glyXtool^{MS}: An Open-Source Pipeline for Semi-Automated Analysis of Glycopeptide Mass Spectrometry Data

Markus Pioch[1], Marcus Hoffmann[1], Alexander Pralow[1,2], Udo Reichl[1,3], Erdmann Rapp[1,2]
1. Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, Magdeburg, Germany
2. glyXera GmbH, Magdeburg, Germany
3. Otto-von-Guericke University, Bioprocess Engineering, Magdeburg, Germany

Keywords: *open-source, bioinformatics, OpenMS, glycopeptide, mass spectrometry, glycoproteomics, N-/O-glycosylation*

Abbreviations: *Immunoglobulin gamma (IgG), mass spectrometry (MS), precursor ion scan (MS1), fragment ion scan (MS2), normalized collision energy (NCE), collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), The OpenMS Proteomics Pipeline Assistant (TOPPAS), reversed phase (RP), liquid chromatography (LC), electrospray ionization (ESI), orbitrap (OT), mass spectrometry (MS), tandem mass spectrometry (MS/MS), N-acetylhexosamine (N or HexNAc), N-acetylglucosamine (GlcNAc), hexose (H or Hex), N-acetylneuraminic acid (Sa or NeuAc), N-glycolylneuraminic acid (Sg or NeuGc), fucose (F or dHex), false-discovery rate (FDR)*

ABSTRACT: For glycoproteomic analyses several web tools and standalone software packages have been developed over the recent years. These tools try to support or replace the time-consuming, cumbersome and error-prone manual spectra analysis and glycopeptide identification. However, existing software tools are usually tailored to one fragmentation technique and only present the final analysis results. This makes manual inspection and correction of intermediate results difficult or even impossible. We solved this problem by dividing the analysis tasks into modular tools with defined functions, which are executed within a software pipeline with a graphical editor. This gives users a maximum of flexibility and control over the progress of analyses. Here, we present the open-source python software suite glyXtool^{MS}, developed for the semi-automated analysis of *N*- and *O*-glycopeptide fragmentation data. glyXtool^{MS} is built around the pipeline engine of OpenMS (TOPPAS) and provides a glycopeptide analysis toolbox for the analysis, interpretation and visualization of glycopeptide spectra. The toolbox encompasses (a) filtering of fragment spectra using a scoring scheme for oxonium ions, (b) in-silico digest of protein sequences to collect glycopeptide candidates, (c) precursor matching to possible glycan compositions and peptide sequences, and finally (d) an annotation tool for glycopeptide fragment ions. The resulting analysis file can be visualized by the *glyXtool^{MS} Evaluator*, enabling further manual analysis, including inspection, verification, and various other options. Using higher energy collisional dissociation data from human immunoglobulin gamma (IgG) and human fibrinogen tryptic digests, we show that glyXtool^{MS} enables a fast, flexible and transparent analysis of *N*- and *O*-glycopeptide samples, providing the user a versatile tool even for explorative data analysis. glyXtool^{MS} is freely available online on https://github.com/glyXera/glyXtoolMS licensed under the GPL-3.0 open-source license. The test data are available via ProteomeXchange with identifier PXD009716.

Protein glycosylation is one of the most common co-translational modifications of proteins in eukaryotes.[1] Despite consisting of only a small number of monosaccharide building blocks, various topologies, branching and linkage variations can yield a very high number of glycan structures.[2] The presence of such structures on a protein can heavily influence glycan properties and thus their biological role involving intercellular adhesion, cell growth, immune response or the protein folding and protein stability.[3] In case of *N*-glycosylation, glycans are linked to the protein backbone via the amino group of asparagine and site occupation is limited to the consensus sequence of Asn-X-Ser/Thr, with X being any amino acid except proline.[4] *O*-glycosylation occurs on the hydroxyl group of either serine or threonine with no known consensus sequence. *N*-glycans share a common trimannosyl core structure,[5] while for *O*-glycans at least eight core structures have been identified so far.[6]

In contrast to glycomics and proteomics, the analysis of glycopeptides with mass spectrometry allows the simultaneous study of the glycan and peptide moiety, which enables the site-specific analysis of protein glycosylation.[7] Glycopeptides are typically generated through proteolytic cleavage – most commonly by trypsin. In some cases a lack of tryptic cleavage sites in the vicinity of potential glycosylation sites, requires using proteases with a broader cleavage specificity.[8-11] Due to the normally lower abundance of glycopeptides within the digested peptide mix, and the suppression of the glycopeptide signal in presence of non-glycosylated peptides, a glycopeptide enrichment step is often required prior to the chromatographic separation and measurement via mass spectrometry[12]. Various fragmentation techniques are used for the study of glycopeptides, such as collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), and electron transfer dissociation (ETD), which generate different fragments of the glycan (B- and Y-ions) and/or the peptide moiety (a-, b-, c-, x-, y- and z-ions).[13-15]

For the analysis of glycopeptide mass spectrometry data, various software tools have been published over the recent years, referenced within several reviews.[16-18] The majority of tools are available as web tools (GlycoMaster,[19] GlycoMod,[20] GlycoPeakFinder,[21] GlycoPepDetector,[22] GlycoPepGrader,[23] GlycoPep ID,[24] GlycopepDB,[25] GlycopeptideID,[26] Protein Prospector[27]). Some tools can be downloaded as standalone software (GlycoFragWork,[28] GlycoPep Evaluator,[29] GlycopeptideSearch,[30] GlycoWorkbench,[31] GlypID 2.0,[32] IGAP,[33] MAGIC,[34] pGlyco[35]). However, about half of the tools are only available on request, or have been discontinued due to funding issues (Sweet Substitute,[36] Sweet-Heart,[37] Branch-and-Bound,[38] GlycoMiner,[39] GlycosidIQ,[40] GlycoSpectrumScan,[41] GlyDB,[42] GPQuest,[43] I-GPA,[44] Peptoonist[45]). As open-source only five tools are available: SweetSEQer,[46] GPFinder[47] (based on GlycoX[48]), GlycoSeq,[49] XGlyScan[50] and GlycoPAT[51]. As commercial tools Byonic (Protein Metrics Inc., San Carlos, CA, USA),[52] ProteinScape (BRUKER DALTONIK GmbH, Bremen),[53,54] and BiopharmaFinder (Thermo Scientific, Waltham, MA, USA) are available.

Hu et.al.[55] extensively reviewed the state of current glycopeptide analysis software and identified significant areas in data analysis requiring further development. This included peptide and glycan structure confirmation using database-based methods, *de novo* sequencing, spectral libraries, validation methods, and glycan quantification. Since many tools complement each other, a focus on tool extension and integration into larger tool pipelines is emphasized. To provide higher flexibility for each stage of the identification, the authors recommend the modularization of those tools.

