GENProtocols

Protocol

How to Determine Glycan Profiles of Biopharmaceuticals from Peptide Mapping Data

OMICS DRUG DISCOVERY

Stuart Pengelley, Bruker Daltonics GmbH & Co. KG
Eckhard Belau, Bruker Daltonics GmbH & Co. KG
Waltraud Evers, Bruker Daltonics GmbH & Co. KG
Anja Resemann, Bruker Daltonics GmbH & Co. KG
Detlev Suckau, Bruker Daltonics GmbH & Co. KG

Jul 27, 2021

Abstract

Glycosylation is a common critical quality attribute (CQA) of therapeutic proteins and needs to be characterized during product development. Typically, this analysis is conducted after enzymatic release and tagging of the glycans followed by fluorescence and MS detection. However, that approach loses information about the glycan position which is important for complex biologics such as fusion proteins; and it requires dedicated method setups and experience far beyond the classical peptide mapping analysis.

Here we describe a workflow to identify glycan compositions directly from tryptic peptide maps [1] acquired with the timsTOF Pro and PASEF® in combination with analytical LC and the new VIP-HESI ion source. It employs a glycan search method from peptide mapping data and uses previously identified aglycons as mass tags. Using this setup, we identified 36 different glycan compositions from NISTmAb tryptic digest which correlate well with those reported in the literature [2]. The results described here were generated directly from peptide mapping data without the need for a dedicated glycan laboratory workflow, thereby retaining the information about which peptide is glycosylated.

Introduction

Characterization of therapeutic proteins such as monoclonal antibodies requires a broad range of analyses including the full confirmation of the protein sequence as well as the detection and identification of post-translational modifications such as protein glycosylation. A common workflow to confirm the entire protein sequence is the peptide mapping approach, which combines pro-teolytic digests with RP-LC-MS/MS analysis in which, e.g., tryptic peptides are separated by reversed phase chromatography and further analyzed using high-resolution tandem mass spectrometry. N-linked glycans are subsequently analyzed in a second experiment comprising enzymatic glycan release, chemical labeling, LC separation on graphitized carbon or HILIC columns and MS analysis [2].

Glycopeptide analysis is typically applied in glycoproteomic experiments to identify the protein as well as the glycan using a specific fragmentation involving a collision energy stepping method [1]. The resulting glycopeptide MS/MS spectra provide both glycan and peptide fragments and allow to assign the peptide sequence as well as the corresponding glycan structure in one spectrum. Glycopeptides from therapeutical antibodies like NISTmAb, however, are derived from a tryptic peptide of known sequence and therefore can be assigned with reasonable confidence based on accurate mass alone.

When applied to glycopeptides, standard PASEF conditions used for peptide mapping preferentially cleave the glycosidic bond

between the carbohydrate units, and subsequently peptide fragments are not observed. Such spectra are dominated by y-type glycan fragments attached to the intact peptide and lower molecular weight b-type glycan fragments (without peptide moiety – i.e., the aglycon - attached). Similar fragmentation patterns are observed in CID spectra from released glycans carrying a fluorescence label like RapiFluor at the reducing end (Figure 1). In the approach reported here, the aglycon of a glycopeptide is treated in a glycan database search like a mass tag at the reducing end. The peptide tag serves the same purpose for the ionization of the glycan as fluorescent labels like RapiFluor in the analysis of labelled glycans – the peptide has an even stronger ionization propensity providing a high analysis sensitivity.

