Vol. 8, No. 4, 2017
ISSN 2233-4203/ e-ISSN 2093-8950
REVIEW www.msletters.org | Mass Spectrometry Letters

MS-Based Technologies for the Study of Site-Specific Glycosylation

Unyong Kim[1,2,3], Myung Jin Oh[1,3], Jua Lee[1,3], Hee Yeon Hwang[1,3], and Hyun Joo An[1,3]*

[1]*Graduate School of Analytical Science and Technology, Chungnam National University, Daejeon, Korea*
[2]*GLYCAN Co., Ltd., Healthcare Innovation Park, 172 Dolma-ro, Bundang-gu, Seongnam 13605, Korea*
[3]*Asia-Pacific Glycomics Reference Site, Daejeon, Korea*

Received November 17, 2017; Revised December 18, 2017; Accepted December 18, 2017
First published on the web December 31, 2017; DOI: 10.5478/MSL.2017.8.4.69

Abstract : Glycosylation, which is one of the most common post-translation modification (PTMs) of proteins, plays a variety of crucial roles in many cellular events and biotherapeutics. Recent advances have led to the development of various analytical methods employing a mass spectrometry for glycomic and glycoproteomic study. However, site-specific glycosylation analysis is still a relatively new area with high potential for technologies and method development. This review will cover current MS-based workflows and technologies for site-specific mapping of glycosylation ranging from glycopeptide preparation to MS analysis. Bioinformatic tools for comprehensive analysis of glycoprotein with high-throughput manner will be also included.

Keywords : Glycosylation, Glycopeptide, Mass Spectrometry, Site-specific mapping, Bioinformatics

1. Introduction

Protein glycosylation, one of the most common post-translational modification (PTMs) of proteins,[1] is an important biological process that plays a vital role in cell signaling, cell adhesion, and the regulation of biochemical pathways.[1-3] In particular, glycosylation on a therapeutic protein affects drug's efficacy and safety by modulating wide range of therapeutic functions.[4-7] In order to better understand biological functions of specific glycoproteins including biotherapeutics, comprehensive studies to determine glycosylation site with glycan micro-heterogeneities should be performed. However, unlike other post translational modifications (PTMs) such as acetylation and phosphorylation, the analysis of protein glycosylation is more challenging due to its inherent complexity and heterogeneity derived from glycan structural diversity.[7-8]

The analysis of protein glycosylation can be achieved in several levels of detail.[8-10] In general, protein glycosylation is characterized by three different approaches; 1) glycomic approach using released glycans, 2) top down approach by intact glycoprotein analysis, and 3) glycoproteomic approach through glycopeptides analysis - individual analytical approaches provide slightly different information on protein glycosylation. The most simple and direct analysis is to explore glycans released enzymatically or chemically from a glycoprotein by mass spectrometry (MS) or HPLC coupled to fluorescence detection (HPLC/FLD).[8-9] Glycomic-based methods provide in-depth information on only glycan level including composition, structure, and relative abundance. When N-glycans are enzymatically released by PNGase F from glycopeptide, Asparagine (Asn) residue linked to a glycan is hydrolyzed to Aspartic acid (Asp) residue and glycosylation site can be determined by detecting 1 Da mass difference. Top down approach using intact glycoprotein analysis can rapidly provide whole picture of glycosylation on a glycoprotein with a minimum sample amount. However, it has several limitations; 1) low sensitivity (low abundant glycoforms are not detected), 2) low ionization efficiency owing to negatively charged glycans, and 3) difficulty of data interpretation caused by complicated MS spectrum. Finally, glycoproteomic approach using glycopeptides provides an unique platform to directly obtain glycan heterogeneities for each occupied glycosylation site in a glycoprotein although sample preparation for glycopeptides and MS analysis takes more time and efforts compared with other two methods.

Recently, mass spectrometry (MS)-based analytical platforms are considered as the most powerful approach to characterize site-specific glycosylation.[11-13] It gains more

*Reprint requests to Hyun Joo An
E-mail: hjan@cnu.ac.kr

and more interests with advanced MS technologies for glycoproteomics for understanding exact functional relevance of protein glycosylation. Various separation techniques such as capillary electrophoresis (CE)[14-16] and liquid chromatography (LC)[17-19] are commonly hyphenated to MS to comprehensively characterize glycopeptides bearing high complexity and structural diversity. Additionally, structural elucidation of glycopeptides and determination of glycosylation sites can be achieved using several tandem mass spectrometry techniques including collision-induced dissociation (CID),[18-23] electron transfer dissociation (ETD),[18,24-26] and higher energy collisional dissociation (HCD).[18,22,24-25,27]

Although recent advances from sample preparation to MS readily demonstrate the ability to determine site-specific glycosylation,[12,20,28-30] it is still a relatively new area with high potential for technologies and method development. Several groups in the last decades have developed various analytical strategies to overcome current problems for glycopeptides analysis such as poor ionization efficacy of glycopeptides and relatively low abundance in a given sample with the goal of analyzing site-specific glycosylation in a sensitive and comprehensive manner. In this review, we will discuss current MS-based workflows and technologies for site-specific mapping of glycoprotein(s) with topics covering sample preparation by proteolytic digestion and glycopeptide analysis by MS as well as their identification with tandem MS technique. In addition, we will cover the bioinformatic tools for comprehensive analysis of glycoprotein with high-throughput manner.