OpenMS[56] is an open-source software that employs tool modularization as a strategy for proteomics experiments. It provides a variety of small tools with a defined proteomics function, which can be linked to form complex analysis pipelines. These pipelines can be executed with the provided OpenMS Proteomics Pipeline Assistant (TOPPAS) engine.[57] To gain insights into each intermediate analysis step and to ensure suitable tool parameters, all MS data can be visualized using the TOPPView software,[58] together with analysis results provided by each tool. The use of a pipeline engine allows the storage of steps performed, thus enabling data re-analysis if necessary. Another focus of OpenMS is the use of open formats for data exchange.

Here we present the open-source software suite glyXtool^{MS}, which provides a glycopeptide analysis toolbox that can be integrated into an OpenMS pipeline and run within the TOPPAS engine using the "Generic Wrapper" functionality. All tools are written in python, since OpenMS gives access to its native functions through the pyOpenMS library[59] making the code more accessible to other developers. Additionally, the software provides the *glyXtool^{MS} Evaluator* – a tool for the visual inspection, verification and validation of results obtained by the analysis pipeline. In addition, it offers the possibility to check the results of each intermediate step.

To demonstrate the functionality and usability of glyXtool^{MS}, the analysis of *N*-glycopeptides derived from human immunoglobulin gamma (IgG) as well as *O*-glycopeptides from human fibrinogen is shown exemplarily. The data have been deposited to the ProteomeXchange Consortium via the PRIDE[60] partner repository with the dataset identifier PXD009716. glyXtool^{MS} is available online on https://github.com/glyXera/glyXtoolMS licensed under the GPL-3.0 open-source license.

MATERIAL AND METHODS

General Software Setup

For the automated glycopeptide analysis an OpenMS pipeline was created and extended with new glycopeptide analysis tools. To enable the visual inspection of analysis results from the OpenMS pipeline and for further manual data/spectra inspection, verification and validation, the *glyXtool*^{MS} *Evaluator* was developed. The general software setup is depicted in Figure 1. The OpenMS pipeline (A) combines native OpenMS tools which provide basic mass spectrometry data analysis functions, with glyXtool^{MS} tools (python scripts) which supply additional glycopeptide analysis functionality. The pipeline and each tool parameter can be adapted by the user according to the analysis problem. After the pipeline has generated an analysis file, *the glyXtool*^{MS} *Evaluator* (B) can be used to inspect, verify, and validate identification results as well as to review the parameters used for the analysis. The *glyXtool*^{MS} *Evaluator* provides in-depth analysis functionalities as it enables the manual annotation of fragment ion spectra along with the addition of new identifications (manual de novo sequencing). It also allows to remove false-positive identifications. Parts of the pipeline can be run separately by adapting the necessary input nodes (e.g. substituting the preprocessing part with an input node). The glyXtool^{MS} python library (C) provides glycopeptide functionality to the pipeline tools (A), the *glyXtool*^{MS} *Evaluator* (B), and for future glycopeptide analysis tools provided for the OpenMS TOPPAS engine.

Experimental Data

As example data sets tryptic digests of human IgG for the analysis of *N*-glycosylation and human fibrinogen for *O*-glycosylation have been measured by nano reversed phase liquid chromatography coupled online to an electrospray ionization orbitrap mass spectrometer (nano RP-LC ESI- OT-MS/MS; LTQ Orbitrap Elite, Thermo Scientific, Waltham, MA, USA) with HCD fragmentation. For more details on the measurement, please refer to the Supplementary Material Section 1.

Data Preparation

The measured raw data-files have been converted into mzML format using *msconvert* included in ProteoWizard (Version 3.0.7408).[61]

Additionally, a general database of *N*-glycan compositions was generated by parsing structures in GlycoCT format from glycomeDB (downloaded from http://www.glycome-db.org/getDownloadPage.action?page=structure_glycoct, date: 2015-12-16). To distinguish *N*-glycans from *O*-glycans, the presence of the trimannosyl core structure for *N*-glycans was required. Afterwards the remaining compositions have been manually assessed for plausibility. To ensure the inclusion of human IgG and human fibrinogen glycan compositions within the database, a list of compositions was compiled from Selman et al.[62] and Mimura et al.[63] for human IgG. In case of human fibrinogen, *N*- and *O*-glycan compositions reported by Zauner et al.[64] have been used as cross-reference. The test data are available via ProteomeXchange with identifier PXD009716

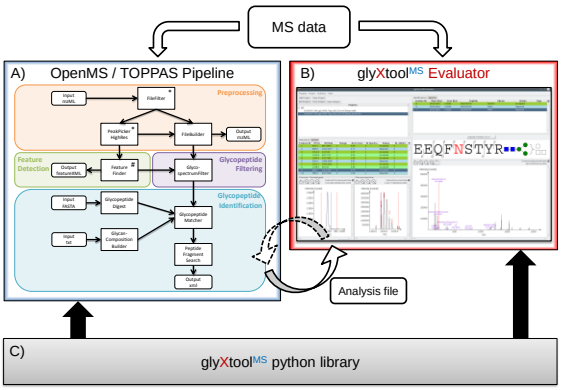


Figure 1. Software. The software suite glyXtoolMS consists of three parts: A) a set of glycopeptide tools that can be used within an OpenMS pipeline for the automated generation of an analysis file from the mass spectrometry data, B) a visual interface for further manual analysis of the analysis file, and C) the glyXtoolMS python library, which contains reusable functions for the glycopeptide analysis.

Software Libraries

For the development of glyXtool^{MS}, python™ 2.7.3 was used (www.python.org). The OpenMS software (Version 1.11) was cloned from the development branch and compiled according to the installation notes. Additionally, the generation of the pyOpenMS package was included during compilation, in order to use the OpenMS functionality within python scripts.

Multiple packages were used to extend the python functionality, most of which were installed via PIP version 1.5.4 (pypi.python.org/pypi/pip). For mathematical operations and matrix calculations numPy 1.6.1 (www.numpy.org) was the general choice. For presentation of results in an excel spreadsheet, package xlwt version 1.0.0 (pypi.python.org/pypi/xlwt) was included. For parsing and writing files in xml format, the software lxml 2.3.2 (www.lxml.de) provided the necessary python bindings to interface with the C library libxml2. For reading config-files, the utility configparser 3.3.0r2 (pypi.python.org/pypi/configparser) has been included. In order to parse command line options, the package argparse 1.2.1 (pypi.python.org/pypi/argparse) was used.

glyXtool^{MS} Python Package

All written software code has been compiled into one python package. The package is available under https://github.com/glyXera/glyXtoolMS under the GPL-3.0 open-source license.

Software Requirements and Limitations

The software has been tested on Linux and Windows machines. The setup currently needs a working version of OpenMS together with pyOpenMS as well as a python 2.7.x installation.

Software Comparison

The performance of glyXtool^{MS} was compared to the open-source software MAGIC-web (http://magic.iis.sinica.edu.tw/index.html)[34] and the commercially available Byonic software[52] (Protein Metrics Inc., San Carlos, CA, USA) version v2.11.0

together with Byologic version v2.7-29. To this end a human IgG dataset was analyzed with each software.

