



Published in final edited form as:

Analyst. 2013 May 21; 138(10): 2793–2803. doi:10.1039/c2an36042j.

Software for Automated Interpretation of Mass Spectrometry Data from Glycans and Glycopeptides

Carrie L. Woodin, Morgan Maxon, and Heather Desaire*

Department of Chemistry, University of Kansas, Lawrence, KS

Abstract

The purpose of this review is to provide those interested in glycosylation analysis with the most updated information on the availability of automated tools for MS characterization of *N*-linked and *O*-linked glycosylation types. Specifically, this review describes software tools that facilitate elucidation of glycosylation from MS data on the basis of mass alone, as well as software designed to speed the interpretation of glycan and glycopeptide fragmentation from MS/MS data. This review focuses equally on software designed to interpret the composition of released glycans and on tools to characterize *N*-linked and *O*-linked glycopeptides. Several websites have been compiled and described that will be helpful to the reader who is interested in further exploring the described tools.

INTRODUCTION

Overview of Glycosylation

The addition of monosaccharide residues onto a protein or lipid, known as glycosylation, serves an important function in many cellular signaling and communication events, including those involving host-pathogen interactions.^{1–4} It has long been understood that protein-carbohydrate interactions play a participatory role in many processes affecting disease progression.^{1, 4–7} Furthermore, experimental evidence demonstrates that the identity of the attached glycans change during these events.^{6, 8–10} For example, aberrant glycosylation is often present in individuals experiencing cancer, diabetes, and inflammation.^{4, 8–11} As such, accurate characterization of a glycoprotein's glycan substituents has been shown to be crucial in the development of potential biomarkers, protein-based vaccine candidates, and pharmaceutical treatments.^{5, 9, 12, 13}

In studies involving protein glycosylation, mass spectrometry (MS) is currently considered the most utilized instrument for the characterization of glycan substituents, both attached or released.¹¹ Accordingly, this review focuses on those automated analysis tools that can be applied to more efficiently analyze glycosylation profiles through interrogation by MS and MSⁿ methods.

*To whom correspondence should be addressed. University of Kansas, 2030 Becker Drive, Lawrence, KS 66047. Phone: 785-864-3015. hdesaire@ku.edu.

Glycosylation Heterogeneity

Unlike DNA replication and protein transcription, glycosylation is a “non-template”-driven process,^{5,14} where the sugar residues form a multitude of arrangements.^{5,12} The monosaccharides that comprise the glycan may be long or short, branched or linear, and linked in a variety of ways, creating a large degree of variability.^{3,5,12} This heterogeneity is described in two ways: Glycan differences at different sites of attachment (macroheterogeneity), or within the same site (microheterogeneity).¹² The large amount of heterogeneity presents a challenging obstacle to researchers attempting to elucidate structural, as well as compositional, information on a protein’s glycan population, especially when samples are mixtures of proteins.

Types of Protein Glycosylation

Over half of all proteins expressed are predicted to be glycosylated.¹⁵ In addition to the established forms of protein glycosylation, including *N*-linked, *O*-linked, and *C*-linked forms,^{1, 16–18} rarer configurations such as *S*-linked glycosylation, are also being discovered.¹⁹ Although a variety of types exist, the two most common types of glycosylation are *N*-linked and *O*-linked.^{1,7,20} In *N*-linked glycosylation, the addition of a glycan may occur at the asparagine residue when the consensus sequence Asn-Xaa-Ser/Thr occurs, where Xaa is any amino acid except proline.^{12,18,21} The inclusion of this pattern is a fundamental requirement for *N*-linked glycosylation to occur, though it is not a guarantee that a glycosylation site will be occupied.²¹ With *O*-linked glycosylation, the glycan addition may occur at any Ser or Thr residue within the protein sequence,^{4,11} though a very low percentage of these sites are actually occupied.¹⁵ In contrast to *N*-linked glycans, *O*-linked glycans have less defined sequence patterns,^{12,18,22} and may consist of several distinct core arrangements.¹¹ For these reasons, both the prediction and determination of *O*-linked glycosylation characteristics has advanced slower than *N*-linked glycosylation analysis.¹¹

Glycosylation Analysis

In the determination of glycan structural information, mass spectrometry (MS) has shown to be a powerful tool, as a large part of the successful interrogation of glycan composition and structures has been accomplished utilizing MS experiments.^{18,23,24} There are two main strategies for elucidating glycosylation information using MS techniques: 1) Characterization of a protein’s glycans after they are released from glycoproteins and 2) Characterization of glycopeptides after proteolytic digestion of a glycoprotein.^{11,18,23} The study of released glycans is particularly useful when rapid analysis of glycan composition is desired. Though *N*- and *O*-linked glycan populations can be studied independently through the use of different cleavage procedures,^{11,12} no information on where the individual glycans were attached along the protein is obtained when the glycans are cleaved *a priori*. In order to obtain glycosylation site-specific information for individual glycoforms, the second method, glycopeptide analysis, which requires digestion of the protein using a protease such as trypsin, is necessary.^{18,23,25} This method is generally advantageous because it provides information about both glycan composition and the site of the glycan’s attachment.²³ Despite the challenges mentioned in this review, techniques that allow for complete

profiling of a peptide's glycan population have advanced greatly in the past decade, especially with respect to site-specific glycopeptide analysis.

In the study of protein glycosylation by either of these techniques, a number of resources, including databases providing information on known glycosylation structures or site occupancy, as well as collections of experimental data, are currently available.^{26–33} For instance, researchers needing to identify occupied *N*-linked glycosylation sites on a specific protein can access UniProtKB,²⁶ while those wanting statistics specific to proteins modified by *O*-GlcNAc could visit dbOGAP.³³ Although this repertoire of information is greater for proteins modified by *N*- and *O*-linked glycan types, databases that contain entries on *C*-glycosylated proteins, such as dbPTM,²⁷ are available as well. A current list and description of these database resources are provided in Table 1.

AUTOMATED ANALYSIS OF RELEASED GLYCANS

Characterization of High Resolution MS data for Oligosaccharides

Often, the easiest way to identify a protein's glycan population is by enzymatically cleaving the glycan substituents and analyzing the monosaccharide residues directly.²⁵ *N*- and *O*-linked glycans from the same protein can be independently characterized in this manner, as in the method described by Goetz *et al.* where β -elimination is used to release *O*-linked glycans, which are simultaneously permethylated.³⁴ Once cleaved, automated analysis tools that determine glycan composition from MS data may be used.

One such tool developed to analyze MS data of glycans is Cartoonist, as described by Goldberg *et al.*³⁵ Cartoonist works to increase the speed of compositional determination in permethylated *N*-linked glycans from matrix laser desorption/ionization-mass spectrometry (MALDI-MS) data through identification and annotation of MALDI-TOF spectra.³⁵ The most likely glycan compositions are selected using precursor mass data.³⁵ The program automatically labels MALDI peaks with cartoons of the most probable oligosaccharide structure, as determined by the program's algorithm, from a library of 300 generated mammalian *N*-linked glycans.³⁵ Recently, Goldberg *et al.* extended this concept by developing an automated tool for the analysis of *O*-linked glycans from MS/MS data,³⁶ as described below in *MS/MS approaches for glycan analysis* section of this review. To date, neither program is publicly available.

