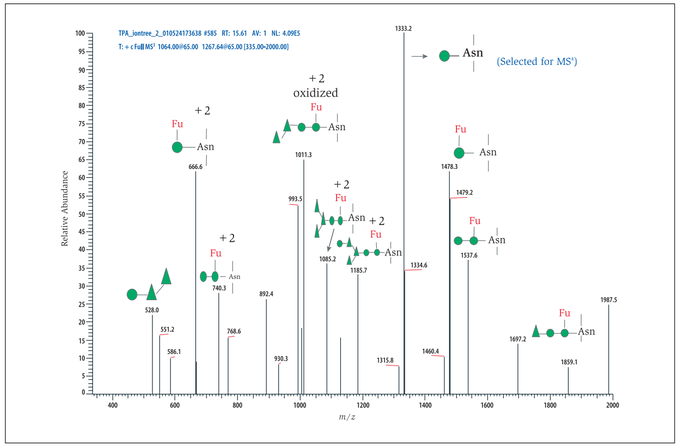

Figure 6. MS^3 on the 210-G, GlcN ion. The dominant ion in the spectrum, the T45 peptide+GlcN at *m/z* 1333, was selected for MS^4.

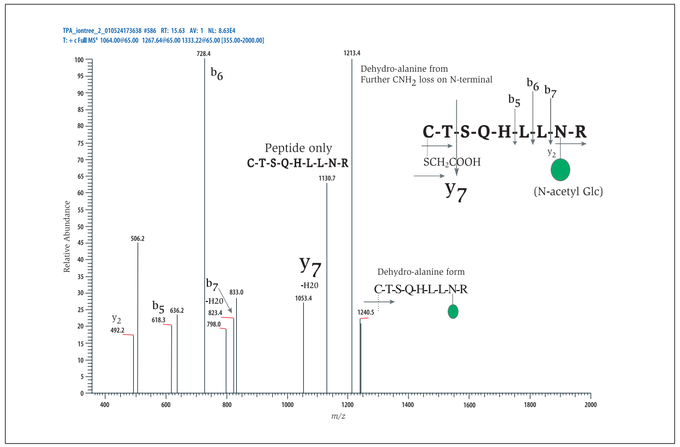

Figure 7. MS^4 on the T45+GlcN ion at *m/z* 1333.

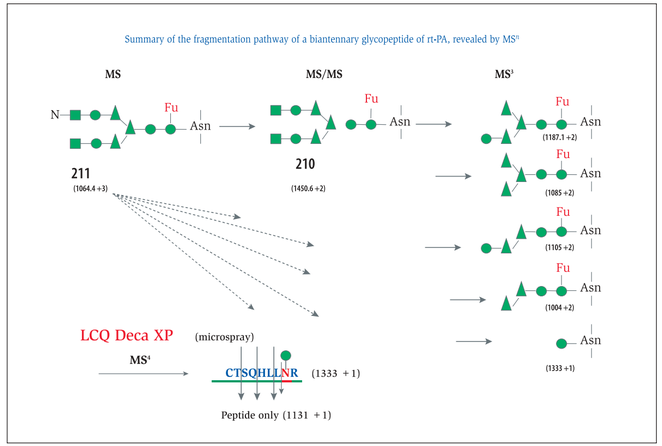

Figure 8. MS^n on the LCQ Deca XP ion trap mass spectrometer confirms predicted sequence, linkage site and glycostructure of an rt-PA peptide. The precursor ion in glycostructure 211 was determined using MS/MS. To determine oligosaccharide structure, glycostructure 210 was selected for further fragmentation using MS^3. To further confirm the linked peptide sequence and the site of attachment, ion *m/z* = 1333 was trapped and fragmented further (MS^4).

Conclusions

The Finnigan LCQ Deca XP ion trap mass spectrometer coupled with TurboSEQUEST protein identification software is an extremely powerful tool for the elucidation of post-translational modifications of proteins. In this case, the combination of Data Dependent MS^n and Dynamic Exclusion have been used to quickly generate results which are more comprehensive than those provided by traditional approaches to glycoprotein analysis. While MS/MS can give limited oligosaccharide structure information, MS^n can unambiguously confirm complete structure and the sites of attachment. This is yet another example of why the LCQ is an essential tool for every proteomics laboratory.

Acknowledgment: rt-PA graciously supplied by Genentech, Inc.

Application Note: 300

In addition to these offices, Thermo Electron Corporation maintains a network of representative organizations throughout the world.

