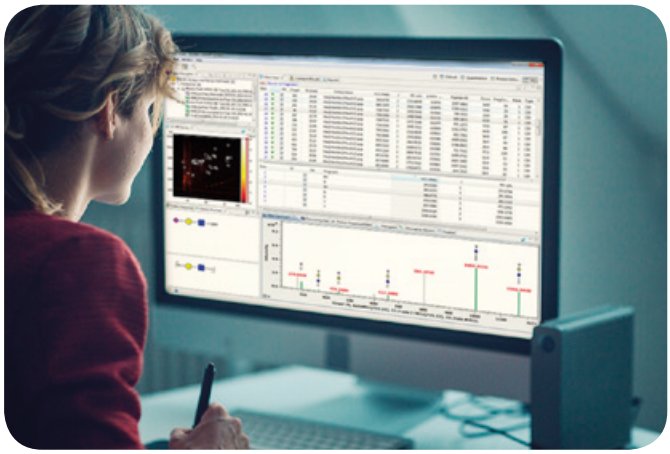

Application Note LCMS-118

Unraveling the Human Blood Plasma *O*-Glycoproteome

amaZon speed ETD ion trap MS and ProteinScape 4.0 – a powerful combination

Introduction

Glycosylation as one of the most common forms of posttranslational modification of proteins is believed to be involved in virtually every biological event in eukaryotes, as well as many prokaryotes and viruses. The addition of complex carbohydrates, so-called glycans, to specific sites of the nascent polypeptide chain is determined by a tightly regulated enzyme-catalyzed glycosylation machinery, and affects either directly or indirectly physicochemical properties of proteins or entire cellular processes. Several clinical studies could for instance link alterations in protein glycosylation to the onset and progression of numerous diseases, including infectious and autoimmune diseases as well as cancer (1, 2). In addition, also the stability, efficacy, half-life and antigenicity of proteins has been shown to be influenced by their glycosylation status. Therefore, understanding and taking advantage of the cellular glycosylation machinery and its biological implications is of high relevance for clinical research as well as for quality control in life sciences and biopharmaceutical industry (3, 4).

Authors

Marcus Hoffmann[1], Kristina Marx[2], Udo Reichl[1,3], Erdmann Rapp[1,4],

1) Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, 39106 Magdeburg, Germany
2) Bruker Daltonics, 28359 Bremen, Germany
3) Otto von Guericke University Magdeburg, Chair of Bioprocess Engineering, 39106 Magdeburg, Germany
4) glyXera GmbH, Leipziger Strasse 44 (Zenit), 39120 Magdeburg, Germany

Keywords	Instrumentation and Software
O-Glycopeptides	amaZon ETD
CID-MS²/MS³	ProteinScape
ProteinScape	GlycoQuest
GlycoQuest	
Ion Trap	

There are two major forms of protein glycosylation, namely *N*- and *O*-glycosylation, which differ in terms of the attached glycan structures as well as the glycosylation sites. The first one refers to the attachment of glycans to the amide group of asparagine according to the consensus sequence N|X|S/T/C (X ≠ P, C rarely found), while the latter refers to glycans bond to the hydroxyl group of either serine or threonine without following a fixed consensus motif. Frequently, *O*-glycans have been described to be attached in regions with a high content of serine, threonine and proline.

The mass spectrometric analysis of glycosylated proteins is quite challenging and exceeds the technical and bioinformatic demands of general proteomic studies. While the glyco-proteomic analysis of protein *N*-glycosylation is already quite advanced, this is not true for *O*-glycosylation analysis (5). This is primarily due to the lack of a conserved consensus motif for the attachment of *O*-glycans, but also because an universal enzyme, such as PNGase F, is missing, which allows for the release of all *O*-glycans from the protein for a separate glycan analysis.

From a clinical and biopharmaceutical point of view, proteins present in the human blood plasma hold an enormous diagnostic and therapeutic potential. However, there are still many human blood plasma proteins that have not been fully characterized – in particular with respect to their *O*-glycosylation.

Marcus Hoffmann *et al.* recently published an explorative study on the human blood plasma *O*-glycoproteome (6). To this end, an *O*-glycoproteomic workflow has been developed, which effectively combines the multistage-fragmentation capabilities (CID-MS^2/MS^3) of the Bruker amaZon ETD ion trap mass spectrometer, with the powerful glycoproteomic features offered by the Bruker ProteinScape 4.0 software suite.