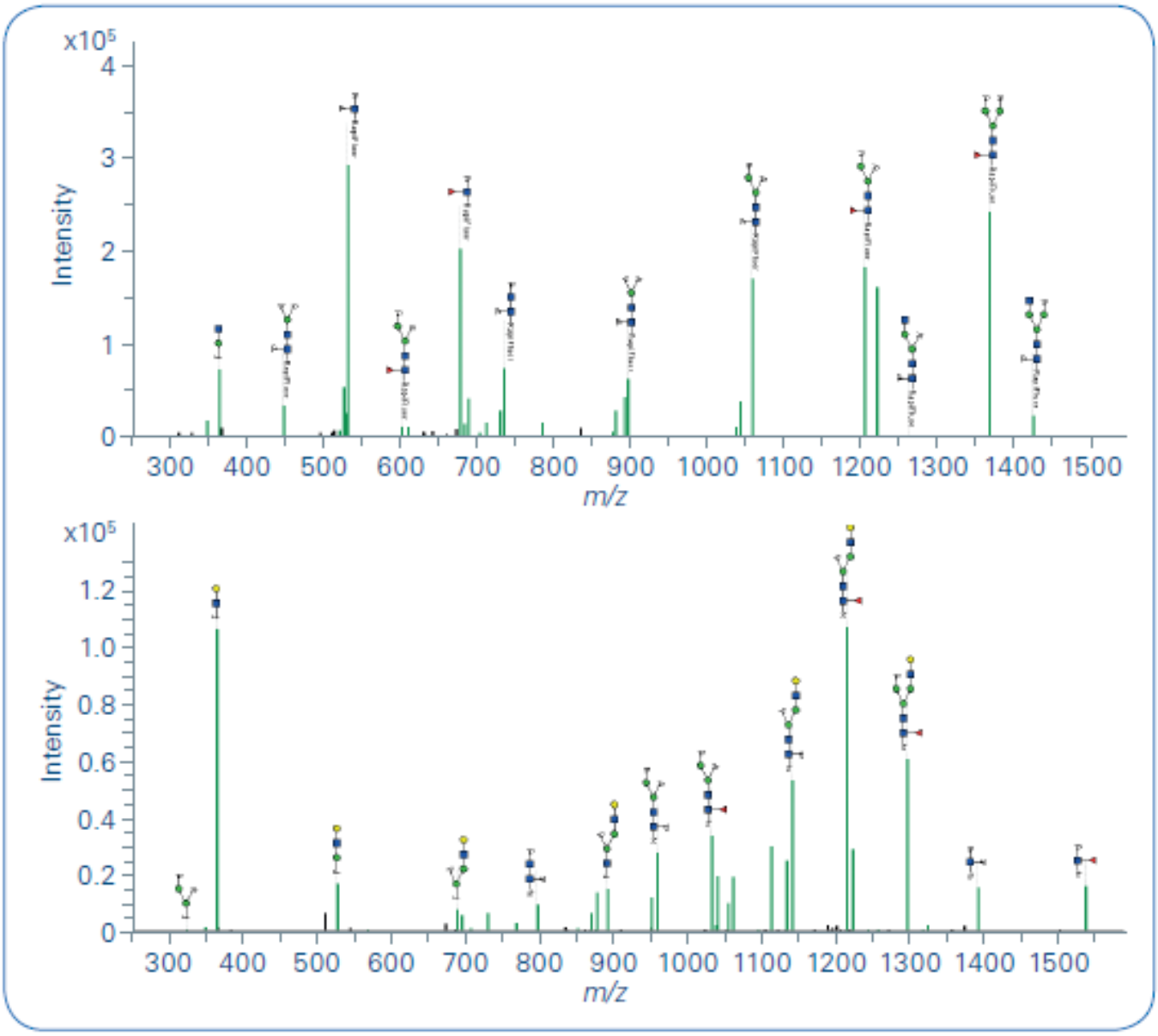

Figure 1. CID MS/MS spectra of G1F glycan from NISTmAb both dominated by y-ions. Top: G1F with RapiFluor label (plus 311.175 Da) attached to the reducing end. Bottom: G1F with the peptide EEQYNSTYR (plus 1670.47 Da) assigned as "fc" attached to the reducing end.

The timsTOF Pro with PASEF [3] further increases the capability to acquire high quality MS/MS spectra from glycopeptides with high dynamic range despite the almost non-existent reverse phase separation of the glycopeptides with identical aglycon. In this work we evaluated the identification characteristics of the glycopeptide approach on the timsTOF Pro linked to the VIP-HESI ion source (Vacuum Insulated Probe – Heated ESI) – a heated ion source optimized for applications requiring high sensitivity.

Reagents and Equipment

Mass spectrometer: timsTOF Pro mass spectrometer via a VIP-HESI ion source

Procedure

Sample Preparation

NIST Monoclonal Antibody Material 8671 (Merck) was reduced using DTT and TFE (25%) for 150 min at 56 °C and alkylated using IAA. Trypsin (Promega) digestion was performed overnight.

LC-MS Data Acquisition

Twenty µg tryptic digest were separated in a 60 min gradient on an Acquity CSH C18 2.1 x 100 mm 1.7 µm column (Waters) using an Elute UHPLC (Bruker). The UHPLC was interfaced with a timsTOF Pro mass spectrometer via a VIP-HESI ion source and peptides were analyzed by PASEF using the standard proteomics acquisition method, adapted by lowering the precursor intensity threshold to 400 counts.