2. Technologies for site-specific analysis of glycoproteins

2.1 Experimental strategies to generate site-specific glycopeptides

MS-based site-specific glycosylation analyses begin with protein digestion with specific/non-specific protease, enrichment and separation of generated glycopeptides, and detection. Various strategies currently in use to characterize site-specific glycosylation were summarized in **Figure 1**. Generation of glycopeptides by protein digestion by single or cocktail protease represents the most crucial element in MS-based site-specific glycosylation mapping. The selection of protease depends on chemical and physical characteristic of target glycoprotein(s). Up to now, specific protease digestion has been predominantly performed with trypsin due to its high specificity (on the C-terminal side of Lys and Arg), easy sequence prediction *in silico*, and widespread availability.[31-32] However, trypsin, especially for glycoprotein analysis, has certain limitations, for instance, occurrence of missed cleavages caused by steric hindrance of glycans, signal suppression of glycopeptides by non-glycosylated peptides, and the presence of multiple glycosylation sites on single glycopeptide. Lately, other specific-proteases such as chymotrypsin, LysC, AspN and GluC[31-33] have been used as an alternative enzyme to complement trypsin. One interesting protease is a non-specific or multi-specific protease to digest a glycoprotein. In particular, pronase which is one of popular class of non-specific protease, has multiple specificities and activities, thus by enabling a glycoprotein to break down to several

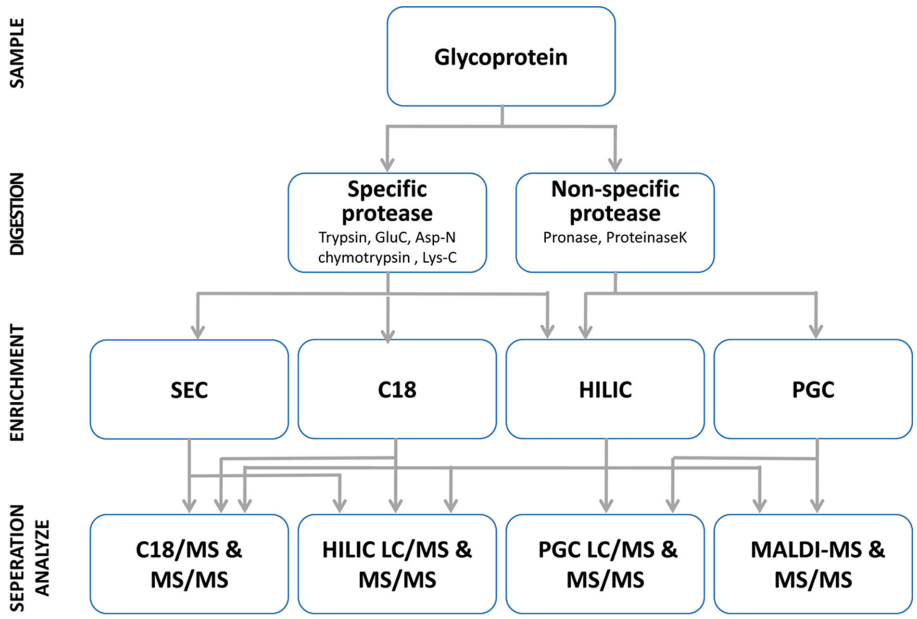

Figure 1. Experimental strategies for site-specific glycosylation analysis.

glycopeptides bearing generally-shorter peptide moieties.[34-35] Other non-specific proteases including proteinase K and subtilisin, and thermolysin were also widely used to produce informative glycopeptides suitable for site-specific glycosylation analyses.[21,35-36]

Enrichment of glycopeptides is a key element with the aim of selectively obtaining glycopeptides from complicated mixture containing non-glycopeptides, enzymes, salts, and so on prior to MS-analysis. For enhancing the MS signal of glycopeptides and ensuring data quality, non-glycopeptides must be carefully removed. Glycopeptides are less ionized than non-glycopeptides because most glycosylation sites carry a multitude of glycans giving rise to different glycoforms of glycopeptides thus glycopeptides present at a relatively minor portion in the total peptide mixture.[28] Therefore, various enrichment tools using physical and chemical properties of glycopeptides have been developed. Size exclusion chromatography (SEC) is often used for enrichment of glycopeptides,[37-38] because most tryptic glycopeptides have relatively higher molecular weight than peptide counterparts. Recently, hydrophilic interaction liquid chromatography (HILIC)-based glycopeptide enrichment has been used in popular due to high specificity for glycopeptides.[39] Reversed phase purification using C18 is one of the common approaches for tryptic glycopeptide enrichment.[40-42] In particular, porous graphitized carbon (PGC) is well applicable for enrichment of non-tryptic glycopeptides with smaller peptide portions.[34-35] In parallel, online LC separation hyphenated with MS is a powerful platform for glycopeptide analysis, allowing detection of glycosylated and non-glycosylated peptides in complex mixtures within single LC run. Chromatographic separation by reversed phase (RP) stationary phase (C18 stationary phase) is applicable for long peptide/ glycopeptide chains commonly produced by tryptic digestion.[32] While PGC is effective for isomer-specific separation of N- and O-glycopeptides modulated by both peptide and glycan structure.[34-35]

Note that conventional tryptic glycopeptides may yield less information on the glycan structure i.e. monosaccharide composition and a few topology/ branching information due to limited separating capacity by RP-LC.[28] On the contrary, pronase-digested glycopeptides with the shorter peptide moiety provides more abundant information about glycoform structures on other stationary phases including PGC.[43-44] Another separation method, HILIC stationary phases are also commonly used for site-specific glycopeptide analysis taking advantage of the hydrophilic characteristic of glycopeptides.

Glycopeptide analysis often employs MS as the final detection step. Initially, possible compositions of glycopeptides are assigned by accurate masses and their structures are further confirmed by tandem MS. LC/MS has been widely used for site-specific glycosylation mapping. Two types of LC/MS are typically preferred in site-specific glycosylation analysis,[45] namely, Orbitrap type mass spectrometry[32] and Q-TOF type hybrid mass spectrometry[34-35] providing high mass accuracy and high resolution. Matrix-assisted laser desorption/ionization (MALDI)-TOF MS and MALDI-Fourier transform-ion cyclotron resonance (FT-ICR) MS can be used for site-specific glycosylation analysis. MALDI-MS has several advantages; ease of sample preparation, relatively high tolerance to salts and other contaminants. LC/MS analysis of glycopeptides allows on-line chromatographic separation of (glyco)peptides. In contrast, MALDI analysis of glycopeptides is generally achieved by targeted enrichment techniques to enable the detection of glycopeptides with minimizing signal suppression caused by non-glycopeptides. Recently, most analysts now prefer LC hyphenated with high-resolution MS for glycopeptide analysis to obtain quantitative information and to reduce false assignments. Indeed, high mass-accuracy and isotope-resolution capabilities of high-resolution MS allow accurate detection of predicted glycopeptide masses, leading to confident assignment of composition.[46]

In order to elucidate structure of glycopeptides including sequence of peptide backbone, the glycan structure and the site of glycan attachment, tandem MS analysis is required. Depending on the configuration of the instruments, various distinct fragmentation techniques in tandem MS analysis can be used. This point will be discussed in detail in the next section named "*Glycopeptide identification by tandem MS*".