For the analysis with MAGIC, the targeted approach (MAGC+) as well as the untargeted approach (MAGIC + Mascot + Results Integrator) were used. The results of both approaches were combined. The necessary mgf file was generated from the mzML file using the *FileConverter* tool within OpenMS. Due to parsing errors during the file upload into MAGIC+, the scan titles within the mgf file required renaming to 'scan' + scan number, which corrected the parsing error. For the analysis, the suggested standard parameters were used, except for the case that the monosaccharides "Pentose" and "Neu5Gc" were unselected.

In case of the Byonic/Byologic analysis, the parameters were set to a tryptic digest with full digestion specificity, a mass tolerance of 10 ppm and CID low energy were chosen as a fragmentation type. As amino acid modifications carbamidomethyl on cysteine and oxidation on methionine as "common1" modifications were included. For the glycan composition database, the provided "*N*-Glycan 50 common biantennary" file was used.

RESULTS AND DISCUSSION

A software package has been developed for the flexible and transparent analysis of glycopeptide mass spectrometry data. It consists of two major parts: a collection of glycopeptide specific tools for the automated processing within an OpenMS TOPPAS engine, and a graphical user interface called *glyXtool*^{MS} *Evaluator* for the assisted inspection, verification and validation of the results. The use of the OpenMS TOPPAS engine enables the user to create flexible analysis pipelines, since it supports rearrangement, addition or removal of tools as well as control over the analysis parameters. Here the purpose of each processing step and its tools shown in Figure 2 will be illustrated using human IgG and human fibrinogen as example data sets.

Preprocessing

The preprocessing steps are used to generate uniform input mass spectrometry data for the glycopeptide analysis tools. Required are profile precursor (MS^1) scans together with centroided fragment (MS^2) scans, sorted by increasing retention time. Sorting the spectra is handled by the *FileFilter* tool of OpenMS, while the *PeakPicker* and the *FileBuilder* tools handle the spectra type conversion. Depending on the type of mass spectrometer or the vendor format, the requirements can already be fulfilled and the preprocessing steps can be removed from the pipeline.

Feature Detection

As a data reduction measure, a FeatureFinder tool is used that establishes a link between fragment spectra and individual analytes by using the feature border and precursor positions (see Supplementary Figure S-2). Additionally, the charge state and the monoisotopic masses are corrected, which can be useful in case the mass spectrometer reports only an average monoisotopic precursor mass.

OpenMS provides a wide range of FeatureFinders, for example the *FeatureFinder Centroided* tool which has been tested within our analysis pipeline. However, the focus of the *FeatureFinder Centroided* is on quantification. Thus, it removes features with low intensity or poor chromatographic peak shape, ultimately leading to fragment spectra without a feature. To ensure a high coverage of fragment spectra with features, a new FeatureFinder called *FeatureFinderMS* has been developed that uses each precursor position as a starting seed. Theoretical isotopic patterns are used to create a feature even for low-quality precursors. A detailed description of the *FeatureFinderMS* tool can be found in the Supplementary Material Section 2.1 and Supplementary Figure S-3.

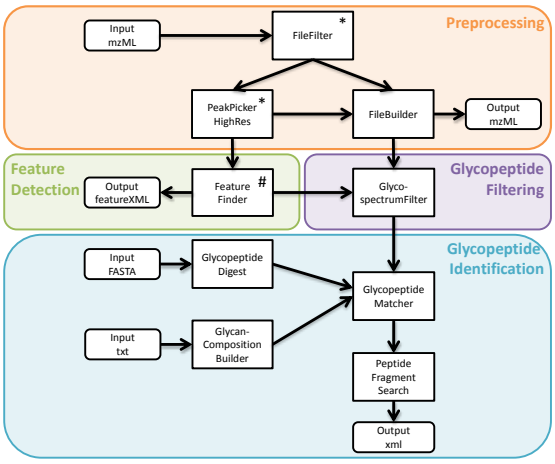


Figure 2. This schematic shows the general TOPPAS pipeline including input nodes, processing tools, output nodes and all connections utilized for the automated glycopeptide analysis. Native OpenMS tools are depicted with a '*'; for the FeatureFinder an alternative FeatureFinder tool was developed which can be used instead, marked by a '#'. The remaining tools are python scripts run with the Generic Wrapper functionality in OpenMS. Inputs are the raw data in mzML format, possible protein sequences in FASTA file format, and a text file with provisional glycan compositions. The output node stores the full glycopeptide analysis results in xml file format. The full TOPPAS pipeline can be found in the Supplementary Figure S-1.

Glycopeptide Filtering

The *GlycospectrumFilter* tool classifies the provided fragment spectra into glycopeptide spectra and non-glycopeptide spectra. The underlying scoring principle is based on the scoring algorithm for *O*-GlcNAc peptides from Hahne et al.[65], and was extended to process *N*- and *O*-glycan compositions. Fragment ions of each spectrum are matched against oxonium ions defined in Table 1 and oxonium losses from the precursor, shown in Supplementary Figure S-4. To minimize false-positive matches, an ion is only matched if its identity is in agreement with other already identified ions (these dependency rules are defined in Table 1). Consideration of certain oxonium ions can be excluded manually within the tool parameters, e.g. presence of a specific type of sialic acid (NeuAc or NeuGc) or fucosylation. Based on the peak intensity (normalized to the total spectrum intensity) and peak ranking of all matched ions, a spectrum score

is calculated. Due to the log-negative scoring, a lower score signifies a higher probability of the spectrum to be a glycopeptide spectrum.

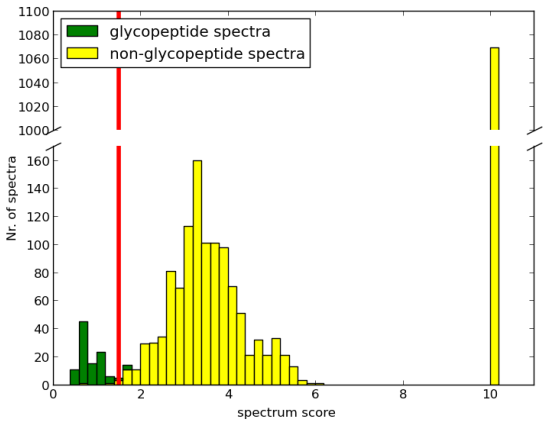


Figure 3. Glycopeptide score distribution of the MS^2 spectra. The stacked bar chart shows the score distribution of the individual fragment spectra from the human IgG sample and the threshold used for classification. Scoring has been performed with a ±0.05 Da mass accuracy setting. Each spectrum in the sample has been manually assessed as glycopeptide/non-glycopeptide. Visible are two distinct populations between the glycopeptides (green, 174 spectra) and the non-glycopeptides (yellow, 839 spectra). For non-glycopeptide spectra that do not contain masses which match oxonium ions or oxonium losses, no score could be calculated, thus a default score of 10 has been assigned. With a threshold of 1.5 the glycopeptide population can be separated from the non-glycopeptide spectra (red line).