Australia
+61 2 9898 1244

Austria
+43 1 333 50340

Belgium
+32 2 482 30 30

Canada
+1 800 532 4752

China
+86 10 5850 3588

France
+33 1 60 92 48 00

Germany
+49 6103 4080

Italy
+39 02 950 591

Japan
+81 45 453 9100

Latin America
+1 512 251 1503

Netherlands
+31 76 587 98 88

Nordic
+46 8 556 468 00

South Africa
+27 11 570 1840

Spain
+34 91 657 4930

Switzerland
+41 61 48784 00

UK
+44 1442 233555

USA
+1 800 532 4752

www.thermo.com

Thermo Finnigan LLC, San Jose, CA USA is ISO Certified.

ANXXXXX_E 1/04S

Key Words

- Finnigan™ LCQ™ series
- Proteomics
- Peptide sequencing
- Structural elucidation
- Glycosolated peptides
- PTMs

Structural Analysis of Glycosylated Peptides in Complex Mixtures with Ion Trap MS^n

Chromatography and Mass Spectrometry Application Note

Shiaw-Lin Wu, Pavel Bondarenko, Tom Shaler, Paul Shieh, and William S. Hancock

The data presented here can be acquired using any Finnigan LCQ with MS^n capability.

Introduction

In humans, glycoproteins on cell surfaces are important for communication between cells, maintaining cell structure and self-recognition by the immune system. Of particular clinical concern is how viruses, bacteria and parasites bind to cell-surface glycoproteins and use them as portals of entry into cells. Understanding the detailed structure of glycoproteins at the molecular level may provide insight to aid in combating glycoprotein-mediated induction of disease.

A variety of mass spectrometry-based approaches have been applied to the fundamental study of glycoproteins. Typically, a glycoprotein is enzymatically digested, the resulting peptide fragments are separated using HPLC and the peptides are identified by on-line MS analysis using electrospray ionization. Alternatively, fractions may be collected after the HPLC separation and analyzed off-line by matrix-assisted laser desorption ionization (MALDI) or nanospray ionization.

Peptides that do not correspond to predicted masses may be present in a glycosylated form. These putative glycopeptides are treated with a glycosidase to cleave the bond between the peptide and the oligosaccharide. The difference in mass following cleavage is used to infer the carbohydrate constituents of the cleaved glycoform. Generally, neither the oligosaccharide structure nor the exact site of attachment to the peptide can be determined. This technique is labor-intensive, time-consuming, and it requires a large amount of sample—all of which severely limit its general utility.

Goal

In this report, we describe a novel, high-throughput technique that will allow researchers to determine:

1) The amino acid sequence of glycopeptides
2) The exact site of attachment of the oligosaccharide
3) Accurate, detailed structures of the attached oligosaccharide

Starting materials for this technique are tryptic fragments of the protein of interest. Treatment with glycosidase is not required, and very little sample is needed. Analysis can be accomplished in a single, automated process using any LCQ ion trap mass spectrometer with Data Dependent™ MS^n—in this case up to MS^4—and Dynamic Exclusion™ to successively isolate, fragment, and analyze the peptide and oligosaccharide structures. In this case, as is the case for most glycoprotein analyses, *MS/MS is simply not enough.*

Experimental Conditions

Finnigan LCQ Deca XP mass spectrometer fitted with ESI probe (positive ion mode)

Ion capillary tube temp: 140 °C

Needle voltage: +3.8 kV

Sheath gas: 9 arbitrary units

Normalized Collision Energy: 65%

Dynamic Exclusion duration: 3 min

Surveyor™ Autosampler with Surveyor MS Pump: 120 µL/min before splitting, 1 µL/min after splitting

0.15×100 mm C18 column (MicroTech Scientific, Vista, CA)

HPLC gradient: 2% acetonitrile, 0.1% formic acid for 3 minutes, ramp to 60% acetonitrile in 90 minutes, ramp to 80% acetonitrile in 5 minutes, hold at 80% for 20 minutes.

Discussion

Recombinant human tissue plasminogen activator (rt-PA) was expressed in Chinese hamster ovary cells and purified. This glycoprotein has an approximate molecular weight of 64 kDa and a heterogeneous distribution of N-linked glycoforms. The purified protein was reduced and alkylated at two sites, then digested with trypsin. A 2-µg aliquot of the digest mixture was analyzed. Automated Data Dependent LC/MS/MS was used to generate the chromatogram shown in Figure 1.

