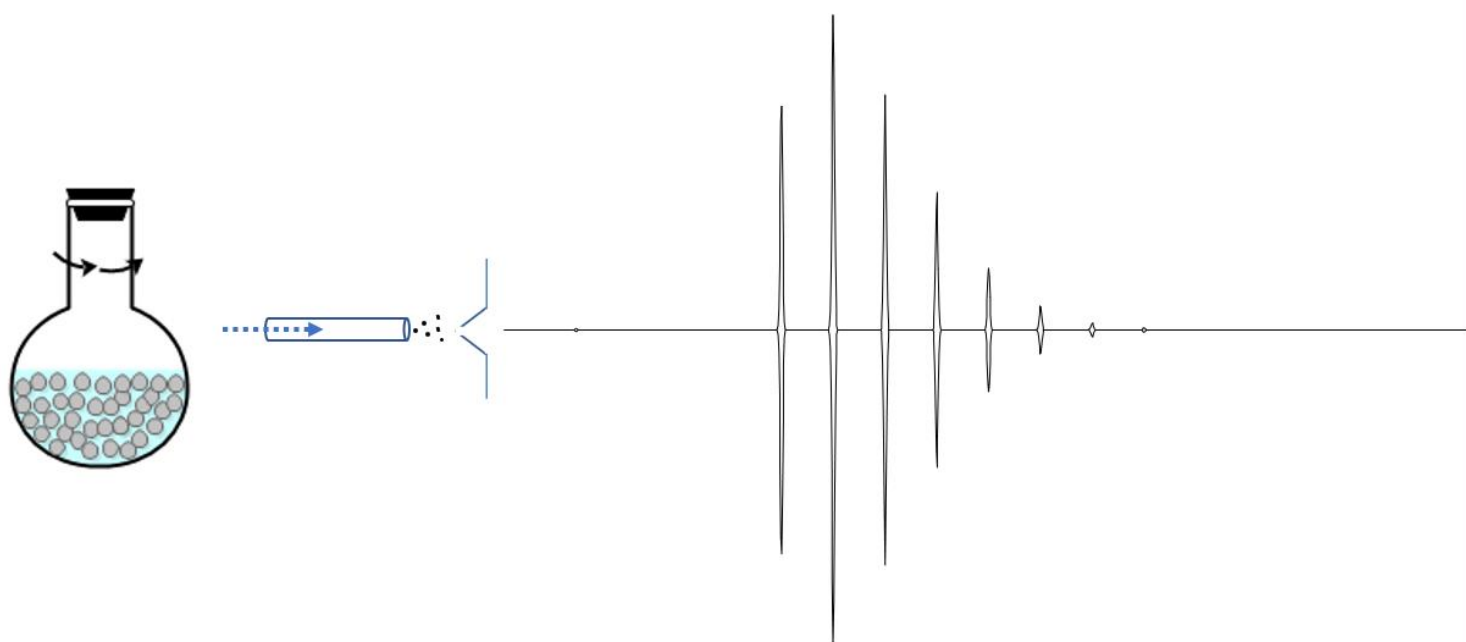


An improved strategy for deep and quantitative *N*-glycomics



Yudong Guan

**An improved strategy for deep and quantitative
N-glycomics**

**Dissertation with the aim of achieving a doctoral degree at the Faculty of
Mathematics, Informatics and Natural Sciences**

Department of Chemistry

University of Hamburg

submitted by

Yudong Guan

from Henan, China

Hamburg

2020

The following evaluators recommend the admission of the dissertation:

Prof. Dr. Hartmut Schlüter

Prof. Dr. Dr. Christian Betzel

Date of disputation: 24.06.2020

This work was done from 2016 until 2020 in the department of clinical chemistry at Universitätsklinikum Hamburg-Eppendorf (UKE) in the research group of Prof. Dr. Hartmut Schlüter.

Dedicated to my lovely parents...

Abstract

In this study, an improved strategy for deep *N*-glycomics that employs an optimized glycan sample preparation involving enhanced (glyco)protein recovery, purification and permethylation efficiency, newly-developed R-scripts matching experimental high-accuracy MS1 data to theoretical monosaccharide compositions in order to enhance the coverage and identification accuracy of protein *N*-glycome with data quality control and a novel bundled sequencing algorithm characterizing the *N*-glycan structures at MS2 level was developed. By this strategy, 57 monosaccharide compositions (133 *N*-glycans) from chicken ovalbumin, 90 monosaccharide compositions (162 *N*-glycans) from etanercept, 133 monosaccharide compositions (230 *N*-glycans) from erythropoietin, 245 monosaccharide compositions (398 *N*-glycans) from human acute promyelocytic leukemia cells and 343 monosaccharide compositions (832 *N*-glycans) from corpus callosum of an adult mouse are identified. The identified *N*-glycans are verified by pGlyco software. This strategy is also applicable to *O*-glycomics. Besides, stable isotopic labeling and relative quantification are performed for *N*-glycome. Finally, this study provides a novel pathway for *N*-glycomics to realize deep identification and biomarker discovery.

Zusammenfassung

Im Rahmen der vorliegenden Studie wurde eine verbesserte Strategie zur tiefgründigen Analyse von Glykanen etabliert. Diese schließt die Optimierung der Probenvorbereitung und der Glykoproteinrückgewinnung, -reinigung und -permethylierungseffizienz ein. Neu entwickelte R-basierte Skripte und Sequenzierungsalgorithmen für komplexe Glykanstrukturen ermöglichen die Analyse experimentell generierter, hochauflösender MS-N-Glykan-Daten unter der Verwendung von theoretischen Glykan-Datenbanken mit verbesserter Abdeckung des Glykoms und einer erhöhten Identifikationsgenauigkeit unter Einschluss von Datenqualitätskontrollen. Darüber hinaus wurde ein neuartig gebündelter Sequenzierungsalgorithmus entwickelt, um N-Glycane auf MS2-Ebene zu identifizieren. Unter Verwendung der beschriebenen Strategie konnten 57 Monosaccharid-Kombinationen (133 *N*-Glykane) aus Ovalbumin vom Huhn, 90 Monosaccharid-Kombinationen (162 *N*-Glykane) aus Etanercept, 133 Monosaccharid-Kombinationen (230 *N*-Glykane) aus Erythropoietin, 245 Monosaccharid-Kombinationen (398 *N*-Glykane) aus menschlichen akuten promyelozytischen Leukämiezellen und 343 Monosaccharid-Kombinationen (832 *N*-Glykane) aus dem Corpus Callosum adulter Mäuse identifiziert werden. Identifizierte Glykane wurden unter Verwendung der Software pGlyco verifiziert. Die hier etablierte Strategie konnte ebenfalls erfolgreich für die *O*-Glykan-Analyse angewendet werden. Eine

Quantifizierung von *N*-Glykanen wurde entweder unter der Verwendung stabiler Isotopenlabel oder im Rahmen Label-freier Quantifizierungsmethoden durchgeführt. Zusammenfassend zeigt die hier durchgeführte Studie eine neue Strategie zur *N*-Glykomanalyse auf, welche auch für die hochsensitive Identifikation von Glykanbiomarkern genutzt werden kann.

List of published papers:

1. Yang F[#], **Guan Y**[#], Feng X, Rolfs A, Schlüter H, Luo J. 2019. Proteomics of the corpus callosum to identify novel factors involved in hypomyelinated Niemann-Pick Type C disease mice. *Mol Brain* 12(1): 17. (***Co-first author***)
2. Fazel R[#], **Guan Y**[#], Vaziri B, Krisp C, Heikaus L, Saadati A, Hidayah SN, Gaikwad M, Schlüter H. 2019. Structural and in vitro functional comparability analysis of altebrel™, a proposed etanercept biosimilar: focus on primary sequence and glycosylation. *Pharmaceuticals (Basel)* 12(1):14. (***Co-first author***)
3. Xie S, **Guan Y**, Zhu P, Li F, Yu M, Linhardt RJ, Chi L, Jin L. 2018. Preparation of low molecular weight heparins from bovine and ovine heparins using nitrous acid degradation. *Carbohydr Polym* 197: 83-91.
4. **Guan Y**, Xu X, Liu X, Sheng A, Jin L, Linhardt RJ, Chi L. 2016. Comparison of low-molecular-weight-heparins prepared from bovine lung heparin and porcine intestine heparin. *J Pharm Sci* 105(6): 1843-1850.
5. Bai X, Li D, Zhu J, **Guan Y**, Zhang Q, Chi L. 2015. From individual proteins to proteomic samples: characterization of O-glycosylation sites in human chorionic gonadotropin and human-plasma proteins. *Anal Bioanal Chem* 407(7): 1857-1869.

Conference Papers:

-
1. Cui W, **Guan Y**, Xydous M, Schlüter H, Sternsdorf T. 2019. Using Acute Promyelocytic Leukemia to study the H3.3 Histone chaperone system and its role in pediatric malignancies. *Klinische Pädiatrie* 231(03): 163-163.

Poster presentations:

1. **Guan Y**, Krisp C, Schlüter H. (March 2019) A comprehensive strategy for deep glycomics. 52nd Annual Meeting of the German Society of Mass Spectrometry (DGMS), Rostock, Germany.
2. **Guan Y**, Krisp C, Thaysen-Andersen M, Nicolle Packer NH, Yang P, Schlüter H. A novel strategy for deep N-glycomics. HUPO 2019-18th Human Proteome Organization World Congress, Adelaide, Australia.
3. Cui W, **Guan Y**, Holdhof D, Modemann F, Schlüter H, Sternsdorf T. (October 2019) Identification of novel interactors of the H3.3 histone chaperone subunit Daxx using proximity-mediated biotin identification (BioID), reveals unexpected crosstalk between seemingly different epigenetic processes. EMBO Workshop-Histone chaperones: Structure, function and role in development and disease, Crete, Greece.
4. Zhang M, **Guan Y**, Schlüter H. (March 2020) Analysis of recombinant erythropoietin by mass spectrometry. 53rd Annual Meeting of the German Society of Mass Spectrometry (DGMS), Münster, Germany.

Table of contents

1. Introduction	1
1.1 Structures of glycoproteins	1
1.2 Disease related glycoproteins	3
1.3 Mass spectrometry-based glycoproteomics	4
2. Aim of this study.....	11
3. Materials and methods.....	12
3.1 Materials	12
3.2 Cell culture and protein extraction from cells and tissues.....	13
3.3 Comparison of different protein precipitation methods for APL cells derived protein recovery	14
3.4 NanoLC-MS/MS analysis and raw data processing for differential proteomics	16
3.5 <i>N</i> - and <i>O</i> -glycan release, purification and permethylation.....	18
3.6 The development of optimized solid-phase permethylation.....	21
3.7 Parallel comparison of 4-aminobenzoic acid butyl ester-based reductive amination and OSPP.....	25
3.8 Derivative <i>N</i> -glycan analysis using PGC-LC-QTOF-MS, nanoC ₁₈ -LC-MS/MS and MALDI-MS	25
3.9 Newly designed workflow of glycome analysis.....	28
3.10 Isotope-based FDR analysis of <i>N</i> -glycans.....	36
3.11 Glycopeptide enrichment by zwitterionic hydrophilic interaction liquid chromatography	36
3.12 NanoLC-MS/MS and pGlyco analyses of enriched glycopeptides	37
3.13 Stable isotopic labeling and relative quantification of <i>N</i> -glycome.....	39
4. Results and discussions	42
4.1 The APL cells derived proteome identification by different precipitation methods using nanoLC-MS/MS	42
4.2 The comparison of different <i>N</i> -glycan purification approaches	43
4.3 The development of optimized solid-phase permethylation.....	45

4.4 The elimination of co-reactions during permethylation	47
4.5 The application of optimized glycan preparation workflow.....	57
4.6 The comparison between OSPP and reductive amination for <i>N</i> -glycan analysis.....	59
4.7 Developed bundled sequencing algorithm for <i>N</i> -glycan structure identification at MS2 level	61
4.8 FDR analysis for <i>N</i> -glycan identification.....	66
4.9 The identification of <i>N</i> -glycans from etanercept, chicken ovalbumin and erythropoietin	67
4.10 The identification of <i>O</i> -glycans from etanercept.....	72
4.11 The analysis of <i>N</i> -glycome from APL cells	74
4.12 The analysis of <i>N</i> -glycome from mouse corpus callosum.....	76
4.13 Post-glycosylational modification analysis of <i>N</i> -glycans.....	76
4.14 The comparison between the OSPP-based <i>N</i> -glycan identification and glycopeptide analysis	79
4.15 Batch-to-batch comparison of erythropoietin derived <i>N</i> -glycans.....	81
4.16 Stable isotopic labeling quantification of APL and APL-6 cells derived <i>N</i> -glycans....	83
4.17 Relative quantification of APL and APL-6 cells derived <i>N</i> -glycome by nanoLC- MS/MS.....	87
5. Conclusions	91
6. References	93
7. Risks and safety statements	127
8. Acknowledgements	132
9. Declarations	133

List of abbreviations

MS	Mass spectrometry
MS/MS (MS2)	Tandem mass spectrometry
ESI	Electrospray ionization
MALDI	Matrix-assisted laser desorption ionization
HPLC	High performance liquid chromatography
HCD	Higher-energy collision dissociation
CID	Collision-induced dissociation
TOF	Time of flight
ACN	Acetonitrile
DMSO	Dimethyl sulfoxide
FA	Formic acid
DDT	Dithiothreitol
IAA	Iodoacetamide
PTMs	Post-translational modifications
ABC	Ammonium bicarbonate
ATRA	All-trans retinoic acid
SDC	Sodium deoxycholate
TEAB	Tetraethylammonium bromide
mL	Millilitre
μL	Microliter
mM	milli Molar
mg	milli gram
kDa	kilo Dalton
rpm	rotations per minute

1. Introduction

1.1 Structures of glycoproteins

The life science research has stepped into an “omics” age since the end of last century. The nomenclature “omics” was derived from “genomics”, firstly coined by Thomas H. Roderick in 1986 (1). Currently, the advanced scientific techniques have promoted great improvements of the genomics and proteomics research (2, 3). Meanwhile, the other “omics” research, such as transcriptomics, metabolomics, glycomics and lipidomics, have become the hot fields widely explored (4-6). In a single organism, the different omics-based biomolecules are closely interacted with each other and contribute to the homeostasis. Furthermore, the “omics” research always involves big-data analysis. For example, protein-coding genes are in a range from 19,587 to 20,245 in human by estimation (7). However, the proteoforms are more complicated due to the various post-translational modifications (PTMs) attached onto the polypeptide backbone (<http://www.unimod.org>). By far, hundreds of PTMs have been identified including glycosylation, acetylation, nitrosylation, phosphorylation, sulfation and ubiquitination, of which glycosylation is the most abundant, complex and heterogeneous PTM. By investigation from the SWISS-PROT database, the glycoproteins account for more than one half in natural proteins (8). Various glycosylation patterns have been identified in eukaryotic cells, including the *N*-glycosylation on amino group of Asn

residue (9), the *O*-glycosylation on hydroxyl groups of Ser, Thr, Hyp, Hyl, Tyr and hLys residues (10-14), *C*-glycosylation at *C*-2 site of Trp residue (15), *S*-glycosylation on thiol group of Cys residue (16), phosphoglycosylation linked to phosphodiester and then to Ser/Thr (-PO₄-Ser/Thr) (17) and glypiation linked to phosphoethanolamine and then to the carboxyl group at the terminal of protein backbone (18). In addition, there are also post-glycosylational modifications on glycans, such as acetylation, phosphorylation, methylation and sulfation, which further increase the heterogeneity and diversity of glycosylation (19). The structural micro-heterogeneity (heterogeneity of the attached glycans on the same site) and macro-heterogeneity (the presence or absence of glycans on glycosylation sites) of glycosylation make glycoproteins challenging for structural characterization. The biosynthesis and modifications of glycans involve about 1% of mammalian genome and over 400 proteins (20, 21). Different with polypeptides and polynucleotides, the biosynthesis of glycans is a non-template driven process and mostly non-linear, which means that glycosylation is less controllable and predictable. As a kind of multi-functional biomolecules, glycoproteins are involved into the recognition interactions between cells and pathogens, cellular migration and cellular signalling transduction (22). According to the previous publications, the *N*- and *O*-glycosylated proteins are mostly studied and proved to be closely related to diseases (23).

1.2 Disease related glycoproteins

Diseased glycoproteins, expressed by diseased cells, break the cellular homeostasis. The glycoproteins with discrepant structures or abundance identified from patients have the potentials to be applied on clinic as the diagnosis biomarkers or therapeutic targets. For examples, the glycosylation of the glial fibrillary acidic protein (GFAP) increases in Alzheimer's disease (AD) (24); the sialylation and fucosylation levels of *N*-glycans rise in breast cancer patients (25); the expression of GP73, a Golgi membrane protein, raises in patients with cirrhotic liver diseases (26); CD147 highly expressed by hepatocellular carcinoma (HCC) cells is developed as the therapeutic target (27).

Cancer, as the deadliest disease, is a hot spot of research for scientists worldwide. According to the cancer statistics in 2016, approximately 1,600 deaths caused by cancer occur in America per day (28). The correlation between cancer and alteration of protein glycosylation has been extensively validated by the previous research (29, 30) although cancer is highly heterogeneous with the various clinical symptoms and still not understood well at present. Based on the central dogma from molecular biology, the mutant genes in cancerous cells translate into changed proteins and some of them, including glycosyltransferases, further regulate the glycosylation process. Currently, the methodological research has experienced great improvements in sensitivity and throughput to

investigate the glycoproteins. Glycans, derived from glycoproteins, are potentially more sensitive than protein biomarkers in the clinical application (31). When individual glycoproteins are discovered definitely as biomarkers, the associated biomolecules including proteases and glycosidases even mutant genes can be also traced along the central dogma. However, the transformation from the mutational glycoproteins to officially approved valid cancer biomarkers is sophisticated and time-consuming (Figure 1), which includes glycoprotein biomarker discovery, targeted glycoprotein verification phase, clinical test and final approval of valid biomarkers (32). Although the publications about glycoprotein biomarker research are greatly increasing, only a few of them are approved by the US Food and Drug Administration (FDA) for clinical test (33).

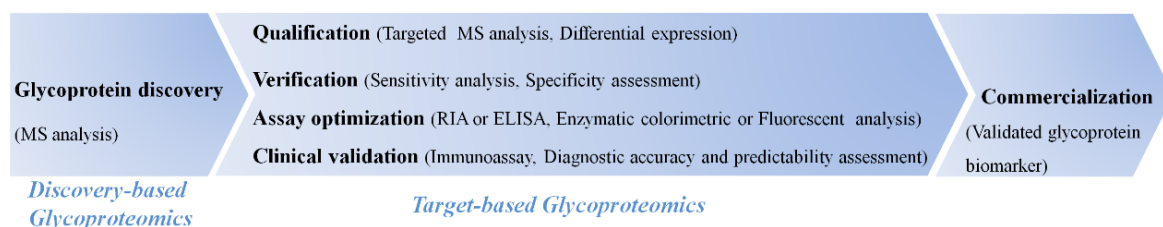


Figure 1. The systematic workflow of approved glycoprotein biomarkers by US FDA.