Figure 3 shows the resulting histogram for the spectrum score distribution of all fragment spectra acquired for the IgG sample. To determine false positives and false negatives, each spectrum has been manually assessed for its glycopeptide identity, which allows the color coding of the distributions: the glycopeptide spectra distribution is shown in green; the broader distribution of non-glycopeptide spectra is shown in yellow. For non-glycopeptide spectra without any ion matches, no score can be calculated. Therefore, a default score of 10.0 was assigned to these spectra. The populations can be distinguished via a threshold of 1.5, leading to 11 false-positive, 1 false-negative, 2168 true-negative, and 102 true-positive glycopeptide identifications within the IgG dataset. Example spectra of each group can be found in the Supplementary Figure S-5. From the receiver operating characteristic curve (ROC plot) shown in Supplementary Figure S-6, the highest accuracy of 0.996 was achieved with a threshold of 1.41.

Scored fragment spectra sharing the same precursor are then grouped into one feature (visualized in Supplementary Figure S-2). By using the grouping information, a consensus spectrum of the fragment spectra is created for each glycopeptide feature by merging up to 300 peaks from each fragment spectrum. Afterwards, the resulting consensus spectrum is used in the subsequent glycopeptide identification steps. The tool reports results in xml format, containing information about the fragment spectra and the features.

Table 1. Oxonium ions used for glycopeptide scoring

Type	Oxonium ion	Mass [M+H+]	Depends on
General	N_1	204.0867	H_1 - H_2O
	N_1 - H_2O	186.0761	N_1
	H_1	163.0601	H_1 - H_2O
	H_1 - H_2O	145.0495	H_1
	N_1H_1	366.1395	-
	N_1H_2	528.1923	-
Contains NeuAc	Sa_1	292.1027	Sa_1- H_2O
	Sa_1- H_2O	274.0921	Sa_1
	$N_1H_1Sa_1$	657.2349	Sa_1
	N_1H_2Sa1	819.2877	Sa_1
Contains NeuGc	Sg_1	308.0976	Sg_1 - H_2O
	Sg_1 - H_2O	290.0870	Sg_1
	N_1Sg_1	511.1770	Sg_1
	$N_1H_1Sg_1$	673.2298	Sg_1
	$N_1H_1Sg_2$	980.3201	Sg_1
Contains NeuAc and NeuGc	$N_1H_1Sa_1Sg_1$	964.3252	NeuAc1 Sg_1
Contains fucose	F_1	147.0652	F_1 - H_2O
	F_1 - H_2O	129.0546	F_1
	$N_1H_1F_1$	512.1974	F_1
Contains NeuAc and fucose	$N_1H_1Sa_1F_1$	803.2928	Sa_1F_1

Here, the masses of all oxonium ions used for the scoring of the glycopeptide spectra are listed. The oxonium ions are classified into types to allow the selective scoring according to the analysis. From manually annotated glycopeptide spectra a list of dependencies has been created, which reduces false-positive assignments. The oxonium ions are only used to calculate the final score if the dependent oxonium ion(s) as listed above were also detected within the spectrum. For example, HexNAc1 and HexNAc1 - H2O are only scored if both ions are present within the spectrum.

Glycopeptide Identification

After the search for glycopeptide features, a precursor mass matching to all possible combinations of peptide and glycan masses is performed by the *GlycopeptideMatcher* tool. For the glycan composition input either a file containing possible glycan compositions can be supplied, or the *GlycanCompositionBuilder* tool generates a set of possible compositions *in silico*. The theoretical peptide masses are generated by the *GlycopeptideDigest* tool, which uses protein sequences provided by the user to identify peptides containing an *N*- and/or *O*-glycosylation site. The output of the *GlycopeptideDigest* tool is a file in xml format, containing all possible peptide sequences, relying on to the number of missed cleavages, fixed and variable modifications. Also reported are the positions of the glycosylation sites and the peptide mass.

Peptide Fragment Search

In case of the example IgG dataset, HCD has been used as a fragmentation technique that generates a number of peptide fragments, which are predominantly single charged but also double charged y- and b-ions, together with the peptide (Y_0), the peptide with ammonia loss (Y_0-NH_3), and the peptide plus HexNAc (Y_1) ions). The existence of these ions can be used to reduce false-positive identifications from the

Glycopeptide Matcher tool. The *Peptide Fragment Search* tool generates the theoretical a, b-, c- and x-, y-, z-ion series based on the peptide sequence and its modifications, and matches it against the fragment ions of the consensus spectrum with a mass tolerance defined by the user.

glyXtool^{MS} Evaluator

The results from each analysis tool are collected within an analysis file in xml format, which is the input for the *glyXtool^{MS} Evaluator* for further manual evaluation. The *glyXtool^{MS} Evaluator* visualizes the results of each tool, provides the opportunity to screen and edit possible false-positive results (partially also false-negative results), and enables further manual analysis. A screenshot of the software tool with the loaded IgG analysis is shown in Figure 4.

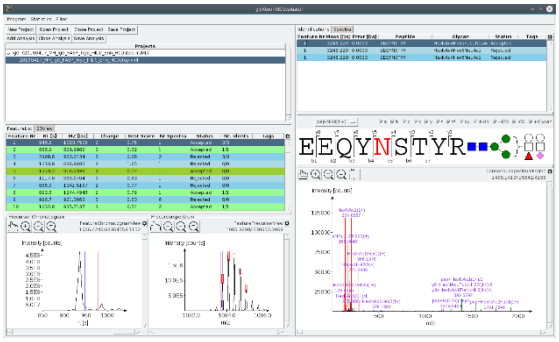


Figure 4. This screenshot shows the results after loading the mass spectrometry raw data file and an analysis file from the OpenMS pipeline into the *glyXtool^{MS} Evaluator*. On the lower left side all features classified as glycopeptide features are shown. For the selected feature the extracted ion chromatogram of the precursor and its isotopic pattern are visualized, together with its consensus fragment spectrum on the lower right side. Within the spectrum, the oxonium ions found by the *GlycospectrumFilter* tool are marked in red. The identifications tab shows all identifications for the selected feature, suggested by the *Glycopeptide Matcher* tool. By selecting an identification, the fragment spectrum is updated with the y- and b- ion series of the peptide sequence and other glycopeptide fragments calculated by the *Peptide Fragment Search* tool.

The "Projects" tab shows the loaded project consisting of one mass spectrometry file (in mzML format, created by the *FileBuilder* tool). This file is used by the *glyXtool^{MS} Evaluator* to access the raw data in order to show isotopic patterns and elution peaks. To the mass spectrometry file, the corresponding analysis file containing the spectra scoring results and glycopeptide identifications is loaded. Loading multiple additional analysis files is also supported, which is helpful to investigate the influence of different parameter settings during the analysis. The "Analysis" tab in Figure 4 visualizes the different tool results (shown here is the "identification" tab where the *GlycopeptideMatcher* and the *PeptideFragmentSearch* results are visualized). In addition, the fragment spectrum and the peptide sequence coverage are shown for the selected identification. A manual status ("Accepted", "Rejected" and "Unknown") can be set after manual review of the identification. On the upper right of Figure 4, the mass deviation of each identification is plotted, which supports verification of the selected accuracy in the *GlycopeptideMatcher* tool.