2.2 Methods for site-specific mapping: Specific proteases vs Non-specific protease

Site-specific characterization of a glycoprotein is a key to understand the functional relevance of protein glycosylation at the protein-specific as well as global level, e.g, in the context of antigenicity, pathogenesis and disease progression. Nevertheless, site-specific mapping is still significantly under-defined due to structural complexity of glycoprotein. In general, site-specific mapping workflow starts with the purification or fractionation of a single glycoprotein (or mixture of glycoproteins), followed by generating glycopeptides/non-glycopeptides using protease, subsequently. Here, we introduce representative but different strategies based on specific proteases and non-specific protease, focusing on two representative enzymes (trypsin and pronase, respectively) for the purpose of comprehensive site-specific mapping.

Trypsin is one of the most popular specific protease thanks to its specificity especially for basic amino acid residues. Employment of trypsin enables an extensive and global shotgun glycoproteomics approach through profiling of highly complex mixture of glycoproteins[47] (e.g. bio-mixtures such as complex cell extracts,[48] tissues,[49]

and bio-fluids[50-51]). Therefore, microheterogeneity (qualitative and quantitative distribution of glycans at defined glycosylation site) glycoform profiling and total glycan occupancy called as macroheterogeneity (quantitative characterization of glycans at different glycosylation sites) can be yielded from a given glycoprotein mixture.[28] And, glycoprotein database based on well-established protein database (e.g. uniprot) enables large scale study for comprehensive analysis of glycosylation signatures in a whole system. Furthermore, by utilizing tools to examine functional interactions as well as functional annotation of proteins (e.g. DAVID and STRING), functional significance can be examined including the Gene Ontology biological processes, cellular components, and functional glycoprotein interaction network.[52]

Based on the global mapping of whole glycoproteome pool, comparative glycoproteomic mapping to monitor the differences of protein glycosylation in the different biological context such as specific pathogenesis has been performed.[53-54] The comparative glycoproteomics usually allows useful assessments as follows: qualitative comparison by the presence/absence of glycopeptides from all proteins, relative quantitative comparison of specific glycopeptide, and absolute quantitative comparison of specific glycopeptide by employing labeled internal glycopeptide analog useful for in-depth characterization, e.g therapeutics analysis.

In quantitative aspect, despite of great interests of researchers on reliable quantitative assessment, quantitative analysis using glycoproteomic approach still bear the analytical challenges due to great differences in ionization efficiency between tryptic-glycopeptides with largely variable peptide lengths.[55-56] This results in the majority of published qualitative studies[49,51,57-61] rather than quantitative studies which still have high barriers to entry.[17,50,61-63]

While specificity of trypsin yields well-defined peptides, limited glycosylation site information could be yielded depending on the protein sequences. Multiple glycosylation sites on a single tryptic peptide avoid confident glycan assignment by tandem MS. In addition, large peptide produced by missed cleavages would not readily be characterized by collision induced dissociation (CID). On the other hands, one popular nonspecific protease, pronase usually provides more glycosylation site information and glyco-isoform information.[64] It can give better understanding to glycan micro-heterogeneity and glyco-profiling of a specified single glycoprotein rather than complex glycoprotein mixture. Thus, some researchers favor the use of pronase for the site-specific characterization of defined proteins with smaller yet rather highly heterogeneous glycopeptides such as second generation EPO, haptoglobin, alpha-2 macroglobulin, and etc.[13,65]

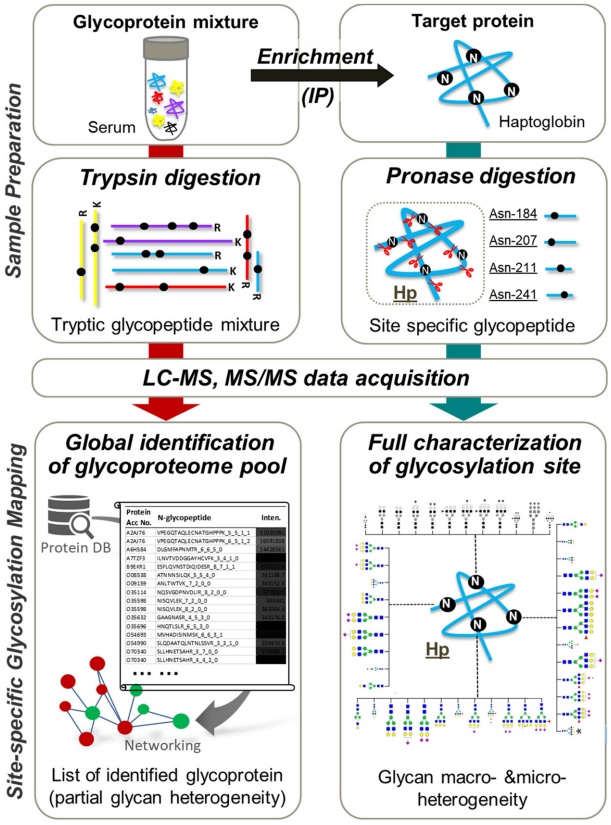

Figure 2. Schematic comparison of site-specific glycosylation analysis: trypsin based specific protease (left) vs pronase based non-specific protease (right) approach.

