Glycopeptide identification using the ZenoTOF 7600 system with EAD fragmentation and PEAKS GlycanFinder software

Remco van Soest[1], Patrick Pribil[2], Zhengwei Chen[1] and Kristina Jurcic[3]
[1]SCIEX USA; [2]SCIEX Canada; [3]Bioinformatics Solutions Inc.

This technical note describes the identification of N-linked glycopeptides using nanoflow liquid chromatography (LC) separation with electron activated dissociation (EAD) fragmentation on the ZenoTOF 7600 system. >1,000 N-glycosylated peptides were identified in glycopeptide-enriched samples from human plasma digests with an EAD-based data-dependent acquisition (DDA) approach. EAD on the ZenoTOF 7600 system is a tunable electron capture-based fragmentation technique that produces unique peptide fragment ions, allowing for the unambiguous assignment of the types and sites of glycosylation post-translational modifications. Data processing using a combination of PEAKS GlycanFinder software and SCIEX OS software enables robust qualitative and quantitative analysis of the rich spectra generated by the ZenoTOF 7600 system.

Key features of N-glycopeptide identification using the ZenoTOF 7600 system with EAD fragmentation

- EAD MS/MS on the ZenoTOF 7600 system is fast, user-tunable, and generates unique fragment ions compared with CID, preserving side chain information for the identification and localization of labile glycan PTMs

- PEAKS GlycanFinder software is designed to use all information from EAD spectra for the identification of the glycopeptides, including confirmation of glycan structures

- SCIEX OS enables easy setup of data-dependent acquisition EAD MS/MS methods, along with industry-standard workflows for quantitative data processing
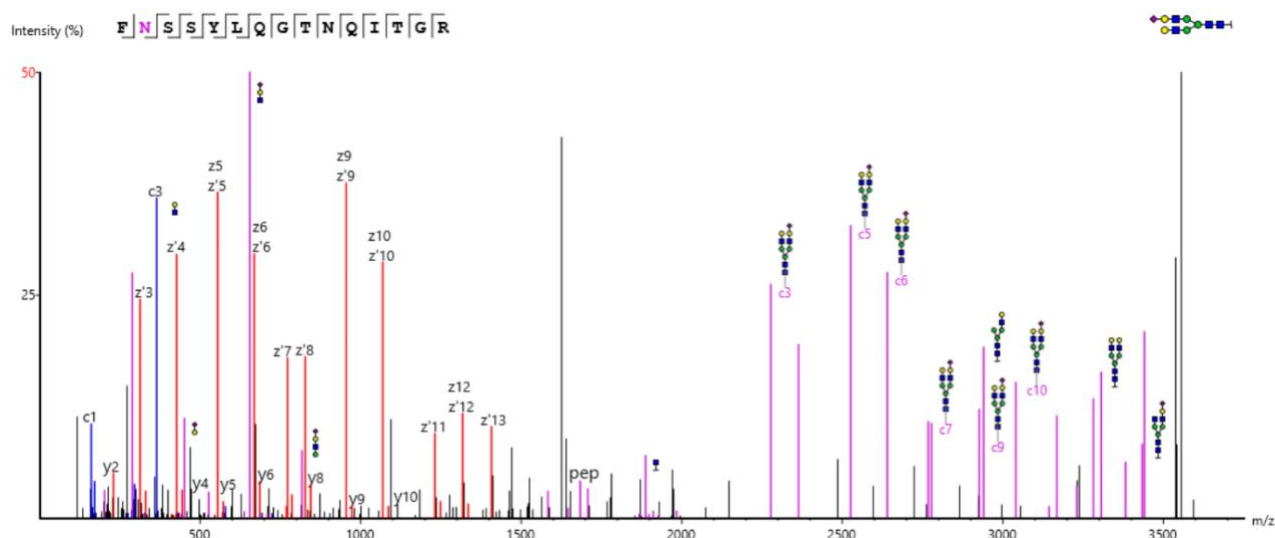

Figure 1. EAD MS/MS spectrum of glycopeptide FN[(HexNAc)$_4$(Hex)$_5$(NeuAc)$_1$]SSYLQGTNQITGR from Apolipoprotein B-100, processed and annotated using PEAKS GlycanFinder software. The peptide backbone is fully sequenced, while several peptide-glycan fragments are evidence for the proposed glycan structure.