Multiple functions are provided for further in-depth manual analysis. A filter function allows the selection of the results through various properties, like precursor mass, existence of fragment ions, glycosylation site, and more. Additionally, regular expressions can be used on glycan and peptide names. The fragmentation spectra can be annotated manually with a mass distance ruler. Identifications can be added or modified by assigning a glycan composition and peptide sequence to an existing feature. The analysis file can also be reanalyzed by the automated pipeline engine, if for example features or other parameters have been manually changed.

Table 2. Results of *N*-glycopeptide analysis for human IgG

Peptide/ Glycan	IgG 1 EEQYNSTYR	IgG 2 EEQFNSTFR	IgG 3/ IgG 4 EEQYNSTFR/ EEQFNSTYR
N_4H_3	830.0^{3+} [14.8] 1244.5^{2+} [14.9]		
N_4H_4	1325.5^{2+} [14.7]		
N_4H_5	938.0^{3+} [14.5]		
N_5H_4		941.0^{3+} [17.1]	
$N_3H_3F_1$*	1216.0^{2+} [14.7]	1200.0^{2+} [17.0]	
$N_3H_4F_1$*	1297.0^{2+} [14.5]	1281.0^{2+} [16.9]	
$N_4H_3F_1$	878.7^{3+} [14.7] 1317.5^{2+} [15.1]	868.0^{3+} [17.0] 1301.5^{2+} [16.7]	873.4^{3+} [15.8] 1309.5^{2+} [15.8]
$N_4H_4F_1$	932.7^{3+} [15.4] 1398.6^{2+} [14.9]	922.0^{3+} [16.6] 1382.6^{2+} [17.5]	1390.6^{2+} [15.8]
$N_4H_5F_1$	1479.6^{2+} [14.8]	976.1^{3+} [16.9] 1463.6^{2+} [16.9]	
$N_4H_6F_1$*	1040.7^{3+} [14.2]	1030.1^{3+} [16.8]	
$N_5H_3F_1$	946.4^{3+} [14.9]	935.7^{3+} [17.2] 1403.1^{2+} [17.2]	
$N_5H_4F_1$	1000.4^{3+} [13.2] 1000.4^{3+} [15.0]	989.7^{3+} [17.1] 1484.1^{2+} [17.0]	
$N_5H_5F_1$	1054.4^{3+} [14.9]	1043.8^{3+} [16.9]	
$N_4H_4F_1Sa_1$		1019.1^{3+} [19.4] 1528.1^{2+} [19.4]	
$N_4H_5F_1Sa_1$	1083.8^{3+} [15.8]	1073.1^{3+} [19.9]	
$N_5H_5F_1Sa_1$	1151.4^{3+} [15.6]		

The table shows the peptide/glycan matches found by the glycopeptide matcher tool. Each cell shows the detected ion masses, the charge states, and - in brackets - the retention time in minutes. Glycan names are abbreviated as "N": HexNAc, "H": Hexose, "Sa": *N*-acetylneuraminic acid, "F": Fucose. All glycan compositions except the three marked with an asterix "*" were reported by Selman et al.[62] and Mimura et al.[63]

Analysis Results

The *reporter* tool from the analysis pipeline can convert the content of the xml formatted result file into an excel spreadsheet. This shows the scores of each MS^2 spectrum of the *GlycospectrumFilter* tool, along

with the identified glycopeptide features, and the corresponding identifications. As an example, results for IgG is shown in Table 2. In addition, the identified *O*-glycopeptides from human fibrinogen are shown in Table3A+B. For each peptide and glycan composition the detected precursor mass and charge state is depicted.

Thirteen of twenty-four *N*-glycan compositions reported for human IgG by Selman et al.[54] and Mimura et al.[55] were identified by glyXtool^{MS}, each supported by the existence of oxonium ions and peptide fragments. Fourteen *N*-glycopeptides were found for IgG1 and twelve glycopeptides for IgG2. The peptide fragments were not sufficient to distinguish the peptide sequence isomers of IgG3 and IgG4, thus the two glycopeptides were assigned to both IgG species. Due to an isomeric mass equivalence between IgG1 containing fucosylated glycans and Ig3/IgG4 containing glycans with a hexose instead of a fucose, 33 false-positive assignments had to be manually removed in the *glyXtool^{MS} Evaluator* by filtering the fragment spectra for the existence of the peptide and peptide-NH_3 ions. Overall, eight false-positive identifications remained with 40 true-positive identifications. Some of the glycan compositions reported for human IgG could not be detected within the sample (N_3H_3, N_5H_3, N_5H_5, the single sialylated species $N_4H_4Sa_1$, $N_4H_5Sa_1$, $N_5H_4Sa_1$, $N_5H_4F_1Sa_1$, $N_5H_5Sa_1$ and the double sialylated species $N_4H_5Sa_2$, $N_5H_5Sa_2$, $N_4H_5F_1Sa_2$). These glycan compositions were either too low abundant within the sample or were not fragmented during the analysis. However, three new *N*-glycan compositions ($N_3H_3F_1$, $N_3H_4F_1$, and $N_4H_6F_1$) were detected on IgG1 and IgG2 (marked with an asterisk Table 2) that were previously not reported. The compositions $N_3H_3F_1$ and $N_3H_4F_1$could be clearly assigned using glyXtool^{MS} via the corresponding peptide fragments. However, the fragment ion spectra identified for the composition $N_4H_6F_1$ in IgG1 and IgG2 show a mass deviation in the higher m/z range (> 30 ppm). Accordingly, glycopeptide fragment ions could not be assigned automatically, but could be identified manually. The same behavior was also observed for the IgG1 glycopeptide fragment ion spectrum identified with the *N*-glycan composition N_4H_3 (precursor m/z 830.0^{3+}).

Table 3. Results of N- and O-glycopeptide analysis for human fibrinogen

A) sp|P02671|FIBA_HUMAN	**N_1H_1**	**$N_1H_1Sa_1$**	**$N_1H_1Sa_2$**
$_{253}$GGSTSYGTGSETESPR$_{268}$	969.4^{2+} [14.4]	1115.0^{2+} [15.7]	
$_{494}$HPDEAAFFDTASTGK$_{508}$		1125.5^{2+} [21.4] 750.7^{3+} [21.4]	847.7^{3+} [24.2]
$_{540}$ESSSHHPGIAEFPSR$_{554}$			862.4^{3+} [19.5]
$_{562}$QFTSSTSYNR**573**		923.9^{2+} [16.6]	
$_{584}$MADEAGSEADHEGTHSTK$_{601}$		843.7^{3+} [13.3]	