Another software solution useful for the identification of glycans from MS data is SysBioWare, described by Vakhrushev *et al.*³⁷ SysBioWare takes raw MS¹ data that a user uploads and performs baseline adjustment and denoising, wavelet analysis, and peak detection before grouping isotopes of detected peaks.³⁷ The isotopic grouping is also performed automatically, which enables the program to deduce monoisotopic *m/z* values and precursor charge states without the need of manual input by the user.³⁷ The software then determines monosaccharide compositions on the basis of mass.³⁷ Currently, the SysBioWare program is being updated to include analysis of MS/MS data for glycans as well.³⁸ SysBioWare is not freely available to the public at this time.

Similar to SysBioWare is GlycoWorkbench. GlycoWorkbench evaluates glycan compositions (which are proposed by the user) by searching the spectral peak list of user-input MS data for matches between calculated theoretical glycan masses and corresponding m/z values.³⁹ The GlycanBuilder tool, designed to interface with GlycoWorkbench, enables the drawing of glycan structure representations, with all stereochemical information on the monosaccharides depicted as specified by the user.⁴⁰ Both analysis tools, GlycanBuilder and GlycoWorkbench, are available online free of charge, as described in Table 2.

GlycoSpectrumScan, another freely available program, was developed by Deshpande *et al.* and works to identify *N*- and *O*-linked glycoforms using MS¹ data.⁴¹ This software is capable of analyzing both singly or multiply charged ions directly from raw data, and accepts the input of both ESI and MALDI spectra.⁴¹ GlycoSpectrumScan also determines the relative abundance of *N*- and *O*-linked glycoforms that are identified for each glycosylation site.⁴¹ However, the user must enter the *N*- and/or *O*-linked glycan compositions potentially present in the sample, as well as the *in silico* peptide masses of the digested glycoprotein.⁴¹ GlycoSpectrumScan is available online (see Table 2).

MS/MS Approaches for Glycan Characterization

Until recently, when automated software tools and scoring algorithms became available, the identification of accurate glycan or glycopeptide assignments from MS/MS data was a key bottle-neck, due to the need for extensive manual data analysis. STAT, designed by Gaucher *et al.*, is one of the first automated tools for the determination of glycan composition using tandem MS.⁴² STAT is designed for glycans of up to ten monosaccharide residues, and has the ability to quickly analyze relevant *N*-glycan compositions.⁴² STAT also lists the most likely structures in order of probability to provide a ranking system when more than one candidate glycan matches the fragmentation profile of the data being analyzed.⁴² Unfortunately, this program is no longer publicly accessible.

An early analysis tool capable of evaluating *O*-linked glycan fragmentation is the OSCAR algorithm.⁴³ OSCAR, as developed by Ashline *et al.*, is specifically designed for the annotation of permethylated *O*-linked oligosaccharides from MS^{*n*} data.⁴³ OSCAR is part of a collection of software tools termed Glyspy, which is not currently accessible to the public.⁴³ Although innovative, the use of OSCAR is limited to direct infusion experiments, as the software cannot effectively process data from LC-MS methods.⁴³

A program contemporary to OSCAR and also developed to handle glycan MS/MS data is StrOligo.⁴⁴ This instrument-specific program was designed by Ethier *et al.* for the determination of *N*-linked glycan structures from MALDI MS/MS data.⁴⁴ In published research, StrOligo successfully assigned the correct glycan structure in 24 out of 28 cases.⁴⁴ Although the results of these two programs are promising, neither program is freely accessible online.

Several alternative glycan analysis tools are freely available online. One of the earliest of these was reported by Lohmann *et al.* in 2004.⁴⁵ The authors describe the web tools GlycoFragment and GlycoSearchMS, which were developed for glycan structural determination.⁴⁵ The theoretical fragmentation patterns of carbohydrate structures are

calculated using GlycoFragment, which displays theoretical b- and y-fragments as well as c-, z-, a- and x-ions.⁴⁵ GlycoSearchMS works to analyze experimental glycan data by comparing it against a library of theoretical spectra from *N*-linked and *O*-linked glycan fragmentation entries extracted from SweetDB.⁴⁵ The GlycoFragment program has been validated on both *N*-linked and *O*-linked glycan classes, and, used in conjunction with GlycoSearchMS, enables researchers to determine the most probable glycan composition according to the information from the combined algorithms.^{45,46} Both GlycoFragment and GlycoSearchMS are freely available. See Table 2 for more information.

Another heavily used, free, online tool for glycoform analysis is GlycoWorkbench, which has shown to be a resourceful tool not only for analysis of MS¹ data, as mentioned previously, but in the identification of glycans from MS/MS data as well.³⁹ To utilize the glycan fragmentation analysis feature, a user must first input/define the possible glycan compositions and spectral peak list.³⁹ The software then calculates expected glycan fragmentation and relative *m/z* values, and annotates peaks of the uploaded data with the most probable identity (shown in red to distinguish it), of all compositions tested.³⁹ As previously stated, GlycoWorkbench is available for free online.

In addition to the freely available tools mentioned above, several other MS/MS analysis tools for glycans are available for researchers, either for purchase or by special request to the tools' developers. Two of these are GlyCH and Glyquest.^{47,48} GlyCH was developed by Tang *et al.* to perform automated interpretation of oligosaccharide tandem mass spectra.⁴⁷ The algorithm has a scoring function built in to allow researchers to compare compositions when more than one is determined to be possible.⁴⁷ The GlyCH algorithm, which has so far been tested on released *N*-glycans, is also capable of *de novo* analysis, providing no more than 10 monosaccharide residues comprise the glycan chain.⁴⁷ Although not freely accessible online; the program is available upon request from the authors.⁴⁷ More recently, Gao *et al.* developed Glyquest, an automated analysis program that takes a different approach to determine compositions of intact *N*-linked glycans.⁴⁸ This software utilizes a database in conjunction with an integrated search engine to determine the composition of peptide-attached *N*-glycans from CID MS/MS data.⁴⁸ After the program algorithmically determines the molecular weight of the protonated peptide within a given spectrum; candidate *N*-glycan compositions are selected and fragmented *in silico* to generate a theoretical spectrum that is then compared to the experimental spectrum.⁴⁸ The glycan compositions with fragmentation profiles that are most similar to the experimental fragmentation are determined to be the most probable candidates.⁴⁸ Glyquest is not freely available to the public.

SimGlycan is another tool that can be used to increase throughput of glycan analysis.^{49,50} More information is available online (<http://www.premierbiosoft.com/>). This for purchase tool is useful for determining glycan structures from MS/MS data obtained from many different mass spectrometers, once an acquisition file is converted into mzXML format.⁵⁰ A user uploads an MS/MS data file, and the software utilizes a built-in database with theoretical fragmentation profiles of nearly 10,000 glycan structures to provide the most likely structural candidates.⁴⁹ One unique feature of SimGlycan is that no filtering of biologically relevant structures is provided, which can be advantageous for identifying novel

glycan structures, but disadvantageous in that it returns a user many structures which are not relevant.⁴⁹ However, commercially available software such as SimGlycan are expensive, which potentially limit their use.