Introduction

Protein glycosylation is a critical post-translational modification (PTM) that affects protein folding and stability. It is also essential for cell-cell adhesion and, as such, plays a role in immune response, cancer, and numerous other diseases. Mass spectrometry (MS) instrument sensitivity is one of the significant limitations when analyzing glycopeptides due to the heterogeneity of glycan structures, resulting in multiple peptide isoforms with much lower abundances than their non-glycosylated forms. The challenges associated with low glycopeptide abundances can be overcome with strategies like enrichment of glycopeptides, a lower flow rate LC separation regime (i.e., nanoflow LC separation), and high-performance MS systems for glycopeptide detection and characterization.

Another challenge for glycopeptide characterization comes from the labile nature of the glycosylation PTM. CID fragmentation of glycosylated peptides often provides more limited peptide backbone information and typically results in fragments lacking the labile side chain modifications. Alternative fragmentation methods, such as EAD-based MS/MS, have been shown to yield more complete peptide backbone information. Additionally, EAD fragmentation provides site-specific PTM localization due to the retention of these modifications on the resulting fragment ions.[1]

In this technical note, nanoflow LC separation and EAD fragmentation were used to demonstrate the identification of N-glycopeptides in enriched, digested human plasma.

Methods

Samples and reagents: Human pooled plasma K_2EDTA was acquired from BioIVT. Top 14 Abundant Protein Depletion Midi spin columns from Thermo Fisher were used for plasma depletion. Trypsin/Lys-C protease mix was purchased from Promega. PolyHYDROXYETHYL A 12 µm 300 Å HILIC resin from PolyLC was used for glycopeptide enrichment.

Sample preparation: After depletion of the top 14 most abundant proteins using the depletion spin columns (using the manufacturer's protocol), human plasma was digested following a filter-aided sample preparation (FASP) protocol described in the literature.[2] After digestion and solid phase extraction clean-up, the sample was enriched for glycopeptides using a PolyHYDROXYETHYL A Hydrophilic Interaction Chromatography (HILIC)-cotton column.[3] The resulting extract was evaporated to dryness and redissolved in water with 0.1% formic acid for analysis by LC-MS. Based on an estimated human plasma protein concentration of 80 mg/mL, and the manufacturer's estimate of removal of 95% of the top 14 abundant proteins, the depleted extract had an assumed peptide concentration equivalent to 7.5 µg protein/µL before digestion and enrichment.

Chromatography: The samples were analyzed using a Waters ACQUITY M-Class system in trap and elute nanoflow LC mode. A Waters nanoEase M/Z Symmetry C18 100 Å, 5 µm, 180 µm x 20 mm trap column was used in combination with a Phenomenex Biozen Peptide XB-C18 100 Å, 2.6 µm, 75 µm x 25 cm nanoLC column. Injection volumes of 1-10 µL sample were loaded on the trap from a 20 µL loop using 4 minutes of loading at 10 µL/min of 0.1% formic acid in water. A 60-minute gradient at 300 nL/min from 1-26% mobile phase B was run for the separation, using 0.1% formic acid in water as mobile phase A and 0.1% formic acid in acetonitrile as mobile phase B. The column and trap were washed at 80% mobile phase B for 5 minutes and re-equilibrated at 1% mobile phase B for 25 minutes. The column temperature was maintained at 50°C.

Mass spectrometry: The ZenoTOF 7600 system was used with an OptiFlow Turbo V ion source in nanoflow mode, using the OptiFlow nanoflow interface. Source parameters used are listed in Table 1. EAD-DDA data, unless otherwise specified, were acquired using the parameters listed in Table 2. Replicate injections were performed for each sample as indicated.

Table 1. Source and gas parameters.

Parameter	Value
Polarity	Positive
Ion spray voltage	3,000 V
Nebulizing gas 1	10 psi
Curtain gas	35 psi
CAD gas	7
Interface heater temperature	100 °C

Table 2. MS method parameters for EAD-DDA acquisition.

Parameter	Value TOF MS	Value TOF MS/MS
Mass range	400 – 2,000 m/z	100 – 2,000 m/z
Accumulation time	0.1 s	0.2 s
Declustering potential	80 V	80 V
Collision energy	10 V	15 V
Time bins to sum	8	8
Electron KE	n/a	7 eV
ETC	n/a	Dynamic
Electron beam current	n/a	5,500 nA
Reaction time	n/a	10 ms
EAD RF	n/a	150 Da
Zeno pulsing	n/a	on
Zeno threshold	n/a	100,000
Q1 resolution	n/a	Low
Maximum candidate ions	9	
Scan time	2.0 s	
Intensity threshold	100 cps	
Dynamic background subtraction	Off	
Exclude former candidate ions	For 9 s after 2 occurrences	
Charge state	2 to 8	
Isotope to select	Monoisotopic	

Table 3. PEAKS GlycanFinder software search parameters.