B) sp|P02675|FIBB_HUMAN	**$N_1H_1Sa_1$**	**$N_4H_4Sa_1$**	**N_4H_5**	**$N_4H_5Sa_1$**	**$N_4H_5Sa_2$**
$_{24}$EEAPSLRPAPPPISGGGYR$_{42}$	869.7^{3+} [22.3]				
$_{347}$GTAGNALMDGASQLM(MSO)GENR$_{365}$			1177.8^{3+} [21.3]	956.4^{4+} [24.7]	
$_{347}$GTAGNALM(MSO)DGASQLM(MSO)GENR$_{365}$				960.4^{4+} [21.1] 1280.2^{3+} [22.0]	1377.2^{3+} [26.6]
$_{347}$GTAGNALM(MSO)DGASQLMGENR$_{365}$		1220.8^{3+} [24.8]	1177.8^{3+} [22.1]	1274.8^{3+} [25.8]	1371.9^{3+} [30.5] 1029.2^{4+} [29.5]
$_{347}$GTAGNALMDGASQLMGENR$_{365}$			1172.5^{3+} [24.9]	1269.5^{3+} [25.5] 1269.5^{3+} [27.8] 952.4^{4+} [27.7]	1025.2^{4+} [31.4] 1025.2^{4+} [29.0] 1366.5^{3+} [31.9]

Several glycopeptides could be found within the human fibrinogen dataset for the alpha and the beta chain. Table A shows the detected glycopeptide ions for the alpha chain, which contains only *O*-glycosylation. Table B shows the glycosylation of the beta chain, containing one *O*-glycopeptide and multiple *N*-glycans for the peptide $_{347}$GTAGNALMDGASQLMGENR$_{365}$, which exhibits different degrees of oxidation of methionine.

The human fibrinogen sample shown in Table3A+B contains both *N*- and *O*-glycopeptides. The provisional peptides used in the analysis were generated from all three human fibrinogen chain sequences. Two searches with different glycan composition databases were performed: a) only with mucin-type *O*-glycans, and b) with the full *N*-glycan database derived from glycomeDB plus the mucin-type *O*-glycans. The automated analysis with the pipeline using only the *O*-glycan database resulted in nine correctly identified glycopeptides together with 15 wrong assignments. Fourteen of the wrong assignments could be removed during manual analysis with the *glyXtool^{MS} Evaluator* using the existence of the peptide ion as a filter. For the analysis with the *N*- and *O*-glycan database, the pipeline generated a large number of false-positives (29 true-positives and 76 false-positives). By using the filter function within the *glyXtool^{MS} Evaluator* for the selection of identifications containing the peptide and peptide-NH_3 ions, the large number of false-positives could be reduced to seven false-positives, while also losing one true-positive assignment.

The *O*-glycosylation of the fibrinogen alpha chain in Table3A contained three different *O*-glycan compositions: N_1H_1, $N_1H_1Sa_1$ and $N_1H_1Sa_2$. Four identified peptides were within the glycosylated regions reported by Zauner et al.,[56] while the glycopeptide $_{584}$MADEAGSEADHEGTHSTK$_{601}$-$N_1H_1Sa_1$ has not been reported, yet. The glycosylation of the beta chain is shown in Table3B. It comprises one *O*-glycopeptide

and four different *N*-glycans on the *N*-glycosylation site N_{364}. The three *N*-glycan compositions N_4H_5, $N_4H_5Sa_1$ and $N_4H_5Sa_2$ were also reported in Zauner et al.; additionally, we could detect $N_4H_4Sa_1$ in the sample, which can be explained as a sub-composition of $N_4H_5Sa_1$, missing one hexose. The reported *N*-glycosylation N_{52} on the gamma chain could not be detected in our sample.

Software Comparison

We compared the performance of glyXtool^{MS} with the commercially available software Byonic/Biologic as well as the recently published MAGIC software[31] on the results of the human IgG sample. The results are shown within Table 4.

28 of a total of 30 glycopeptides were detected by glyXtool^{MS}, 24 with Byonic, and 19 using MAGIC. Of overall sixteen glycopeptides for IgG1, fourteen were detected by glyXtool^{MS}, thirteen by Byonic, and ten by MAGIC. All twelve glycopeptides of IgG2 were detected by glyXtool^{MS}, nine by Byonic and eight by MAGIC. The two glycopeptides of IgG3/IgG4 were detected by glyXtool^{MS} and Byonic, while the composition $N_4H_4F_1$ was missed by MAGIC.

The limitation of MAGIC is mainly due to its approach regarding the detection of peptide sequences. In particular, the assumption that the triplet fragment ion pattern (Y_0, Y_0-NH_3, Y_1) of the peptide and the trimannosyl core must exist within the fragment spectra. Additionally, Mascot searches did not result in correct peptide assignments in several cases due to unknown reasons.

All three tools reported additional unique glycopeptides, which were not detected previously by the other tools. The glycan composition $N_4H_6F_1$ on IgG1, and the glycan compositions N_5H_4 and $N_4H_4F_1Sa_1$ on IgG2 were only reported by glyXtool^{MS}. The reason why both Byonic and MAGIC did not identify these glycopeptides is unclear. Byonic reported the composition N_4H_6 on IgG1 – this identification could be rejected within glyXtool^{MS}, due to a wrong assignment of the triplet peptide pattern. Later manual analysis showed that the glycopeptide most likely was a sodium adduct of $N_4H_5F_1$ on IgG1. MAGIC also identified the composition $N_3H_2F_1$ on IgG1. This composition could be explained as an in-source decay of other glycopeptides. Thus, only glyXtool^{MS} was able to correctly identify all 28 true glycopeptide candidates. Furthermore, the possibility of glyXtool^{MS} to show intermediate and not only the final results was helpful to evaluate the source of differences.

Table 4. Performance comparison between glyXtool^{MS}, Byonic and MAGIC

Peptide/ Glycan	IgG 1 EEQYNSTYR	IgG 2 EEQFNSTFR	IgG 3/ IgG 4 EEQYNSTFR/ EEQFNSTYR
N_4H_3	830.0^{3+} [G,B] 1244.5^{2+} [G,B,M]		
N_4H_4	1325.5^{2+} [G,B]		
N_4H_5	938.0^{3+} [G,B]		
N_4H_6	992.1^{3+} [B]		
N_5H_4		941.0^{3+} [G]	
$N_3H_2F_1$	1135.0^{2+} [M]		
$N_3H_3F_1$	1216.0^{2+} [G,B,M]	1200.0^{2+} [G,B,M]	
$N_3H_4F_1$	1297.0^{2+} [G,B,M]	1281.0^{2+} [G,B,M]	
$N_4H_3F_1$	878.7^{3+} [G,B,M] 1317.5^{2+} [G,B,M]	868.0^{3+} [G,B,M] 1301.5^{2+} [G,B,M]	873.4^{3+} [G,B] 1309.5^{2+} [G,B,M]
$N_4H_4F_1$	932.7^{3+} [G,B,M] 1398.6^{2+} [G,B]	922.0^{3+} [G,B,M] 1382.6^{2+} [G,B,M]	927.4^{3+} [B] 1390.6^{2+} [G,B]
$N_4H_5F_1$	1479.6^{2+} [G,B,M]	976.1^{3+} [G,B,M] 1463.6^{2+} [G,B]	
$N_4H_6F_1$	1040.7^{3+} [G]	1030.1^{3+} [G,B]	
$N_5H_3F_1$	946.4^{3+} [G,M]	935.7^{3+} [G] 1403.1^{2+} [G,M]	
$N_5H_4F_1$	1000.4^{3+} [G,B]	989.7^{3+} [G,B,M] 1484.1^{2+} [G,B,M]	
$N_5H_5F_1$	1054.4^{3+} [G,B,M]	1043.8^{3+} [G,B,M]	
$N_4H_4F_1Sa_1$		1019.1^{3+} [G] 1528.1^{2+} [G]	
$N_4H_5F_1Sa_1$	1083.8^{3+} [G,B]	1073.1^{3+} [G,B]	
$N_5H_5F_1Sa_1$	1151.4^{3+} [G,B,M]		