Data Analysis

The raw data were initially processed using BioPharma Compass® 2021b (Bruker) with the Peptide Mapping method Tutorial NIST_mAb with small modifications resulting in classical peptide maps with comprehensive sequence coverage (not shown). Typically, the Fc-glycopeptides are eluting in a narrow retention time range early in the gradient. To reduce the number of MS/MS spectra submitted to the glycan search and computing time, the Rt range for data computing was limited to the glycopeptide elution range (5-15 min, Figure 4).

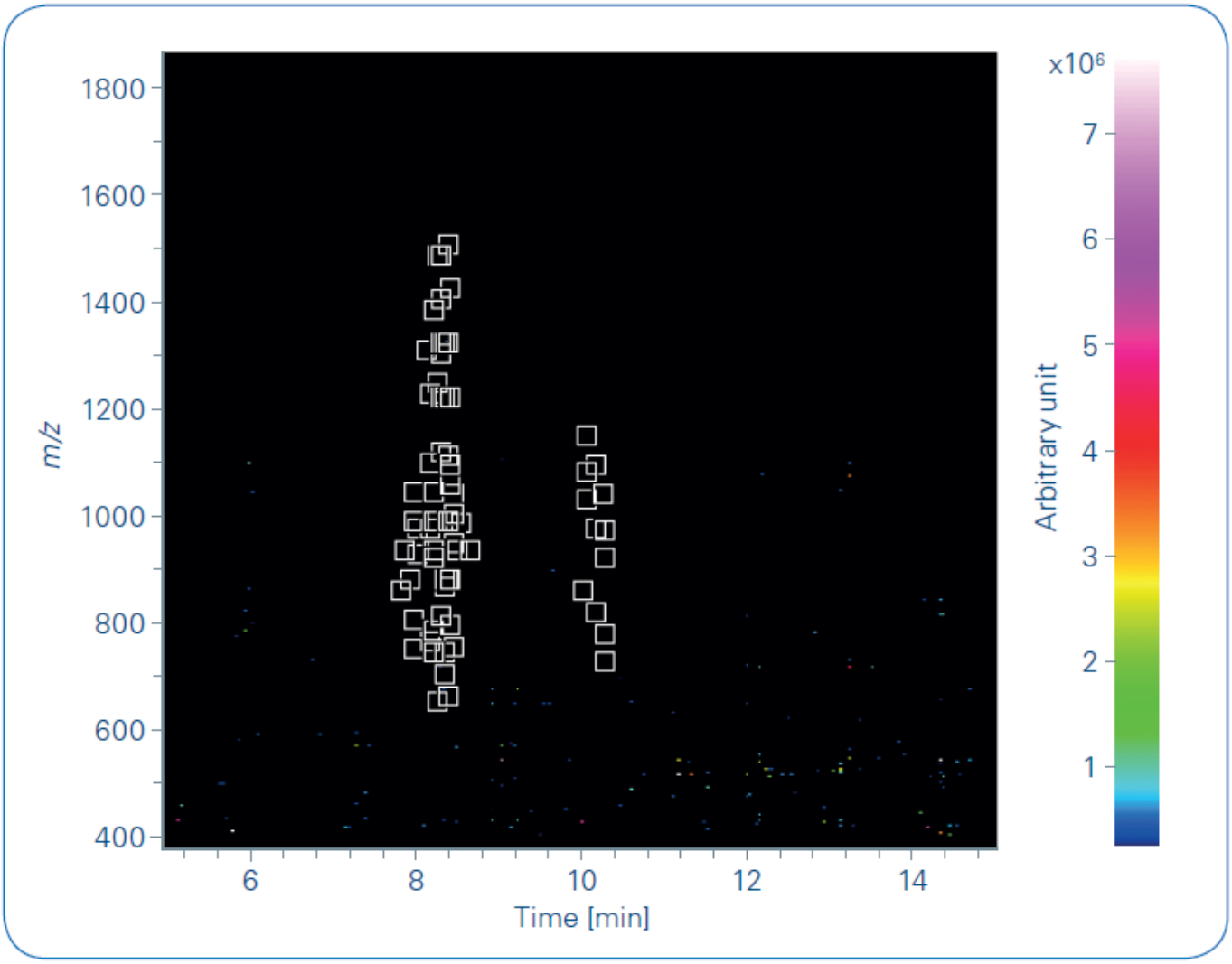

Figure 4. Survey view of identified glycopeptides at two different retention time ranges.

In a second processing step, all MS/MS spectra in the Rt range were submitted to a glycan search using the GlycoQuest search engine included within BioPharma Compass.