Parameter	Value
Error tolerance	Precursor 10 ppm; Fragment ion 0.05 Da; Glycan fragment 20 ppm
Digest mode	Specific
Max. missed cleavages	2
PTMs fixed	Carbamidomethylation
PTMs variable	Deamidation (NQ), Oxidation (HWM)
Max. variable PTMs	2
Peptide length	6 to 45
Isotopic shift	-1 to 3
Glycan score FDR	1 %
Unique peptides	≥1
Protein -10LgP	≥15

Data processing: Glycopeptide identification was done using PEAKS GlycanFinder 2.0 software (Bioinformatics Solutions Inc).[4] This software identifies peptides and glycopeptides as described in Figure 2. A UniProt reviewed human plasma protein database downloaded on April 2, 2024 was used, in combination with the structural glycan database included in PEAKS GlycanFinder software consisting of 1,867 N-linked glycan structures. Search parameters used, unless otherwise specified, are listed in Table 3. Additional quantitative data processing was done using SCIEX OS Analytics software.
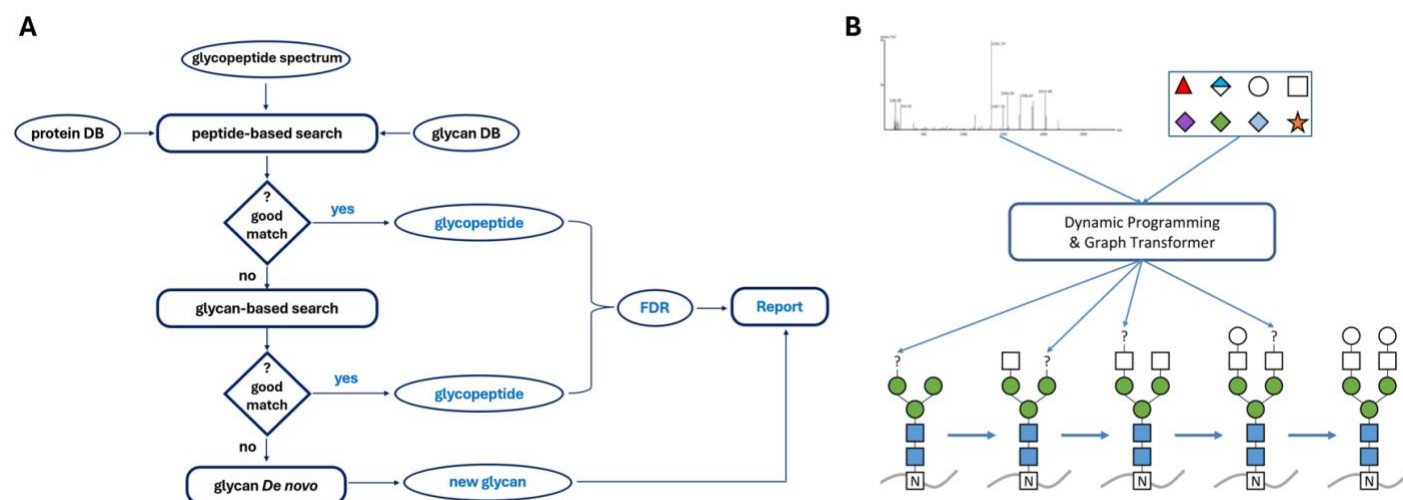
A

B

Figure 2. Workflow for glycopeptide database search and de novo sequencing with PEAKS GlycanFinder software. (A) Glycopeptide spectra can be processed in three stages: peptide-based search, glycan-based search, and de novo sequencing. If spectra cannot be identified in one stage, it proceeds to the next stage. Once a candidate glycopeptide is identified, the false discovery rates (FDRs) at both peptide and glycan levels are calculated; (B) An N-linked glycan tree is constructed from the N-linked core by iteratively adding monosaccharides to the tree. At each iteration, a dynamic programming algorithm coupled with a Graph Transformer neural network is used to predict the next monosaccharides based on the input spectrum and the partial tree obtained from the previous iteration.