1.3 Mass spectrometry-based glycoproteomics

In glycoproteomics, mass spectrometry (MS) have been the central analytical technique with high accuracy, sensitivity and throughput (34-36). Besides, two-dimensional electrophoresis (2-DE) (37), high-performance liquid chromatography (HPLC) (38) and capillary electrophoresis (CE) (39) are commonly utilized as on- or off-line separation techniques coupled

with MS instruments. Mostly, the combination of above technologies can show sufficient details for structural elucidation of glycoproteins (40-42). However, the structural characterization of glycans is more challenging than peptides. Substantially, the various monosaccharides can be linked by α - or β -bond at any sites, generating variable isomers, sequences and branches with post-glycosylational modifications (43). By estimating, the nine monosaccharides commonly derived from human can theoretically obtain more than 15 million tetrasaccharide structures (44). *N*-glycans generally have core pentasaccharide structure $\text{Man}\alpha 1-6(\text{Man}\alpha 1-3)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}$. Enzymatically, *N*-glycans are released from polypeptides by Peptide-N-Glycosidase F (PNGase F) except the species with $\alpha 1-3$ fucosylated GlcNAc at the reducing end, which is commonly found from insects or plants and cleaved by PNGase A (45). Chemically, *N*-glycans can be released by hydrazinolysis, while it is not controllable due to the cleavages of the amide bonds in the polypeptide backbones and disruption of acyl groups of sugars (46). Unlike *N*-glycans, the intact *O*-glycans are mainly released by chemical approaches including hydrazinolysis and β -elimination. Relatively, the reductive β -elimination approach is more commonly used to cleave the *O*-glycans (47). Furthermore, Huang *et al.* have proposed a modified reductive β -elimination approach, which can release both of *N*- and *O*-glycans simultaneously (48). For glycan analysis, various LC-MS-based strategies

have been developed, such as porous graphitic carbon (PGC)-LC-MS (49), hydrophilic interaction liquid chromatography (HILIC)-LC-MS (50, 51) and reversed-phase (RP)-LC-MS (52). In addition, CE-MS is also a powerful platform in glycan analysis (53). Compared with LC-electrospray ionization (ESI)-MS, matrix-assisted laser desorption ionization (MALDI)-MS possesses unique superiority for glycan analysis due to high sensitivity and rapid data acquisition. However, it has been demonstrated that the native glycans are unstable and generate fragments during the laser desorption process, which can be solved by derivative approaches prior to MALDI-MS analysis (54). Moreover, the derivatization increases ionization efficiency and lowers limit of detection (LOD) of glycans, such as permethylation and reductive amination. Permethylation has been the most popular derivative approach for MS analysis, in which all the hydrogen atoms of hydroxyl, imino and carboxyl groups are replaced by methyl groups (Figure 2) with several recognized benefits including: 1.) the stabilization of the glycan structures to prevent sample handling- and MS-induced decomposition, 2.) enhancement of their hydrophobicity for easier desalting and chromatographic separation on standard RP-LC columns and better MS properties including ionization in positive polarity mode, 3.) an option for multiplexing and quantitatively comparing glycan samples via isotopic labeling within a single analysis.

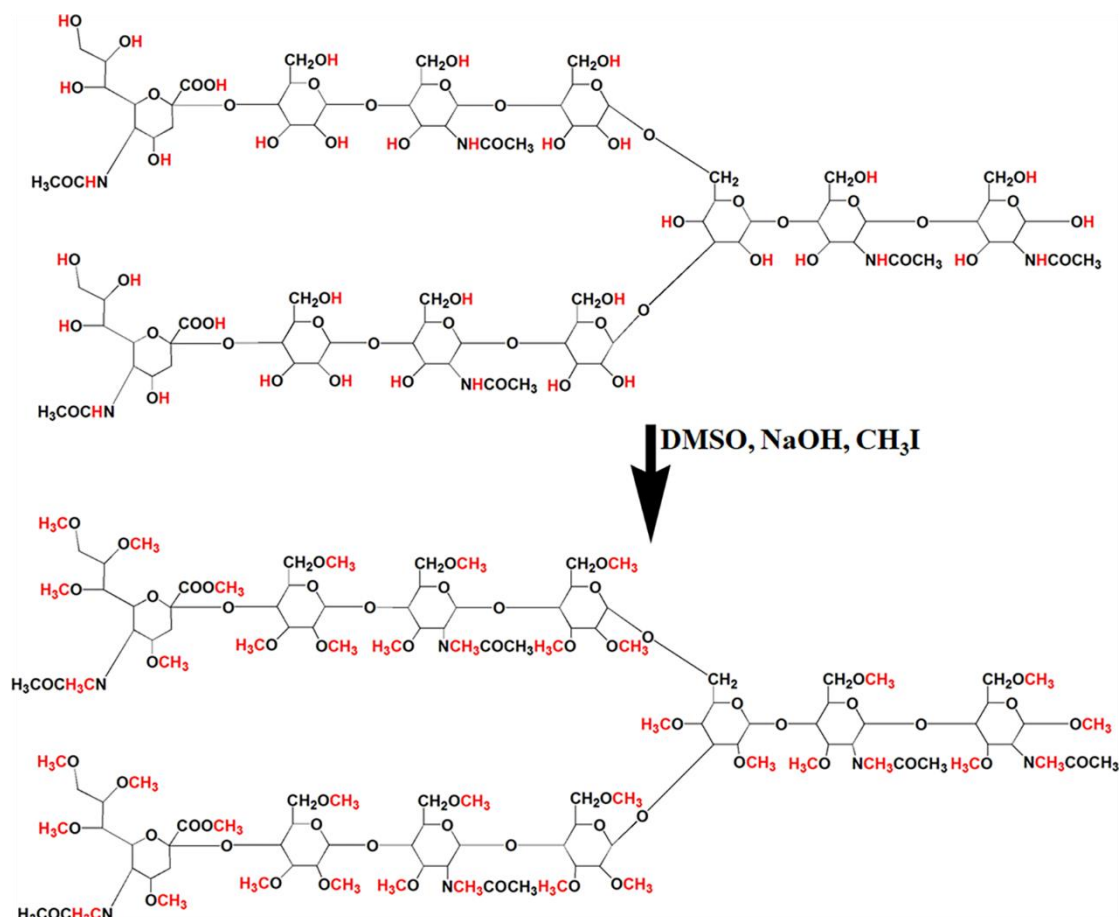


Figure 2. The schematic of glycan-based permethylation.

Top-down approach is mostly applied to the glycan characterization, in which ESI- or MALDI-MS coupled with MS/MS module provides numerous informative fragments of glycans (55, 56). Isomeric glycan structures have been easily identified by nano-LC-Chip/Time of flight (TOF)-MS system (57) and ion mobility spectrometry (IMS) instrument (58). Like proteomics, glycan related informatics is essential to perform MS-based data analysis. Glycoinformatics has experienced a great development in the last two decades (59). For example, GlycoWorkbench is an efficient platform to facilitate the illustration of glycan structures with MS1 and MS2 data (60). In addition, more other MS-based glycodatabases

have been developed such as UniCarb-DB (61), GlycoPattern (62), GALAXY (63) and GlycomeDB (64). However, it still lags behind genomics and proteomics with several drawbacks, such as diversiform standards, lack of efficient data interpretation software and limited connection with other biomolecules. Besides the structural illustration of glycan, the quantitative analysis is also important especially in the screening of glycan biomarkers from diseases. The isotope-based derivative approaches, such as permethylation and reductive amination, have been the most effective quantitative strategies by attaching with the stable isotopic groups (65-69), in which the identical glycans from different samples will be distinguished with the mass shift detected by MS.

Glycopeptides are also heterogeneous due to the attached different glycan patterns (70). According to the known structures of glycoproteins, about one-third glycosylation sites hide inside the folded proteins (71), which means that the glycosylation analysis at native protein level is challenging. However, the analysis of glycopeptides, generally obtained from tryptic digestion of glycoproteins, is able to provide the information of peptide sequences, monosaccharide compositions and glycosylation sites. It has been suggested that the digestion increases the solubility of peptides and glycopeptides account for 2-5% in tryptic pool (72). Glycopeptides are detected with low signal intensities at about 10-50% of the equivalent non-glycosylated peptides during MS analysis (73). To improve the

concentration of glycopeptides, several enrichment strategies have been developed. Sun *et al.* have performed the glycopeptide enrichment by hydrazide chemistry improving the selectivity of glycopeptides 19-45 folds conservatively (74). Lectin affinity approach is also efficient to enrich the glycopeptides from digested mixtures of glycoproteins (75), while the different lectins capture specific sugar structures. In addition, the glycosylated peptides harbor more hydroxyl groups than non-glycosylated species, which enhance their hydrophilicity and enable the separation or enrichment by HILIC column (76). Moreover, the size exclusion chromatography (SEC), as a non-selective and rapid approach, has been demonstrated that there are at least three-fold increases in all the identified glycopeptides using LC-MS/MS system (77). Recently, the combination of C₁₈ and PGC columns in LC-MS instrument is also performed for more comprehensive analysis of glycopeptides (78). To further characterize glycopeptides, MS/MS even MSⁿ are commonly utilized to obtain informative fragments. Collision-induced dissociation (CID) mainly fragments the glycan chain and produces B- and Y-ions annotated in the MS/MS spectra. Furthermore, the glycopeptides with higher *m/z* need higher-energy CID (HCD) for MS/MS analysis. HCD also generates b- and y-ions of the peptide backbones. The electron transfer dissociation (ETD) and electron capture dissociation (ECD), as the complementary MS/MS fragment strategies, are also efficient to characterize the glycopeptides with

producing c- and z-ions of peptide backbones and leaving the intact glycan chains on the amino acid residues (79). However, ETD approach is performed at a limited m/z range of less than about 1,400 (80). In addition, the infrared multiphoton dissociation (IRMPD) can also cleave the glycan and peptide bonds of glycopeptide (81). Also, higher-energy C-trap dissociation shows different types of fragment ions including b- and y-ions of peptides, glycan derived oxonium ions and peptides attached with fragmented glycans (82). The MS-based glycopeptide analysis should be assisted by the database software, such as pGlyco, Byonic, ProteinScape, GlycoQuest and GlycomeDB (83, 84).

The advanced strategies for glycan preparation and data analysis are pivotal to obtain substantive structural information of glycoproteins. By another subjective illustration, as Anne Dell emphasized in the meeting entitled “The Frontiers in Glycomics; Bioinformatics and Biomarkers in Disease”, a technically skilled operator is always indispensable (85). To understand the disease-associated glycome, the Human Disease Glycomics/Proteome Initiative (HGPI) focuses on two complementary approaches, which are functional and MS-based glycomics (86). Meanwhile, the MS instrument vendors have updated the hardware and software with higher sensitivity, faster identification and broader application, promoting the development of a new glycomics age.

2. Aim of this study

For *N*-glycome analysis, most scientific research focuses on the analysis of highly abundant species due to the lack of efficient strategy to identify *N*-glycans at “omics” level by MS.

To solve this lack, I developed an improved strategy for both isotope-based and relative quantitative *N*-glycomics, that employed the optimized *N*-glycan sample preparation, newly-developed R-scripts matching experimental high-accuracy MS1 data to theoretical monosaccharide compositions with isotope-based data quality control and a bundled sequencing algorithm simplifying the structural characterization of *N*-glycans with MS2 fragments. To test the developed workflow, it was applied on purified glycoproteins including chicken ovalbumin, etanercept and erythropoietin and glycoproteome deriving from human acute promyelocytic leukemia cells and corpus callosum of an adult mouse. In addition, the identified monosaccharide compositions were orthogonally verified using the software pGlyco to evaluate this improved strategy. This strategy was also extended to be applicable to *O*-glycomics. The aim of this study was to develop an improved strategy, enabling deep and accurate glycome profiling for glycobiology research and biomarker discovery.

3. Materials and methods

3.1 Materials

Phosphate-buffered saline (PBS) was purchased from Thermo Fisher Scientific (Bremen, Germany). Chicken ovalbumin, sodium hydroxide, dimethylsulfoxide (DMSO), iodomethane, iodomethane- d_3 (CD_3I), iodomethane- ^{13}C ($^{13}CH_3I$), sodium deoxycholate (SDC), tetraethylammonium bromide (TEAB) and all-trans retinoic acid (ATRA) were purchased from Sigma (Darmstadt, Germany). All other chemicals were also purchased from Sigma unless otherwise stated. Two different lots of Enbrel® (G30909, H17609) (etanercept) were purchased from Pfizer (Istanbul, Turkey), which were herein named Enbrel-G and Enbrel-H. Two different batches of erythropoietin (epoetin beta: RDF9729003 and RDF9729004) were purchased from CinnaGen (Tehran, Iran), which were herein named EPO-3 and EPO-4. Sequencing grade modified trypsin, chymotrypsin and PNGase F were obtained from Promega (Madison, WI, USA). 0.5 mL centrifugal filters (3 k and 10 k devices) were purchased from Merck KGaA (Darmstadt, Germany). Solid-phase-extraction (SPE) columns containing RP materials (C_{18} Sep-Pak cartridges) were obtained from Waters (Miford, MA, USA). Mice were purchased from the Charles River Laboratories (Wilmington, MA, USA). All experiments were approved by the local ethical committee and conducted according to the guidelines for the Care and Use of Laboratory Animals.

3.2 Cell culture and protein extraction from cells and tissues

The acute promyelocytic leukemia (APL) cells are distinguished from other forms of acute myeloid leukemia (AML) by its responsiveness to ATRA therapy. Treatment with ATRA allows DNA transcription and differentiation of the immature leukemic promyelocytic into mature granulocytes by targeting the oncogenic transcription factor, promyelocytic leukemia-retinoic acid receptor α (PML-RAR α) fusion transcript. The NB4 APL-derived cell line (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany) was cultured in RPMI 1640 supplemented with 10% FCS in a 95/5% air/CO₂ atmosphere. For neutrophil-like differentiation, cells were treated with 2 μ M ATRA in DMSO for 6 days, hereafter “APL-6 cells”. The control group was treated with DMSO, hereafter “APL cells”. The cells were washed twice using PBS solution. 1 mL SDC buffer (1% SDC in 0.1M TEAB) was added to each cell sample followed by incubation at 99°C for 10 min and sonication with 25% of normal energy at 3 cycles for 30 sec on ice. For corpora callosa homogenization, the isolated forebrain of an adult mouse was first transected. The corpora callosa was dissected from adjacent tissues and frozen by liquid nitrogen immediately. Then it was homogenized in 150 μ L 8 M urea and 50 μ L SDC buffer and then sonicated on ice (87). The protein concentration was estimated using the PierceTM BCA Protein Assay Kit, following the manufacturer’s instructions (Thermo Fisher Scientific,

Bremen, Germany).

3.3 Comparison of different protein precipitation methods for APL cells derived protein recovery

After lysis of APL cells with SDC buffer, the proteins (250 µg) were denatured by 6 M urea, reduced by 20 mM dithiothreitol (DTT) at 60°C for 30 min and alkylated by 40 mM iodoacetamide (IAA) at room temperature for 30 min in dark. DTT and IAA were prepared in 100 mM ammonium bicarbonate (ABC) buffer. Then the sample was digested by trypsin (1/100, w/w) at 37°C for 20 h and then quenched by adding 2% (v/v) formic acid (FA). The tryptic digestion was desalted by RP-SPE C₁₈ cartridge according to previous protocol (88). Briefly, RP-SPE C₁₈ cartridge was conditioned by 9 mL acetonitrile (ACN) and equilibrated by 5 mL 0.1% (v/v) trifluoroacetic acid (TFA), which was followed by loading sample and desalting by 3 mL 0.1% (v/v) TFA. Finally, the peptides were eluted by 3 mL 90% (v/v) ACN in 0.5% (v/v) acetic acid. The eluted peptides were evaporated by a SpeedVacTM vacuum concentrator and dissolved into 0.1% (v/v) FA for nanoLC-MS/MS analysis. This approach, protein precipitation-independent tryptic digestion (further referred to as SDC-based digestion), was compared with protein precipitation-dependent tryptic digestion.

Protein precipitation was performed in order to separate proteins from other biomolecule species, such as free glycans (89). Here, precipitated

proteins derived tryptic digestion were compared to SDC-based digestion to estimate the highest protein recovery efficiency, prior to PNGase F based *N*-glycan cleavage for glycome analysis. Starting with equal amount of proteins with the SDC-based digestion approach, the proteins were recovered by different precipitation approaches after alkylation including chloroform/methanol and trichloroacetic acid (TCA) precipitation (90, 91). For chloroform/methanol precipitation, four-fold volume of methanol was firstly added into the solution of alkylated proteins and then one volume of chloroform and three-fold volume of water was added. After centrifugation at 12,000 rpm for 10 min by Centrifuge 5424 (Eppendorf AG, Hamburg, Germany), the precipitated proteins suspended at the boundary between the supernatant mixture solution and subnatant chloroform. The supernatant solution was discarded and three-fold volume of methanol was added and the supernatant was discarded. The precipitated proteins were dried by a SpeedVacTM vacuum concentrator. For the parallel comparison, the TCA precipitation was performed. Briefly, the solution containing 250 µg proteins was mixed with an equivalent volume of 20% (v/v) TCA after alkylation and followed by the incubation at -20°C for 1 h. Then the sample thawed at room temperature and the supernatant was removed after centrifugation at 12,000 rpm by Centrifuge 5424 for 10 min at a temperature of 4°C. 0.5 mL ice-cold acetone was added and centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant was removed, and the

precipitated proteins were dried by a SpeedVacTM vacuum concentrator. Above proteins obtained from methanol/chloroform and TCA precipitation were redissolved into 100 mM ABC buffer, digested by trypsin (1/100, w/w) at 37°C for 20 h and finally dried by a SpeedVacTM vacuum concentrator. The peptides were resuspended into 0.1% (v/v) FA in equivalent volume with the peptide sample from SDC-based digestion approach.

3.4 NanoLC-MS/MS analysis and raw data processing for differential proteomics

The recovery efficiency of proteins precipitated by chloroform/methanol and TCA was compared with SDC-based approach after trypsin digestion using nanoLC-MS/MS by three replicates. The one, which obtained and identified more proteins, would be used to perform the consequent *N*-glycan preparation. The tryptic peptides were separated using a nanoLC system (DIONEX UltiMate 3000, Thermo Fisher Scientific, Bremen, Germany). The solvent A was 0.1% (v/v) FA and solvent B was 0.1% (v/v) FA in ACN. The sample was injected through the loading pump to a C₁₈ trapping column (Thermo ScientificTM Acclaim PepMapTM, 100 µm×2 cm, 5 µm, 100Å) at a flow rate of 15 µL/min with 1% solvent B. Then the trap column switched to an analytical C₁₈ column (Thermo ScientificTM Acclaim PepMapTM RSLC, 75 µm×50 cm, 2 µm, 100Å) at a flow rate of 0.275 µL/min with 2% solvent B. Then the solvent B increased to 20% in 115.5 min, to 32% in 135.5 min and finally to 95% in 136.5 min. The

nanoLC system was coupled with a hybrid quadrupole-orbitrap mass spectrometer (Q Exactive, Thermo Fisher Scientific, Bremen, Germany). For MS1 scanning, the maximum injection time was 60 ms and AGC target was 3×10^6 ; m/z scan range was set from 375 to 1,600 with an orbitrap resolution of 70,000 full width at half maximum (FWHM) at m/z 200 for data acquisition. Data dependent acquisition was performed by top N mode. For HCD-MS/MS, the 10 highest abundant precursor ions were selected for fragmentation with a normalized HCD collision energy of 27%. Fragment spectra were recorded using an orbitrap mass analyzer with an orbitrap resolution of 17,500 FWHM at m/z 200; the maximum injection time was 50 ms and AGC target was 1×10^5 . Obtained nanoLC-MS/MS raw data was visualized using the Xcalibur software in the version 4.0.27.13 (Thermo Fisher Scientific, Bremen, Germany) and processed with Proteome Discoverer in the version 2.0 (Thermo Fisher Scientific, Bremen, Germany) using the SEQUEST algorithm. For peptide identification, obtained MS2 spectra were searched against theoretical fragment spectra of tryptic peptides, generated from the reviewed SWISSProt FASTA database, containing 20,239 entries, obtained in October 2018. For protein identification by Proteome Discoverer, the following parameters were used: the precursor mass tolerance was set to 10 p.p.m.; the fragment mass tolerance was set to 0.02 Da; variable modifications including oxidation of methionine (M, +15.995 Da) and acetyl (Protein N-term, +42.011 Da) were

considered and the carbamidomethylation of cysteine (C, +57.021 Da) was set as a fixed modification; minimum length of considered peptides was set to 6 amino acids; 2 missed cleavages were tolerated. A false discovery rate (FDR) value threshold less than 0.01, using a reverted decoy peptide databases approach, was set for peptide identification.