A more recent program developed specifically for the compositional interpretation of *O*-linked glycan fragmentation is CartoonistTwo, as described by Goldberg *et al.*³⁶ CartoonistTwo was designed using CID data acquired on an FTICR-MS, and validated using data from a test set of 34 spectra acquired from *Xenopus* egg jelly.³⁶ Unfortunately, the program is not freely accessible to the public.

AUTOMATED ANALYSIS OF GLYCOPEPTIDES

For researchers performing site-specific glycosylation analysis, the initial step toward accomplishing the characterization of attached glycoforms at unique sites within a digested protein is to identify potential glycosylation sites within that protein; the tools to facilitate this step are described in Table 1 of this review. In addition to these, programs that utilize algorithms to predict the likelihood of site-occupancy by examination of the amino acid residues surrounding the potential glycosylation site have also been developed.^{51–59} These prediction tools, along with a description and link to each tool, are provided in Table 3.

Experimental Data Requirements

After the resultant glycopeptides are obtained from the proteolytic digest, two types of data are generally used to accurately characterize the identity of a glycopeptide. First, high resolution MS data of the glycopeptide is used to infer possible glycopeptide compositions; second, tandem MS data is acquired to distinguish between isomers and isobars.⁶⁰ Figure 1 provides a schematic of this analysis process.

N-LINKED GLYCOPEPTIDES

Although *N*-linked glycoforms share a common core structure, the rest of the glycan follows one of three distinct arrangements. Based on the arrangement pattern, *N*-linked glycans compositions are classified into three main types, those with: 1) High mannose type glycans 2) Complex type glycans and 3) Hybrid type glycans.^{12,18} This information is useful when deciphering glycopeptide compositions from MS experiments, specifically from CID MS/MS data.⁶¹

N-linked Glycopeptide MS Data

A variety of automated and semi-automated analysis tools have been created to aid in the interpretation of *N*-linked glycopeptide MS data. The key objective of these tools is to provide glycopeptide compositions that are consistent with the high resolution MS data. Researchers then typically use MS/MS analysis, described later in this review, to determine which of the compositions is correct for each given ion. Three of these tools are accessible to the public: GlycoMod (<http://web.expasy.org/glycomod/>), GlycoPep DB (<http://hexose.chem.ku.edu/glycop.htm>), and the previously mentioned GlycoSpectrumScan.^{41,62,63} GlycoMod, the earliest and most heavily used tool, accepts a protein sequence, possible monosaccharide building blocks, and experimental mass data as inputs, and it calculates all

possible glycopeptide compositions that fall within the mass tolerance.⁶² One restriction in the capacity of GlycoMod to analyze glycopeptide data is the inability to handle multiply-charged precursors.⁶²

Programs such as GlycoPep DB and GlycoSpectrumScan were designed to overcome some of the limitations in GlycoMod. GlycoPep DB, developed by Go *et al.* limits its output by restricting the potential glycans in the glycopeptide to a database of biologically relevant glycoforms that have been previously identified in MS data.⁶³ It also accepts precursor ions in multiple charge states.⁶³ The disadvantage of using this approach, however, is that if the glycan in the spectrum is not in the GlycoPep DB database, then the software will not be effective at providing the correct assignment for the peak.⁶³ GlycoSpectrumScan is a more recent program, developed by Deshpande *et al.* that also interprets MS data on both *N*- and *O*-linked glycopeptides.⁴¹ Like GlycoPep DB, this program has the ability to handle input for both singly- and multiply-charged data.⁴¹ GlycoSpectrumScan is described in detail below for *O*-linked linked MS data analysis. Regardless of which tool is used for assigning the high resolution data, these assignments must be supported by MS/MS data, to provide high confidence assignments.⁶⁰

N-linked Glycopeptide MS/MS Data

Each common *N*-linked glycan type (complex, hybrid, or high mannose) has a signature fragmentation profile that is present when a glycopeptide is subjected to MS/MS experiments.^{25,61} These characteristic fragmentation profiles are useful for determining the correct composition of an *N*-linked glycopeptide when isobaric candidate compositions are possible.⁶¹ However, as manual interpretation of these data are challenging, software is required to speed analysis time.

Many of the available tools to analyze glycopeptides are software expansions of tools that have been developed previously to analyze released glycans. One disadvantage of expanding glycan analysis tools to glycopeptides is that these tools generally lack capabilities for analyzing and scoring the peptide component of the glycopeptides. SimGlycan is one such example. Available for purchase, SimGlycan has been updated to perform fragmentation analysis for glycopeptides, in addition to glycans.^{49,50} As stated previously, SimGlycan uses a database of over 9,000 glycan structures that could be consistent with the MS/MS data to identify the most appropriate composition for the acquired spectrum.⁵⁰ SimGlycan may be purchased online (<http://www.premierbiosoft.com/>).

Many other publicly available tools to elucidate glycosylation profiles of glycopeptides have also emerged out of glycan analysis software. GlycoWorkbench and Glyco-Peakfinder both can annotate glycan fragmentation in glycopeptide data, although the peptide portion of the glycopeptide must be determined by some means other than the use of these tools.^{39,64} On the positive side, Glyco-Peakfinder is useful for *de novo* calculation and annotation of glycan fragment ions within tandem mass spectra.⁶⁴ Users may allow constraints on the oligosaccharide such as size and attachment of other substituents (such as acetate, phosphate, and sulfate), and the program is capable of annotating multiply-charged ions (– 4 to + 4).⁶⁴ Additionally, glycan fragmentation is analyzed across multiple charge states, and across multiple charge carriers (cationic carriers), within the same spectrum.⁶⁴

A completely different approach is used in GlycoPep ID.⁶⁵ GlycoPep ID is a web-based tool developed by Go *et al.* to interpret MS/MS data of glycopeptides and to identify the peptide component of glycopeptides through analysis of expected product ions.⁶⁵ The URL to access this program is listed in Table 4. Although this program is useful for identification of the peptide portion of the glycopeptide in complex LC-MS samples, it does not contain a scoring algorithm to identify the most probable glycopeptide match.⁶⁵

Software with the ability to score potential compositions is especially useful to researchers. Often, more than one glycan or glycopeptide composition could correspond to a given spectrum within the accepted range of mass tolerance. Therefore, programs that have a scoring function to evaluate each of those possible matches, and return which of them is the most likely structure, greatly improves the efficiency of the analysis. For programs that lack this feature, a user must spend time manually determining which of the mathematically possible predictions is the best match for the data

Some alternative, unique strategies have been developed with the goal of scoring MS/MS data against potential glycopeptide compositions, such as those described using Peptonist, Medcel Integrator, the Branch-and-Bound algorithm, GlycoMaster, Sweet Substitute, and GlyDB.^{66,67,68,69,70,71} Unfortunately, none of these programs are currently publicly available.