Table 4: Optimization of maximum number of candidates for the EAD-DDA method. The total scan time was kept constant at 2 seconds. The amount injected was the equivalent of 37.5 µg depleted plasma.

Method	# MS/MS spectra	# GlycoPSMs *	# Glycopeptides
50 ms - 36 candidates			
Replicate injection 1	45,760	1,048	491
Replicate injection 2	46,385	970	449
100 ms - 18 candidates			
Replicate injection 1	26,625	1,400	569
Replicate injection 2	26,811	1,457	574
200 ms - 9 candidates			
Replicate injection 1	14,589	1,949	710
Replicate injection 2	14,563	1,816	661

* Glycopeptide spectral match.

Optimization of EAD parameters and amount of sample loaded on-column

The EAD method parameters were mainly as described in a previous technical note on the analysis of glycopeptides using EAD.[5] The maximum number of candidate precursor ions and MS/MS accumulation times were optimized for the highest number of glycopeptides identified, keeping the scan time constant at 2.0 s to ensure enough data points across the precursor peaks. The results are summarized in Table 4. While many more spectra were acquired using the methods that allowed for higher numbers of candidates, the best results were achieved using a combination of the highest accumulation time and lowest maximum number of candidate ions. In addition, due to the data files having fewer overall numbers of spectra to be processed, data analysis using PEAKS GlycanFinder software was also approximately 3 times faster for the 9 precursor candidates method (20 minutes instead of 54 minutes for the 36 candidates method). Sample load was optimized by injecting varying volumes of glycopeptide-enriched sample.
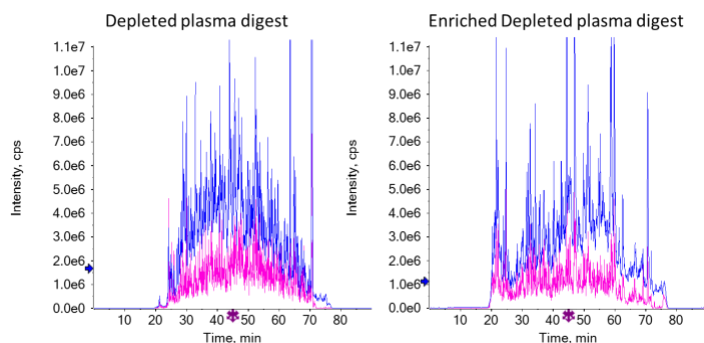
Figure 3. Total ion chromatograms (TICs) of TOF MS (blue) and MS/MS (pink). The panel on the left shows the depleted sample (the equivalent of 0.5 μg protein on-column), while the panel on the right shows the enriched depleted sample (the equivalent of 37.5 μg protein on-column).

Injections of sample above the amount equivalent to 37.5 μg protein before digestion and enrichment did not result in additional glycopeptide identifications (data not shown).

Effect of glycopeptide enrichment

Samples taken before enrichment were analyzed using the same EAD-DDA method to see the effect of the glycopeptide enrichment on the HILIC-cotton columns. An amount equivalent to 0.5 μg protein before digestion was
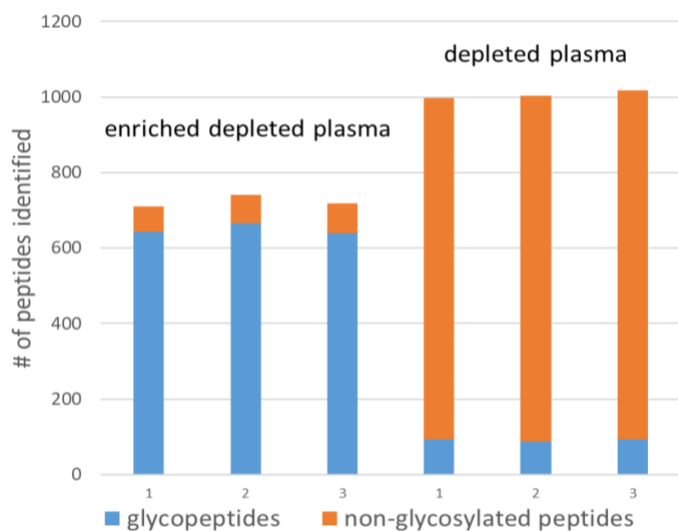
Figure 4. Fraction of identified glycopeptides versus non-glycosylated peptides in the enriched (left) and non-enriched depleted plasma samples.

compared with the equivalent of 37.5 μg protein before digestion and enrichment. The intensities of the XICs of the TOF MS scans for these injections were similar, indicating similar amounts of overall peptides injected. Figure 3 shows a comparison of the total ion chromatograms (TICs) of both samples. Figure 4 shows that the enrichment worked efficiently; approximately 90% of all identified peptides in the enriched sample were glycopeptides, whereas in the non-enriched sample, this was only about 10%.