3.5 *N*- and *O*-glycan release, purification and permethylation

Etanercept (Enbrel-G) was used as model glycoproteins to optimize the different purification approaches. 200 µg Enbrel-G was denatured, reduced and alkylated as mentioned above. Then buffer exchange was performed by 100 mM ABC buffer using 3 kDa centrifugal filters. 1:30 (v/w) PNGase F was added and incubated at 37°C for 24 h. Then the PNGase F digestion of Enbrel-G was divided into three aliquots with different purification approaches, including 90% (v/v) ethanol precipitation, filter separation and tryptic digestion coupled to RP-SPE C₁₈ cartridge clean-up (Figure 3).

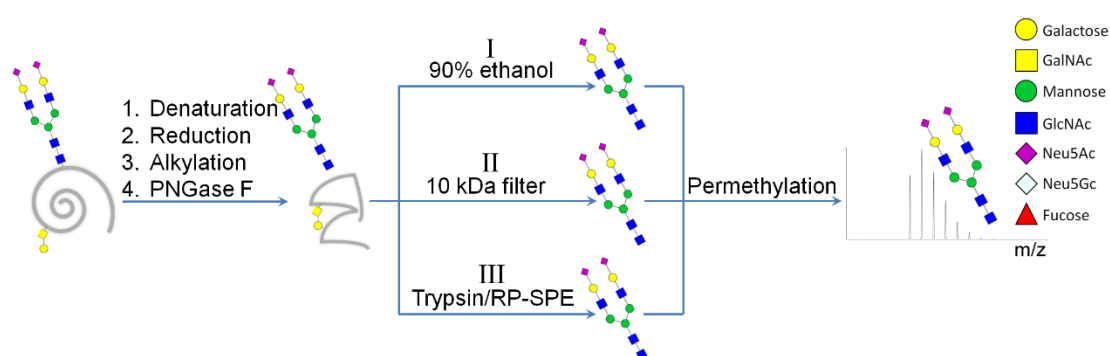


Figure 3. The workflows of three different *N*-glycan purification approaches after PNGase F digestion of Enbrel-G.

For 90% (v/v) ethanol precipitation, the above PNGase F digestion was evaporated to remove the contained ABC salt due to its effect on protein precipitation. Then *N*-glycans and de-*N*-glycosylated proteins were

separated by adding 500 μ L 90% (v/v) ethanol (92), in which de-*N*-glycosylated proteins were precipitated and removed after centrifugation at 12,000 rpm by Centrifuge 5424. For filter separation, a 10 kDa centrifugal filter was utilized to separate the *N*-glycans (filtrate fraction) from the larger de-*N*-glycosylated proteins (retentate fraction) by centrifugation at 12,000 rpm (93). For trypsin digestion with RP-SPE C₁₈ cartridge clean-up approach, the PNGase F digestion was firstly quenched by boiling at 99°C for 5 min. After cooling at room temperature, trypsin (1/100, w/w) was added and incubated at 37°C for 20 h. Then, the digested sample was evaporated by a SpeedVacTM vacuum concentrator and redissolved into 100 μ L 5% (v/v) acetic acid. Then *N*-glycans and tryptic peptides were separated with a modified approach according to the previous protocol using RP-SPE C₁₈ cartridge (94). Briefly, the RP-SPE C₁₈ cartridge was conditioned with 5 mL methanol and equilibrated with 10 mL 5% (v/v) acetic acid respectively. Then the digested sample was loaded into the cartridge and *N*-glycans were eluted with 5 mL 5% (v/v) acetic acid. Next, peptides were eluted by 5% (v/v) acetic acid in 90% (v/v) ACN and evaporated by a SpeedVacTM vacuum concentrator, for subsequent *O*-glycan analysis (48). The eluted peptides were resuspended in 300 μ L 28% ammonium hydroxide with 40 mg ammonium carbonate and incubated at 60°C for 30 h. Then repeated evaporation with additions of water, using a SpeedVacTM vacuum concentrator, was performed to

remove ammonium carbonate. Afterwards, 0.5 M boric acid was added and incubated at 37°C for 30 min, which was removed by repeated evaporation with three additions of 300 µL methanol. Finally, *O*-glycans were released from the peptides. Through above three workflows, the *N*-glycans and de-*N*-glycosylated proteins or peptides were separated. The *N*-glycans were dried by a SpeedVacTM vacuum concentrator and then three purified *N*-glycan samples and one blank sample were permethylated by classical solid-phase permethylation approach (95) (Figure 4). Briefly, sodium hydroxide beads were packed into the spin column and washed by DMSO twice. Purified *N*-glycan and blank samples were redissolved into 50 µL DMSO with 1 µL water and 30 µL iodomethane. The mixed solution was loaded into the spin column and flowed through the sodium hydroxide beads with slight pressure. Reloading was performed five times and then 100 µL DMSO was loaded to wash the beads. The reaction solution was added by 150 µL 5% (v/v) acetic acid to quench the permethylation. Then 200 µL chloroform was added to extract the permethylated *N*-glycans and chloroform-water extraction was repeated ten times. Finally, the chloroform containing permethylated *N*-glycans was evaporated by a SpeedVacTM vacuum concentrator, which was stored in -20°C until further use.

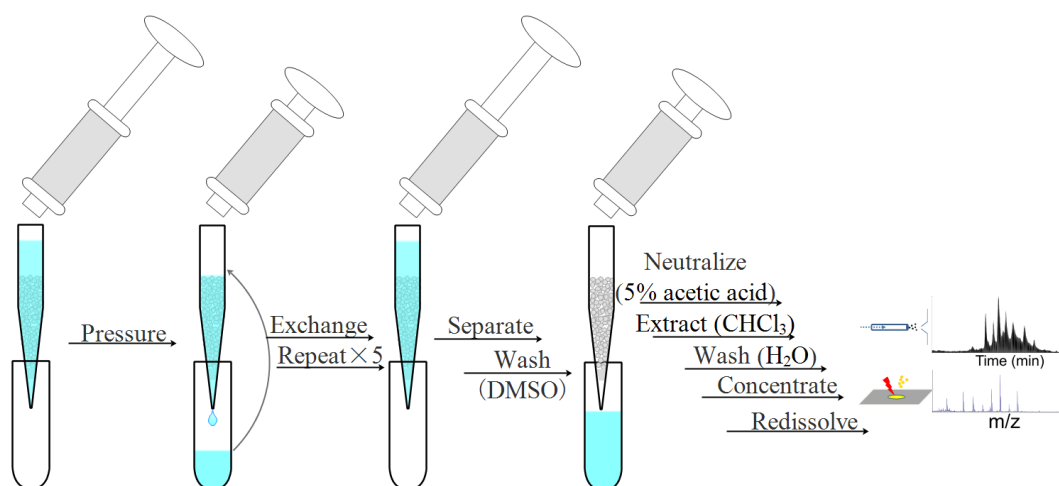


Figure 4. The workflow of the classical solid-phase permethylation for MS analysis.

3.6 The development of optimized solid-phase permethylation

The classical “solid-phase” permethylation approach meant that sodium hydroxide powders or beads was packed inside a capillary or spin-column, as a stationary phase (96). The approach is limited by complicated operation, limitation of water accessibility in the solid sodium hydroxide and impracticality for multiple samples. In our optimization, a solution including water, DMSO and iodomethane was added to submerge sodium hydroxide beads in a glass vial, to prevent the absorption of moisture from air during the rotation (Figure 5). Briefly, 200 mg sodium hydroxide beads were weighed in a glass vial. Dried, PNGase F cleaved *N*-glycans were dissolved in a solution, containing water and DMSO, and 100 μ L iodomethane was added. Resuspended *N*-glycans were transferred to the glass vial containing sodium hydroxide beads. Samples were incubated in a Thermomixer compact (Eppendorf AG, Hamburg, Germany) using a rotation speed of 1,300 rpm. This optimized solid-phase permethylation (OSPP) preparation was developed using cleaved and purified *N*-glycans

from Enbrel-G.

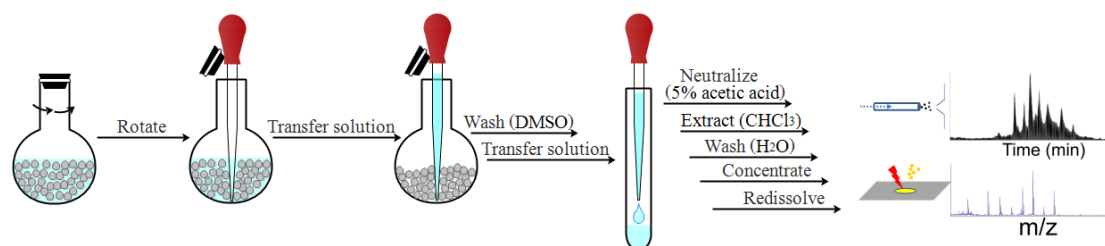


Figure 5. The schematic of OSPP.

Currently, glycan permethylation is mainly performed by two different approaches including slurry and solid-phase permethylation. For slurry permethylation, sodium hydroxide in DMSO was grounded into slurry with a dry mortar, added to dried glycans and mixed with 100 μ L iodomethane (97). For the assessment of the effects of different sodium hydroxide concentrations in DMSO, *N*-glycans cleaved from 20 μ g Enbrel-G were dissolved with different sodium hydroxide concentrations, 10, 20, 30, 60, 120 and 200 μ g/ μ L in 200 μ L DMSO, and then 100 μ L iodomethane was added (Experiment A). Samples were shaken for 10 min at 1,300 rpm using a Thermomixer compact and 200 μ L 5% (v/v) acetic acid was added to quench the permethylation reaction. Permethyated *N*-glycans were extracted using 300 μ L chloroform by chloroform-water extraction. For OSPP development, purified, dried *N*-glycans, cleaved from 20 μ g Enbrel-G, were dissolved in 110 μ L water/DMSO (10/100, v/v) and 100 μ L iodomethane was added into the mixture. Redissolved *N*-glycans were transferred to a glass vial containing 200 mg sodium hydroxide beads and shaken at 1,300 rpm for 10 min by Thermomixer compact, which was compared to the slurry permethylation performed as described above, using

120 $\mu\text{g}/\mu\text{L}$ of sodium hydroxide (Experiment B). OSPP was performed using different ratios of water/DMSO (0/100, 5/100, 10/100, 15/100, 20/100) (v/v) with a duration of 10 min, referred to as Experiment C. Furthermore, different reaction time, 10 min and 30 min, was compared in OSPP at a water/DMSO ratio of 10/100 (v/v) (Experiment D). Also, the workflows by addition of iodomethane before ($\text{CH}_3\text{I}/\text{NaOH}$) and after transferring to sodium hydroxide beads ($\text{NaOH}/\text{CH}_3\text{I}$) were compared (Experiment E). Finally, *N*-glycans were reduced by borane-ammonia to eliminate anomers followed by permethylation, which is compared to non-reducing *N*-glycan permethylation (Experiment F). Here, 10 $\mu\text{g}/\mu\text{L}$ borane-ammonia was added to dried *N*-glycans and incubated at 60°C for 1 h (92). Borane-ammonia was removed by evaporation with three additions of 300 μL methanol prior to permethylation. After above permethylation, reaction solution was transferred into a new vial. 150 μL DMSO was added to further wash the sodium hydroxide beads and transferred into this new vial. 200 μL 5% (v/v) acetic acid were added to permethylated glycan solution to quench the reaction and eliminate oxidation reactions (98). Then 300 μL chloroform were added and chloroform-water extraction was repeated ten times. Chloroform was evaporated using a SpeedVacTM vacuum concentrator. From the experiments described above, the optimal experimental parameters for OSPP were confirmed and used for further experiments (Compare Table 1).

Table 1. Optimized parameters for solid-phase permethylation.

Experiment	Optimization								
	Slurry permethylation	Solid-phase permethylation	Concentration of NaOH (μg/μL)	Water/DMSO (v/v)	Reaction time (min)	Iodomethane before sodium hydroxide	Iodomethane after sodium hydroxide	Non-reducing glycans	Reducing glycans
A	√		10	0/100	10		√	√	
			20						
			30						
			60						
			120						
			200						
B	√		120	0/100	10	√		√	
		√		10/100					
C		√		0/100	10	√		√	
				5/100					
				10/100					
				15/100					
				20/100					
D		√		10/100	10	√		√	
					30				
E		√		10/100	10	√	√	√	
F		√		10/100	10	√		√	√

3.7 Parallel comparison of 4-aminobenzoic acid butyl ester-based reductive amination and OSPP

Reductive amination is another effective derivative method to improve glycan ionization efficiency for MS analysis and specially obtains fluorescent labels for HPLC analysis (99). However, less research was performed to compare the reductive amination and permethylation using nanoLC-MS/MS system. Herein, 4-Aminobenzoic acid butyl ester (ABBE) was utilized as the reductive amination reagent following the classical protocol (100, 101). Briefly, 30 μ L 0.35 M ABBE, prepared into acetic acid/DMSO (3/7, v/v) solution, was added into the purified and dried *N*-glycans released from 20 μ g Enbrel-G. Then 30 μ L 1M 2-picoline borane DMSO solution was added into the mixture. This reaction was incubated at 65°C for 1 h, quenched by adding 900 μ L water and finally evaporated by a SpeedVacTM vacuum concentrator. Meanwhile, the *N*-glycans from 20 μ g Enbrel-G were also permethylated by OSPP with optimal parameters.

3.8 Derivative *N*-glycan analysis using PGC-LC-QTOF-MS, nanoC₁₈-LC-MS/MS and MALDI-MS

The ABBE aminated *N*-glycans were compared with the native *N*-glycans by PGC-LC-QTOF-MS. Both samples were dissolved into 0.1% (v/v) FA, injected into Waters Acquity UPLC (Waters Corporation, USA) and separated by the PGC351 column (Thermo ScientificTM, 50 \times 2.1 mm, 5 μ m). The solvent A was 10 mM ammonium acetate and the solvent B was 10

mM ammonium acetate in 90% (v/v) ACN. The solvent B started with 2% for a duration of 5 min and then increased to 60% in 60 min to separate the native or ABBE aminated *N*-glycans at flow rate of 0.15 mL/min. The HPLC was coupled with the Waters Micromass Q-ToF Premier mass spectrometer (Waters Corporation, USA). The MS parameters were set as follows: positive voltage was set at 3.2 kV; source temperature was 100°C; desolvation gas was 600 L/h; *m/z* scan range was set from 400 to 2,000; the detector voltage was 2 kV.

In addition, the permethylated and ABBE aminated *N*-glycans were dissolved into 0.1% (v/v) FA for nanoLC-MS/MS analysis. The labeled methyl groups or ABBE tags resulted into hydrophobic *N*-glycan derivatives, which were retained on the C₁₈ column strongly. The samples were separated by DIONEX UltiMate 3000 UHPLC system. The solvent A was 0.1% (v/v) FA and solvent B was 0.1% (v/v) FA in ACN. For comparison, equal amount of *N*-glycans, permethylated by OSPP and amidated with ABBE, were injected using the loading pump to the trap column (Thermo Scientific™ Acclaim PepMap™, 100 μm×2 cm, 5 μm, 100Å) at the flow rate of 3 μL/min with 2% solvent B. Then the trap column switched to analytical C₁₈ column (Thermo Scientific™ Acclaim PepMap™ RSLC, 75 μm×50 cm, 2 μm, 100Å) at a flow rate of 0.2 μL/min. The LC gradient was different between the permethylated and ABBE aminated *N*-glycans. For the permethylated *N*-glycans, the solvent B

started with 10% and increased to 30% in 5 min, to 75% in 70 min and finally to 95% in 80 min. For the ABBE aminated *N*-glycans, solvent B started with 5% for 10 min and then increased to 80% in 90 min. The UHPLC was coupled with a tribrid quadrupole-orbitrap-ion trap mass spectrometer (Fusion, Thermo Fisher Scientific, Bremen, Germany). For MS1 scanning, an orbitrap mass analyzer was used with an orbitrap resolution of 120,000 FWHM at m/z 200; the maximum injection time was 120 ms and AGC target was 2×10^5 ; m/z scan range at MS1 was set from 450 to 2,000. Data dependent acquisition mode was performed by top speed mode. For CID-MS/MS, the most intense precursor ions were selected for fragmentation and isolated using an isolation window of 3; the normalized collision energy of CID was set to 35%; fragment ions were injected to an ion trap with a maximum injection time of 20 ms and an AGC target of 1×10^5 . The parallelization of MS1 and MS2 acquisition was performed. The data were visualized and analyzed using Xcalibur.

MALDI-MS is commonly utilized to measure the permethylated glycans due to its high sensitivity and throughput. To enable the permethylated *N*-glycans to attach with sodium ions during MALDI-MS analysis, the permethylated *N*-glycans were resuspended into 10 mM sodium chloride in 50% (v/v) methanol. Then the samples were dropped on the target followed by dropping saturated α -cyano-4-hydroxycinnamic acid (CHCA) in 50% (v/v) methanol. The MALDI-MS/MS system (Ultraflextreme,

Bruker, Bremen, Germany) was used to perform the sample analysis. The MALDI laser energy was set at 30% in the positive-reflectron mode; MALDI-MS spectra were acquired in the m/z range from 500 to 8,000 for *N*-glycan analysis. The data were analyzed using the flexAnalysis software in the version 3.3 (Build 80) (Bruker, Bremen, Germany).

3.9 Newly designed workflow of glycome analysis

The MS-based glycomics is challenging due to the lack of efficient computer-assisted software (102). In this study, a computational workflow for glycome analysis of MS raw data based on the optimized glycan preparation was developed (Figure 6).

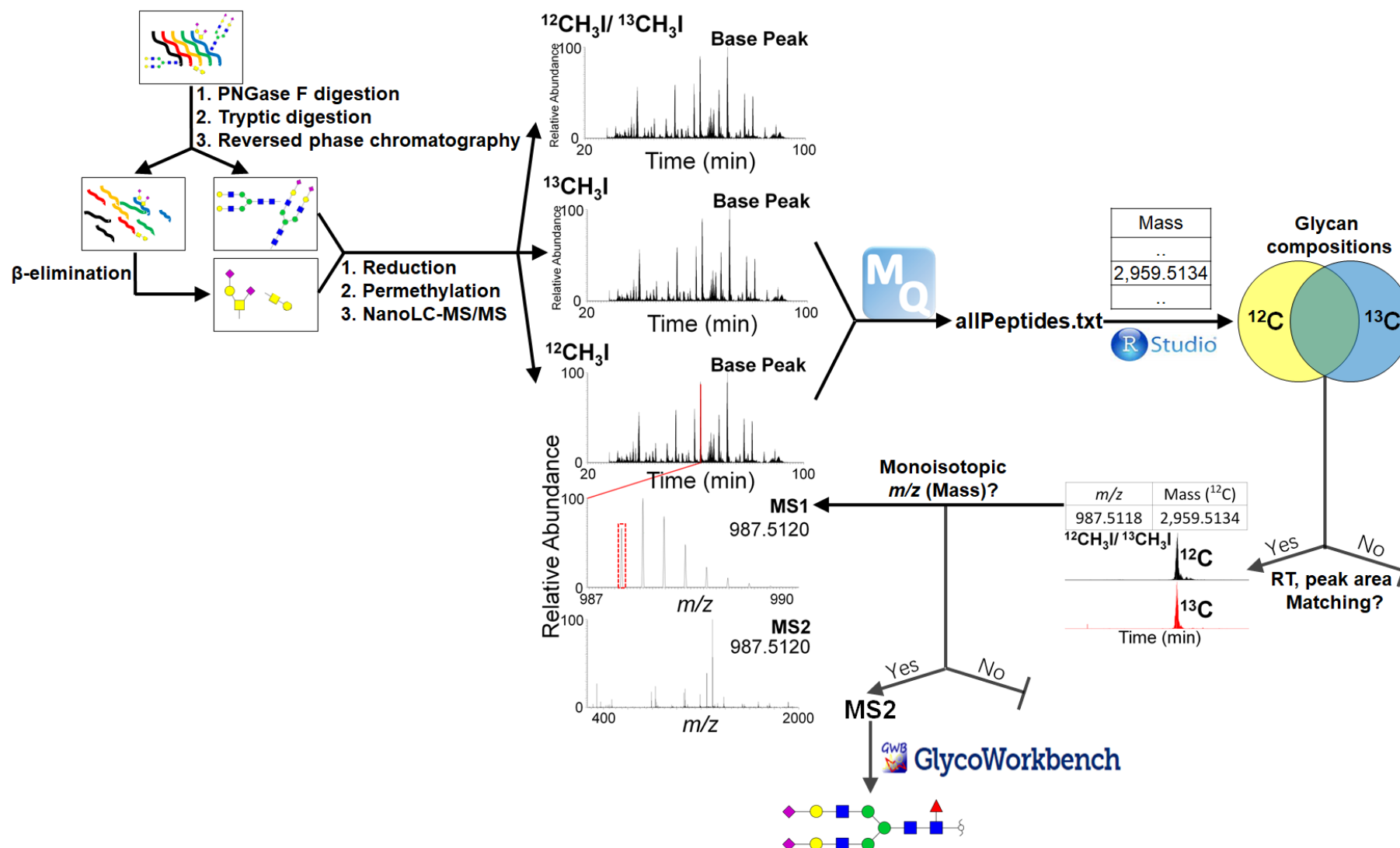


Figure 6. The novel analytical workflow of glycan identification with MaxQuant, home-made R-scripts, GlycoWorkbench and isotope-based data quality control, designed based on the optimized glycan preparation.