This application note summarizes the workflow described by Marcus Hoffmann for the site-specific *O*-glycoproteomic analysis of nonspecifically digested human blood plasma (glyco)peptides after HILIC enrichment. It focuses on the acquisition of CID MS^2 and MS^3 *O*-glycopeptide spectra, and highlights the strengths of ProteinScape 4.0 with respect to the automatic classification, annotation and identification of such spectra.

Experimental

Sample preparation and RP-nanoHPLC-ESI-ion trap (IT) MS^n

Pooled human blood plasma of healthy donors (Affinity Biologicals Inc.) was reduced, carbamidomethylated, and proteolytically digested using Proteinase K (Sigma Aldrich). Protein-ase K cleaves proteins between amino acids X and Y (X-↓-Y), with X = an aliphatic, aromatic, or hydrophobic amino acid, and Y = any amino acid. After precipitation with acetonitrile, glycopeptides were enriched and fractionated via HILIC-HPLC (ACQUITY UPLC BEH HILIC Column, 130Å, 1.7 µm, 2.1 x 100 mm; Waters, Manchester, UK; see table 1 for more details). Enriched *O*-glycopeptide fractions were subjected to mass spectrometric analysis using reversed-phase liquid chromatography (see table 2 for more details) coupled to an amaZon ETD ion trap mass spectrometer. Enhanced resolution fragment spectra in positive ion mode were obtained using data-dependent CID-MS^2 acquisition (details are given in table 3). From these data precursor ions for MS^3 experiments were determined by the user and analyzed in a separate run.

Table 1: Glycopeptide enrichment and fractionation via HILIC-HPLC.

LC settings				
System	UltiMate™ Nano HPLC (Thermo Scientific)			
Analytical column	HILIC Column: ACQUITY UPLC BEH HILIC column, 130Å, 1.7 µm, 2.1 x 100 mm (Waters)			
Oven temperature	40 °C			
Eluents	A: acetonitrile B: 50 mM ammonium formate in water (pH 4.4)			
Gradient	#	Time [min]	Flow rate [µl/min]	B [%]
	1	0	250	20
	2	5	250	20
	3	30	250	50
	4	31	150	90
	5	40	150	90
	6	40.1	150	20
	7	50	150	20
	8	50.1	250	20
	9	60	250	20
Fractions	Collection every 2 min (from 0-34 min), in total 17 fractions			

Table 2: HPLC settings

LC settings						
System	UltiMate™ Nano HPLC (Thermo Scientific)					
Trap column	Acclaim PepMap100, C18, 5 µm, 100 Å, 300 µm i.d. x 5 mm					
Analytical column	Acclaim PepMap RSLC, C18, 2 µm, 100 Å, 75 µm i.d. x 15 cm					
Oven temperature	40 ˚C					
Valve Switching	2 min and 42 min					
Eluents	Loading Pump Solvent 1: 98% water, 2% acetonitrile, 0.05% trifluoroacetic acid Loading Pump Solvent 2: 20% water , 80% acetonitrile, 0.05% trifluoroacetic acid Nano Pump Eluent A: 98% water, 2% acetonitrile, 0.1% formic acid Nano Pump Eluent B: 10% water , 10% 2,2,2-trifluoroethanol, 80% acetonitrile, 0.1% formic acid					
Gradient			Loading Pump		Nano Pump	
	#	Time [min]	Flow rate [µl/min]	B [%]	Flow rate [µl/min]	B [%]
	1	0	7	0	0.3	4
	2	4	-	-	0.3	4
	3	5	7	0	-	-
	4	6	30	0	-	-
	5	30	30	100	-	-
	6	33	-	-	0.3	30
	7	34	-	-	0.3	90
	8	38	30	100	0.3	90
	9	39	-	-	0.3	4
	10	40	7	100	-	-
	11	45	7	100	-	-
	12	46	7	0	-	-
	13	60	7	0	0.3	4

Table 3: MS and CID-MS^2 parameters.