The Fragmentation Type CID byi4Cl was initially defined in "Admin Preferences": Protocols/ProteinScape/Glycomics/GlycoFragmentationType. Only b-, y- and internal ions are selected. Maximum cleavages are set to 4, max crosslinks to 0.

The tutorial glycan search method N-glycan QTOF CID RapiFluor was adapted subsequently to the analysis of the previously known/established tryptic NISTmAb Fc glycopeptide EEQYNSTYR using the following parameters (Figure 2):

1. Use GlycO as database as it is a rather condensed database containing all relevant glycans expected on a humanized IgG1
2. Define the Reducing end: Reducing end mass: 1170.494166 (peptide mass of EEQYNSTYR -H2O was calculated using the Sequence Editor in BioPharma Compass) Reducing end name "fc" for Fc tryptic peptide
3. Select Fragmentation Type CID byi4Cl
4. Thresholds for result compilation were stringent to reduce the glycan list for confident search results: Score > 40, Fragmentation coverage [%] > 40, Intensity coverage [%] > 40

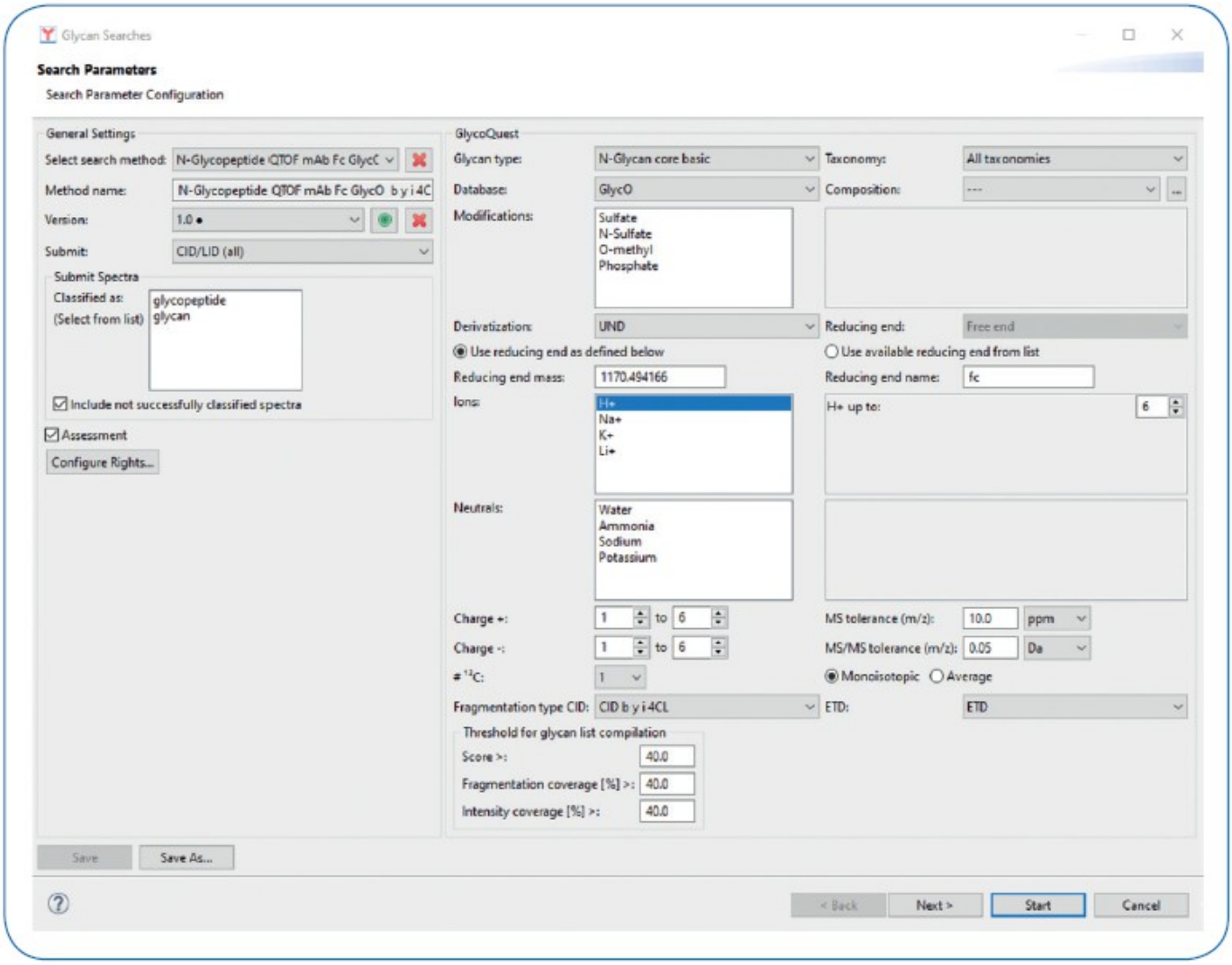

Figure 2. Glycan search method adapted for NISTmAb tryptic digest glycopeptide analysis. The monoisotopic reducing end mass is calculated from the tryptic glycopeptide sequence EEQYNSTYR. A new "Fragmentation type CID" is created, and high thresholds are used to reduce the number of false positives drastically listing only glycan compositions identified with high confidence. Especially the Intensity coverage value is important for the reliable identification of low abundant glycans.

Results

The glycan search of the MS/MS spectra yielded 36 specific glycan compositions of different glycan classes like complex, hybrid, and high mannose structures (Figure 3).

Info | Glycans & Fragments

Row		Composition	Int.	Score	IntCov. [%]	FragCov. [%]	Rt [min]	m/z meas.	z	Mr calc.	Accession
1	☑	Hex3HexNAc4dHex1-fc	161013	89.6	94	86	8.38	878.6905	3	2633.0386	2620
2	☑	Hex4HexNAc4dHex1-fc	149800	85.3	96	76	8.30	1398.5562	2	2795.0914	13972
3	☑	Hex4HexNAc3dHex1-fc	27556	68.5	90	52	8.30	1297.0142	2	2592.0121	13971
4	☑	Hex3HexNAc3dHex1-fc	24036	83.4	92	76	8.30	1215.9866	2	2429.9592	14458
5	☑	Hex5HexNAc2-fc	21720	50.8	46	56	7.97	802.6506	3	2404.9276	234