Firstly, glycan samples obtained from above optimized preparation were permethylated with iodomethane- ^{12}C and ^{13}C by OSPP respectively and then mixed by ratio of 1:1, which were analyzed by nanoLC-MS/MS with triplicate. MaxQuant (Version No. 1.6.2.3, <http://www.maxquant.org>), an efficient software for proteome analysis, is used to extract all the detected masses from MS raw data of ^{12}C - and ^{13}C -permethylated glycans in the “allpeptides.txt” file. In that, the “Mass” column is extracted and calculated by home-made R-scripts based on the monoisotopic molecular weight (monoMW) of each monosaccharide permethylated by iodomethane- ^{12}C and ^{13}C respectively, within a narrow deviation. The exported table files show the composition of each ^{12}C - or ^{13}C -permethylated glycan and absolute value of deviations between theoretical and experimental masses. The data quality is controlled by defined deviation threshold and isotope-based intersection. In this study, the analytical workflow is described using *N*-glycome from human and mouse, whose building blocks are different in the designed R-scripts (<https://github.com/guan181992/Glyco-informatics>).

For humans, the building blocks are N-acetylneuraminic acid (Neu5Ac), N-acetylhexoseamine (HexNAc) including N-acetylglucosamine (GlcNAc) and N-Acetylgalactosamine (GalNAc), Hexose (Hex) including glucose (Glc), galactose (Gal) and mannose (Man), Fucose (Fuc) and reduced N-acetylhexoseamine (Red-HexNAc). They are summarized with monoMWs

after iodomethane- ^{12}C and ^{13}C permethylation in Figure 7.

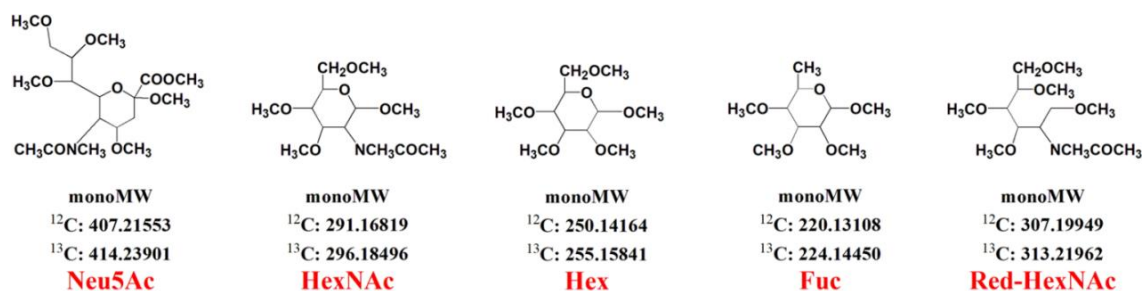


Figure 7. The structure and monoMW of each *N*-glycan building block from humans after $^{12}\text{CH}_3\text{I}$ and $^{13}\text{CH}_3\text{I}$ permethylation.

To investigate the applicable deviation used in the matching algorithm of R-script, human derived ^{12}C -permethylated *N*-glycan library is established. This *N*-glycan library covers 10,618 species from HexNAc₁Hex₃Red-HexNAc₁ (trimannosylchitobiose core) to 6,000 Da, including all the ^{12}C -permethylated *N*-glycan compositions in mathematical possibilities (each monosaccharide composition means one glycan species). All the masses are ranked from minimum to maximum and each deviation between contiguous masses is calculated (Figure 8), showing that only 237 species (2.2%) are lower than 0.59 p.p.m., one species has a deviation of 1.43 p.p.m. and all the other species (97.8%) are higher than 1.86 p.p.m. (Figure 9). In this matching algorithm, 1.5 p.p.m. is utilized as deviation threshold in the designed R-script, for human derived ^{12}C -permethylated *N*-glycan analysis. The experimental masses extracted from MS raw data by MaxQuant are matched with the theoretical monosaccharide compositions, in which multiple experimental masses are matched with one *N*-glycan species in 1.5 p.p.m.. The number of positively matched experimental masses shows the abundance of this *N*-glycan species (Figure 10).

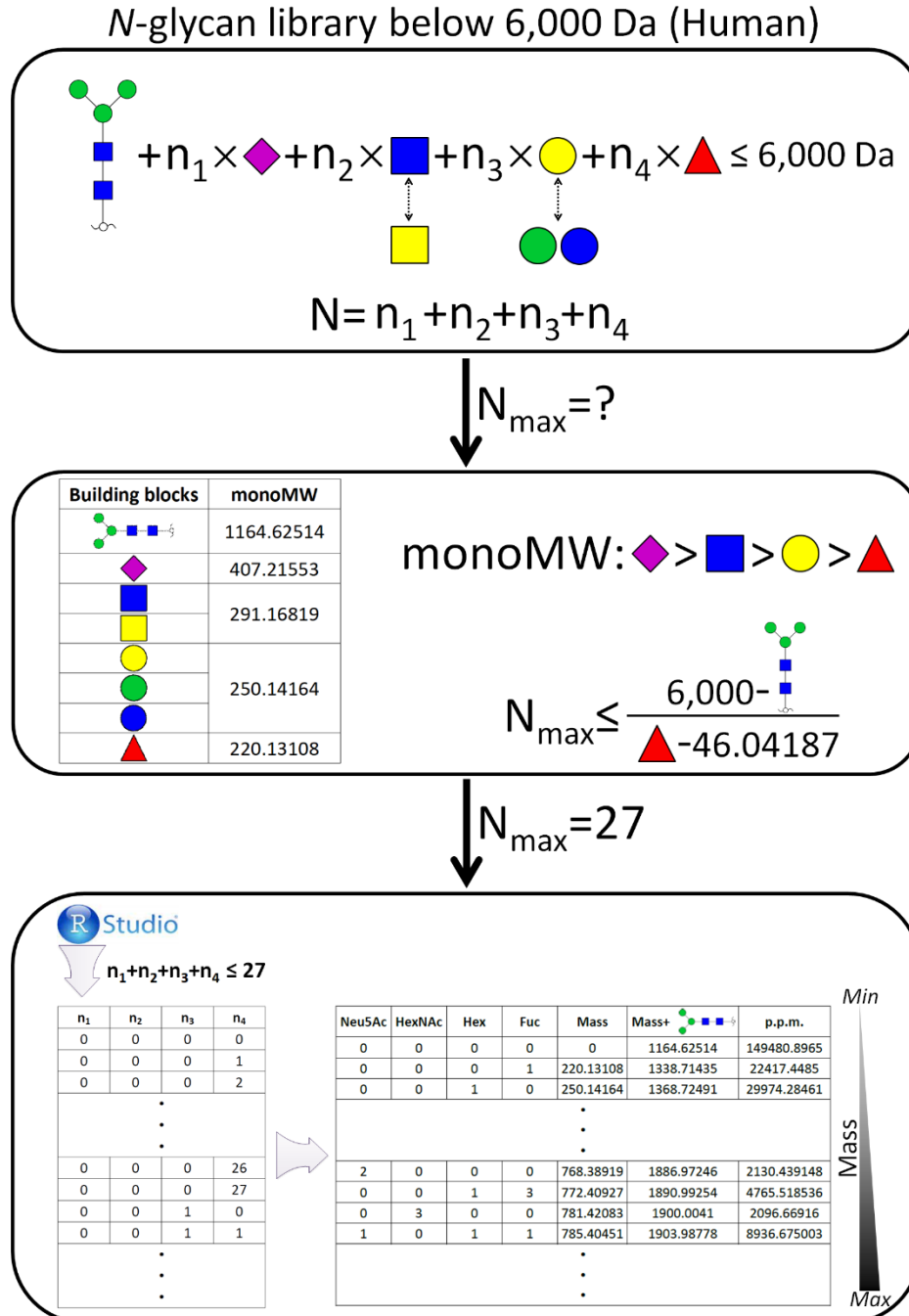


Figure 8. The development of human derived ^{12}C -permethylated N-glycan library (monosaccharide compositions) below 6,000 Da.

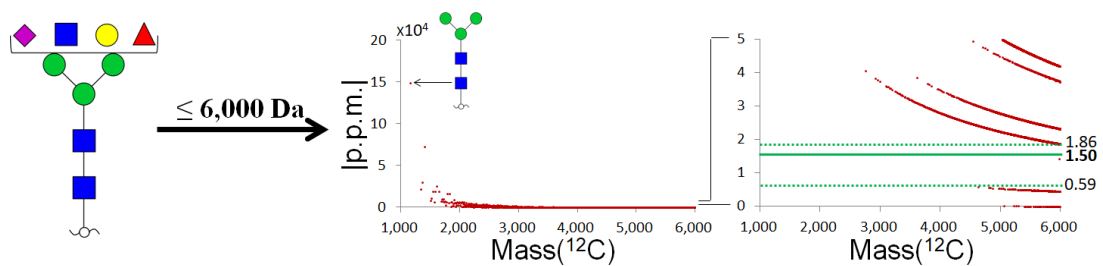


Figure 9. The deviation analysis for human derived ^{12}C -permethylated N-glycan library from HexNAc₁Hex₃Red-HexNAc₁ to 6,000 Da.

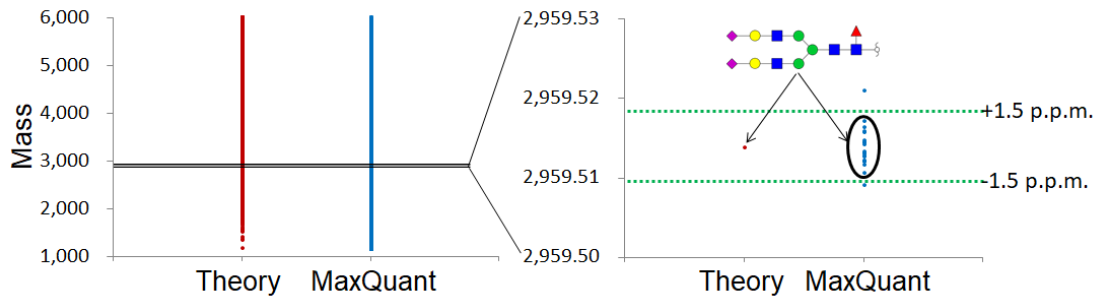


Figure 10. The matching algorithm between the theoretical *N*-glycan species and experimental masses extracted from MS raw data by MaxQuant within defined deviation threshold (± 1.5 p.p.m.).

```

mass<-read.csv("E:/file_name.csv")
row_mass=nrow(mass)
mass_out=cbind(mass,Neu5Ac=rep(0,row_mass),HexNAc=rep(0,row_mass),Hex=rep(0,row_mass),Fuc=rep(0,row_mass),Red_HexNAc=rep(0,row_mass),deviation=r
ep(0,row_mass))

identify_glycans<-function(q) {

  Neu5Ac=361.17366
  HexNAc=245.12632
  Hex=204.09977
  Fuc=174.08921
  Red_HexNAc=261.15762

  t=q%/%Neu5Ac
  j=q%/%HexNAc
  m=q%/%Hex
  n=q%/%Fuc
  w=q%/%Red_HexNAc

  for(Neu5Ac_n in 0:t)
    for( HexNAc_n in 0:j)
      for( Hex_n in 0:m)
        for( Fuc_n in 0:n)
          for(Red_HexNAc_n in 0:w)

            {
              total_1=sum(407.21553*Neu5Ac_n,291.16819*HexNAc_n,250.14164*Hex_n,220.13108*Fuc_n,307.19949*Red_HexNAc_n)
              total_2=total_1-sum(Neu5Ac_n,HexNAc_n,Hex_n,Fuc_n,Red_HexNAc_n,1)*46.04187
              deviation=abs(q-total_2)/total_2*1000000

              if(Red_HexNAc_n==1&HexNAc_n>0&Hex_n>2&deviation<1.5)
              { y=c(Neu5Ac_n,HexNAc_n,Hex_n,Fuc_n,Red_HexNAc_n,deviation)
                return(y)}
            }

  }

  for(n_record in 1:row_mass)
  {
    print(n_record)
    pass=lapply (mass[n_record,1],identify_glycans)
    pass_2=unlist(pass)
    if(length(pass_2)>2)
    {mass_out[n_record,2]=pass_2[1]
      mass_out[n_record,3] =pass_2[2]
      mass_out[n_record,4] =pass_2[3]
      mass_out[n_record,5] =pass_2[4]
      mass_out[n_record,6] =pass_2[5]
      mass_out[n_record,7] =pass_2[6]}
  }

  write.table(mass_out,row.names=F,col.names=T,file="E:/file_name_out.csv",sep=",")

```

Filter parameters:

- I) Including trimannosylchitobiose core structure (HexNAc₁Hex₃Red-HexNAc₁)
- II) Deviation limitation (from 1 to 5 p.p.m.)
- III) Changed as “if (Red_HexNAc==1&deviation<1.5)” for masses lower than 1,150 Da to identify *N*-glycans without trimannosylchitobiose core structure

Figure 11. The ¹²C-based R-script for human derived *N*-glycome identification at MS1 level.

The ^{12}C -based script is shown with detailed parameters for human derived *N*-glycome identification in Figure 11, including the core structure HexNAc₁Hex₃Red-HexNAc₁. Also, deviations of human derived ^{13}C -permethylated *N*-glycan library from HexNAc₁Hex₃Red-HexNAc₁ to 6,000 Da are also analyzed (Figure 12). Most of the deviations, 9,950 from 9,995 species (99.5%), are above 1.12 p.p.m. and 1.1 p.p.m. is used as the deviation threshold in the designed ^{13}C -based R-script. To cater for the MS data with different accuracies, the deviation thresholds of matching experimental masses in ^{12}C - and ^{13}C -based R-scripts are tested successfully from 1 to 5 p.p.m. and higher deviation thresholds generate more false positives.

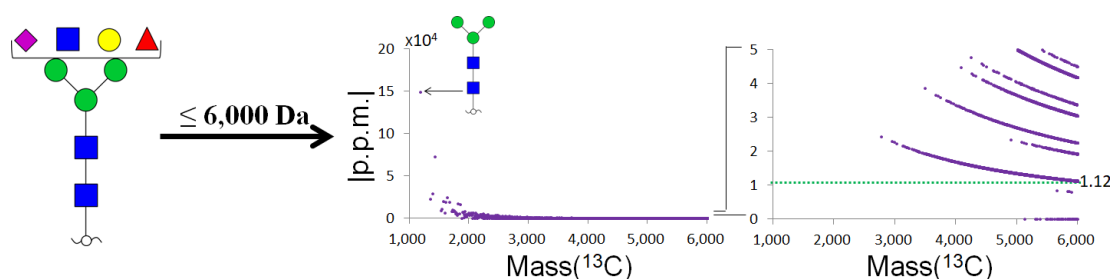


Figure 12. The deviation analysis for human derived ^{13}C -permethylated *N*-glycan library from HexNAc₁Hex₃Red-HexNAc₁ to 6,000 Da.

For mice, *N*-glycolylneuraminic acid (Neu5Gc) (Figure 13a), not biosynthesized in humans (103), is utilized as an additional building block for the designed R-script. In addition, the total molecular weight of Neu5Ac and Hex is equal to the total molecular weight of Neu5Gc and Fuc, which is 657.3572 Da. The ^{12}C -permethylated *N*-glycan library from HexNAc₁Hex₃Red-HexNAc₁ to 6,000 Da consists of 34,404 *N*-glycan species and 20,127 species (58.5%) have the deviations lower than 0.01

p.p.m. (Figure 13b). 1.5 p.p.m. is also used as deviation threshold in the ^{12}C -based R-script and 1.1 p.p.m. in ^{13}C -based R-script, for mouse derived *N*-glycan identification.

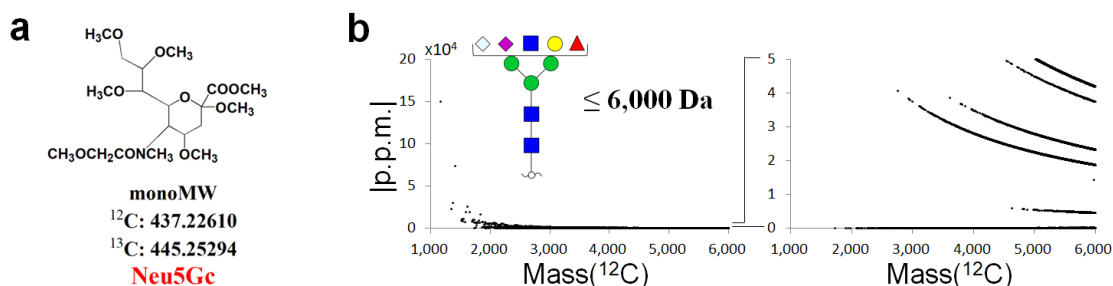


Figure 13. The structure of permethylated Neu5Gc and deviation investigation of mouse derived ^{12}C -permethylated *N*-glycan library. (a) The structure and monoMW of Neu5Gc, deriving from $^{12}\text{CH}_3\text{I}$ and $^{13}\text{CH}_3\text{I}$ permethylation. (b) The deviation analysis for mouse derived ^{12}C -permethylated *N*-glycan library from HexNAc₁Hex₃Red-HexNAc₁ to 6,000 Da.

To remove the false positives, the pairs of ^{12}C - and ^{13}C -permethylated monosaccharide compositions, matched by retention time (RT) and peak areas (roughly 1:1) from MS raw data of the mixture of ^{12}C - and ^{13}C -permethylated *N*-glycans, are kept. Next, the positively matched *N*-glycan species are continued to track back to match the monoisotopic *m/z* from the MS raw data of ^{12}C -permethylated *N*-glycans. Finally, the MS2 fragments, derived from positively matched precursor, are utilized to characterize the *N*-glycan structures by GlycoWorkbench (Version No. 2.1 build 146, https://download.cnet.com/GlycoWorkbench-64-bit/3000-2383_4-75758804.html). For ^{12}C -permethylated *O*-glycans, 16.0313 Da is utilized as the extra building block, replacing Red-HexNAc in *N*-glycan analysis, in which 16.0313 Da is the increased mass of monosaccharide at the reducing end after reduction compared to the non-reducing species.