While the use of pronase generally provide better site information, they may increase the risk of ambiguous identification by generating peptide moiety which is too small and may not designate the site. Thus, optimal digestion time is required be optimized depending on the character of target protein to generate appropriately scaled down peptide moiety. The digestion time need to be short enough to unambiguously identify site-specificity of peptide sequence as well as easy separation and detection by PGC, yet still long enough to break down non-glycosylated peptides into the level of amino acid to reduce signal suppression by them.[56]

Taken together, it is critical to designate appropriate analytical strategy, whether specific or non-specific protease approach, depending on the character of researcher's target glycoprotein (or glycoprotein mixture) and purpose of comprehensive site-specific mapping which allows useful insight how molecular, cellular, or systemic alterations affect protein glycosylation. Employing both enzymes can allow more complete site-specific glycan mapping.[65] **Figure 2** finally shows schematic comparison of trypsin-based and pronase-based glycoproteomic approach in parallel.

3. Glycopeptide identification by tandem MS

Theoretically, tandem MS analysis of a glycopeptide can provide information both on glycosylation site and structure of attached glycan. In general, fragmentation of glycosidic bond is more favorable than peptide backbone fragmentation in collision induced dissociation (CID) and high-energy collisional dissociation (HCD).[12,66] In other words, glycan moiety on a glycopeptide is dissociated at lower vibrational energies than does the peptide backbone.[67] Consequently, the fragmentation on glycosidic bond occurs before peptide backbone fragmentation.[68] For this reason, typical fragment ions in CID or HCD tandem MS spectra are produced by the loss of glycan moiety, thus the information on glycosylation site is often minimal.[12] Alternative techniques for tandem MS analysis are electron-activated dissociation (EAD) including electron-capture dissociation (ECD) and electron-transfer dissociation (ETD) as well as ultraviolet photodissociation (UVPD). EAD, which takes advantages of ion-electron or ion-ion reaction, can cleave peptide backbone of glycopeptide while glycosidic bond remained intact, thus by generating fragment ion from which amino acid sequence and finally provide information on glycosylation site.[69-71] Recently, hybrid techniques for tandem MS analysis of glycopeptides such as electron-transfer/higher-energy collision dissociation (EThcD) and electron-transfer/higher-energy collision induced dissociation (ETciD) have been developed.[25,72-73] Both methods can simultaneously provide information on the peptide sequence and the glycosylation site as well as glycan structure.[67,74] Moreover, UVPD, which takes advantage of ion-photon interaction, has recently been applied to glycopeptide analysis and showed promising results in determination of glycosylation sites.[75-76] In this review we focused on CID fragmentation which is an essential MS tool to extract structural information on both glycan moiety and peptide sequence of a glycopeptide.[73]

Figure 3-A shows the representative N-glycopeptide CID tandem MS spectrum of the precursor ion at *m/z* 891.338 ($[M+3H]^{3+}$) consisting of mono-fucosylated mono-sialylated bi-antennary glycan with a peptide backbone, QYNST, obtained from human immunoglobulin G (IgG). Existence of the fragment ions at *m/z* 274.093 $[NeuAc\text{-}H_2O+H]^+$, *m/z* 292.107 $[HexHexNAc+H]^+$ and *m/z* 657.229 $[HexHexNAcNeuAc+H]^+$ readily indicates that glycopeptide

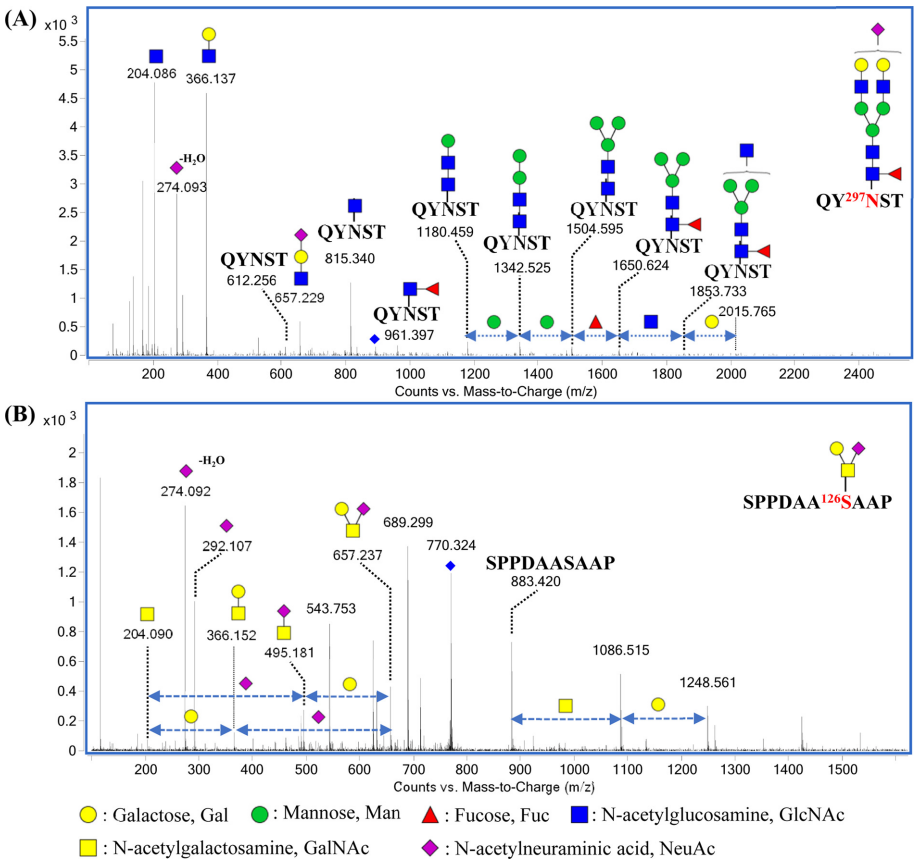

Figure 3. Representative tandem MS spectra of (A) N-glycopeptide and (B) O-glycopeptide obtained from IgG and recombinant human erythropoietin, respectively.

was decorated with NeuAc.[12,77-78] Furthermore, the loss of HexHexNAcNeuAc residue from the precursor ion (*m/z* 2015.765, $[M+H]^+$) and sequential losses of monosaccharide residues (*m/z* 1853.733, $[QNYST+Hex_3HexNAc_3Fuc_1+H]^+$; *m/z* 1650.624, $[QNYST+Hex_3HexNAc2Fuc_1+H]^+$; *m/z* 1504.595 $[QNYST+Hex_3HexNAc2+H]^+$; *m/z* 1342.525, $[QNYST+Hex_2HexNAc_2+H]^+$; *m/z* 1180.459, $[QNYST+Hex_2HexNAc_1+H]^+$) provided the composition and structure of glycan moiety. In addition, the presence of ion at *m/z* 961.397 ($[QNYST+HexNAcFuc+H]^+$) definitely indicated the core fucosylation. Finally, the peptide moiety has been identified by the presence of protonated peptide backbone $[QYNST+H]^+$ (*m/z* 612.256).