6	☑	Hex5HexNAc4dHex1-fc	20358	81.9	92	73	8.22	986.7229	3	2957.1442	13970
7	☑	Hex4HexNAc3NeuGc1dHex1-fc	19296	75.9	88	65	10.29	967.3748	3	2899.1024	38160
8	☑	Hex6HexNAc4dHex1-fc	16997	72.3	85	61	8.22	1040.7406	3	3119.1971	13422
9	☑	Hex3HexNAc2dHex1-fc	9804	74.4	91	61	8.30	1114.4468	2	2226.8799	1643
10	☑	Hex3HexNAc3-fc	7498	67.9	72	64	8.22	762.3089	3	2283.9013	1865
11	☑	Hex7HexNAc4dHex1-fc	7076	69.7	84	57	8.18	1094.7607	3	3281.2499	13423
12	☑	Hex5HexNAc4NeuGc1dHex1-fc	6800	66.8	85	52	10.17	1089.0838	3	3264.2346	13922
13	☑	Hex5HexNAc3dHex1-fc	6316	75.0	87	65	8.22	919.0319	3	2754.0649	14736
14	☑	Hex5HexNAc3dHex1-fc	6316	78.1	92	66	8.22	919.0319	3	2754.0649	1248
15	☑	Hex4HexNAc2-fc	5789	66.8	59	76	7.97	748.6333	3	2242.8748	2827
16	☑	Hex5HexNAc4NeuGc2-fc	5564	54.3	73	40	10.05	1143.1029	3	3426.2757	13325
17	☑	Hex4HexNAc4NeuGc1dHex1-fc	4869	78.4	86	72	10.25	1035.0672	3	3102.1818	13924
18	☑	Hex3HexNAc5dHex1-fc	4835	64.6	57	73	8.46	946.3814	3	2836.1180	1549
19	☑	Hex4HexNAc5dHex1-fc	4667	52.1	49	56	8.46	1000.3991	3	2998.1708	13904
20	☑	Hex6HexNAc5dHex1-fc	3290	56.3	55	58	8.38	1108.4347	3	3322.2764	13909
21	☑	Hex5HexNAc5dHex1-fc	3283	67.6	77	59	8.42	1054.4152	3	3160.2236	13392
22	☑	Hex6HexNAc4NeuGc1dHex1-fc	2963	62.9	75	53	10.01	857.5796	4	3426.2874	13941
23	☑	Hex6HexNAc3-fc	2484	71.5	84	61	8.01	924.3597	3	2770.0598	1710
24	☑	Hex6HexNAc3NeuGc1-fc	2235	67.8	82	56	10.05	1026.7246	3	3077.1501	38099
25	☑	Hex3HexNAc4-fc	2043	64.5	55	76	8.26	1244.4968	2	2486.9807	1857
26	☑	Hex4HexNAc3-fc	1998	64.5	68	61	8.18	1223.9848	2	2445.9541	13944
27	☑	Hex6HexNAc3NeuGc1dHex1-fc	1512	56.8	70	46	10.05	1075.4088	3	3223.2080	38101
28	☑	Hex6HexNAc3dHex1-fc	1496	54.3	73	40	8.18	973.0463	3	2916.1177	1247
29	☑	Hex6HexNAc2-fc	1280	69.2	77	62	7.81	856.6681	3	2566.9804	6296
30	☑	Hex3HexNAc2-fc	1185	70.4	88	57	8.22	1041.4188	2	2080.8219	773
31	☑	Hex5HexNAc3NeuGc1-fc	1133	65.6	75	58	10.17	972.7046	3	2915.0973	38094
32	☑	Hex5HexNAc4dHex2-fc	1085	43.9	45	43	8.46	1035.4107	3	3103.2022	10137
33	☑	Hex6HexNAc4dHex2-fc	1062	45.0	48	43	8.42	1089.4280	3	3265.2550	14487
34	☑	Hex4HexNAc4dHex2-fc	971	63.7	67	61	8.54	981.3913	3	2941.1493	13206
35	☑	Hex4HexNAc3NeuGc1-fc	743	59.4	76	46	10.29	918.6874	3	2753.0445	38143
36	☑	Hex5HexNAc3-fc	574	51.4	48	55	8.14	1305.0067	2	2608.0070	13711