Therefore, this workflow realizes the in-depth glycome identification confidently and is also applicable to both *N*- and *O*-glycomics. For the data analysis, the glycan fragment annotations of MS2 spectra follows the nomenclature proposed by Domon *et al.* (104). This analytical workflow is firstly tested by purified glycoproteins including etanercept (Enbrel-H), chicken ovalbumin and erythropoietin (EPO-3 and EPO-4), which is further applied on the glycoproteome from APL cells, APL-6 cells and mouse corpus callosum.

3.10 Isotope-based FDR analysis of *N*-glycans

Firstly, the MS data from ^{12}C - and ^{13}C -permethylated *N*-glycans are matched with deviation thresholds of 1.5 p.p.m. and 1.1 p.p.m. in R-scripts respectively. Additionally, the mixture of ^{12}C - and ^{13}C -permethylated *N*-glycans were analyzed in one nanoLC-MS/MS run. The performance with the pairs of ^{12}C - and ^{13}C -permethylated *N*-glycans at MS1 level significantly decreases the false positives by RT matching and peak areas comparison (roughly 1:1). Then monoisotopic *m/z* matching and MS2 fragment identification further remove all the false positives. The FDR of *N*-glycan data from Enbrel-H and chicken ovalbumin is calculated with the number of false matched precursors divided by total number of the matched precursors.

3.11 Glycopeptide enrichment by zwitterionic hydrophilic interaction liquid chromatography

Enbrel-H and proteins extracted from APL and APL-6 cells were digested by trypsin as mentioned in section 3.3. EPO-3 and EPO-4 were digested by chymotrypsin, replacing the trypsin in the protocol as mentioned in section 3.3. The glycopeptides were enriched from tryptic digestion of 400 µg Enbrel-H, 1 mg proteins extracted from APL and APL-6 cells and 200 µg EPO-3 and EPO-4, resuspended into 100 µL 80% (v/v) ACN with 1% (v/v) TFA following the previous approaches (83). Briefly, the micro-column was packed with 30 mg zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) particles obtained from a HILIC column (SeQuant® ZIC®-cHILIC 3µm, 100Å 250×4.6 mm, Merck KGaA, Darmstadt, Germany) on the top of a C₈ disk (3M, Eagan, MN, USA). The micro-column was equilibrated by 600 µL 80% (v/v) ACN with 1% (v/v) TFA and followed by loading tryptic peptide samples. 600 µL 80% (v/v) ACN with 1% (v/v) TFA was utilized to remove the low-hydrophilic peptides and then glycopeptides were eluted by 300 µL 0.1% (v/v) TFA. Finally, the enriched glycopeptides were dried by a SpeedVacTM vacuum concentrator and redissolved into 0.1% (v/v) FA for nanoLC-MS/MS analysis.

3.12 NanoLC-MS/MS and pGlyco analyses of enriched glycopeptides

The solvent A was 0.1% (v/v) FA and solvent B was 0.1% (v/v) FA in ACN. Samples were injected into the EASY-nano-LC system (Thermo Fisher

Scientific, Bremen, Germany) without trap column and separated by C₁₈ column (Thermo Scientific™ Acclaim PepMap™ RSLC, 75 μm×50 cm, 2 μm, 100Å). Nano-pump started with 1% solvent B at the flow rate of 0.2 μL/min. The gradient for tryptic glycopeptides of Enbrel-H lasted for 1 h. Solvent B increased to 20% in 40 min, to 30% in 47 min and to 90% in 50 min. The gradient for chymotryptic glycopeptides of EPO-3 and EPO-4 lasted for 140 min. Solvent B increased to 20% in 90 min, to 30% in 120 min and to 90% in 130 min. The gradient for tryptic glycopeptides of APL and APL-6 cells lasted for 6 h in total. Solvent B increased to 20% in 300 min, to 30% in 342 min and to 90% in 345 min. The LC was coupled with Orbitrap Fusion Tribrid mass spectrometer. The MS parameters were set as follows: for MS1 scanning, m/z scan range was set from 350 to 2,000; an orbitrap mass analyzer was utilized with an orbitrap resolution of 120,000 FWHM at m/z 200; the maximum injection time was 50 ms and AGC target was 5×10^5 ; precursor ions with a charge state between 2 and 6 were included. Data dependent acquisition mode was performed in top speed mode. For HCD-MS/MS, the most intense precursor ions were selected for fragmentation and isolated using an isolation window of 4; dynamic exclusion of selected precursor ions was performed for a duration of 15s; isolated precursor ions were fragmented using a stepped HCD gradient with normalized collision energies of 20%, 30% and 40%; fragment ions were analyzed in an orbitrap mass analyzer at an orbitrap resolution of

15,000 FWHM at m/z 200; the maximum injection time was 250 ms and AGC target was 5×10^5 .

Glycopeptides were identified from MS raw data by pGlyco (Version No. 2.1.2, <http://pfind.ict.ac.cn/software/pGlyco/index.html>) with a protein database containing 20,239 entries from SWISSProt obtained in October 2018. The glycan database was extracted from GlycomeDB (www.glycome-db.org) with the total *N*-glycan entries of 7,884 (83). The parameters were set as follows: the precursor tolerance was 5 p.p.m.; the fragment mass tolerance was 20 p.p.m.; as a variable modification, the oxidation of methionine was considered; carbamidomethylation of cysteine was set as a fixed modification.

3.13 Stable isotopic labeling and relative quantification of *N*-glycome

The *N*-glycome from APL cells was also permethylated by iodomethane, iodomethane- d_3 and iodomethane- ^{13}C , named as APL-H, APL-D and APL- ^{13}C respectively. The *N*-glycome from APL-6 cells was permethylated by iodomethane, named as APL-6-H. All the *N*-glycan samples were prepared, starting with same amount of proteins. Then APL-6-H was mixed with APL-D and APL- ^{13}C by ratio of 1:1 respectively and quantification was performed by MALDI-MS. Furthermore, APL-H and APL-6-H were also measured by nanoLC-MS/MS for relative quantification (Figure 14). Briefly, the list of identified *N*-glycans would be transformed as a new table

including monosaccharide compositions, m/z , charges and masses. This table was input into the Skyline software (Version No. 20.1, <http://skyline.maccosslab.org>) and the peak area of each *N*-glycan species was integrated. Then the exported table, mainly including monosaccharide compositions and Total Area MS1, was input into the Perseus software (Version No. 1.6.2.1, <http://www.perseus-framework.org>). Log₂(x) transformation, normalization and two-sample student's T-test were performed to compare the *N*-glycans from APL-H and APL-6-H, in which p-value less than 0.05 and a fold change (FC) more than 2 were defined as the minimum requirements to reach statistical significance. To evaluate the production reproducibility, the *N*-glycans of different batches of erythropoietin, EPO-3 and EPO-4, were also performed with relative quantification as described above.

List of identified *N*-glycans

<i>m/z</i>	Mass	Neu5Ac	HexNAc	Hex	Fuc	Red-HexNAc	Deviation (p.p.m.)
1,165.6331	1,164.6258	0	1	3	0	1	0.58
915.4914	1,828.9683	0	3	3	1	1	0.73
..
..
987.5118	2,959.5134	2	3	5	1	1	0.15
..



Composition	<i>m/z</i>	Charge	Mass
HexNAc ₁ Hex ₃ Red-HexNAc ₁	1,165.6331	1	1,164.6258
HexNAc ₃ Hex ₃ Fuc ₁ Red-HexNAc ₁	915.4914	2	1,828.9683
..
..
Neu5Ac ₂ HexNAc ₃ Hex ₅ Fuc ₁ Red-HexNAc ₁	987.5118	3	2,959.5134
..



Skyline (integration of EICs)

List of *N*-glycans showing their abundance



Perseus (statistical analysis)

List of *N*-glycans showing significant changes in their abundance

Figure 14. Workflow of relative quantification for *N*-glycome using Skyline and Perseus software.

4. Results and discussions

4.1 The APL cells derived proteome identification by different precipitation methods using nanoLC-MS/MS

After lysis of APL cells with SDC buffer, the other biomolecules were also existed in the extraction solution including free glycans (89), which complicated and misguided the glycoproteome derived *N*-glycome analysis. For this problem, TCA and chloroform/methanol precipitation methods were tested to compare the recovery efficiency and identification number of proteins. Based on the nanoLC-MS/MS proteome data, the total protein amount of lysate performed by SDC-based digestion was normalized as 100% and 3,086 identified proteins were yielded. The TCA precipitation yielded 65% total protein amount and 4,590 identified proteins. With the chloroform/methanol precipitation, 45% total protein amount and 4,253 identified proteins were obtained (Figure 15). Therefore, TCA precipitation was utilized for subsequent *N*-glycome analysis of APL and APL-6 cells.

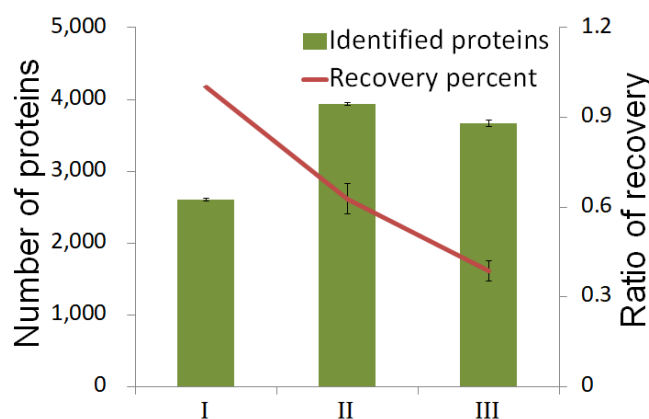


Figure 15. The recovery percent and identified proteins from the three different approaches. I) The SDC-based non-precipitated approach derived proteome analysis.

II) The TCA precipitation derived proteome analysis. III) The chloroform/methanol precipitation derived proteome analysis.

4.2 The comparison of different *N*-glycan purification approaches

Efficient *N*-glycan purification is indispensable for permethylation followed by MS-based structural characterization. 90% (v/v) ethanol precipitation, filter separation and tryptic digestion coupled to RP-SPE C₁₈ cartridge clean-up approaches were utilized to purify the *N*-glycans cleaved from Enbrel-G. After permethylation, total ion chromatograms (TICs) of above three purified *N*-glycan samples and one blank sample were compared and *N*-glycans were quantified based on the peak area of HexNAc₄Hex₃Fuc₁ (Figure 16), the most abundant *N*-glycan on etanercept (51).

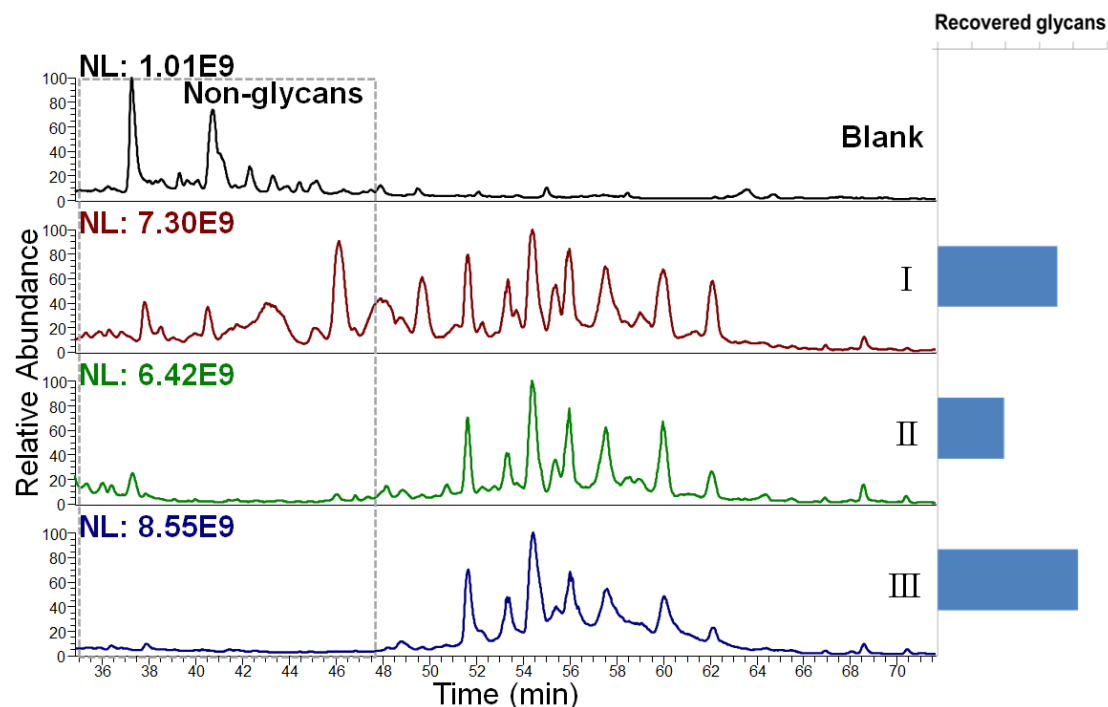


Figure 16. The TIC comparison and *N*-glycan recovery of different purification approaches after classical solid-phase permethylation. I) TIC of 90% (v/v) ethanol precipitation derived permethylated *N*-glycans. II) TIC of filter separation derived permethylated *N*-glycans. III) TIC of tryptic digestion coupled to RP-SPE C₁₈ cartridge

clean-up derived permethylated *N*-glycans.

The blank permethylation mainly showed two interferential peaks in the chromatogram, which were also observable in the other three permethylated *N*-glycan samples. 90% (v/v) ethanol precipitation derived permethylated *N*-glycans showed the highest abundance of non-glycans. The filter purification approach was more efficient than the 90% (v/v) ethanol precipitation in the separation of de-*N*-glycosylated proteins and *N*-glycans but with lowest *N*-glycan recovery efficiency among the three purification approaches. The tryptic digestion coupled to RP-SPE C₁₈ cartridge clean-up approach showed the highest *N*-glycan purity and recovery efficiency compared with the 90% (v/v) ethanol precipitation and filter separation approaches. The above samples were also compared using MALDI-MS (Figure 17). Especially, 90% (v/v) ethanol precipitation approach showed complicated mass spectrum due to the low removal efficiency of de-*N*-glycosylated proteins (Figure 17c). The tryptic digestion coupled to RP-SPE C₁₈ cartridge clean-up approach was the optimal purification approach and further applied to the subsequent *N*-glycan purification for OSPP development. During this digestion, the order of adding trypsin and PNGase F could be exchanged for etanercept, whose amino acid sequence was known clearly (105). However, PNGase F digestion was often performed prior to the addition of trypsin, as it cannot release the *N*-glycans from the terminal amino acid residues of the tryptic peptides (106).

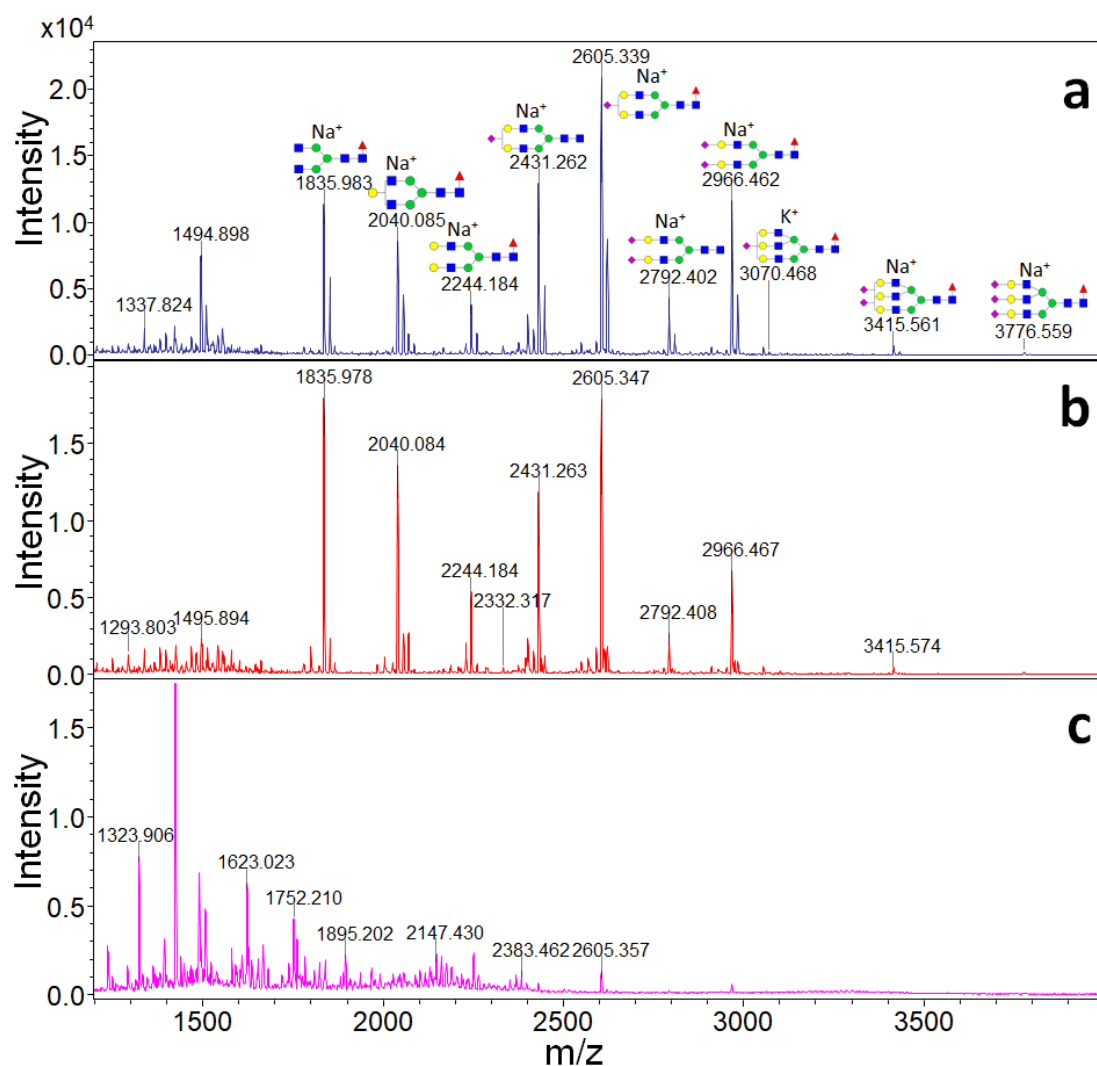


Figure 17. The MALDI mass spectrum comparison of permethylated *N*-glycans from different sample preparation approaches. (a) MALDI mass spectrum of permethylated *N*-glycans, obtained by initial trypsin digestion, followed by RP-SPE C_{18} cartridge purification. (b) MALDI mass spectrum of permethylated *N*-glycans, derived from the filter separation. (c) MALDI mass spectrum of permethylated *N*-glycans yielded by 90% (v/v) ethanol precipitation in the initial step, followed by centrifugation.

4.3 The development of optimized solid-phase permethylation

The permethylation happens at alkaline condition including two steps: (i) the deprotonation of proton-donating groups (hydroxyl, imino and carboxyl groups) (ii) the deprotonated groups are methylated by the methylating reagent (CH_3I) (107) (Figure 18). The stepwise reactions happen until all the hydrogen atoms of proton-donating groups are replaced

by the methyl groups.

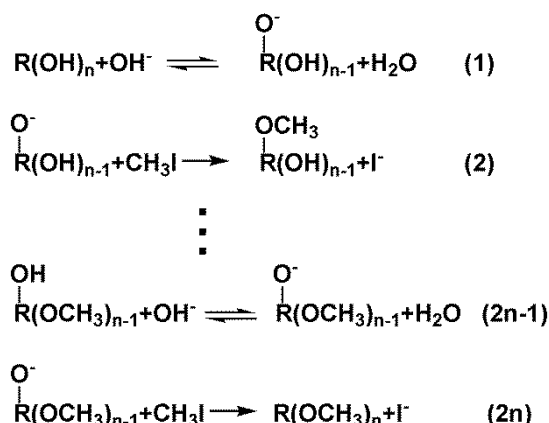


Figure 18. The schematic of glycan-based permethylation with hydroxyl groups representing proton-donating groups.