Examples of glycopeptide identifications in GlycanFinder software

As an example, Figure 1 shows the processed spectrum of a glycopeptide from Apolipoprotein B-100. This particular N-glycosylation at position 1523 with $(HexNAc)_4(Hex)_5(NeuAc)_1$ has been previously described in the literature.[6] A series of c- and z-ion fragments fully confirmed the peptide sequence, while several peptide-glycan fragments confirmed the glycan identification. Figure 5 shows the raw spectrum of Figure 1, illustrating the depth of information generated by EAD fragmentation, with fragments from both the peptide backbone and glycan. Figure 6A shows the processed spectrum of a glycopeptide from the approximately 20x lower abundant protein Kallikrein B, with N-glycosylation at position 308.[7,8] Glycosylation at this site has been previously described.[9] As before, abundant c- and z-ion fragments allow for sequencing of the peptide backbone. Figure 6B shows the CID spectrum of the same peptide. While some glycan fragment ions are seen, CID fragmentation typically generates b- and y-ion fragments instead of c- and z-ion fragments. Many of the b- and y-ions generated have lower intensity and yield only partial sequence coverage for this peptide backbone. In addition, although high-intensity oxonium ions are generated with CID MS/MS that can identify specific monosaccharide species present, the higher-energy nature of CID fragmentation often precludes observation of the complex branched glycan structures. EAD fragmentation enables fuller characterization of the glycans due to the resulting fragments retaining these complex branched structures.
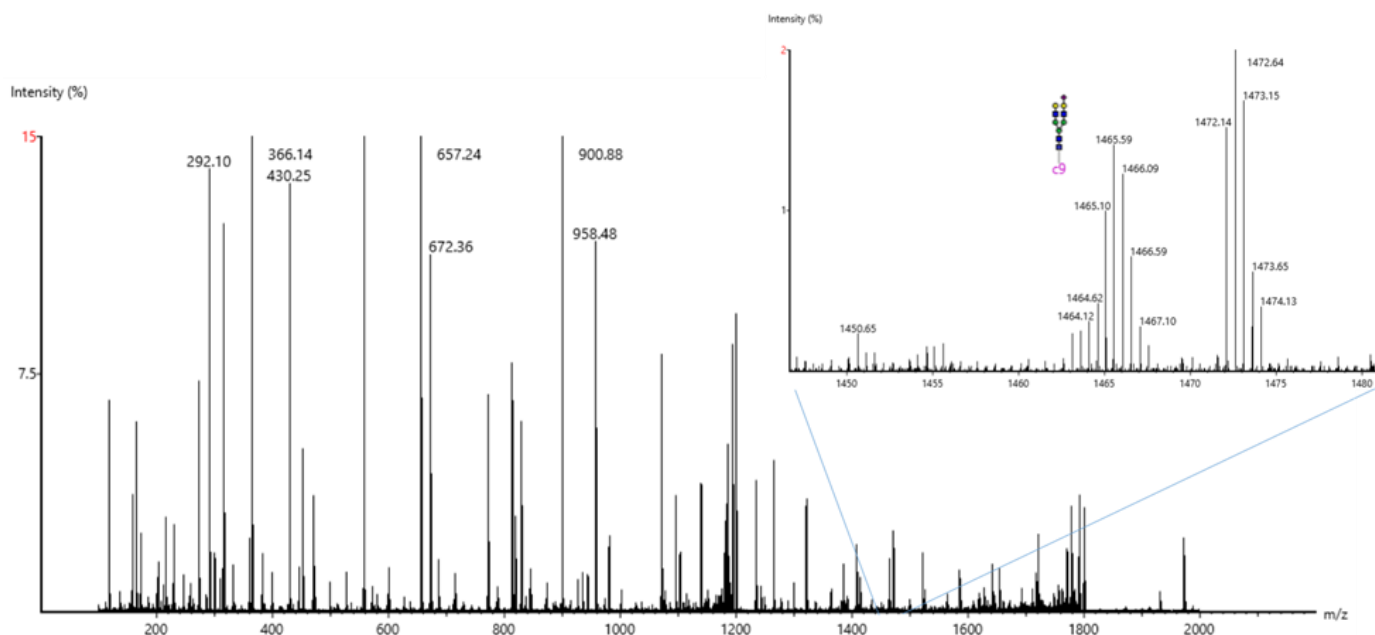
Figure 5. Raw MS/MS spectrum from Figure 1. EAD fragmentation provides information-rich spectra for glycopeptides, with fragments of both the peptide backbone and glycan. As an example, the inset shows the isotope cluster for the +2 charge state of the c9-glycan fragment.
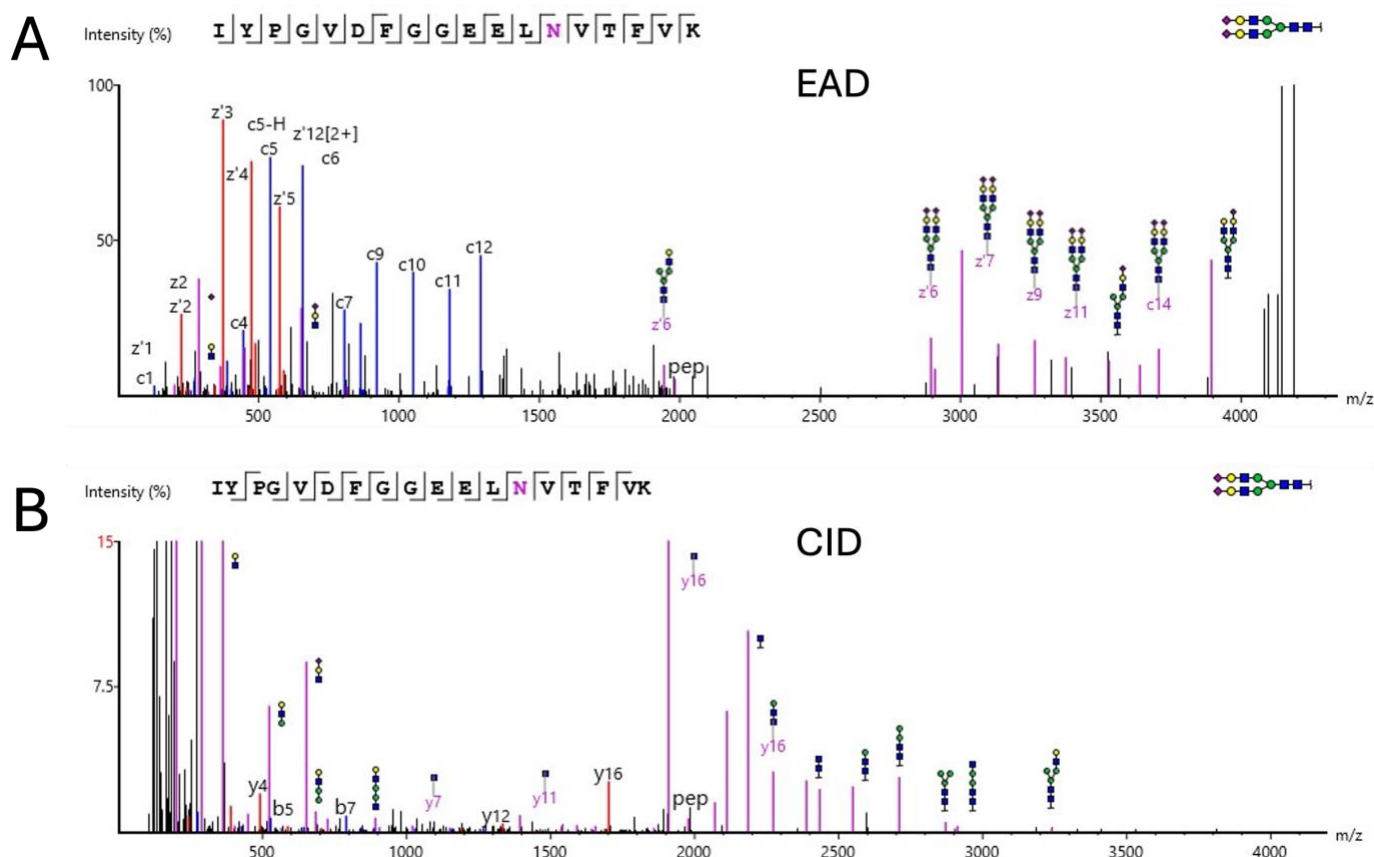

Figure 6. Comparison of EAD versus CID MS/MS spectra of a glycopeptide from Kallikrein B. (A) EAD spectrum, with the peptide backbone fully sequenced with c- and z-ion fragments, while several peptide-glycan fragments are evidence for the proposed glycan structure. (B) CID spectrum of the same glycopeptide, which, while showing some evidence of the glycan, lacks many peptide backbone fragments. Spectra were processed and annotated using PEAKS GlycanFinder software.