Acquisition parameters	
Source, Polarity	nanoFlow ESI Sprayer, positive ion mode
MS settings:	
Scan mode	Enhanced Resolution
Scan range	100–2,500 *m/z*
ICC target, Max. Accu. Time	300,000, 200 ms
Spectra averages	5 (Rolling averaging: off)
SPS tuning mass	850 *m/z*
MS^2 settings:	
Scan mode	Enhanced Resolution mode
Scan range	100-2500 *m/z*
ICC target, Max. Accu. Time	300,000, 200 ms
No. of precursor ions	4 (active exclusion after 2 spectra, release after 0.15 min; reconsider precursor, if current intensity/previous intensity > 5)
Isolation width	4
Spectra averages	3
Fragmentation amplitude (CID)	1.2 V (SmartFrag 30-200%)
Fragmentation time	40 ms
Cut-Off	27%

Data processing and *O*-glycopeptide identification

Raw data were processed in DataAnalysis 4.4 and subsequently exported to Protein-Scape 4.0 for semi-automated glycoproteomic analysis. For protein identification CID-MS^2 spectra were first searched against the human subset of the UniProtKB/SwissProt database using MASCOT. Results are not shown here.

For the identification of the glycan as well as the glycopeptide amino acid sequence, corresponding CID-MS^2 and MS^3 spectra of *O*-glycopeptides were merged. These merged spectra contain glycan (from MS^2 fragmentation) and peptide fragments (from MS^3 fragmentation). In the classification step, ProteinScape calculates glycan and peptide mass of each glycopeptide based on the glycan fragments. For the glycan identification, the classified *O*-glycopeptide spectra were searched against the CarbBank database using the GlycoQuest search engine, which is embedded in ProteinScape 4.0 (see table 4).

Furthermore, the merged MS^2/MS^3 spectra can be searched through MASCOT using the peptide mass obtained from the classification step. If the peptide sequence is already known, the spectrum can be matched with the sequence (no protein search performed) to proof the annotation. Afterwards, glycan and peptide annotation can be combined within one spectrum.

Table 4: ProteinScape 4.0 GlycoQuest search parameters.

Parameter	Value
Submitted to search	Classified glycopeptide CID spectra
Glycan type	-
Taxonomy	No restriction
Database	CarbBank
Composition restriction	Hex 0-10; HexNAc 0-7; Neu5Ac 0-5; Fuc 0-3
Derivatization	Underivatized
Ionization	H+ up to 4
MS tolerance	± 0.3 Da
MS/MS tolerance	± 0.3 Da
#13C	1
Fragmentation	B-ions, Y-ions; max. 2 cleavages

Results

Results shown here represent a short extract of the study published by Hoffmann *et al.* (6). The aim of this study was to analyze the human blood plasma *O*-glycoproteome in a site-specific manner. To this end, an analytical workflow has been optimized for the glycoproteomic analysis of nonspecifically digested *O*-glycopeptides after HILIC enrichment. Central to this workflow is the beneficial combination of the multistage-fragmentation capabilities (CID-MS^2/-MS^3) of the amaZon ETD ion trap mass spectrometer.

Low-energy CID-MS^2 fragmentation of *O*-glycopeptides primarily generates fragment ions that allow the identification of the glycan composition. This piece of information however, needs to be complemented by the identification of the corresponding peptide moiety. For this purpose CID-MS^3 experiments were conducted on putative peptide masses, that have been manually selected from CID-MS^2 spectra.

To reduce time consuming *de novo* annotation the obtained data were processed in a semi-automated manner using the glycoproteomic data analysis features of ProteinScape 4.0. Figure 1 gives an overview over the workflow.

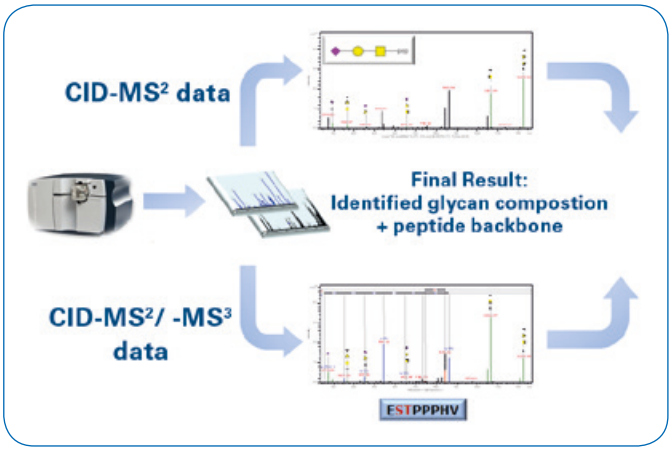

Figure 1: Mass spectrometric workflow for the identification and characterization of *O*-glycopeptides: Identification of the glycan moiety (CID-MS^2) and sequencing of the peptide backbone (CID-MS^2/-MS^3).

The ProteinScape user interface with some of the main features is illustrated in figure 2.