To explore the effect of sodium hydroxide on the proton-donating groups, the different concentrations of sodium hydroxide were compared and more permethylated *N*-glycans were obtained with the increased concentrations of sodium hydroxide in DMSO for 10 min (Figure 19a). The OSPP was compared to the slurry permethylation, showing that the OSPP-based approach harvested about 23% more permethylated *N*-glycans than the slurry permethylation (Figure 19b). The different ratios of water/DMSO were also compared in permethylation. With equivalent amount of *N*-glycans, the water/DMSO ratios of 10/100, 15/100 and 20/100 (v/v) obtained almost same amounts of permethylated *N*-glycans (Figure 19c). However, the ratio of 10/100 (v/v) was optimal due to the introduction of much less sodium hydroxide into the solution compared with ratios of 15/100 and 20/100 (v/v), while incomplete permethylation were resulted from water/DMSO ratios of 0/100 and 5/100 (v/v) in 10 min. Then different permethylation time, 10 min and 30 min, was compared (Figure 19d),

showing that almost equivalent amount of permethylated *N*-glycans were obtained. Above optimization proved that 10/100 (v/v) of water/DMSO and 10 min of incubation time were optimal experimental parameters when 100 μ L iodomethane and 200 mg sodium hydroxide beads were provided for glycan permethylation.

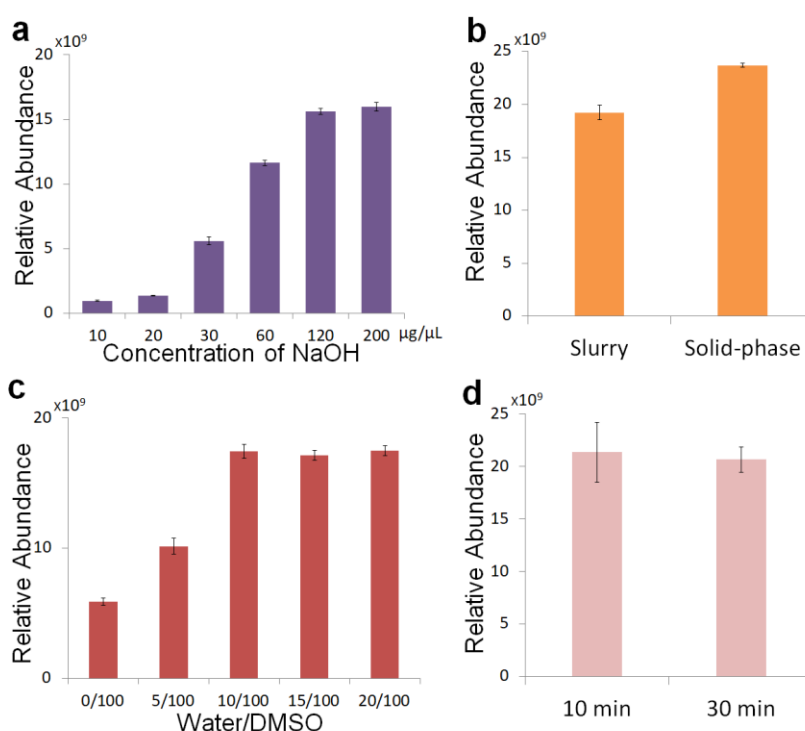


Figure 19. The quantitative comparisons of permethylated *N*-glycan yielded with different parameters for OSPP development. (a) The quantitative comparison of permethylated *N*-glycans derived from different concentrations of sodium hydroxide in slurry permethylation. (b) The quantitative comparison of slurry permethylation and OSPP. (c) The quantitative comparisons of different ratios of water/DMSO solution during OSPP. (d) The quantitative comparison of different reaction time for OSPP (10 min and 30 min). *The permethylated *N*-glycan yields were compared based on HexNAc₄Hex₃Fuc₁, the most abundant *N*-glycan of etanercept (51).

4.4 The elimination of co-reactions during permethylation

Although glycan permethylation is the main reaction in this system, there are also other co-reactions, including the oxidation degradation, peeling reaction and interconversion of α -, β - and γ -anomers. The oxidation

degradation occurs under anhydrous permethylation (Figure 20a) (98). The reaction 1 is an equilibrium process with the electron-acceptor character of DMSO. The generated sulfonium ion is attacked by carbohydrate hydroxyl ($R-CH_2-OH$), which is also equilibrium reaction (Reaction 2). Finally, the addition of base (OH^-) converts the alkoxy-sulfonium salt to dimethyl sulfide and aldehyde irreversibly (Reaction 3). The oxidation degradation can be totally eliminated by addition of water (1%, v/v) in DMSO. In OSPP, water was added and oxidation degradation would not be discussed in this study. Also, peeling reaction happens simultaneously, which is a base-catalyzed elimination reaction (108). Generally, peeling reaction starts with the generation of aldehyde monosaccharide at the reducing end. The double-bond of this saccharide is rearranged from C_1-C_2 till C_3-C_4 . Finally, the peeled *N*-glycans are generated in alkaline condition and start next peeling reaction (Figure 20b).

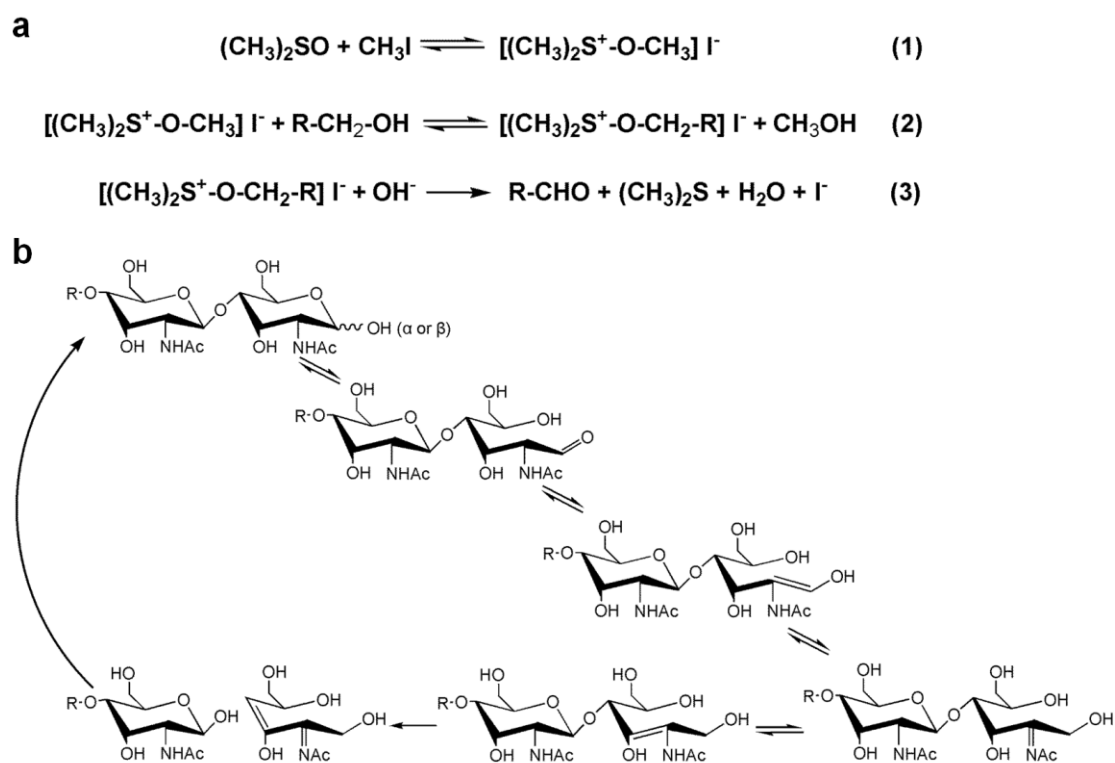


Figure 20. The oxidative degradation and peeling reaction during glycan permethylation. (a) The schematic of oxidative degradation during anhydrous permethylation. (b) The schematic of peeling reaction in alkaline condition.

Interestingly, after the glycans dissolved into the water/DMSO solution, the optimization by first addition of iodomethane ($\text{CH}_3\text{I}/\text{NaOH}$) or sodium hydroxide ($\text{NaOH}/\text{CH}_3\text{I}$) affected the peeling reaction significantly. Neu5Ac₂HexNAc₃Hex₅, peeled from Neu5Ac₂HexNAc₄Hex₅Fuc₀₋₁, was monitored. Each *N*-glycan species showed different anomers and were characterized by the MS2 spectra (Figure 21a, b, c). Neu5Ac₂HexNAc₃Hex₅ was confirmed by the unique fragment ion at m/z 1701.1644. The peeling reaction was quantified by the ratio calculated with peak area of Neu5Ac₂HexNAc₃Hex₅ divided by the total peak areas of Neu5Ac₂HexNAc₄Hex₅Fuc₀₋₁. In $\text{CH}_3\text{I}/\text{NaOH}$ workflow, a peeling rate of 0.55% was observed, while the peeling rate was 2.03% in the $\text{NaOH}/\text{CH}_3\text{I}$

workflow (Figure 21d). Above comparison demonstrated that the methylation and peeling reaction of monosaccharide at the reducing end was competitive in alkaline condition. Due to the low yield, the peeled *N*-glycans were not detectable by MALDI-MS, while it affected the glycan quantification and complicated *N*-glycan identification during nanoLC-MS/MS analysis.

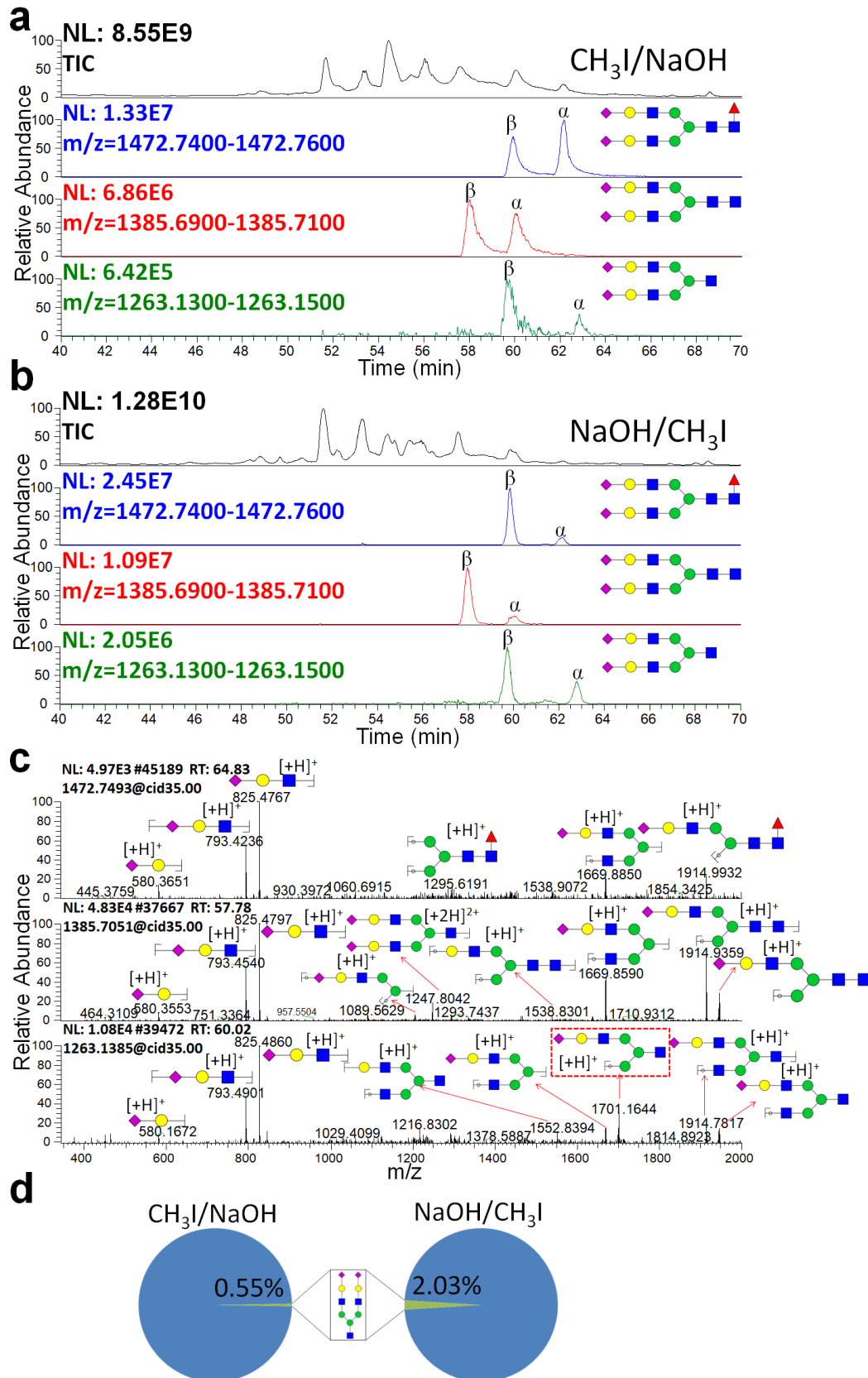


Figure 21. The quantitative comparison of peeling reaction between CH₃I/NaOH and NaOH/CH₃I workflows. (a) TIC of permethylated *N*-glycans by CH₃I/NaOH workflow and extracted ion chromatograms (EICs) of Neu5Ac₂HexNAc₃Hex₅ and

Neu5Ac₂HexNAc₄Hex₅Fuc₀₋₁. (b) TIC of permethylated *N*-glycans by NaOH/CH₃I workflow and EICs of Neu5Ac₂HexNAc₃Hex₅ and Neu5Ac₂HexNAc₄Hex₅Fuc₀₋₁. (c) MS2 spectra of Neu5Ac₂HexNAc₃Hex₅ and Neu5Ac₂HexNAc₄Hex₅Fuc₀₋₁. (d) The peeling ratio comparison of Neu5Ac₂HexNAc₄Hex₅Fuc₀₋₁ between CH₃I/NaOH and NaOH/CH₃I workflows.

Besides peeling reaction, the ratios of α - and β -anomers were also significantly different between CH₃I/NaOH and NaOH/CH₃I workflows. α - and β -anomers converts in dynamical balance as cyclic hemiacetal forms and γ -anomer is the intermediate between α - and β -anomers as a linear and free aldehyde form in solution (Figure 22a) (109). In alkaline medium, the β -D-glucose has been confirmed as the favored form accounting for about 80% compared with α -D-glucose (110), which agreed with the observation of predominant β -GlcNAc in the NaOH/CH₃I workflow. Focusing on the permethylation of HexNAc₄Hex₃Fuc₁, the NaOH/CH₃I workflow resulted in about a two-fold larger amount of β -GlcNAc and about a two-fold smaller amount of α -GlcNAc than CH₃I/NaOH workflow, while the γ -GlcNAc levels remained largely the same amount between two workflows (Figure 22b). In practice, the different anomers sophisticated the glycan identification and quantification.

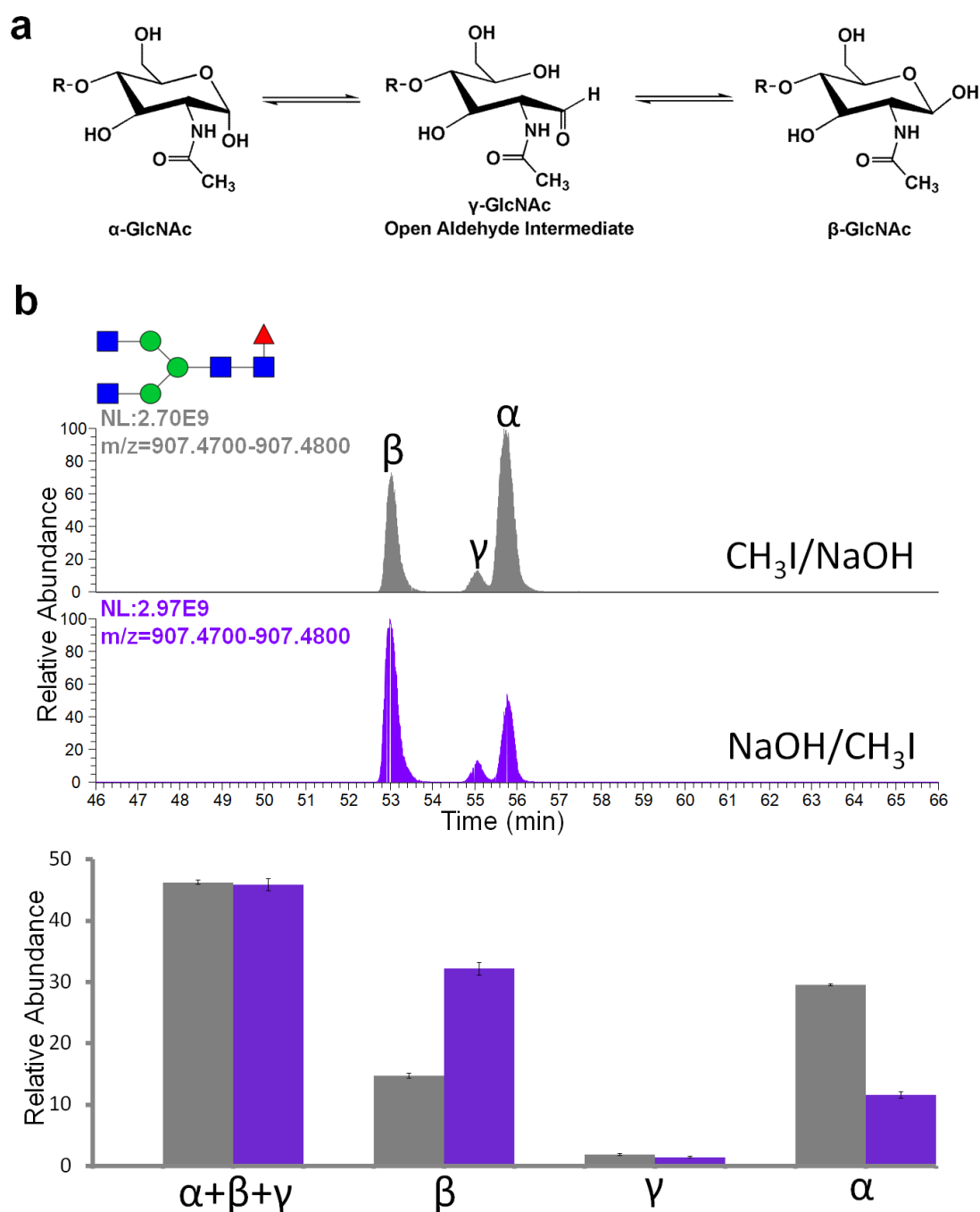


Figure 22. The GlcNAc interconversion and quantitative comparison of α -, β - and γ -GlcNAc anomers between $\text{CH}_3\text{I}/\text{NaOH}$ and $\text{NaOH}/\text{CH}_3\text{I}$ workflows. (a) The schematic of GlcNAc interconversion at the reducing end of the *N*-glycan. (b) Quantitative comparison of $\text{HexNAc}_4\text{Hex}_3\text{Fuc}_1$ with different anomers between $\text{CH}_3\text{I}/\text{NaOH}$ and $\text{NaOH}/\text{CH}_3\text{I}$ workflows.

To eliminate the effects from α -, β - and γ -GlcNAc anomers, the *N*-glycans were reduced by borane-ammonia prior to permethylation, which was compared with the non-reducing *N*-glycans. Permethyated, reduced *N*-

glycans have the uniform GlcNAc alditol on the reducing end, decreasing the complexities of samples by a factor of 3 (Figure 23a). Compared with the non-reducing *N*-glycans, the monoisotopic masses of *N*-glycans were increased by 16.0313 Da after permethylation, demonstrated by MALDI-MS analysis (Figure 23b).