The fragmentation pattern of O-glycopeptides is similar to N-glycopeptide counter parts. **Figure 3-B** shows the representative O-glycopeptide CID spectrum of the precursor ion at *m/z* 770.324 ($[M+2H]^{2+}$) consisting of mono-sialylated *core 1* O-glycan with peptide backbone SPPDAASAAP. The loss of peptide backbone from a precursor ion yielded only glycan moiety corresponding to $Gal_1GalNAc_1NeuAc_1$ (m/z 657.237). Definitely, the presence of fragment ion at *m/z* 495.17 $[GalNAcNeuAc+H]^+$ could be evidence to indicate that the NeuAc residue was attached to GalNAc residue. In addition, fragment ion at *m/z* 366.152 $[GalGalNAc+H]^+$ readily identified as a *core 1*. The fragment ion corresponding to peptide backbone (m/z 883.420, $[SPPDAASAAP+H]^+$) could also provide the information on peptide sequence. However, the glycosylation site was not determined in CID spectrum of glycopeptide. The detailed glycosylation site could be precisely assigned using latest tandem MS techniques such as ETD, EThcD, ETciD and UVPD. Integration of information provided by tandem MS spectra with known glycobiology was unambiguously able to identify the glycosylation site and moiety.

4. Bioinformatic tools for site-specific glycosylation

In the past decade, many bioinformatics tools employing MS such as GlycoWorkbench,[79] Glyco-Fragment,[80] SimGlycan,[81] GlycoMod,[82] GlyPID,[83] and etc., have developed for the prediction of glycopeptide composition in glycoproteomic approach. Numerous problems caused by composition complexity of glycan, various size of glycopeptides, multiplicity of glycosylation at single glycopeptide, multiple peaks from single compound in ESI-MS, and multiple MS/MS data according to different fragmentation energies really make the study of site-specific glycosylation difficult. Here, we briefly introduce representative softwares for high-throughput glycoproteomics in this review (**Table 1**).

Byonic, a commercial PTM-centric search engine, has ability to annotate and identify N- and O-glycopeptide from MS/MS spectra with the high-throughput manner. Various data of LC-MS/MS obtained from Q-TOF CID, Orbitrap CID, HCD, ETD and EThcD can be used for the input of this program. Byonic score represents a sum of positively predicted and observed m/z signals and negatively predicted and not observed m/z signals, using pre-defined or user-defined protein sequence and glycan databases. However, Packer and her co-workers reported that Byonic showed high accuracy and coverage (> 80%) with low FDR (1 <%) using Q-Exactive Orbitrap HCD-MS/MS, they also mentioned that the manual validation for some N-glycopeptides from the limits of the single glycoprotein analysis was required.[84] GPFinder 3 is a python script for the analysis of N- and O-glycopeptides using their MS/MS spectra.[85] In spite of the lack of user friendly interface, it provides novel decoy generation strategy of no shuffled peptide sequence with 11 Da plus in each glycan composition. Lebrilla and his co-workers demonstrated the annotation of site-specific glycosylation with glycan micro-heterogeneity in control mixture sample, providing less than 5% FDR with high confidence. GPS was reported to show similar performance (5.58% FDR) with GPFinder 3 in single standard protein of Haptoglobin, but limited with tryptic digested N-glycopeptides, where it used decoy database of non-motif peptides with glycan composition in the analysis.[86] I-GPA is another commercial search program for identification and quantification of N- and O-glycopeptides using three scoring system and a decoy method with reversed peptide sequence and mixed glycan composition.[87] From the mixture standard proteins of IgG and AGP, and human plasma, tremendous number of N-glycopeptides was identified and quantified with < 1.0 FDR with high-throughput manner. MAGIC reported a novel algorithm named Trident that detects a triple cores of Y series ions obtained from tandem MS/MS spectra.[88] The *in silico* spectra including the original precursor m/z of the naked peptide and removing all of the glycan-related ions for the analysis was generated by the Trident for high-throughput identification of N-glycopeptides.[88] Finally, the MAGIC computes and ranks the glycan compositions, and uses Mascot search engine for identification of naked peptide fragment ions. Using this method, 36 N-glycopeptides from 26 glycoproteins were reported from HeLa cell data set.[89] pGlyco 2.0 was recently released for identification of N-glycopeptide from various types of tandem MS spectra. Support vector machine (SVM), one of the machine learning method, was used for calculation of FDR. From the yeast data set, this method showed the acceptable result with < 1% FDR, whereas 2.0% of FDR was calculated from the mouse database. Yang and his co-workers identified 995 N-glycoproteins from five kinds of mouse tissues with high-throughput manner.[90]

Table 1. List of representative informatic tools for high-throughput glycopeptide analysis.