The table shows the detected glycopeptides of glyXtool^{MS} [G], Byonic [B] and MAGIC [M]

CONCLUSION

To our knowledge we build the first pipeline engine-based *N*- and *O*-glycopeptide analysis platform, which supports users in developing their own analysis pipelines based on the data they want to analyze. For this purpose, we provide several glycopeptide-specific tools, which we used to as an example to analyze HCD fragmentation data of human IgG and human fibrinogen. The modularity of the pipeline enables fast, flexible, and transparent glycopeptide analysis. Our general approach comprising feature finding, fragment spectrum scoring, and identification should be applicable to most basic glycoproteomic analysis pipelines, and support the implementation of new tools to extend the functionality of glyXtool^{MS}. A main strength of the software is the flexible implementation of new tools within the TOPPAS engine. This should allow other research groups to further improve or to tailor the analysis pipeline to their experimental needs by addition or replacement of tools.

Additional tools are needed to provide functionality like false-discovery rate (FDR) calculation, spectral matching, and additional scoring algorithms, as described by other groups.[29,43] Due to the open-source license, these algorithms can be implemented and further improved by other workgroups as well.

The software is available via https://github.com/glyXera/glyXtoolMS under the GPL-3.0 open-source license.

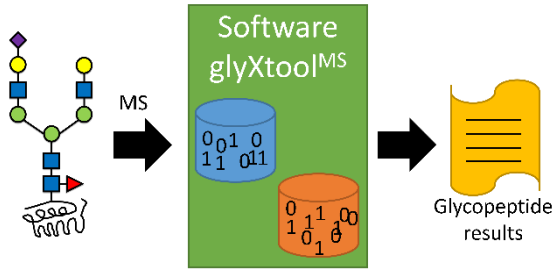
ACKNOWLEDGMENTS

The authors would like to thank Terry Nguyen-Khuong and Alexander Behne for helpful discussions. MH MP and ER acknowledge support by the German Federal Ministry of Education and Research (BMBF) under the project "Die Golgi Glykan Fabrik 2.0" (grant identifier 031A557). ER acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the project "The concert of dolichol-based glycosylation: from molecules to disease models" (grant identifier FOR2509) and support by the European Union (EC) under the project "IMforFuture" (grant no. 721815).

LITERATURE

(1) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1999**, *1473*, 4-8.
(2) Werz, D. B.; Ranzinger, R.; Herget, S.; Adibekian, A.; Lieth, C.-W.; Seeberger, P. H. *ACS Chemical Biology* **2007**, *2*, 685-691.
(3) Varki, A. *Glycobiology* **1993**, *3*, 97-130.
(4) Kaplan, H. A.; Welply, J. K.; Lennarz, W. J. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **1987**, *906*, 161-173.
(5) Kornfeld, R.; Kornfeld, S. *Annual Review of Biochemistry* **1985**, *54*, 631-664.
(6) Hanisch, F. G. *Biological Chemistry* **2001**, *382*, 143-149.
(7) Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. *Journal of Chromatography B* **2007**, *849*, 115-128.
(8) Geyer, H.; Geyer, R. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764*, 1853-1869.
(9) Zauner, G.; Koeleman, C. A. M.; Deelder, A. M.; Wuhrer, M. *J Sep Sci* **2010**, *33*, 903-910.
(10) Hoffmann, M.; Marx, K.; Reichl, U.; Wuhrer, M.; Rapp, E. *Mol Cell Proteomics* **2016**, *15*, 624-641.
(11) Pralow, A.; Hoffmann, M.; Nguyen-Khuong, T.; Rapp, E.; Reichl, U. *Scientific Reports* **2017**, *7*, 11419.
(12) Ongay, S.; Boichenko, A.; Govorukhina, N.; Bischoff, R. *J Sep Sci* **2012**, *35*, 2341-2372.
(13) Cao, L.; Tolic, N.; Qu, Y.; Meng, D.; Zhao, R.; Zhang, Q. B.; Moore, R. J.; Zink, E. M.; Lipton, M. S.; Paga-Tolic, L.; Wu, S. *Anal Biochem* **2014**, *452*, 96-102.
(14) Hinneburg, H.; Stavenhagen, K.; Schweiger-Hufnagel, U.; Pengelley, S.; Jabs, W.; Seeberger, P. H.; Silva, D. V.; Wuhrer, M.; Kolarich, D. *Journal of the American Society for Mass Spectrometry* **2016**, *27*, 507-519.
(15) Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Analytical Chemistry* **1993**, *65*, 877-884.
(16) Woodin, C. L.; Maxon, M.; Desaire, H. *Analyst* **2013**, *138*, 2793-2803.
(17) Dallas, D. C.; Martin, W. F.; Hua, S.; German, J. B. *Briefings in Bioinformatics* **2012**, bbs045.
(18) Li, F.; Glinskii, O. V.; Glinsky, V. V. *PROTEOMICS* **2013**, *13*, 341-354.
(19) Xin, L.; Shan, P. *Journal of Biomolecular Techniques : JBT* **2011**, *22*, S51-S52.
(20) Cooper, C. A.; Gasteiger, E.; Packer, N. H. *PROTEOMICS* **2001**, *1*, 340-349.
(21) Maass, K.; Ranzinger, R.; Geyer, H.; Lieth, C.-W.; Geyer, R. *Proteomics* **2007**, *7*, 4435-4444.
(22) Zhu, Z.; Hua, D.; Clark, D. F.; Go, E. P.; Desaire, H. *Analytical Chemistry* **2013**, *85*, 5023-5032.
(23) Woodin, C. L.; Hua, D.; Maxon, M.; Rebecchi, K. R.; Go, E. P.; Desaire, H. *Analytical Chemistry* **2012**, *84*, 4821-4829.
(24) Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. *Analytical Chemistry* **2007**, *79*, 3065-3074.
(25) Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. *Analytical Chemistry* **2007**, *79*, 1708-1713.
(26) Joenvaara, S.; Ritamo, I.; Peltoniemi, H.; Renkonen, R. *Glycobiology* **2008**, *18*, 339-349.
(27) Darula, Z.; Chalkley, R. J.; Baker, P.; Burlingame, A. L.; Medzihradszky, K. F. *Eur J Mass Spectrom* **2010**, *16*, 421-428.
(28) Mayampurath, A.; Yu, C.-Y.; Song, E.; Balan, J.; Mechref, Y.; Tang, H. *Analytical Chemistry* **2014**, *86*, 453-463.
(29) Zhu, Z.; Su, X.; Go, E. P.; Desaire, H. *Analytical Chemistry* **2014**, *86*, 9212-9219.
(30) Chandler, K. B.; Pompach, P.; Goldman, R.; Edwards, N. *Journal of Proteome Research* **2013**, *12*, 3652-3666.
(31) Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M. *Journal of Proteome Research* **2008**, *7*, 1650-1659.
(32) Mayampurath, A. M.; Wu, Y.; Segu, Z. M.; Mechref, Y.; Tang, H. *Rapid Communications in Mass Spectrometry* **2011**, *25*, 2007-2019.
(33) Wang, H.; Wong, C.-H.; Chin, A.; Taguchi, A.; Taylor, A.; Hanash, S.; Sekiya, S.; Takahashi, H.; Murase, M.; Murase, M.; Murase, M.; Murase, M. *Nature Protocols* **2011**, *6*, 253-269.
(34) 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. *Analytical Chemistry* **2015**, *87*, 2466-2473.
(35) Zeng, W.-F.; Liu, M.-Q.; Zhang, Y.; Wu, J.-Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; Chi, H.; Sun, R.-X.; Wong, C. C. L.; He, S.-M.; Yang, P. *Scientific Reports* **2016**, *6*, 25102.
(36) Clerens, S.; Van den Ende, W.; Verhaert, P.; Geenen, L.; Arckens, L. *PROTEOMICS* **2004**, *4*, 629-632.
(37) Wu, S.-W.; Liang, S.-Y.; Pu, T.-H.; Chang, F.-Y.; Khoo, K.-H. *Journal of Proteomics* **2013**, *84*, 1-16.
(38) Peltoniemi, H.; Joenväärä, S.; Renkonen, R. *Glycobiology* **2009**, *19*, 707-714.
(39) Ozohanics, O.; Krenyacz, J.; Ludányi, K.; Pollreisz, F.; Vékey, K.; Drahos, L. *Rapid Communications in Mass Spectrometry* **2008**, *22*, 3245-3254.
(40) Joshi, H. J.; Harrison, M. J.; Schulz, B. L.; Cooper, C. A.; Packer, N. H.; Karlsson, N. G. *PROTEOMICS* **2004**, *4*, 1650-1664.
(41) Deshpande, N.; Jensen, P. H.; Packer, N. H.; Kolarich, D. *Journal of Proteome Research* **2010**, *9*, 1063-1075.
(42) Ren, J. M.; Rejtar, T.; Li, L.; Karger, B. L. *Journal of Proteome Research* **2007**, *6*, 3162-3173.
(43) Toghi Eshghi, S.; Shah, P.; Yang, W.; Li, X.; Zhang, H. *Analytical Chemistry* **2015**, *87*, 5181-5188.
(44) 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. *Scientific Reports* **2016**, *6*, 21175.
(45) Goldberg, D.; Bern, M.; Parry, S.; Sutton-Smith, M.; Panico, M.; Morris, H. R.; Dell, A. *Journal of Proteome Research* **2007**, *6*, 3995-4005.