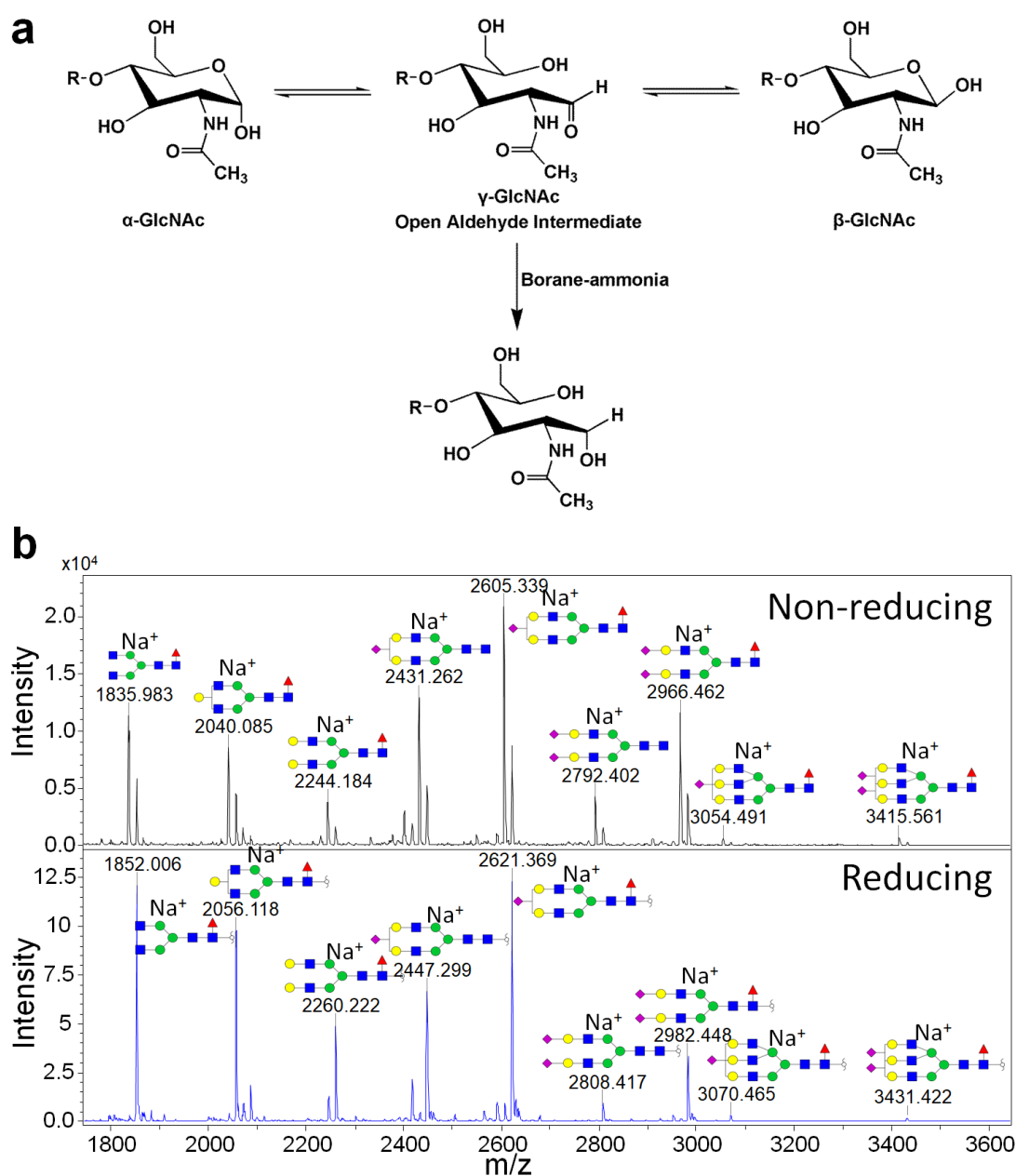


Figure 23. The schematic of reduction from GlcNAc anomers to GlcNAc alditol, with a mass increase of 16.0313 Da after permethylation. (a) The schematic of reduction from GlcNAc anomers to GlcNAc alditol by borane-ammonia. (b) MALDI-MS

analysis for non-reducing and reducing *N*-glycans after permethylation. Furthermore, the reducing and non-reducing *N*-glycans were compared after OSPP using nanoLC-MS/MS (Figure 24). The permethylated HexNAc₄Hex₃Fuc₁ and HexNAc₃Hex₃Fuc₁Red-HexNAc₁ were extracted from each sample (Figure 24a). In the permethylated non-reducing *N*-glycans, three peaks from α -, β - and γ -GlcNAc anomers of HexNAc₄Hex₃Fuc₁ were observed, while only one peak from HexNAc₃Hex₃Fuc₁Red-HexNAc₁ was present in the permethylated reducing species (Figure 24b). In addition, no permethylated HexNAc₄Hex₃Fuc₁ was found in the permethylated reducing *N*-glycans, showing complete reduction of *N*-glycans. Besides, the reduction also eliminated the peeling reaction totally during the permethylation due to the removal of hemiacetal at the reducing end of *N*-glycans. Here, Neu5Ac₂HexNAc₃Hex₅Fuc₀₋₁Red-HexNAc₁ and their peeled product, Neu5Ac₂HexNAc₃Hex₅, were extracted from the permethylated reducing *N*-glycans (Figure 25). Neu5Ac₂HexNAc₃Hex₅ was not detected, which demonstrated that peeling reaction was eliminated during the permethylation of reduced *N*-glycans.

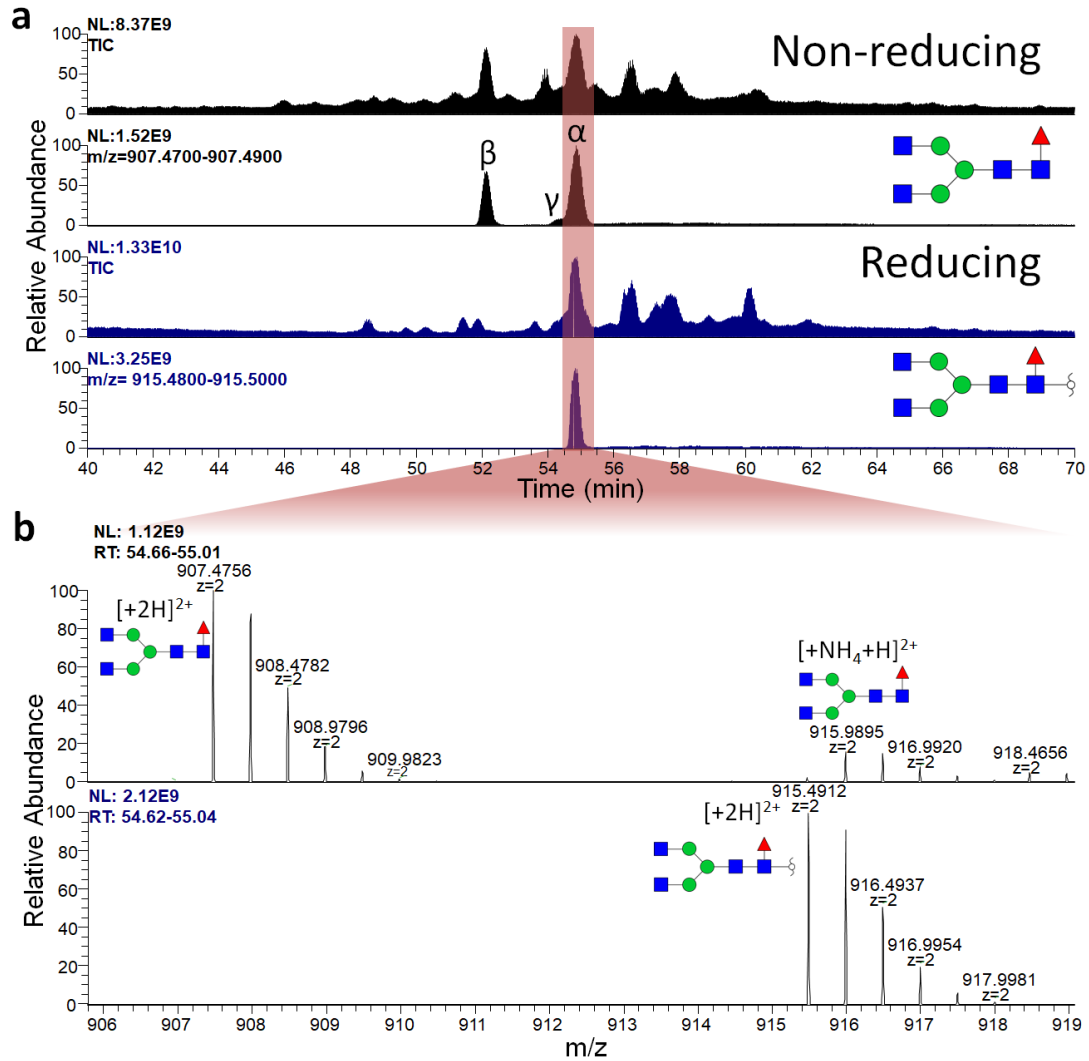


Figure 24. The nanoLC-MS/MS comparative analysis between permethylated non-reducing and reducing *N*-glycans. (a) The EICs of HexNAc₄Hex₃Fuc₁ and HexNAc₃Hex₃Fuc₁Red-HexNAc₁. (b) MS full scan from the EICs of HexNAc₄Hex₃Fuc₁ and HexNAc₃Hex₃Fuc₁Red-HexNAc₁ during the retention time from 54.5 to 55.5 min.

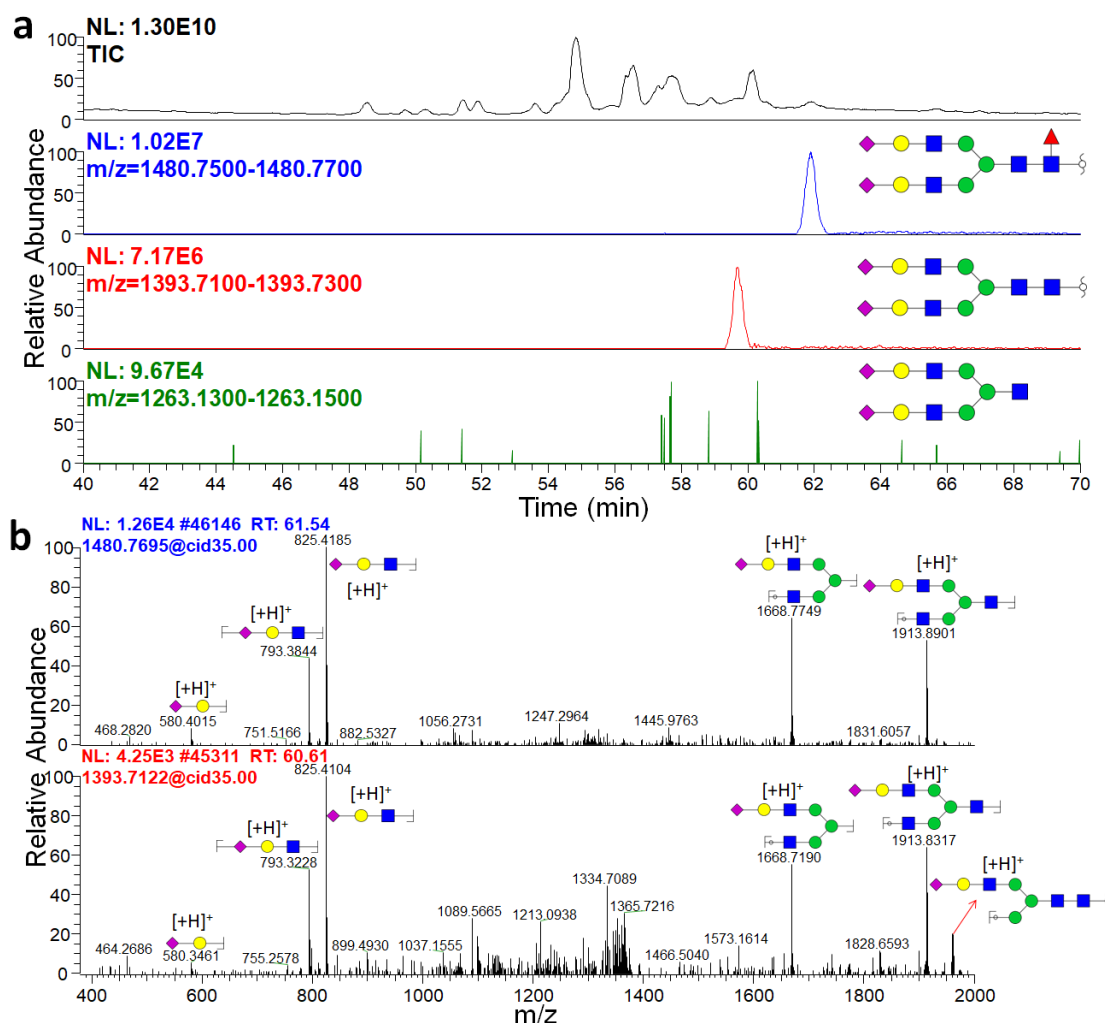


Figure 25. The assessment of peeling reaction from the permethylated reducing *N*-glycans released from Enbrel-G. (a) The TIC of permethylated reducing *N*-glycans and EICs of Neu5Ac₂HexNAc₃Hex₅Fuc₀₋₁Red-HexNAc₁ and Neu5Ac₂HexNAc₃Hex₅. (b) The MS2 structural elucidation of Neu5Ac₂HexNAc₃Hex₅Fuc₀₋₁Red-HexNAc₁.

4.5 The application of optimized glycan preparation workflow

After the optimization, the optimal workflows of *N*-glycan preparation were developed, which were tested using Enbrel-G as the example. Briefly, Enbrel-G was denatured, reduced, alkylated and then exchanged to 100 mM ABC buffer using a 3 kDa centrifugal filter. For *N*-glycan preparation, Enbrel-G was digested by PNGase F at 37°C for 24 h and then by trypsin for another 20 h, which was followed by RP-SPE C₁₈ cartridge purification as mentioned above. Next, the OSPP with optimal experimental parameters

was also performed (Figure 26). Briefly, purified *N*-glycans were firstly reduced by borane-ammonia. After evaporation by a SpeedVacTM vacuum concentrator, 110 μ L water/DMSO (10/100, v/v) and 100 μ L iodomethane was added and transferred into sodium hydroxide beads (200 mg) in a glass vial by rotation for 10 min. The solution was transferred into a new vial with one more addition of 150 μ L DMSO washing the sodium hydroxide beads. The permethylated glycans was purified using chloroform-water extraction and dried by a SpeedVacTM vacuum concentrator.

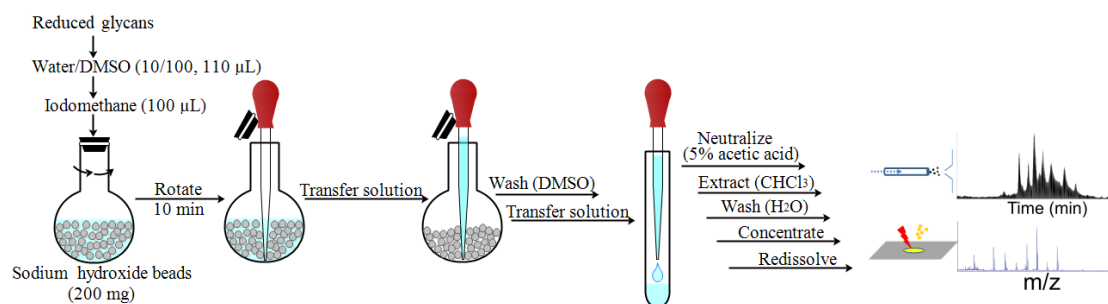


Figure 26. The optimal experimental parameters of OSPP workflow coupled with MS analysis.

To further test the optimal workflows, *N*-glycans were prepared from more glycoprotein and glycoproteome samples. Enbrel-H, chicken ovalbumin, EPO-3, EPO-4 and extracted proteins from corpus callosum tissue of an adult mouse were denatured, reduced, alkylated and then exchanged to 100 mM ABC buffer using 3 kDa centrifugal filters as described above. The proteins, precipitated by TCA from APL and APL-6 cell lysates after alkylation, were directly dissolved into 100 mM ABC buffer. Then optimal workflows of *N*-glycan preparation and OSPP were applied to each sample. Additionally, the *O*-glycans were released from Enbrel-H by β -elimination

and prepared with OSPP as described above.

4.6 The comparison between OSPP and reductive amination for *N*-glycan analysis

The native and ABBE aminated *N*-glycans were firstly measured by PGC-LC-QTOF-MS with the identification of a mass increase of 177.11536 Da (Figure 27).

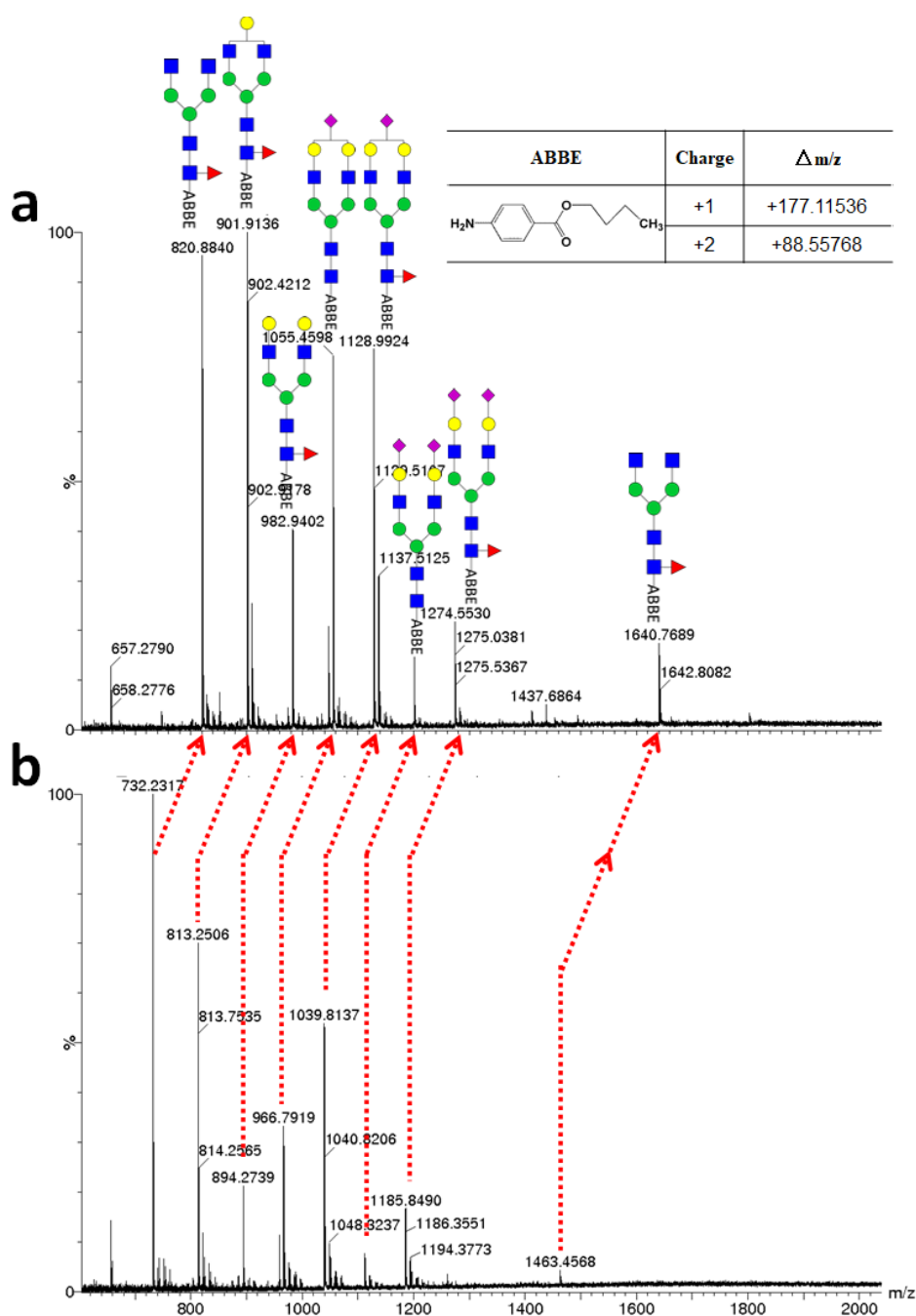


Figure 27. PGC-LC-QTOF-MS analysis for *N*-glycans before and after ABBE-based reductive amination. (a) ABBE aminated *N*-glycans identification. (b) The native *N*-glycans identification.