To address the need for publicly accessible tools specifically designed to interpret and score fragmentation of glycopeptides, GlycoMiner was developed by Ozohanics *et al.*⁷² In the analysis of 3132 spectra, the program was reported to have found 338 that corresponded to MS/MS data of glycopeptides (versus peptides).⁷² Designed using qTOF data, the software is capable of assigning glycopeptide compositions when both the peptide and glycan components portions are unknown.⁷² However, the program is only capable of performing compositional analysis when the spectra are of good quality.⁷² The program fails when spectral quality is low, as evidenced by the program's identification of glycan composition in only 196/338 glycopeptide spectra.⁷² Although this tool is a great advancement towards automated interpretation of glycopeptide MS/MS data, GlycoMiner often generates multiple plausible compositions and fails to rank the correct glycopeptide as the top candidate.⁷² In addition, the program requires spectra containing a low S/N ratio as well as the presence of low-mass marker ions, which are not typically present in data collected on ion trap instruments.⁷² Available online, GlycoMiner is free to download and use; see Table 4.

Similar to GlycoMiner, GlycoPeptide Search (GPS) is a recently developed program by Chandler *et al.* for the determination of glycopeptide composition from CID data.⁷³ Designed for purified glycoprotein samples analyzed by LC-MS/MS, GPS utilizes GlycomeDB, a glycan database in conjunction with the peptide file, which is supplied by the user, to generate an Excel file of glycopeptide matches based on fragmentation evidence.⁷³ To generate the peptide-glycan pairs, GPS must find both low mass oxonium, and *N*-glycan core-containing, product ions.⁷³ GPS is freely available online as well.⁷³ For further information, see Table 4.

The targeted MS/MS approach utilizing the computational tool GlypID recently described by Wu *et al.* aims to characterize *N*-linked glycopeptides through the combined use of MS¹ and MS² information extracted from LC-MS/MS experiments.⁷⁴ One of the benefits to the method is that no prior knowledge of the potential glycosylation or identity of the glycopeptide is necessary.⁷⁴ Instead, GlypID assigns a cluster of glycopeptides in the “same family” (microheterogeneities) based on observed mass.⁷⁴ In addition, the approach utilizes an isotope deconvolution algorithm to assign ion charges along with monoisotopic ions.⁷⁴ This information is then added to the inclusion list of “prioritized precursor ions” for the MS/MS analysis that follows.⁷⁴ Next, the resultant CID data is searched for the longest series of glycosidic bond cleavage series.⁷⁴ These product ions are used to determine the *oligosaccharide sequence tag*, which is used to verify whether or not the spectrum is from a glycopeptide.⁷⁴ A score is assigned to the CID spectrum based on this sequence tag.⁷⁴ MS data is used to evaluate and score the relative probability of a glycopeptide by examining the clusters of peptide glycoforms, or those glycopeptides with the same peptide backbone that co-elute within a specific time range.⁷⁴ The glycoform is then identified using the mass of the attached *N*-linked glycan, though the most current version of GlypID allows the entry of user-defined glycan compositions as well.⁷⁴ A limitation to the program is that when low resolution data is used, there is a significant increase in the number of false-positive identifications of glycopeptide microheterogeneities within a cluster. Although the new targeted MS/MS approach has been optimized for FT-MS instrumentation and data, the original GlypID algorithm was designed using LC-MS ion trap data.⁷⁵ A publicly accessible version of the computational tool is currently available online, free of charge to users (see Table 4).

Mayampurath *et al.* have now modified the GlypID algorithm with a scoring function that works to determine glycopeptide composition from high-energy C-trap dissociation (HCD) MS/MS data.⁷⁶ The new software tool, GlypID 2.0, uses high resolution MS¹ data along with CID and HCD scan information to improve the accuracy of *N*-linked glycopeptide identification.⁷⁶ Like the original GlypID, GlypID 2.0 also works to score CID spectra independently on MS systems that do not contain the HCD instrument option.⁷⁶ GlypID 2.0 is freely available to download, as listed in Table 4.

Woodin *et al.*, have also developed a freely accessible web-based tool, GlycoPep Grader (GPG), to assign a glycopeptide composition to MS/MS data in an automated fashion.⁶¹ This tool is specifically designed for data collected in an ion trap mass spectrometer, and it features a novel algorithm that enables users to identify the correct glycopeptide composition from a pool of candidate compositions of the same nominal mass.⁶¹

GPG utilizes the MS/MS data by calculating, scoring, and searching for the expected product ions of potential glycopeptide candidate compositions.⁶¹ The algorithm scores the glycopeptide candidate composition through detection of two types of product ions: 1) Ions that contain the peptide portion and some portion of the pentasaccharide core, or [peptide + core component] ions, and 2) Ions formed via neutral loss of monosaccharide residues from the precursor ion, or [precursor – monosaccharide] ions.⁶¹ The algorithm that powers GPG has been shown to assign the correct glycopeptide candidate after performing the MS/MS peak list search with a very high degree of accuracy.⁶¹

One advantage to the algorithm behind GPG is that the precursor ion's charge state is included in the input data, so all product ions can be searched for in a charge state specific manner.⁶¹ Secondly, no spectral transformation (to singly charged ions) needs to be performed prior to using the program as GPG automatically searches for product ions in a charge-specific fashion, bypassing the need for additional processing software.⁶¹ A disadvantage of the program is that the user must utilize a separate program, such as GlycoMod, to obtain potential matches for the high resolution MS data, prior to assigning the MS² data with GPG.⁶¹ GPG can be found online, and is free to use.

O-LINKED GLYCOPEPTIDES

The analysis of *O*-linked glycoforms is particularly challenging, as no single consensus sequence exists to predict the site of glycan attachment.^{7,77} Further adding to the difficulty of analysis, factors that affect the efficiency of glycosylation at *N*-linked sites are different than those affecting *O*-glycosylation efficiency. For example, the presence of aromatic residues near an *O*-linked site inhibits glycosylation; whereas the presence of an aromatic residue near an *N*-linked site increases the likelihood of site-occupancy.²²

Mucin-type O-linked Glycosylation

The most prevalent form of *O*-linked glycosylation to occur in eukaryotic organisms is mucin-type *O*-glycosylation, which occurs where glycans are attached to a protein by the addition of α -*N*-acetylgalactosamine (GalNAc) residues to the hydroxyl group of Ser/Thr side chains (commonly referred to as the Tn antigen).^{2,7} Though still in the infancy stage, analysis tools have recently been created to assist researchers in the determination of *O*-linked glycoforms, many of which are mucin in type, from MS data.