(46) Serang, O.; Serang, O.; Muntel, J.; McDowell, G.; Steen, H.; Lee, R. S.; Steen, J. A. *Molecular & cellular proteomics: MCP* **2013**, *12*, 1735-1740.
(47) Nwosu, C. C.; Strum, J. S.; An, H. J.; Lebrilla, C. B. *Analytical Chemistry* **2010**, *82*, 9654-9662.
(48) An, H. J.; Tillinghast, J. S.; Woodruff, D. L.; Rocke, D. M.; Lebrilla, C. B. *Journal of Proteome Research* **2006**, *5*, 2800-2808.
(49) Yu, C.-Y.; Mayampurath, A.; Zhu, R.; Zacharias, L.; Song, E.; Wang, L.; Mechref, Y.; Tang, H. *Analytical Chemistry* **2016**, *88*, 5725-5732.
(50) Aiyetan, P.; Zhang, B.; Zhang, Z.; Zhang, H. *MOJ proteomics & bioinformatics* **2014**, *1*.
(51) Liu, G.; Cheng, K.; Lo, C. Y.; Li, J.; Qu, J.; Neelamegham, S. *Mol Cell Proteomics* **2017**, *16*, 2032-2047.
(52) Bern, M.; Kil, Y. J.; Becker, C. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **2012**, *CHAPTER*, Unit13.20-Unit13.20.
(53) Neue, K.; Kiehne, A.; Meyer, M.; Macht, M.; Schweiger-Hufnagel, U.; Resemann, A.; Bruker Daltonik GmbH: http://bruker.poznan.pl/images/stories/Daltonics/noty/lcms66.pdf, 2012.
(54) Thiele, H.; Glandorf, J.; Korting, G.; Chamrad, D.; Suckau, D.; Stephan, C.; Bluggel, M. *Mol Cell Proteomics* **2006**, *5*, S221-S221.
(55) Hu, H.; Khatri, K.; Zaia, J. *Mass Spectrometry Reviews* **2016**, n/a-n/a.
(56) Sturm, M.; Bertsch, A.; Gröpl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. *BMC Bioinformatics* **2008**, *9*, 163.
(57) Junker, J.; Bielow, C.; Bertsch, A.; Sturm, M.; Reinert, K.; Kohlbacher, O. *Journal of Proteome Research* **2012**, *11*, 3914-3920.
(58) Sturm, M.; Kohlbacher, O. *Journal of Proteome Research* **2009**, *8*, 3760-3763.
(59) Röst, H. L.; Schmitt, U.; Aebersold, R.; Malmström, L. *Proteomics* **2014**, *14*, 74-77.
(60) Vizcaino, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q. W.; Wang, R.; Hermjakob, H. *Nucleic Acids Res* **2016**, *44*, D447-D456.
(61) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534-2536.
(62) Selman, M. H. J.; Derks, R. J. E.; Bondt, A.; Palmblad, M.; Schoenmaker, B.; Koeleman, C. A. M.; Geijn, F. E.; Dolhain, R. J. E. M.; Deelder, A. M.; Wuhrer, M. *Journal of Proteomics* **2012**, *75*, 1318-1329.
(63) Mimura, Y.; Ashton, P. R.; Takahashi, N.; Harvey, D. J.; Jefferis, R. *Journal of Immunological Methods* **2007**, *326*, 116-126.
(64) Zauner, G.; Hoffmann, M.; Rapp, E.; Koeleman, C. A. M.; Dragan, I.; Deelder, A. M.; Wuhrer, M.; Hensbergen, P. J. *Journal of Proteome Research* **2012**, *11*, 5804-5814.
(65) Hahne, H.; Kuster, B. *Journal of The American Society for Mass Spectrometry* **2011**, *22*, 931-942.



For Table of Contents only