Software Names	Availablity	High-Throuput	User Friendly Interface	Library Construction Reuirement	Analyzable Tandem MS	Multimodal Tandem MS Usage for Scoring	FDR Esitmation with Decoy Method	Quantifi-cation	Applicalble to N- and O- Glycoep-tides	Analyzable Enzymatic Digestion	Works on Platforms	Link	Reference
Byonic	Commercial	Yes	Yes	No	CID/HCD /ETD	No	No	No	both	Trypsin	Windows	https://www.pro-teinmetrics.com/ get-started/ #download-free-trial	Lee et al. JPR 2016
GPFinder 3.0	Free	Yes	No	Yes	CID	No	Yes	No	both	Trypsin/ Pronase/ Others	Windows	https:// github.com/ gpfinder/ gpfinder	Strum et al. Anal Chem 2013
GPS	Free	Yes	No	No	CID	No	No	No	N- only	Trypsin	Windows/ Linux	http://edward-slab.bmcb.georg etown.edu/soft-ware/GlycoPep- tideSearch.html	Chandler et al. 2013 J. Proteome Res.
I-GPA 1.0	Commercial	Yes	Yes	No	CID/HCD /ETD	Yes	Yes	Yes	both	Trypsin	Windows	https:// www.igpa.kr/	Park et al. 2016 Sci. Rep.
MAGIC	Free	Yes	Yes	No	CID	No	Yes	No	N- only	Trypsin	Web-based	http:// magic.iis.sinica. edu.tw/ index.html	Lynn et al. 2015 Anal Chem
pGlyco 2.0	Free	Yes	Yes	No	CID/HCD /ETD	Yes	Yes	No	N- only	Trypsin	Windows	http:// pfind.ict.ac.cn/ software/ pGlyco/ index.html	Liu et al. 2017 Nat. Comm.

5. Conclusion

Information on glycosylation site occupancy with glycan micro-heterogeneity is increasingly recognized as crucial to understanding structure and function of glycoprotein(s) in a biological system. In the past decade, significant efforts have been made to comprehensively characterize site-specific glycosylation of various biological samples in clinical and biopharmaceutical field. However, there is no single analytical method to allow high throughput and detailed investigation on individual glycosylation sites. Although MS-based techniques are still the most effective and versatile approach for the analysis of site-specific glycosylation, full characterization including macro- and micro-heterogeneity in a complex biological sample still remains challenging. Nonetheless, there is no doubt that MS-based tools will be valuable for the eventual development of glycoproteomics with continued evolution in separation technology and bioinformatics.

Acknowledgment

This work was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (NRF-2015M3A9B5053960 to H.J. An).

References

1. Apweiler, R.; Hermjakob, H.; Sharon, N. *BBA-Gen. Subjects* **1999,** 1473, 4.
2. James, W. Dennis.; Maria, Granovsky.; Warren, C. E. *BioEssays* **1999,** 21, 421.
3. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. *Nat. Rev. Mol. Cell Biol.* **2012,** 13, 448.
4. Dwek, R. A.; Butters, T. D.; Platt, F. M.; Zitzmann, N. *Nat. Rev. Drug Discov.* **2002,** 1, 65.
5. Jefferis, R. *Trends Pharmacol. Sci.* **2009,** 30, 356.
6. Li, H.; d'Anjou, M. *Curr. Opin. Biotechnol.* **2009,** 20, 678.
7. Lingg, N.; Zhang, P.; Song, Z.; Bardor, M. *Biotechnol. J.* **2012,** 7, 1462.
8. Mariño, K.; Bones, J.; Kattla, J. J.; Rudd, P. M. *Nat. Chem. Biol.* **2010,** 6, 713.
9. Wuhrer, M.; Deelder, A. M.; Hokke, C. H. *J. Chromatogr. B* **2005,** 825, 124.
10. Kolarich, D.; Lepenies, B.; Seeberger, P. H. *Curr. Opin. Chem. Biol.* **2012,** 16, 214.
11. Park, Y.; Lebrilla, C. B. *Mass Spectrom. Rev.* **2005,** 24, 232.
12. An, H. J.; Froehlich, J. W.; Lebrilla, C. B. *Curr. Opin. Chem. Biol.* **2009,** 13, 421.
13. Seo, Y.; Kim, U.; Oh, M. J.; Yun, N. Y.; An, H. J. *Mass Spectrom. Lett.* **2015,** 6, 53.
14. Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. *J. Chromatogr. B* **2007,** 849, 115.
15. Klampfl, C. W. *Electrophoresis* **2006,** 27, 3.
16. Zamfir, A.; Peter-Katalinic, J. *Electrophoresis* **2001,** 22, 2448.
17. Harazono, A.; Kawasaki, N.; Itoh, S.; Hashii, N.; Matsuishi-Nakajima, Y.; Kawanishi, T.; Yamaguchi, T. *J. Chromatogr. B* **2008,** 869, 20.
18. Nilsson, J. *Glycoconj. J.* **2016,** 33, 261.
19. Plematl, A.; Demelbauer, U. M.; Josic, D.; Rizzi, A. *Proteomics* **2005,** 5, 4025.
20. Tajiri, M.; Yoshida, S.; Wada, Y. *Glycobiology* **2005,** 15, 1332.
21. Zauner, G.; Koeleman, C. A.; Deelder, A. M.; Wuhrer, M. *J. Sep. Sci.* **2010,** 33, 903.
22. Singh, C.; Zampronio, C. G.; Creese, A. J.; Cooper, H. J. *J. Proteome Res.* **2012,** 11, 4517.
23. Pai, P. -J.; Hu, Y.; Lam, H. *Anal. Chim. Acta* **2016,** 934, 152.
24. Zhang, Y.; Xie, X.; Zhao, X.; Tian, F.; Lv, J.; Ying, W.; Qian, X. *J. Proteomics* **2018,** 170, 14.
25. Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Hojrup, P.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J. *Mol Cell Proteomics.* **2011,** 10, M000031-MCP201.
26. Wang, D.; Hincapie, M.; Rejtar, T.; Karger, B. L. *Anal. Chem.* **2011,** 83, 2029.
27. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007,** 4, 709.
28. Thaysen-Andersen, M.; Packer, N. H. *BBA-Proteins and Proteom.* **2014,** 1844.
29. Kolarich, D.; Jensen, P. H.; Altmann, F.; Packer, N. H. *Nat. Protoc.* **2012,** 7, 1285.
30. Leymarie, N.; Griffin, P. J.; Jonscher, K.; Kolarich, D.; Orlando, R.; McComb, M.; Zaia, J.; Aguilan, J.; Alley, W. R.; Altmann, F.; Ball, L. E.; Basumallick, L.; Bazemore-Walker, C. R.; Behnken, H.; Blank, M. A.; Brown, K. J.; Bunz, S. -C.; Cairo, C. W.; Cipollo, J. F.; Daneshfar, R.; Desaire, H.; Drake, R. R.; Go, E. P.; Goldman, R.; Gruber, C.; Halim, A.; Hathout, Y.; Hensbergen, P. J.; Horn, D. M.; Hurum, D.; Jabs, W.; Larson, G.; Ly, M.; Mann, B. F.; Marx, K.; Mechref, Y.; Meyer, B.; Möginger, U.; Neusüβ, C.; Nilsson, J.; Novotny, M. V.; Nyalwidhe, J. O.; Packer, N. H.; Pompach, P.; Reiz, B.; Resemann, A.; Rohrer, J. S.; Ruthenbeck, A.; Sanda, M.; Schulz, J. M.; Schweiger-Hufnagel, U.; Sihlbom, C.; Song, E.; Staples, G. O.; Suckau, D.; Tang, H.; Thaysen-Andersen, M.; Viner, R. I.; An, Y.; Valmu, L.; Wada, Y.; Watson, M.; Windwarder, M.; Whittal, R.; Wuhrer, M.; Zhu, Y.; Zou, C. *Mol. Cell. Proteomics* **2013,** 12, 2935.
31. Faid, V.; Denguir, N.; Chapuis, V.; Bihoreau, N.; Chevreux, G. *Proteomics* **2014,** 14, 2460.
32. Jiang, J.; Tian, F.; Cai, Y.; Qian, X. H.; Costello, C. E.; Ying, W. T. *Anal. Bioanal. Chem.* **2014,** 406, 6265.
33. Ohta, M.; Kawasaki, N.; Itoh, S.; Hayakawa, T. *Biologicals* **2002,** 30, 235.