O-linked Glycopeptide Characterization from MS Data

Recently, Deshpande *et al.* advanced the MS data analysis of *N*- and *O*-linked glycopeptides with the advent of the GlycoSpectrumScan program.⁴¹ GlycoSpectrumScan is designed to analyze LC-MS data of intact glycopeptides from proteolytic digests.⁴¹ The program utilizes MS¹ data to determine glycopeptide composition, along with the relative distribution of glycoforms at each of the sites.⁴¹ In addition, the algorithm behind the program offers a few distinct advantages in that it handles multiply charged ions, making it amenable to both MALDI and ESI data, and is currently freely available online (www.glycospectrumscan.org).⁴¹

GlycoX and GlycoMod, described earlier in the analysis of *N*-linked glycopeptides, are capable of *O*-linked glycopeptide data interpretation as well.^{62,78} Unlike GlycoMod (<http://web.expasy.org/glycomod/>), GlycoX is not publically-available, though it is available upon request from the authors.⁷⁸ GlycoWorkbench, also described previously, performs automation of *O*-linked glycopeptide MS¹ data to elucidate the most likely composition from an experimental peak list in the same manner for *N*-linked glycopeptide spectra.³⁹ GlycoWorkbench is freely available online (<http://download.glycoworkbench.org/>).³⁹

O-linked Glycopeptide Characterization from MS/MS Data

Currently, there are no freely available stand-alone programs designed to automate the analysis of *O*-linked glycopeptide CID MS/MS data through evaluation of both unknown portions of a glycopeptide, the peptide and glycan. The GlycoWorkbench program is capable of annotating glycans in CID fragmentation data of glycopeptides.³⁹ However, as described for the MS/MS characterization of *N*-linked glycopeptides, the identity of the peptide portion must already be known, as GlycoWorkbench solely evaluates the fragmentation of the glycan-containing portion of a glycopeptide.³⁹

There are promising advances being made in the compositional determination of glycopeptides using electron transfer dissociation (ETD) fragmentation techniques,^{11,24} or a combination of CID and ETD, particularly in the study of *O*-linked species.⁷⁹ A recent method described by Daracula *et al.* in which MS¹, CID, and ETD data are used in conjunction with Protein Prospector v5.3 for the identification of SA₁₋₁₀GalGalNAc-containing *O*-linked glycopeptides enriched from bovine serum, demonstrates the potential for automated analysis through a combination of these techniques and database searches.⁷⁹ However, this process is only semi-automated and works only in the case of simple carbohydrate structures are present in the sample.⁷⁹ Hopefully, the compositional information gained between the two complementary fragmentation methods of CID and ETD will enable researchers to develop algorithms and gain insight into creating automated programs to speed the analysis of *O*-linked MS/MS glycopeptide data as well.

CONCLUSION

For the study of protein glycosylation, there are two main approaches used by researchers: Glycan analysis and glycopeptide analysis. The less challenging mode of analysis is to release the glycans from a glycoprotein and analyze them independently. However, the most informative approach is to utilize a protease and cleave the glycoprotein into glycopeptides, retaining information on where each glycan is attached within the protein sequence. The generation of automated MS and MS/MS analysis tools to assist in the characterization of both glycans and glycopeptides is emerging as an effort to facilitate more rapid characterization of glycosylation by either method. Without these analysis tools, data interpretation is a difficult and tedious task. Currently available and recently developed programs for various glycan types, along with their associated advantages and disadvantages, are discussed within this review. With access to the automated tools described here, research in glycomics and glycoproteomics can be greatly facilitated.

Acknowledgments

The authors acknowledge financial support from the National Institutes of Health (RO1RR026061) and an NSF Career award (0645120) to HD, an NSF Fellowship (DGE-0742523) and Pfizer Award to CW, and a Seo Scholarship to MM.

References

1. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002; 12:43R–56R.

2. Hang H, Bertozzi CR. The chemistry and biology of mucin-type *O*-linked glycosylation. *Bioorg Med Chem*. 2005; 13:5021–5034. [PubMed: 16005634]
3. Murrell MP, Yarema KJ, Levchenko A. The systems biology of glycosylation. *Chem Biochem*. 2004; 5:1334–1347.
4. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of *O*-linked glycosylation. *Crit Rev Biochem Mol Biol*. 1998; 33:151–208. [PubMed: 9673446]
5. Bertozzi CR, Kiessling LL. Chemical glycobiology. *Science*. 2001; 291:2357–2364. [PubMed: 11269316]
6. Dennis JW, Granovsky M, Warren C. Protein glycosylation in development and disease. *BioEssays*. 1999; 21:412–421. [PubMed: 10376012]
7. Tian E, Ten Hagen KG. Recent insights into the biological roles of mucin-type *O*-glycosylation. *Glycoconjugate J*. 2009; 26:325–334.
8. Wada Y, Tajiri M, Ohshima S. Quantitation of saccharide compositions of *O*-glycans by mass spectrometry of glycopeptides and its application to rheumatoid arthritis. *J Proteome Res*. 2010; 9:1367–1373. [PubMed: 20104905]
9. Dube DH, Bertozzi CR. Glycans in cancer and inflammation Potential for therapeutics and diagnostics. *Nat Rev Drug Discovery*. 2005; 4:477–488.
10. Lefebvre T, Dehennault V, Guinez C, Olivier S, Drougart L, Mir AM, Mortuaire M, Vercoutter-Edouart AS, Michalski JC. Dysregulation of the nutrient/stress sensor *O*-GlcNAcylation is involved in the etiology of cardiovascular disorders, type-2 diabetes and Alzheimer's disease. *J Biochim Biophys Acta*. 2010; 1800:67–79.
11. Jensen PH, Kolarich D, Packer NH. Mucin-type *O*-glycosylation Putting the pieces together. *FEBS J*. 2010; 277:81–94. [PubMed: 19919547]
12. Marino K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: A path through the maze. *Nat Chem Biol*. 2010; 6:713–723. [PubMed: 20852609]
13. Budnik BA, Lee RS, Steen JAJ. Global methods for protein glycosylation analysis by mass spectrometry. *J Biochim Biophys Acta*. 2006; 1764:1870–1880.
14. Raman R, Raguram S, Venkataraman G, Paulson JC, Sasiekhara R. Glycomics: An integrated systems approach to structure-function relationships of glycans. *Nat Methods*. 2005; 2:817–824. [PubMed: 16278650]
15. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *J Biochim Biophys Acta*. 1999; 1473:4–8.
16. Brazier-Hicks M, Evans KM, Gershater MC, Puschmann H, Steel PG, Edwards R. The *C*-glycosylation of flavonoids in cereals. *J Biol Chem*. 2009; 284:17926–17934. [PubMed: 19411659]
17. Hofsteenge J, Muller DR, de Beer T, Loffler A, Richter WJ, Vliegenthart JFG. New type of linkage between a carbohydrate and a protein: *C*-glycosylation of a specific tryptophan residue in human RNase U_s. *Biochemistry*. 1994; 33:13524–13530. [PubMed: 7947762]
18. Morelle W, Canis K, Chirat F, Faïd V, Michalski JC. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics*. 2006; 6:3993–4015. [PubMed: 16786490]
19. Stepper J, Shasti S, Loo TS, Preston JC, Novak P, Man P, Moore CH, Havlik V, Patchett ML, Norris GE. Cysteine *S*-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett*. 2011; 585:645–650. [PubMed: 21251913]
20. Mazola Y, China G, Mussacchio A. Integrating bioinformatics tools to handle glycosylation. *PLOS Comput Biol*. 2011; 7
21. Jones J, Krag SS, Betenbaugh MJ. Controlling *N*-glycan site occupancy. *Biochim Biophys Acta*. 2005; 1726:121–137. [PubMed: 16126345]
22. Christlet THT, Veluraja K. Database analysis of *O*-glycosylation sites in proteins. *Biophys J*. 2001; 80:952–960. [PubMed: 11159462]
23. Dalpathado DS, Desaire H. Glycopeptide analysis by mass spectrometry. *Analyst*. 2008; 133:731–738. [PubMed: 18493671]
24. North SJ, Hitchen PG, Haslam SM, Dell A. Mass spectrometry in the analysis of *N*-linked and *O*-linked glycans. *Curr Opin Struct Biol*. 2009; 19:498–506. [PubMed: 19577919]