Table 5. Result statistics of data as presented by PEAKS GlycanFinder software. Results are from three replicate injections of the equivalent of 37.5 µg of plasma protein before enrichment.

	#Scans					#Features	Identified				#Glycopeptides	#Peptides	#Sequences	#Proteins		
	MS1	MS/MS	#Refined	#Glycan	#Chimera		#GlycoPSMs	#PSMs	#Scans	#Features				Groups	All	Top
Total	42481	42895	108407	101196	39025	517249	5160	353	5433	2363	1002	95	297	98	104	101
Sample 1	14322	14229	35775	33541	12931	171209	1678	105	1752	779	642	69	251	95	100	98
Sample 2	14081	14321	36213	33681	13033	173701	1803	123	1900	815	666	74	254	93	99	96
Sample 3	14078	14345	36419	33974	13061	172339	1679	125	1781	769	639	79	249	91	96	94

Glycopeptide identifications in enriched depleted plasma digests

Table 5 shows the results summary reported by PEAKS GlycanFinder software for three replicate injections of the equivalent of 37.5 µg of plasma protein before enrichment. More than 1,000 glycopeptides were identified in this data set; 57% of these were identified in at least two of the three replicates. An additional 25 glycopeptides were found from spectra with matching peptide and glycan fragments but with no matching precursor m/z using the fixed and modified PTMs used for this search. Ninety-five non-glycosylated peptides were reported. A total of 98 protein groups were identified.

Figure 7. Cumulative % of glycopeptides quantified by Analytics based on the TOF MS data. Data was analyzed using SCIEX OS Analytics software with the Autopeak integration method. The monoisotopic m/z was used, with a precursor ion XIC width of 0.02 m/z. The graph indicates that >90% of the quantified precursor ions had CVs of <20%.