34. Hua, S.; Nwosu, C. C.; Strum, J. S.; Seipert, R. R.; An, H. J.; Zivkovic, A. M.; German, J. B.; Lebrilla, C. B. *Anal. Bioanal. Chem.* **2012,** 403, 1291.
35. Hua, S.; Hu, C. Y.; Kim, B. J.; Totten, S. M.; Oh, M. J.; Yun, N.; Nwosu, C. C.; Yoo, J. S.; Lebrilla, C. B.; An, H. J. *J. Proteome Res.* **2013,** 12, 4414.
36. Plomp, R.; Hensbergen, P. J.; Rombouts, Y.; Zauner, G.; Dragan, I.; Koeleman, C. A.; Deelder, A. M.; Wuhrer, M. *J. Proteome Res.* **2014,** 13, 536.
37. Alvarez-Manilla, G.; Atwood, J.; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M. *J. Proteome Res.* **2006,** 5, 701.
38. Joenvaara, S.; Ritamo, I.; Peltoniemi, H.; Renkonen, R. *Glycobiology* **2008,** 18, 339.
39. Hemstrom, P.; Irgum, K. *J. Sep. Sci.* **2006,** 29, 1784.
40. Groleau, P. E.; Desharnais, P.; Cote, L.; Ayotte, C. *J. Mass Spectrom.* **2008,** 43, 924.
41. Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal. Chem.* **1993,** 65, 877.
42. Cointe, D.; Beliard, R.; Jorieux, S.; Leroy, Y.; Glacet, A.; Verbert, A.; Bourel, D.; Chirat, F. *Glycobiology* **2000,** 10, 511.
43. Thaysen-Andersen, M.; Wilkinson, B. L.; Payne, R. J.; Packer, N. H. *Electrophoresis* **2011,** 32, 3536.
44. Dam, S.; Thaysen-Andersen, M.; Stenkjær, E.; Lorentzen, A.; Roepstorff, P.; Packer, N. H.; Stougaard, J. *J. Proteome Res.* **2013,** 12, 3383.
45. Thaysen-Andersen, M.; Packer, N. H. *BBA-Proteins Proteom.* **2014,** 1844, 1437.
46. Hua, S.; Oh, M. J.; Ozcan, S.; Seo, Y. S.; Grimm, R.; An, H. J. *TrAC- Trends Anal. Chem.* **2015,** 68, 18.
47. Sun, B.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X.; Lin, B.; Hood, L. *Mol. Cell. Proteomics* **2007,** 6, 141.
48. Stadlmann, J.; Taubenschmid, J.; Wenzel, D.; Gattinger, A.; Dürnberger, G.; Dusberger, F.; Elling, U.; Mach, L.; Mechtler, K.; Penninger, J. M. *Nature* **2017,** 549, 538.
49. Parker, B. L.; Thaysen-Andersen, M.; Solis, N.; Scott, N. E.; Larsen, M. R.; Graham, M. E.; Packer, N. H.; Cordwell, S. J. *J. Proteome Res.* **2013,** 12, 5791.
50. Halim, A.; Nilsson, J.; Rüetschi, U.; Hesse, C.; Larson, G. *Mol. Cell. Proteomics* **2012,** 11, 1.
51. Halim, A.; Rüetschi, U.; Larson, G. r.; Nilsson, J. *J. Proteome Res.* **2013,** 12, 573.
52. Jensen, L. J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; Bork, P.; von Mering, C. *Nucleic Acids Res.* **2009,** 37, D412.
53. Ueda, K. *Proteomics Clin. Appl.* **2013,** 7, 607.
54. Tian, Y.; Zhang, H. *Proteomics Clin. Appl.* **2010,** 4, 124.
55. Cancilla, M. T.; Wong, A. W.; Voss, L. R.; Lebrilla, C. B. *Anal. Chem.* **1999,** 71, 3206.
56. Hua, S.; Hu, C. Y.; Kim, B. J.; Totten, S. M.; Oh, M. J.; Yun, N.; Nwosu, C. C.; Yoo, J. S.; Lebrilla, C. B.; An, H. J. *J. Proteome Res.* **2013,** 12, 4414.
57. Mayampurath, A.; Yu, C. -Y.; Song, E.; Balan, J.; Mechref, Y.; Tang, H. *Anal. Chem.* **2014,** 86, 453.
58. Anonsen, J. H.; Vik, Å.; Egge-Jacobsen, W.; Koomey, M. *J. Proteome Res.* **2012,** 11, 5781.
59. Yin, X.; Bern, M.; Xing, Q.; Ho, J.; Viner, R.; Mayr, M. *Mol. Cell. Proteomics* **2013,** 12, 956.
60. Trinidad, J. C.; Schoepfer, R.; Burlingame, A. L.; Medzihradszky, K. F. *Mol. Cell. Proteomics* **2013,** 12, 3474.
61. Thannhauser, T. W.; Shen, M.; Sherwood, R.; Howe, K.; Fish, T.; Yang, Y.; Chen, W.; Zhang, S. *Electrophoresis* **2013,** 34, 2417.
62. Nilsson, J.; Ruetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G. *Nat. Methods* **2009,** 6, 809.
63. Uematsu, R.; Furukawa, J. -i.; Nakagawa, H.; Shinohara, Y.; Deguchi, K.; Monde, K.; Nishimura, S. -I. *Mol. Cell. Proteomics* **2005,** 4, 1977.
64. Nwosu, C. C.; Seipert, R. R.; Strum, J. S.; Hua, S. S.; An, H. J.; Zivkovic, A. M.; German, B. J.; Lebrilla, C. B. *J. Proteome Res.* **2011,** 10, 2612.
65. Hong, Q.; Ruhaak, L. R.; Stroble, C.; Parker, E.; Huang, J.; Maverakis, E.; Lebrilla, C. B. *J. Proteome Res.* **2015,** 14, 5179.
66. Jebanathirajah, J.; Steen, H.; Roepstorff, P. *J. Am. Soc. Mass. Spectrom.* **2003,** 14, 777.
67. Hu, H.; Khatri, K.; Klein, J.; Leymarie, N.; Zaia, J. *Glycoconj. J.* **2016,** 33, 285.
68. Bykova, N. V.; Rampitsch, C.; Krokhin, O.; Standing, K. G.; Ens, W. *Anal. Chem.* **2006,** 78, 1093.
69. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, 9528.
70. Wiesner, J.; Premsler, T.; Sickmann, A. *Proteomics* **2008,** 8, 4466.
71. Kim, M. -S.; Pandey, A. *Proteomics* **2012,** 12, 530.
72. Håkansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Anal. Chem.* **2001,** 73, 4530.
73. Leymarie, N.; Zaia, J. *Anal. Chem.* **2012,** 84, 3040.
74. Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L.-W.; Kent, K. C.; Li, L. *J. Am. Soc. Mass. Spectrom.* **2017,** 28, 1751.
75. Ko, B. J.; Brodbelt, J. S. *Int. J. Mass spectrom.* **2015,** 377, 385.
76. Aboufazeli, F.; Kolli, V.; Dodds, E. D. *J. Am. Soc. Mass Spectrom.* **2015,** 26, 587.
77. An, H. J.; Lebrilla, C. B. *Mass Spectrom. Rev.* **2011,** 30, 560.
78. Oh, M. J.; Hua, S.; Kim, U.; Kim, H. J.; Lee, J.; Kim, J.-H.; An, H. J. *Bioanalysis* **2016,** 8, 711.
79. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M. *J. Proteome Res.* **2008,** 7, 1650.
80. Lohmann, K. K.; von der Lieth, C. W. *Proteomics* **2003,** 3, 2028.
81. Li, J. *Functional Glycomics: Methods and Protocols*. Humana Press: Totowa, NJ, **2010**.
82. Catherine A. Cooper; Elisabeth Gasteiger; Packer, N. H. *Proteomics* **2001,** 1, 340.
83. Wu, Y.; Mechref, Y.; Klouckova, I.; Mayampurath, A.;