25. Nwosu CC, Seipert RR, Strum JS, Hua SS, An HJ, Zivkovic AM, German BJ, Lebrilla CB. Simultaneous and extensive site-specific *N*- and *O*-glycosylation analysis in protein mixtures. *J Proteome Res.* 2011; 10:2612–2624. [PubMed: 21469647]
26. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckermann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. The Universal Protein Resource (UniProt): An expanding universe of protein information. *Nucleic Acids Res.* 2006; 34:D187–D191. [PubMed: 16381842]
27. Lee TY, Huang HD, Hung JH, Huang HY, Yang YS, Wang TH. dbPTM: An informal repository of protein post-translational modification. *Nucleic Acids Res.* 2006; 34:D622–D627. [PubMed: 16381945]
28. Ranzinger R, Herget S, Wetter T, von der Lieth CW. Glycome DB Integration of open-access carbohydrate structure search databases. *BMC Bioinformatics.* 2008; 9
29. Cooper CA, Harrison MJ, Wilkins MR, Packer NH. GlycoSuiteDB: A new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res.* 2001; 29:332–335. [PubMed: 11125129]
30. Gupta R, Birch H, Rapacki K, Brunak S, Hansen JE. O-GLYCBASE version 4.0: A revised database of *O*-glycosylated proteins. *Nucleic Acids Res.* 1999; 27:370–372. [PubMed: 9847232]
31. Zhang H, Loriaux P, Eng J, Campbell D, Keller A, Moss P, Bonneau R, Zhang N, Zhou Y, Wollscheid B, Cooke K, Yi EC, Lee H, Peskind ER, Zhang J, Smith RD, Aebersold R. UniPep A database for human *N*-linked glycosites: a resource for biomarker discovery. *Genome Biol.* 2006; 7
32. Campbell MP, Roylez L, Radcliffe CM, Dwek RA, Rudd PM. GlycoBase and autoGU: Tools for HPLC-based glycan analysis. *Bioinformatics.* 2008; 24:1214–1216. [PubMed: 18344517]
33. Wang J, Torii M, Liu H, Hart GW, Hu ZZ. dbOGAP An integrated bioinformatics resource for protein *O*-GlcNAcylation. *BMC Bioinformatics.* 2011; 12
34. Goetz JA, Novotny MV, Mechref Y. Enzymatic/chemical release of *O*-glycans allowing MS analysis at high sensitivity. *Anal Chem.* 2009; 81:9546–9552. [PubMed: 19874002]
35. Goldberg D, Sutton-Smith M, Paulson J, Dell A. Automatic annotation of matrix-assisted laser desorption/ionization *N*-glycan spectra. *Proteomics.* 2005; 5:865–876. [PubMed: 15693066]
36. Goldberg D, Bern M, Li B, Lebrilla CB. Automatic determination of *O*-glycan structure from fragmentation spectra. *J Proteome Res.* 2006; 5:1429–1434. [PubMed: 16739994]
37. Vakhrushev SY, Dadimov D, Peter-Katalini J. Software platform for high-throughput glycomics. *Anal Chem.* 2009; 81:3252–3260. [PubMed: 19341273]
38. Vakhrushev, SY.; Dadimov, D.; Peter-Katalini, J. SysBioWare: Structure assignment tool for automated glycomics. *Glyco-Bioinformatics*; Postdam, Germany: 2009. p. 141-161.
39. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkBench: A tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res.* 2008; 7:1650–1659. [PubMed: 18311910]
40. Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. GlycoBase and autoGU: Tools for HPLC-based glycan analysis. *Bioinformatics.* 2008; 24:1214–1216. [PubMed: 18344517]
41. Deshpande N, Jensen PH, Packer NH, Kolarich D. GlycoSpectrumScan: Fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J Proteome Res.* 2010; 9:1063–1075. [PubMed: 20030399]
42. Gaucher SP, Morrow J, Leary J. STAT: A saccharide topology analysis tool used in combination with tandem mass spectrometry. *Anal Chem.* 2000; 72:2331–2336. [PubMed: 10857602]
43. Ashline DJ, Lapadula AJ, Liu Y-H, Lin M, Grace M, Pramanik B, Reinhold VN. Carbohydrate structural isomers analyzed by sequential mass spectrometry. *Anal Chem.* 2007; 79:3830–3842. [PubMed: 17397137]
44. Ethier M, Saba JA, Ens W, Standing KG, Perreault H. Automated structural assignment of derivatized complex *N*-linked oligosaccharides from tandem mass spectra. *Rapid Commun Mass Spectrom.* 2002; 16:1743–1754. [PubMed: 12207362]
45. Lohmann KK, von der Lieth CW. GlycoFragment and GlycoSearchMS: Web tools to support the interpretation of mass spectra of complex carbohydrates. *Nucleic Acids Res.* 2004; 32:W261–W266. [PubMed: 15215392]

46. Lohmann KK, von der Lieth CW. GLYCO-FRAGMENT: A web tool to support the interpretation of mass spectra of complex carbohydrates. *Proteomics*. 2003; 3:2028–2035. [PubMed: 14625865]
47. Tang H, Mechref Y, Novotny M. Automated interpretation of MS/MS spectra of oligosaccharides. *Bioinformatics*. 2005; 21:I431–I439. [PubMed: 15961488]
48. Gao HY. Generation of asparagine-linked glycan structure databases. *J Am Soc Mass Spectrom*. 2009; 20:1739–1742. [PubMed: 19556140]
49. Blow N. A spoonful of sugar. *Nature*. 2009; 457:617–620. [PubMed: 19177129]
50. Apte, A.; Meitei, NS. Bioinformatics in glycomics: Glycan characterization with mass spectrometric data using SimGlycan. In: Li, J., editor. *Methods in Molecular Biology*. Vol. 600. Humana Press; 2009. p. 269-281.
51. Caragea C, Sinapov J, Silvescu A, Dobbs D, Honavar V. Glycosylation site prediction using ensembles of support vector machine classifiers. *BMC Bioinformatics*. 2007;8. [PubMed: 17212835]
52. Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res*. 2005; 33:D242–D246. [PubMed: 15608187]
53. Hamby SE, Hirst JD. Prediction of glycosylation sites using random forests. *BMC Bioinformatics*. 2008; 9
54. Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput; Lyngby, Denmark*. 2002. p. 310-322.
55. Julenius K. NetCGlyc 1.0: Prediction of mammalian C-mannosylation sites. *Glycobiology*. 2007; 17:868–876. [PubMed: 17494086]
56. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate J*. 1998; 15:115–130.
57. Chen YZ, Tang YR, Sheng ZY, Zhang Z. Prediction of mucin-type O-glycosylation sites in mammalian proteins using the composition of *k*-spaced amino acid pairs. *BMC Bioinformatics*. 2008; 9
58. Gerken TA, Jamison O, Perrine CL, Collette JC, Moinova H, Ravi L, Markowitz SD, Shen W, Patel H, Tabak LA. Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J Biol Chem*. 2011; 286:14493–14507. [PubMed: 21349845]
59. Gupta R, Jung E, Gooley AA, Williams KL, Brunak S, Hansen J. Scanning the available *Dictyostelium discoideum* proteome for O-linked GlcNAc glycosylation sites using neural networks. *Glycobiology*. 1999; 9:1009–1022. [PubMed: 10521537]
60. Desaire H, Hua D. When can glycopeptides be assigned based solely on high-resolution mass spectrometry data? *Int J Mass Spectrom*. 2009; 287:21–26.
61. Woodin CL, Hua D, Maxon M, Rebecchi KR, Go EP, Desaire H. GlycoPep Grader: A web-based utility for assigning the composition of N-linked glycopeptides. *Anal Chem*. 2012; 84:4821–4829. [PubMed: 22540370]
62. Cooper CA, Gasteiger E, Packer NH. GlycoMod A software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*. 2001; 1:340–349. [PubMed: 11680880]
63. Go EP, Rebecchi KR, Dalpathado DS, Bandu ML, Zhang Y, Desaire H. GlycoPep DB: A tool for glycopeptide analysis using a “smart search”. *Anal Chem*. 2007; 79:1708–1713. [PubMed: 17297977]
64. Maass K, Ranzinger R, Geyer H, von der Lieth CW, Geyer R. “Glyco-peakfinder” *De novo* composition analysis of glycoconjugates. *Proteomics*. 2007; 7:4435–4444. [PubMed: 18072204]
65. Irungu J, Go EP, Dalpathado DS, Desaire H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal Chem*. 2007; 79:3065–3074. [PubMed: 17348632]
66. Goldberg D, Bern M, Parry S, Sutton-Smith M, Panico M, Morris HR, Dell A. Automated N-glycopeptide identification using a combination of single- and tandem-MS. *J Proteome Res*. 2007; 6:3995–4005. [PubMed: 17727280]
67. Joenvaara S, Ritamo I, Peltoniemi H, Renkonen R. N-glycoproteomics An automated workflow approach. *Glycobiology*. 2008; 18:339–349. [PubMed: 18272656]