The product ion spectra from the tryptic digest mixture were used by TurboSEQUEST™ software to search the human protein database. Figure 2 shows that the protein was identified as tPA. Sequence coverage is shown. The peptides highlighted in red were immediately identified using TurboSEQUEST, representing 79% coverage of the protein. The peptides labeled in blue were not initially identified, however each of these "unknown" peptides contains an Asn-X-Ser/Thr motif (with the Asn labeled in green), strongly suggesting that these peptides might be glycosylated.

SYQVICRDEK TQMIYQQHQS WLRPVLRSNR VEYCWCNSGR
AQCHSVPVKS CSEPRCFNGG TCQQALYFSD FVCQCPEGFA
GKCCEIDTRA TCYEDQGISY RGTWSTAESG AECTNWNSSA
LAQKPYSGRR PDAIRLGLGN HNYCRNPDRD SKPWCYVFKA
GKYSSEFCST PACSEGNSDC YFGNGSAYRG THSLTESGAS
CLPWNSMILI GKVYTAQNPS AQALGLGKHN YCRNPDGDAK
PWCHVLKNRR LTWEYCDVPS CSTCGLRQYS QPQFRIKGGL
FADIASHPWQ AAIFAKHRRS PGERFLCGGI LISSCWILSA
AHCFQERFPP HHLTVILGRT YRVVPGEEEQ KFEVEKYIVH
KEFDDDTYDN DIALLQLKSD SSRCAQESSV VRTVCLPPAD
LQLPDWTECE LSGYGKHEAL SPFYSERLKE AHVRLYPSSR
CTSQHLLNRT VTDNMLCAGD TRSGGPQANL HDACQGDSGG
PLVCLNDGRM TLVGIISWGL GCGQKDVPGV YTKVTNYLDW
IRDNMRP

Figure 2. Sequence of rt-PA. Observed peptides labeled in red. Blue highlighted peptides correspond to glycoforms identified after possible glycoform structures were manually added to the database. Small peptides in black, representing 10% of the amino acids, were not observed.

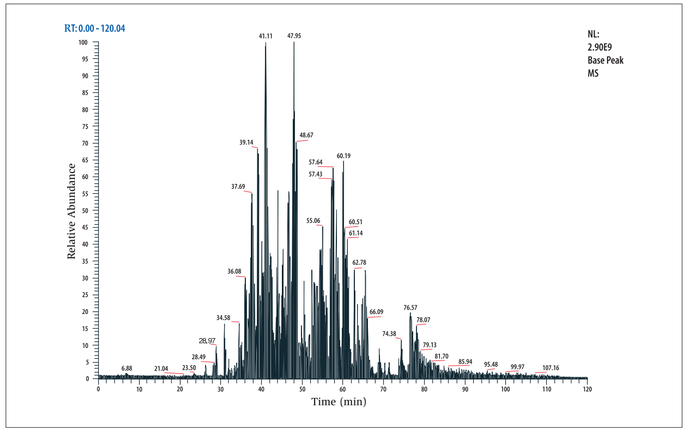

Figure 1. Base peak chromatogram from Data Dependent LC-MS/MS of a rt-PA tryptic digest.

N-linked oligosaccharides in human proteins are generally predictable but cell lines may incorporate uncommon structures making direct confirmation of complete structure with MS^n important. All have the same core structure with extensions and branches composed primarily of sialic acid, galactose and N-acetylglucosamine.

Figure 3 designates the expected mass-to-charge ratios for the possible forms of the glycosylated tryptic peptide T45 from rt-PA. This glycoform was found to have a typical structure including one sialic acid, two galactoses and two N-acetylglucosamines (GlcN) (labeled as structure 211).

Theoretical *m/z* values for all possible cleavage products of the glycoform were added to the database to enable TurboSEQUEST to automatically match glycopeptide structures. From these, TurboSEQUEST was able to identify the 211 form of the T45 peptide (211-T45) at a retention time of 15.57 minutes. As would be expected from such a complex mixture, a number of co-eluting ions were observed. However, the triply-charged 211-T45 ion at *m/z* 1064 was easily identified.

The base peak ion chromatogram and full-scan mass spectrum for the targeted glycopeptide are shown in Figure 4.

The signal intensities for the doubly- and triply-charged glycopeptides (*m/z* 1595 and *m/z* 1064, respectively) were less intense than those for co-eluting peptides because glycosylated peptides typically have lower ionization efficiencies than unmodified peptides. The use of Dynamic Exclusion allowed for automated analysis of these low-intensity glycopeptides. The LCQ Deca XP was operated in Data Dependent Ion Tree mode, where the most abundant ion in the MS^2 spectrum was selected for MS^3, and the most abundant MS^3 product ion was chosen for MS^4.