Figure 3. Result table of glycan compositions identified from NISTmAb tryptic glycopeptides sorted by absolute intensity (Int.). Hex=Hexose; HexNAc=N-Acetyl hexosamine; dHex=desoxyhexose; NeuGc= N-glycolylneuraminic acid; fc=reducing end mass of the tryptic Fc peptide EEQYNSTYR. Besides the Score from the glycan search software GlycoQuest, intensity coverage (IntCov.[%]) and fragmentation coverage (FragCov.[%]) are listed showing the high confidence level of the search results.

The glycopeptides were detected at 2 different Rt ranges: glycopeptides with neutral glycans and those with acidic glycan units (N-glycolyl-neuraminic acid) eluted after approx. 8 and 10 min, respectively (Figure 4).

The most intense glycan structure (G0F, h3n4f1) was detected with an absolute intensity of 161x103, lower abundant compositions were identified with intensities below 2x103 (Figure 5). These were mainly non-fucosylated neutral or acidic glycan structures or doubly fucosylated glycans like G1F2 or G2F2 (h4n4f2 or h5n4f2, Figure 6).

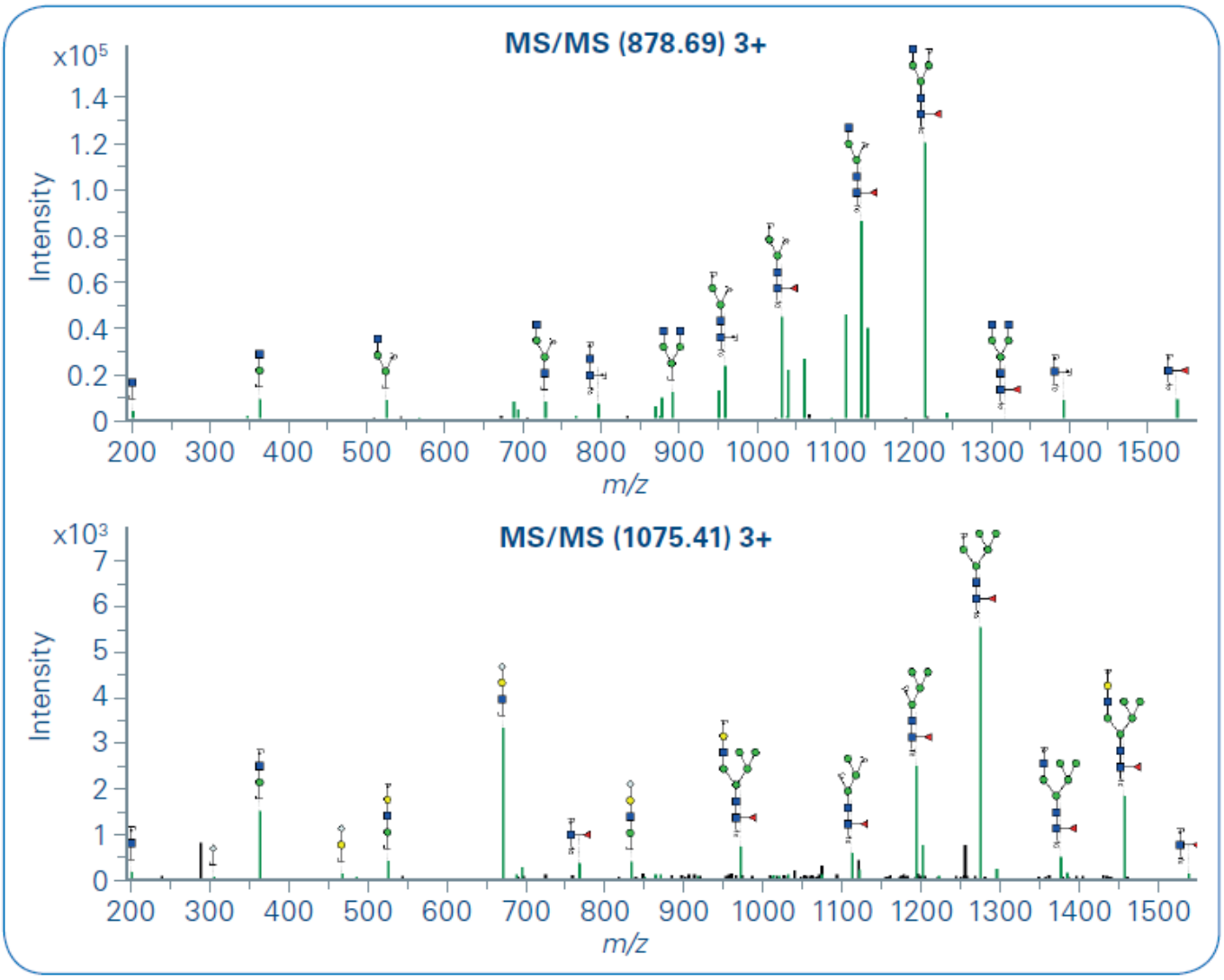

Figure 5. MS/MS spectra of two different glycan structures of different intensities with a dynamic range of 100 to 1. Top: G0F (h3n4f1; Hex3HexNAc4dHex1) attached to EEQYNSTYR with intensity of 161,013 a.u.. Bottom: Hybrid structure (h6n3f1g1, Hex6HexNAc3NeuGc1) with intensity 1,512 a.u..