Quantitation of identified glycopeptides using SCIEX OS Analytics software

PEAKS GlycanFinder software reports peak areas for identified glycopeptides based on MS1 features. If more control over integration parameters is needed, the Analytics quantitation module within SCIEX OS software can alternatively be used. A list of the identified glycopeptides with retention times and (monoisotopic) m/z values was exported from PEAKS GlycanFinder software. Six glycopeptides were removed from this analysis; PEAKS GlycanFinder software analysis indicated they had the same peptide sequence, monosaccharide composition, retention times and precursor m/z values, but differ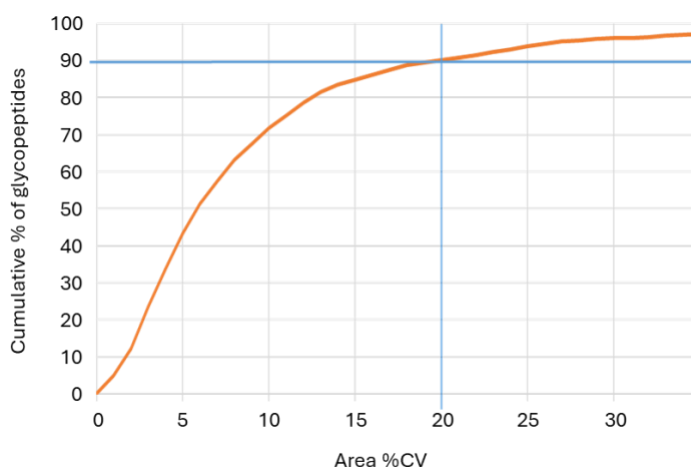ent glycan structures. The remaining 996 peptides were imported into SCIEX OS Analytics software and quantitated with the AutoPeak integration method using the acquired TOF MS data. A total of 987 of the imported glycopeptides were quantified. Figure 7 shows the cumulative fraction of peptides quantified as a function of % coefficient of variation (CV). >90% of the identified glycopeptides were quantified with a CV ≤20%. Figure 8 shows how the quantitation can be further improved by summing multiple isotopes in cases where quantitation using only the monoisotopic m/z precursor peak cannot be performed satisfactorily. It is likely that a targeted MRM method (i.e. MRM^HR) on the ZenoTOF 7600 system or on the SCIEX Triple Quad 7500 System may also provide more sensitivity and selectivity for quantitation.
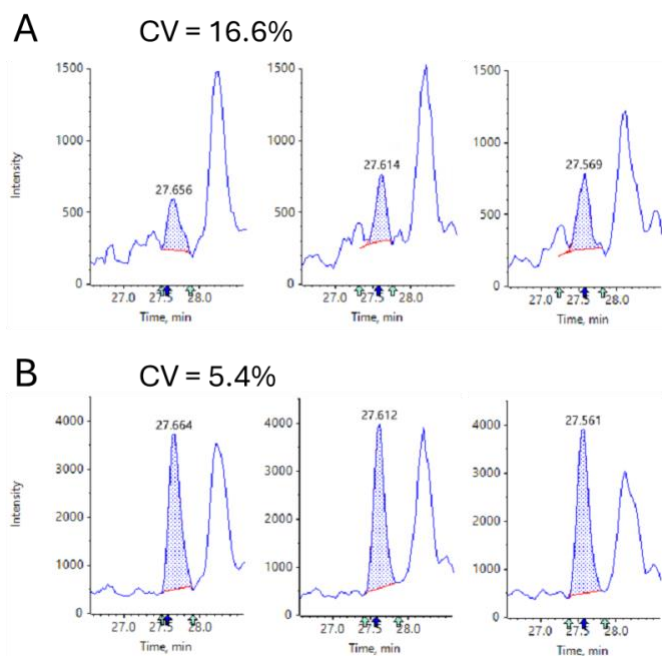
Figure 8. TOF MS XICs for the glycopeptide GIC(+57.02)N(+3007.06)SSDVR from three replicate runs. Data was analyzed using SCIEX OS Analytics software with the Autopeak integration method, with an XIC width of 0.02 m/z. (A) Peak integration using only the monoisotopic precursor ion m/z. (B) Peak integration using the sum of the first 4 precursor ion isotopes. The CV across triplicate runs improved from 16.6% to 5.4% using peak area summing

Conclusions

- EAD allows for the comprehensive structural characterization and localization of the glycan PTM for glycopeptides

- PEAKS GlycanFinder software uses information from EAD spectra for the identification of the glycopeptide, including confirmation and localization of the glycan structure

- >1,000 glycopeptides were identified in a depleted, digested human plasma sample that was enriched for glycosylated peptides

- SCIEX OS software allows for easy setup of data-dependent acquisition (DDA) EAD MS/MS methods, and the Analytics module can be used for highly accurate and precise analyte quantitation

References

1. Analysis of post-translational modifications using fast electron-activated dissociation (EAD). SCIEX Technical Note RUO-MKT-02-14795-A.

2. Chen, Z., et.al. (2021). In-depth site-specific analysis of N-glycoproteome in human cerebrospinal fluid and glycosylation landscape changes in Alzheimer's disease. Mol & Cell. Proteomics, 20:100081.

3. Wang, D., et al. (2022). Boost-DiLeu: enhanced isobaric N,N-dimethyl leucine tagging strategy for comprehensive quantitative glycoproteomic analysis. Anal. Chem. 94:11773-11782.

4. Shan, B., et al. (2023). Glycopeptide database search and de novo sequencing with PEAKS GlycanFinder enable highly sensitive glycoproteomics. Nat. Comm. 14:4046.

5. A new electron activated dissociation (EAD) approach for comprehensive glycopeptide analysis of therapeutic proteins. (2021). SCIEX technical note, RUO-MKT-02-12980-A.

6. Harazono, A., et al. (2005). Site-specific glycosylation analysis of human apolipoprotein B100 using LC/ESI MS/MS. Glycobiology 15:447-462.

7. Proteins detected in human plasma by mass spectrometry. https://www.proteinatlas.org/humanproteome/blood+protein/proteins+detected+in+ms%20from%20v23.0.%20proteinatlas.org.

8. Uhlen, M., et al. (2010). Towards a knowledge-based Human Protein Atlas. Nat. Biotech. 28:1248-1250.

9. Liu, T., et al. (2005). Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. J. Proteome Res. 4:2070-2080.