MS^2 through MS^4 spectra are presented in Figures 5–7. The MS^2 and MS^3 spectra generated a heterogeneous collection of glycosylated product ions. GlcN was not easily cleaved from the peptide, so it serves as a marker to identify the site of glycosylation. The CID energy was sufficient at the MS^4 stage to begin fragmenting the T45 peptide and to confirm the peptide sequence.

When all of the glycosylated peptides were identified and combined with the unmodified peptides, the sequence coverage obtained in the experiment was increased to greater than 90%.

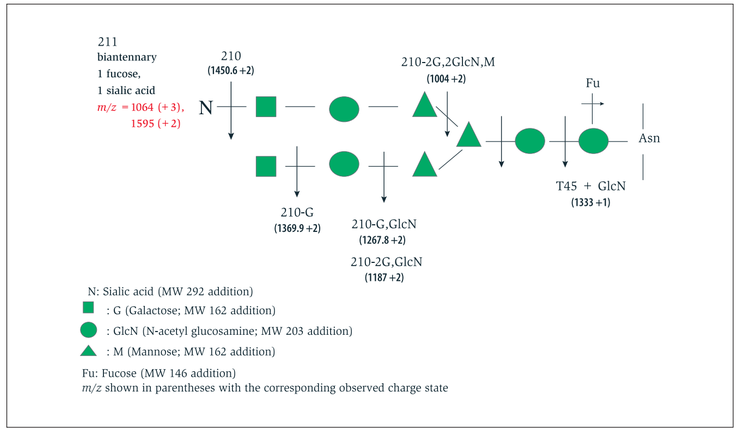

Figure 3. Typical theoretical oligosaccharide sequences for rt-PA peptide T45. Mass-to-charge ratios and charge states for the heterogeneous structures are listed in parentheses. The unglycosylated peptide was determined to have a molecular weight of 1131 amu.

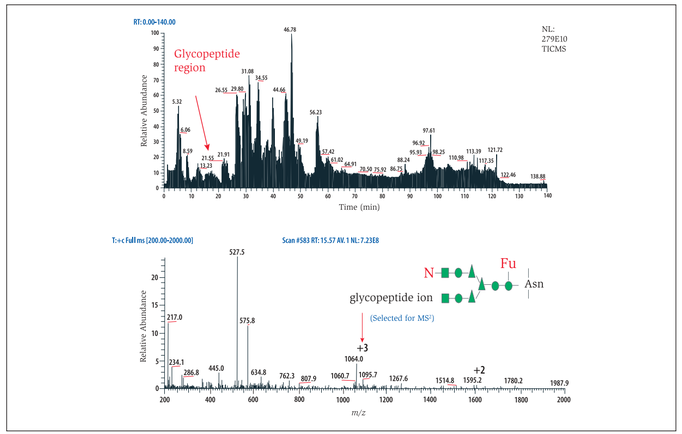

Figure 4. Top panel—Base peak ion chromatogram of tryptic peptides from rt-PA.
Bottom panel—Full-scan mass spectrum of the 211-T45 glycoform eluting at 15.57 minutes.

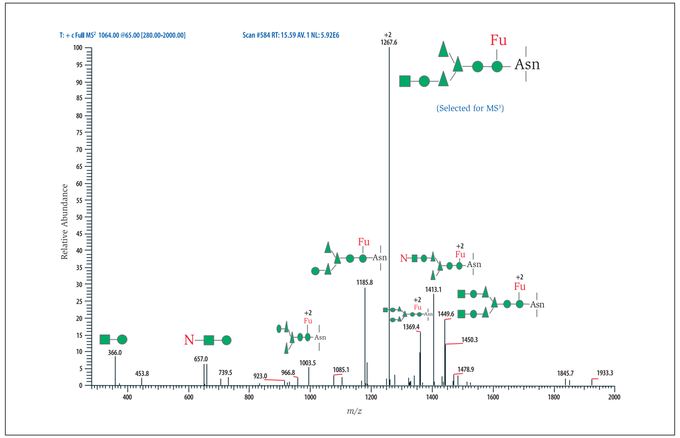

Figure 5. MS/MS on the 211-T45 ion at *m/z* 1064. The glycopeptide fragmented into many different glycoforms.
The 210-G, GlcN ion *m/z* 1267.6 (+2) was selected for MS^3 analysis.