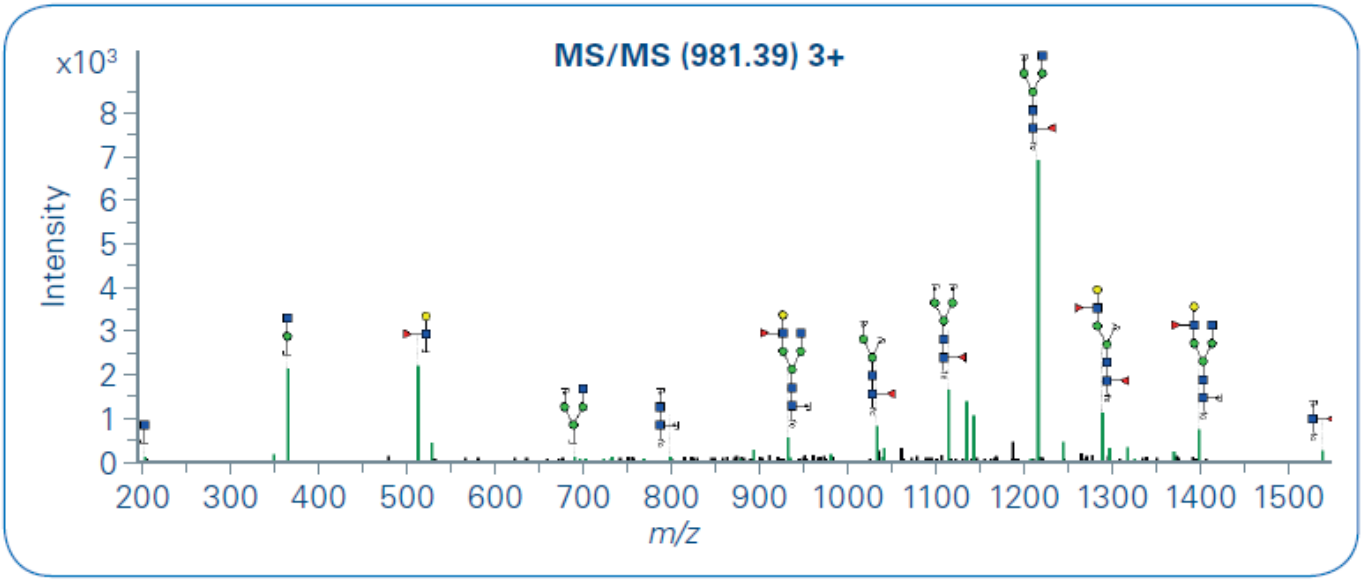

Figure 6. Additionally identified low abundant MS/MS spectrum of the di-fucosylated glycan composition h4n4f2 with intensity 971 a.u.. A score of 63.7 and intensity coverage of 67 underline the high stringency of the search result even for such a low abundant compound.

Compared to the results of a traditional analysis using released labeled glycans and HILIC separation [2], 27 out of 30 different compositions were found using our glycopeptide analysis method. The three missing compositions were h7n2, n7n3 and h8n5f1. The compositions h4n2, h4n3, h4n4f2, h5n3, h5n3g1, h5n4f2, h6n4g1 and h6n4f2 were additionally detected using the method described here (Figure 6). Differentiation of isomeric structures has not been further investigated.

The co-elution of glycopeptides always presents a challenge for LC-MS analysis, which typically results in much lower assignment rates of glycan compositions compared to labelled glycan analysis. Here, however, PASEF contributed the required high acquisition speed (> 100 Hz MS/MS) and sensitivity (ion mobility time and space focusing), whilst sensitivity and signal intensities were given a further boost by the VIP-HESI ion source, which is heated to 400 °C. The combined effect was that 36 glycan compositions were identified from a single aglycon using standard, well established peptide mapping methods, which correlates well with data published by an expert group using a dedicated released glycan approach [2].

Conclusion

- Thirty-six different glycan compositions were identified from a NISTmAb tryptic digest by dedicated glycoanalysis of a selected peptide using a conventional peptide mapping dataset and the GlycoQuest search engine in the BioPharma Compass software.
- The result compares favorably with previous studies using the established labelled-glycan approach [2]. However, a thorough isomer assessment still relies on MS/MS analysis and spectra library-based identification of labelled glycans.