Novotny, M. V.; Tang, H. *Rapid Commun. Mass Spectrom.* **2010,** 24, 965.
84. Lee, L. Y.; Moh, E. S.; Parker, B. L.; Bern, M.; Packer, N. H.; Thaysen-Andersen, M. *J. Proteome Res.* **2016,** 15, 3904.
85. Strum, J. S.; Nwosu, C. C.; Hua, S.; Kronewitter, S. R.; Seipert, R. R.; Bachelor, R. J.; An, H. J.; Lebrilla, C. B. *Anal. Chem.* **2013,** 85, 5666.
86. Chandler, K. B.; Pompach, P.; Goldman, R.; Edwards, N. *J. Proteome Res.* **2013,** 12, 3652.
87. Park, G. W.; Kim, J. Y.; Hwang, H.; Lee, J. Y.; Ahn, Y. H.; Lee, H. K.; Ji, E. S.; Kim, K. H.; Jeong, H. K.; Yun, K. N.; Kim, Y. S.; Ko, J. H.; An, H. J.; Kim, J. H.; Paik, Y. K.; Yoo, J. S. *Sci. Rep.* **2016,** 6, 21175.
88. Lih, T. M.; Choong, W. K.; Chen, C. C.; Cheng, C. W.; Lin, H. N.; Chen, C. T.; Chang, H. Y.; Hsu, W. L.; Sung, T. Y. *Nucleic Acids Res.* **2016,** 44, w575.
89. Lynn, K. S.; Chen, C. C.; Lih, T. M.; Cheng, C. W.; Su, W. C.; Chang, C. H.; Cheng, C. Y.; Hsu, W. L.; Chen, Y. J.; Sung, T. Y. *Anal. Chem.* **2015,** 87, 2466.
90. Liu, M. Q.; Zeng, W. F.; Fang, P.; Cao, W. Q.; Liu, C.; Yan, G. Q.; Zhang, Y.; Peng, C.; Wu, J. Q.; Zhang, X. J.; Tu, H. J.; Chi, H.; Sun, R. X.; Cao, Y.; Dong, M. Q.; Jiang, B. Y.; Huang, J. M.; Shen, H. L.; Wong, C. C. L.; He, S. M.; Yang, P. Y. *Nat. Commun.* **2017,** 8, 438.