68. Peltoniemi H, Joenvaara S, Renkonen R. *De novo* glycan structure search with the CID MS/MS spectra of native *N*-glycopeptides. *Glycobiology*. 2009; 19:707–714. [PubMed: 19270074]
69. Shan B, Ma B, Zhang K. Complexities and algorithms for glycan sequencing using tandem mass spectrometry. *J Bioinform Comput Biol*. 2008; 6:77–91. [PubMed: 18324747]
70. Clerens S, Van den Ende W, Verhaet P, Geenen L, Archens L. Sweet Substitute: A software tool for *in silico* fragmentation of peptide-linked *N*-glycans. *Proteomics*. 2004; 4:629–632. [PubMed: 14997486]
71. Ren JM, Rejtar T, Li L, Karger BL. *N*-glycan structure annotation of glycopeptides using a linearized glycan structure database (GlyDB). *J Proteome Res*. 2007; 6:3162–3173. [PubMed: 17625816]
72. Ozohanics O, Krenyacz J, Ludanyi K, Pollreis F, Vekey K, Drahos L. GlycoMiner: A new software tool to elucidate glycopeptide composition. *Rapid Commun Mass Spectrom*. 2008; 22:3245–3254. [PubMed: 18803335]
73. Pompach P, Chandler K, Lan R, Edwards N, Goldman R. Semi-automated identification of *N*-glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search. *J Proteome Res*. 2012; 11:1728–1740. [PubMed: 22239659]
74. Wu Y, Mechref Y, Klouckova I, Mayampurath AM, Novotny MV, Tang H. Mapping site-specific protein *N*-glycosylations through liquid chromatography/mass spectrometry and targeted tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2010; 24:965–972. [PubMed: 20209665]
75. Wu, Y.; Mechref, Y.; Klouckova, I.; Novotny, MV.; Tang, H. A computational approach for the identification of site-specific protein glycosylations through ion-trap mass spectrometry. In: Ideker, T.; Bafna, V., editors. *Systems Biology and Computational Proteomics*. Vol. 4532. Springer Berlin/Heidelberg; 2007. p. 96-107.
76. Mayampurath AM, Wu Y, Segu ZM, Mechref Y, Tang H. Improving confidence in detection and characterization of protein *N*-glycosylation sites and microheterogeneity. *Rapid Commun Mass Spectrom*. 2011; 25:2007–2019. [PubMed: 21698683]
77. Rakus, JF.; Mahal, LK. New technologies for glycomic analysis: Toward a systematic understanding of the glycome. In: Cooks, RG.; Yeung, ES., editors. *Annual Reviews of Analytical Chemistry*. Vol. 4. Annual Reviews; Palo Alto: 2011. p. 367-392.
78. An HJ, Tillinghast JS, Woodruff DL, Rocke DM, Lebrilla CB. A new computer program (GlycoX) to determine simultaneously the glycosylation sites and oligosaccharide heterogeneity of glycoproteins. *J Proteome Res*. 2006; 5:2800–2808. [PubMed: 17022651]
79. Darula Z, Chalkley RJ, Baker P, Burlingame AL, Medzihradsky KF. Mass spectrometric analysis, automated identification and complete annotation of *O*-linked glycopeptides. *Eur J Mass Spectrom*. 2010; 16:421–428.

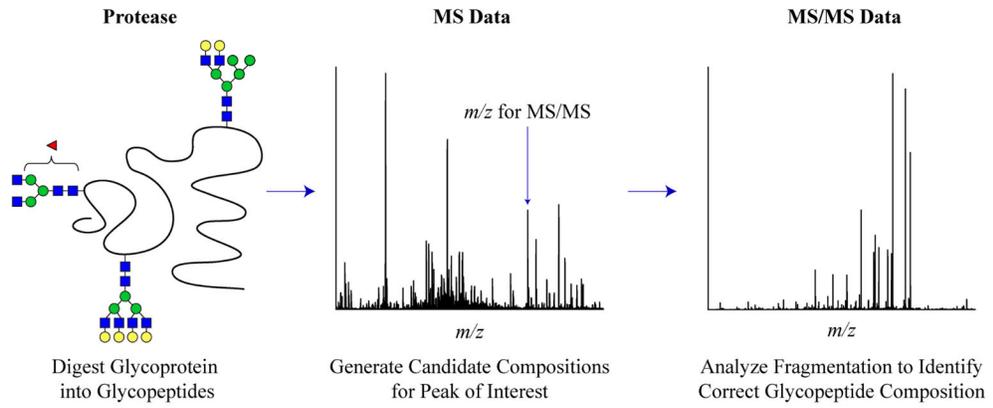


Figure 1.

Flow chart outlining the use of MS and MS/MS data for glycopeptide identification.

Table 1

Glycosylation Databases.