- The PASEF technology was key to acquire high quality MS/MS spectra from coeluting glycopeptides covering a dynamic range of 100:1. The setup included analytical RP-LC separation and the new VIP-HESI ion source. This new approach enabled glycosylation analysis for non-experts.
- For more complex glycoproteins with multiple glycosylation sites, such as the SARS-CoV-2 S-glycoprotein [4], the approach can be repeated peptide-wise to obtain site-specific glycan composition profiles with great sensitivity and specificity.

Notes and Comments

Further reading

Bruker Application Note: Recombinant SARS-CoV-2 Receptor Binding Domain: Comprehensive Top-Down Sequence

Confirmation, Curation and O-Glycosylation Site Determination

Bruker Application Note: VIP-HESI dual source, Next generation in Mass Spectrometry sensitivity

Bruker Application Note: Fully automated glycoform profiling and sequence validation of the NIST reference antibody

References

[1] Hinneburg H, Stavenhagen K, Schweiger-Hufnagel U, Pengelley S, Jabs W, Seeberger PH, Varón Silva D, Wuhrer M, Kolarich D (2016). The art of destruction: Optimizing collision energies in Quadrupole – Time of Flight (QTOF) instruments for glycopeptide based glycoproteomics. J Am Soc Mass Spectrom, DOI: 10.1007/s13361-015-1308-6

[2] Hilliard M, Alley WR Jr, McManus CA, Yu YQ, Hallinan S, Gebler J, Rudd PM (2017). Glycan characterization of the NIST RM monoclonal antibody using a total analytical solution: From sample preparation to data analysis. mAbs, DOI: 10.1080/19420862.2017.1377381

[3] Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA, Bache N, Hoerning O, Cox J, Räther O, Mann M (2018). Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. Mol Cell Proteomics 17(12):2534-2545, doi.org/10.1074/mcp.TIR118.000900

[4] Gstöttner C, Zhang T, Resemann A, Ruben S, Pengelley S, Suckau D, Welsink T, Wuhrer M, Domínguez-Vega E (2021). Structural and Functional Characterization of SARS-CoV-2 RBD Domains Produced in Mammalian Cells. Anal Chem 93(17):6839-6847. doi: 10.1021/acs.analchem.1c00893. Epub 2021 Apr 19.