The Identification of the glycan moiety is based on the fragments generated during MS^2 fragmentation as described in the experimental section. An example for an identified glycan moiety is given in figure 3. To reduce the complexity, only the MS^2 spectrum is shown, although classification and identification can be performed on the merged CID-MS^2/-MS^3 spectra as well. Glycan fragments (Y- as well as oxonium ions) are annotated automatically by the ProteinScape software.

For sequencing of the peptide backbone targeted MS^3 experiments on the putative peptide masses were performed. The resulting merged CID-MS^2/-MS^3 spectra can be submitted to a MASCOT database search using the peptide mass obtained from the classification, or they can be directly matched against the theoretical sequence if already known. The merged CID-MS^2/-MS^3 spectrum of the same *O*-glycopeptide shown in figure 3 is matched with the peptide sequence $_{689}$E**ST**PPPHV$_{696}$ derived from inter-alpha-trypsin inhibitor heavy chain H2 precursor. The annotated CID-MS^n spectrum is shown in figure 4 demonstrating the capabilities of ProteinScape for automatic combined annotation of glycan and peptide sequence in one spectrum.

The complete results of the *O*-glycopeptide identification and characterization were reported by Hoffmann *et al.* (6). They describe careful inspection of the data as well as *de novo* annotation in their publication.

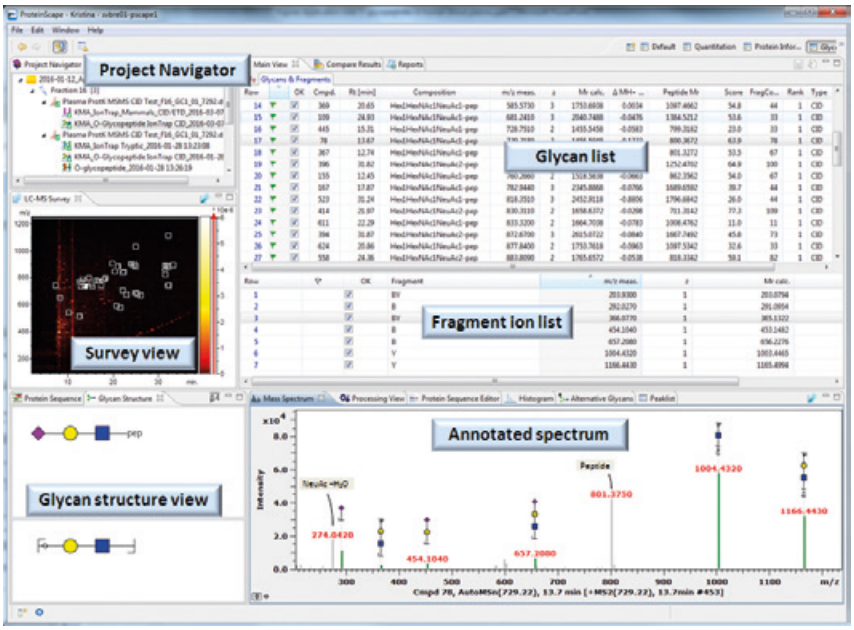

Figure 2: ProteinScape 4.0: GlycoQuest search results of human blood plasma *O*-glycopeptides.

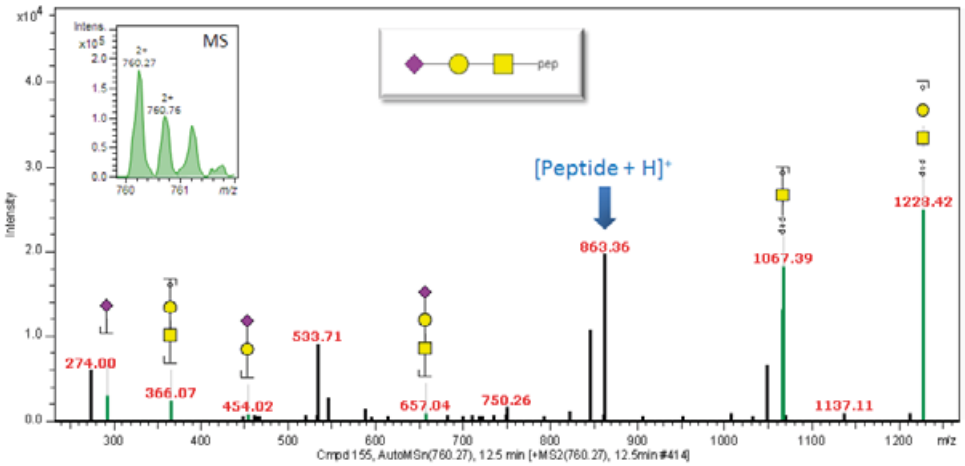

Figure 3: MS^2 CID spectrum of the precursor *m/z* 760.25 (2+). The precursor signal is shown in the upper left.