Database	Link to Database	Type	Description
UniProtKB	http://www.uniprot.org/	<i>N</i> -glycos. <i>O</i> glycos. <i>C</i> -glycos.	Contains annotation of <i>N</i> -, <i>O</i> -, and <i>C</i> -linked glycosylation, as well as glycation. Both mammalian and non-mammalian entries are provided.
dbPTM	http://dbptm.mbc.nctu.edu.tw/	<i>N</i> -glycos. <i>O</i> -glycos. <i>C</i> -glycos.	Contains a combinational repertoire of protein PTMs from other databases, including experimentally obtained data on site of modification.
GlycomeDB	http://www.glycome-db.org/	<i>N</i> -glycos. <i>O</i> -glycos. <i>C</i> -glycos.	Contains over 30,000 carbohydrate structures from all major taxonomies.
GlycoSuiteDB	http://glycosuitedb.expasy.org/glycosuite/glycodb	<i>N</i> -glycos. <i>O</i> -glycos.	Contains over 9400 entries of curated and annotated glycans from various organisms.
O-GlycBase	http://www.cbs.dtu.dk/databases/OGLYCBASE/	<i>O</i> -glycos. <i>C</i> -glycos.	Contains over 2000 entries of protein glycosylation sites, the majority of which are <i>O</i> -linked.
UniPep	http://www.unipep.org/	<i>N</i> -glycos.	Contains over 1500 entries of human protein <i>N</i> -linked glycosylation sites.
GlycoBase	http://glycobase.univ-lille1.fr/base/	<i>N</i> -glycos.	Contains HPLC elution positions for 2-AB labeled <i>N</i> -linked glycans from exoglycosidase sequencing and LC-MS data.
dbOGAP	http://cbsb.lombardi.georgetown.edu/hulab/OGAP.html	<i>O</i> -glycos.	Contains over 1100 entries on <i>O</i> -GlcNAcylation sites.

Table 2

Tools available online to facilitate glycan characterization from MS and MS/MS data.

MS Analysis Tool	Link to Tool	Concept and Data Type
GlycoWorkbench	http://download.glycoworkbench.org/	Identifies and annotates MS and MS/MS data with appropriate glycan compositions or fragments.
GlycanBuilder	http://live.glycanbuilder.org/	Drawing tool interfacing with GlycoWorkbench that displays different stereochemical representations of glycans.
GlycoSpectrumScan	http://www.glycospectrumscan.org.	Quantitatively identifies <i>N</i> - and <i>O</i> -linked glycoforms within a protein using LC-MS Data.
GlycoFragment	http://www.glycosciences.de/tools/GlycoFragments/fragment.php4	Identifies and displays the main product ions expected for oligosaccharide MS/MS data.
GlycoSearchMS	http://www.glycosciences.de/database/start.php?action=form_ms_search	Compares experimental MS/MS data to product ions calculated from an extensive library of <i>N</i> - and <i>O</i> -linked glycans.
SimGlycan	http://www.premierbiosoft.com/glycan/index.html	Predicts the structure of glycans from MS/MS data by matching spectra to a built-in database.

Table 3

Glycosylation Site Prediction Tools.

Database	Link to Prediction Tool	Type	Overview
EnsembleGly	http://turing.cs.iastate.edu/EnsembleGly/	<i>N</i> -glycos. <i>O</i> -glycos. <i>C</i> -glycos.	Uses ensemble learning to predict <i>N</i> -, <i>O</i> -, and <i>C</i> -linked sites, as well as <i>O</i> -linked glycan types.
GlySeq	http://www.glycosciences.de/tools/glyseq/	<i>N</i> -glycos. <i>O</i> -glycos.	Uses the PDB and SwissProt to perform statistical analysis of glycosylation sites.
GPP	http://comp.chem.nottingham.ac.uk/glyco/	<i>N</i> -glycos. <i>O</i> -glycos.	Algorithmically predicts <i>N</i> -linked and <i>O</i> -linked glycosylation.
NetNGlyc	http://www.cbs.dtu.dk/services/NetNGlyc/	<i>N</i> -glycos.	Uses consensus sequence to predict <i>N</i> -glycosylation in human proteins.
NetCGlyc	http://www.cbs.dtu.dk/services/NetCGlyc/	<i>C</i> -glycos.	Predicts sites in mammalian proteins for <i>C</i> -mannosylation attachment.
NetOGlyc	http://www.cbs.dtu.dk/services/NetOGlyc/	<i>O</i> -glycos.	Predicts mucin-type GalNAc <i>O</i> -glycosylation in mammalian proteins.
CKSAAP_OGlysite	http://bioinformatics.cau.edu.cn/zzd_lab/CKSAAP_OGlysite	<i>O</i> -glycos.	Predicts mucin-type <i>O</i> -glycosylation sites in mammalian proteins.
OGPET	http://ogpet.utep.edu/	<i>O</i> -glycos.	Predicts occurrence of mucin-type <i>O</i> -glycosylation in eukaryotic proteins.
YinOYang	http://www.cbs.dtu.dk/services/YinOYang/	<i>O</i> -glycos.	Predicts <i>O</i> - β -GlcNAc attachment sites in eukaryotic proteins.
DictyOGlyc	http://www.cbs.dtu.dk/services/DictyOGlyc/	<i>O</i> -glycos.	Predicts sites for <i>O</i> -GlcNAc attachment in <i>Dictyostelium discoideum</i> proteins.

Table 4

Freely Available *N*-linked Glycopeptide Analysis Tools.

Analysis Tool	Link to Automated Program	Overview
GlycoMod	http://web.expasy.org/glycomod/	GlycoMod determines potential glycopeptide compositions, on the basis of mass, from MS data.
GlycoPep DB	http://hexose.chem.ku.edu/glycop.htm	GlycoPep DB deduces possible biologically relevant glycan compositions from glycopeptide MS data using a "smart search".
GlycoSpectrumScan	http://www.glycospectrumscan.org .	GlycoSpectrumScan searches LC-MS data to identify glycopeptides and determine glycoform location.
GlycoWorkbench	http://download.glycoworkbench.org/	GlycoWorkbench was designed for the annotation of glycopeptide MS/MS data through fragmentation analysis and scoring of only the glycan portion.
Glyco-Peakfinder	http://glyco-peakfinder.org/	Glyco-Peakfinder is capable of <i>de novo</i> glycopeptide analysis through glycan fragmentation profiling for spectra after the peptide sequence is input by the user.
GlycoPep ID	http://hexose.chem.ku.edu/predictiontable2.php	GlycoPep ID works to analyze glycopeptides from MS/MS spectra of complex mixtures by identifying the peptide portion based on expected product ions.
GlycoMiner	http://www.chemres.hu/ms/glycominer/tutorial.html	GlycoMiner is designed to identify glycopeptides from qTOF MS/MS data, and assigns composition for spectra of quality containing specific marker ions.
GPS	http://edwardslab.bmcb.georgetown.edu/software/GlycoPeptideSearch.html	GPS utilizes a glycan database to generate glycopeptide compositions after searching LC-MS/MS data of purified proteins and matching fragmentation patterns.
GlypID	http://www.cbs.dtu.dk/services/DictyOGlyc/	GlypID is designed to identify glycopeptides from LC-MS/MS experiments using a combination of MS ¹ and MS ² data.
GlypID 2.0	http://mendel.informatics.indiana.edu/~chuyu/glypID/software.html	GlypID 2.0 uses CID and HCD information to identify glycan type, monosaccharide composition, and attachment site for <i>N</i> -linked glycopeptide MS/MS data.
GPG	http://glycopro.chem.ku.edu/GPGHome.php	GPG scores glycopeptide candidates after searching MS/MS data for the predicted product ions for each composition tested.