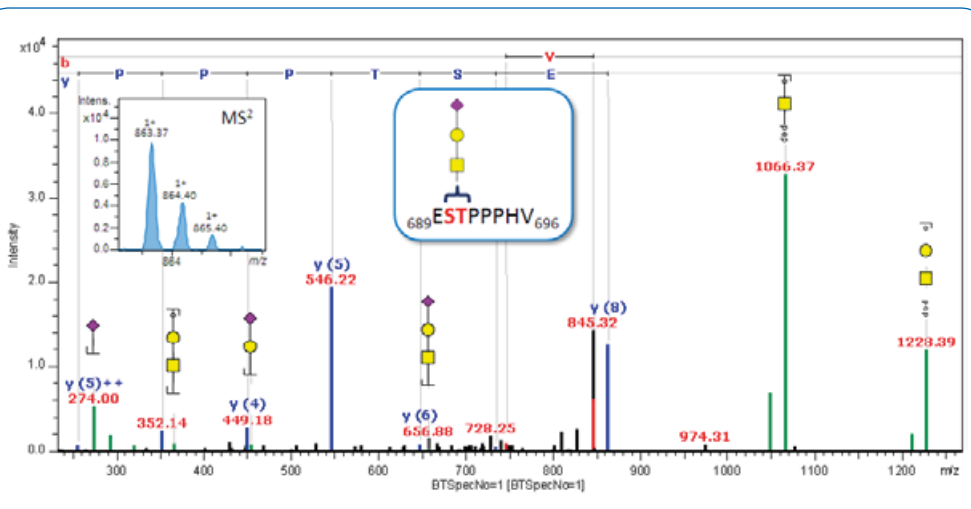

Figure 4: Merged MS^2/MS^3 CID spectrum of MS^2 precursor m/z 760.25 (2+) and MS^3 precursor *m/z* 863.37 (1+). The latter is shown in the upper left.

Conclusion

Hoffmann *et al.* (6) have demonstrated the non-targeted site-specific analysis of the human blood plasma *O*-glycoproteome. The *O*-glycoproteomic analysis of such a complex sample is extremely challenging and therefore requires a dedicated workflow comprising optimized sample preparation, sensitive and fast LC-MS^n measurements, as well as sophisticated bioinformatics tools.

This application note highlights the powerful combination of the amaZon ETD ion trap mass spectrometer together with ProteinScape 4.0 for the *O*-glycoproteomic analysis of human blood plasma.

The multistage-fragmentation capabilities of the amaZon ETD system allow to perform CID-MS^2 and -MS^3 experiments on *O*-glycopeptides enabling the identification of the glycan composition and sequencing of the peptide backbone. The acquired spectra were processed in ProteinScape 4.0. Classification of glycopeptide spectra as well as identification and annotation of glycan and peptide moiety can be both performed using merged MS^2/MS^3 CID spectra.

Incorporation of the ProteinScape bioinformatics platform in the glycoproteomic workflow significantly reduces the time required for the interpretation of glycopeptide spectra.

References

[1] Ceciliani, F., and Pocacqua, V. (2007) The acute phase protein alpha1-acid glycoprotein: a model for altered glycosylation during diseases. *Curr. Protein Pept. Sci.* 8, 91-108

[2] Cazet, A., Julien, S., Bobowski, M., Burchell, J., and Delannoy, P. (2010) Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res.* 12, 204

[3] Berger, M., Kaup, M., and Blanchard, V. (2012) Protein glycosylation and its impact on biotechnology. *Adv. Biochem. Eng. Biotechnol.* 127, 165-185

[4] Rhodes, J., Campbell, B. J., and Yu, L.-G. (2010) Glycosylation and Disease

[5] Jensen, P. H., Kolarich, D., and Packer, N. H. (2010) Mucin-type *O*-glycosylation--putting the pieces together. *FEBS J. 277, 81-94*

[6] Hoffmann, M., Marx, K., Reichl, U., Wuhrer, M., and Rapp, E. (2016) Site-specific *O*-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol. Cell. Proteomics* 15, 624-641

Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com - www.bruker.com