Due to the introduction of ABBE, the *N*-glycans are retained on the C₁₈ column strongly (111). With the injection of equimolar *N*-glycans, the reductive amination with ABBE was compared with OSPP by nanoC₁₈-LC-MS/MS (Figure 28a). The ABBE aminated HexNAc₄Hex₃Fuc₁ showed two isomeric structures, at 53.4 min (Peak I) and 55.2 min (Peak II), while the permethylated HexNAc₃Hex₃Fuc₁Red-HexNAc₁ was detected with only one peak at 54 min (Peak III). It was demonstrated that HexNAc₄Hex₃Fuc₁ was epimerized during the reductive amination. In addition, the amount of permethylated HexNAc₃Hex₃Fuc₁Red-HexNAc₁ was about six-fold larger than ABBE aminated species according to their peak areas. Their structures were elucidated by the MS2 fragments and two isomers from ABBE aminated HexNAc₄Hex₃Fuc₁ shared the same fragments (Figure 28b). Therefore, the OSPP workflow enabled simpler and more quantitative *N*-glycan analysis than ABBE-based reductive amination using nanoLC-MS/MS.

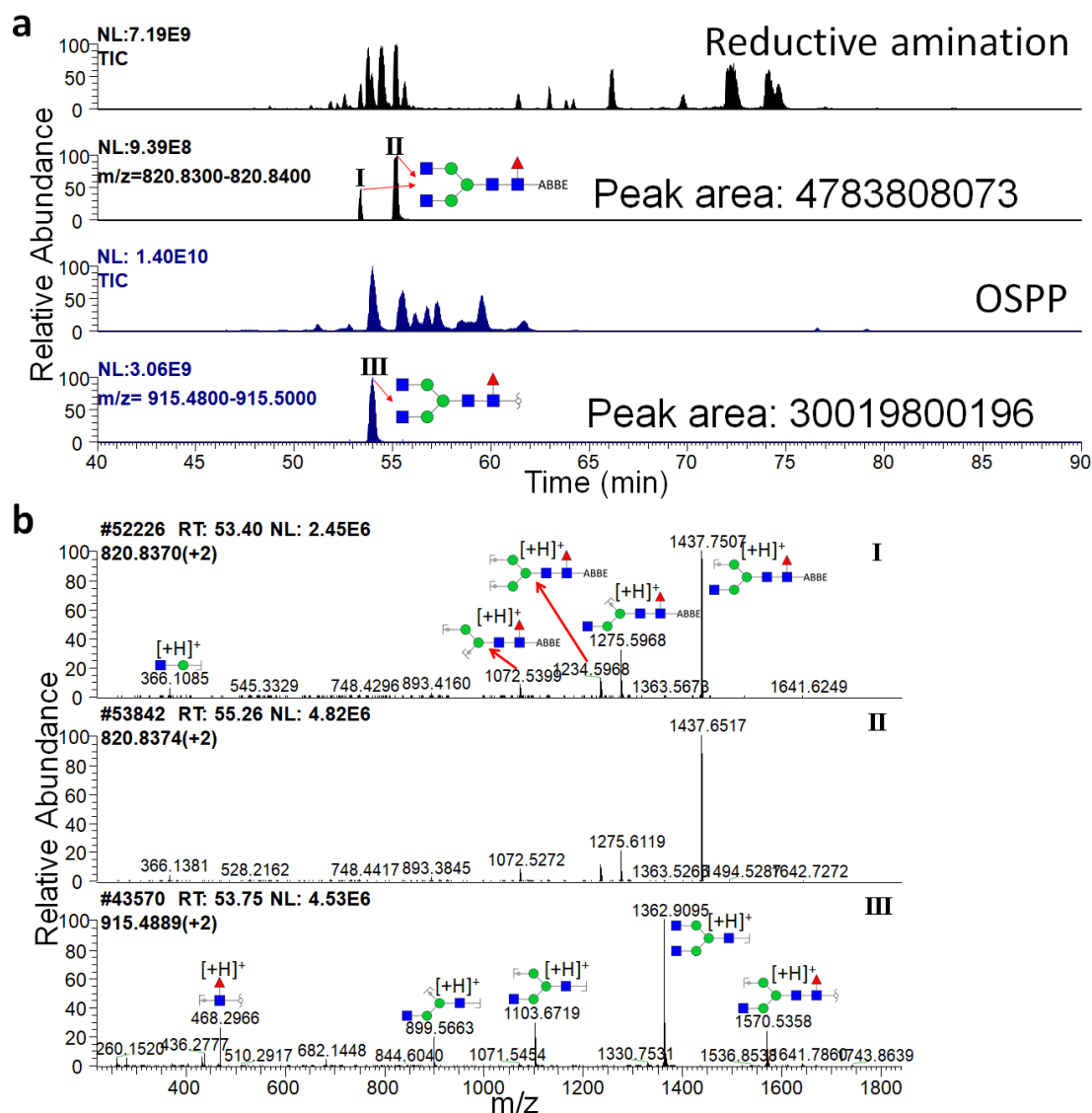


Figure 28. NanoLC-MS/MS analysis of *N*-glycans after reductive amination with ABBE and permethylation. (a) EICs of ABBE aminated HexNAc₄Hex₃Fuc₁ and permethylated HexNAc₃Hex₃Fuc₁Red-HexNAc₁. (b) MS2 spectrum elucidation for ABBE aminated HexNAc₄Hex₃Fuc₁ and permethylated HexNAc₃Hex₃Fuc₁Red-HexNAc₁.

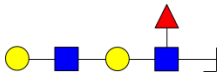
4.7 Developed bundled sequencing algorithm for *N*-glycan structure identification at MS2 level

The diagnostic MS2 ions of permethylated reducing *N*-glycans were utilized to develop a novel algorithm for structural identification. At MS2 level, the native glycans generated the A-, B-, C-, X-, Y- and Z-ion fragments by CID (112), sophisticating the structural identification. For the

permethylated *N*-glycans, B- and Y-ion fragments were mainly generated without any cross-ring cleavage. Different with peptides, most *N*-glycans were characterized with branching structures.

Table 2. The common MS2 diagnostic ions of permethylated reducing *N*-glycans.

MS2 ions	<i>m/z</i> (H ⁺)
	260.1492
	294.1911
	344.1704
	374.1809
	376.1966
	406.2072
	432.2228
	450.2334
	464.2490
	468.2803
	580.2964
	621.3229
	638.3382
	651.3335
	654.3331
	668.3488
	793.3965
	795.4121
	825.4227
	825.4227
	842.4380
	855.4333
	899.4595
	913.4751
	999.5119
	1029.5225

	1087.5643
---	-----------

The diagnostic MS2 ions from CID to interpret the glycan branching sequence are summarized (Table 2). Based on the diagnostic B- and Y-ions, the novel bundled sequencing algorithm is designed for *N*-glycan characterization at MS2 level including four steps. 1) To define the core structure (Fucosylated or not), starting with the number of Fuc (N_{Fuc}) (Figure 29a), for $N_{\text{Fuc}} > 0$, the issue if Fuc is linked on the Red-HexNAc is considered based on the m/z 468.2803 and the paired fragment of $M-466.26522 (+1)$ (M means the mass of the precursor and more charges are identified with increased molecular weight); if m/z 468.2803 and its paired fragment are not found, the fragment of $M-292.17601 (+1)$ is searched for confirmation of the non-fucosylated core structure (m/z 294.1911 derived from non-fucosylated Red-HexNAc is not always found). For $N_{\text{Fuc}} = 0$, the non-fucosylated core structure will be defined directly and Neu5Ac₁HexNAc₃Hex₄Red-HexNAc₁ is used as the example. The fragment of $M-292.17601 (+1)$ is identified (Fragment with +2 is also found) (Red asterisk). 2) To define the branching sequence by bundled groups, their MS2 diagnostic ions are searched. For Neu5Ac₁HexNAc₃Hex₄Red-HexNAc₁, the branching group Neu5Ac₁Hex₁HexNAc₁ is characterized by m/z 793.4504 and 825.4688 (Green asterisk). It is assembled onto the core structure as Neu5Ac₁HexNAc₂Hex₄Red-HexNAc₁. 3) To match the identified

monosaccharide composition with the assembled *N*-glycan structure from step 2, the assembled *N*-glycan will be decided to re-structure or not. For Neu5Ac₁HexNAc₂Hex₄Red-HexNAc₁, one more HexNAc will be assembled onto another branching (Figure 29b). 4) The deduced *N*-glycan structure is characterized with other fragments from the spectrum. This algorithm simplifies the glycan identification significantly, which is applicable to both *N*- and *O*-glycans. This bundled sequencing algorithm with the assistance of GlycoWorkbench will promote the development of advanced glycoinformatics.

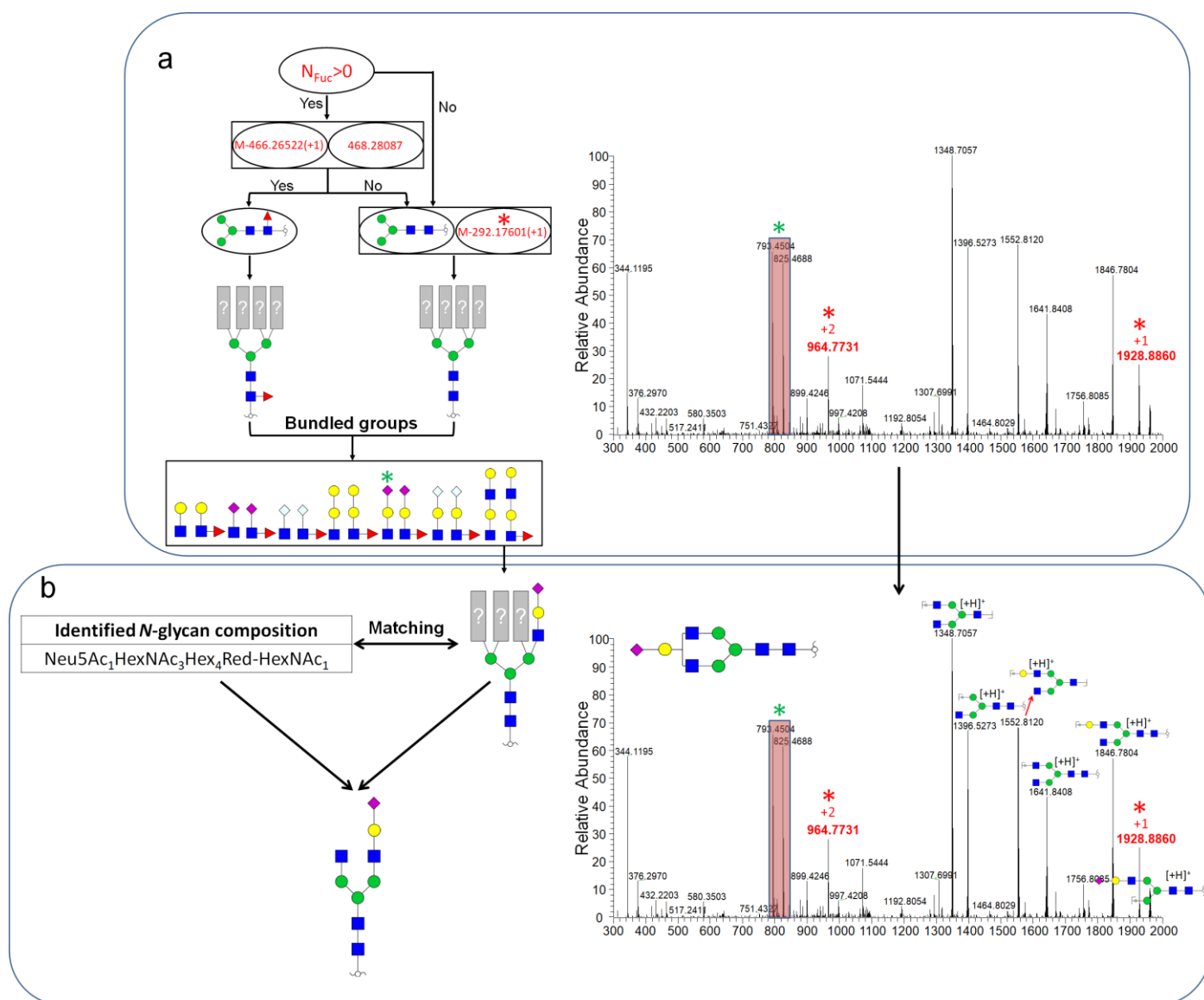


Figure 29. The development of a novel bundled sequencing algorithm for *N*-glycan structural identification at MS2 level. (a) The identification of fucosylated or non-fucosylated core structure and the bundled groups with diagnostic ions in MS2 spectrum. (b) The *N*-glycan structural identification after matching between monosaccharide composition at MS1 level and fragments at MS2 level.

4.8 FDR analysis for *N*-glycan identification

The co-existing of both ^{12}C - and ^{13}C -permethylated *N*-glycans firstly identified the *N*-glycan species at MS1 level (Figure 30a, b, c) and was further confirmed by monoisotopic m/z and its MS2 fragments. Enbrel-H achieved 62.5% for ^{12}C -permethylated *N*-glycans, 44.7% for ^{13}C and 8.6% for their intersection. In comparison, chicken ovalbumin achieved 60.1%, 42.4% and 10.3% respectively (Figure 30d). Therefore, the pairs of ^{12}C - and ^{13}C -permethylated *N*-glycans reduced the FDR significantly. The FDR was affected by the defined deviation threshold, purity of *N*-glycan sample and accuracy of MS1 data.

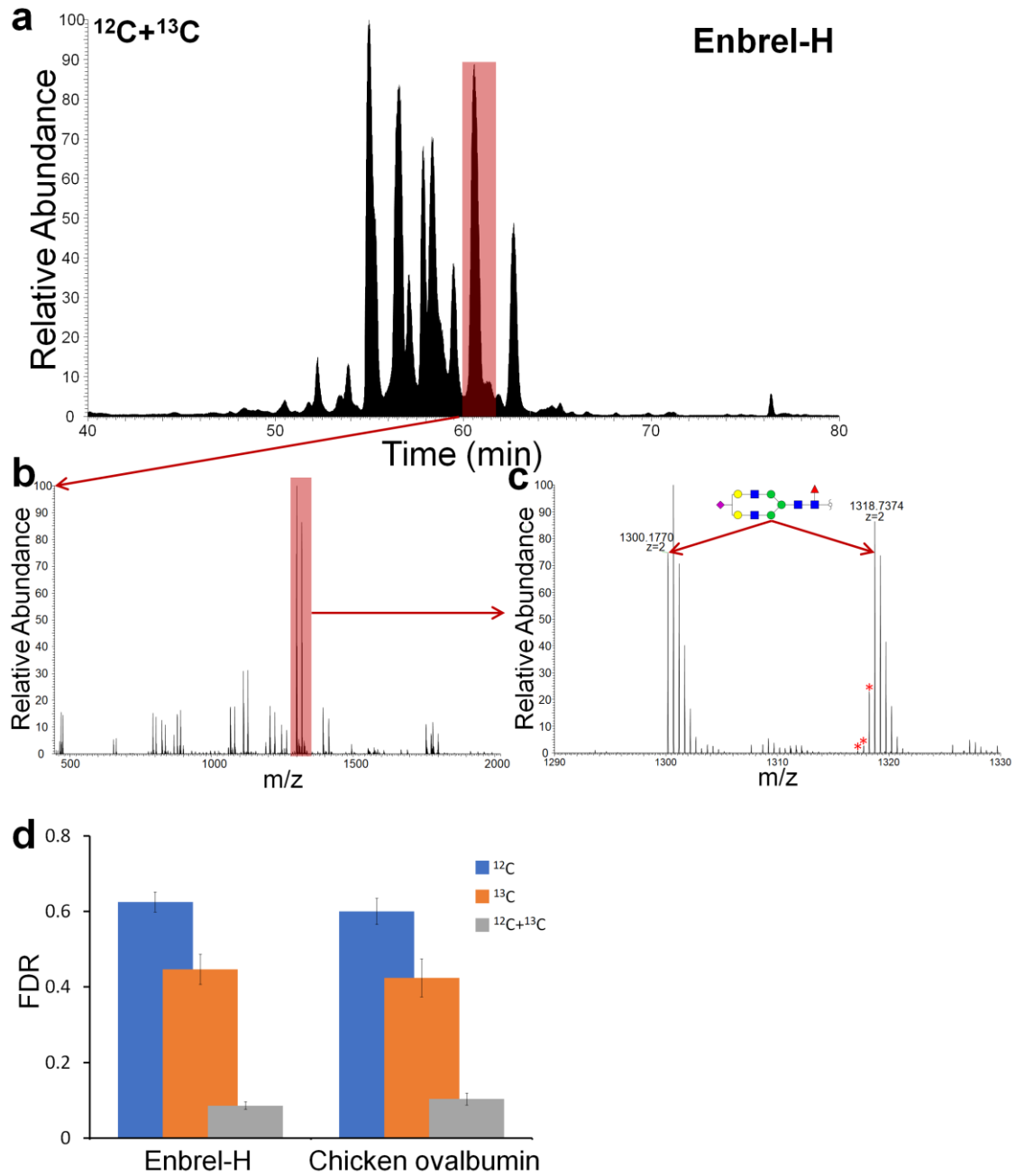


Figure 30. The pairs of ^{12}C - and ^{13}C -permethylated *N*-glycans for data quality control and FDR analysis. (a) TIC of mixture from ^{12}C - and ^{13}C -permethylated *N*-glycans of Enbrel-H. (b) The detected m/z at retention time from 60 to 61.5 min. (c) Neu5Ac₁HexNAc₃Hex₅Fuc₁Red-HexNAc₁ from pairs of ^{12}C - and ^{13}C -permethylation (Because the purity of $^{13}\text{CH}_3\text{I}$ was 99%, the glycans from incomplete ^{13}C -permethylation were also existed and labeled by red asterisks). (d) The isotope-based FDR analysis for the identified *N*-glycan species from Enbrel-H and chicken ovalbumin.

4.9 The identification of *N*-glycans from etanercept, chicken ovalbumin and erythropoietin

Etanercept derived *N*-glycans has been characterized by HPLC-FLR (51)

and MALDI-MS (113), which focused on the highly abundant *N*-glycans. In this study, the OSPP was performed to identify the *N*-glycans released from Enbrel-H, chicken ovalbumin, EPO-3 and EPO-4 by this newly designed strategy for data analysis. The permethylated reducing *N*-glycans were analyzed by MALDI-MS and nanoLC-MS/MS. For Enbrel-H, 32 monosaccharide compositions were identified by MALDI-MS (Figure 31a), while 90 monosaccharide compositions (162 *N*-glycans) were identified by nanoLC-MS/MS and all the *N*-glycans from MALDI-MS were present in the identification by nanoLC-MS/MS (Figure 31b). The number of identified monosaccharide compositions was at least three-fold larger than other recent studies (Figure 31c) and covered all the reported *N*-glycans. For chicken ovalbumin, 23 monosaccharide compositions were identified by MALDI-MS (Figure 31d), while 57 monosaccharide compositions (133 *N*-glycans) were identified by nanoLC-MS/MS (Figure 31e) and at least two-fold larger than other recent studies (Figure 31f). In addition, 133 monosaccharide compositions (230 *N*-glycans) from both of EPO-3 and EPO-4 were identified (data not shown).

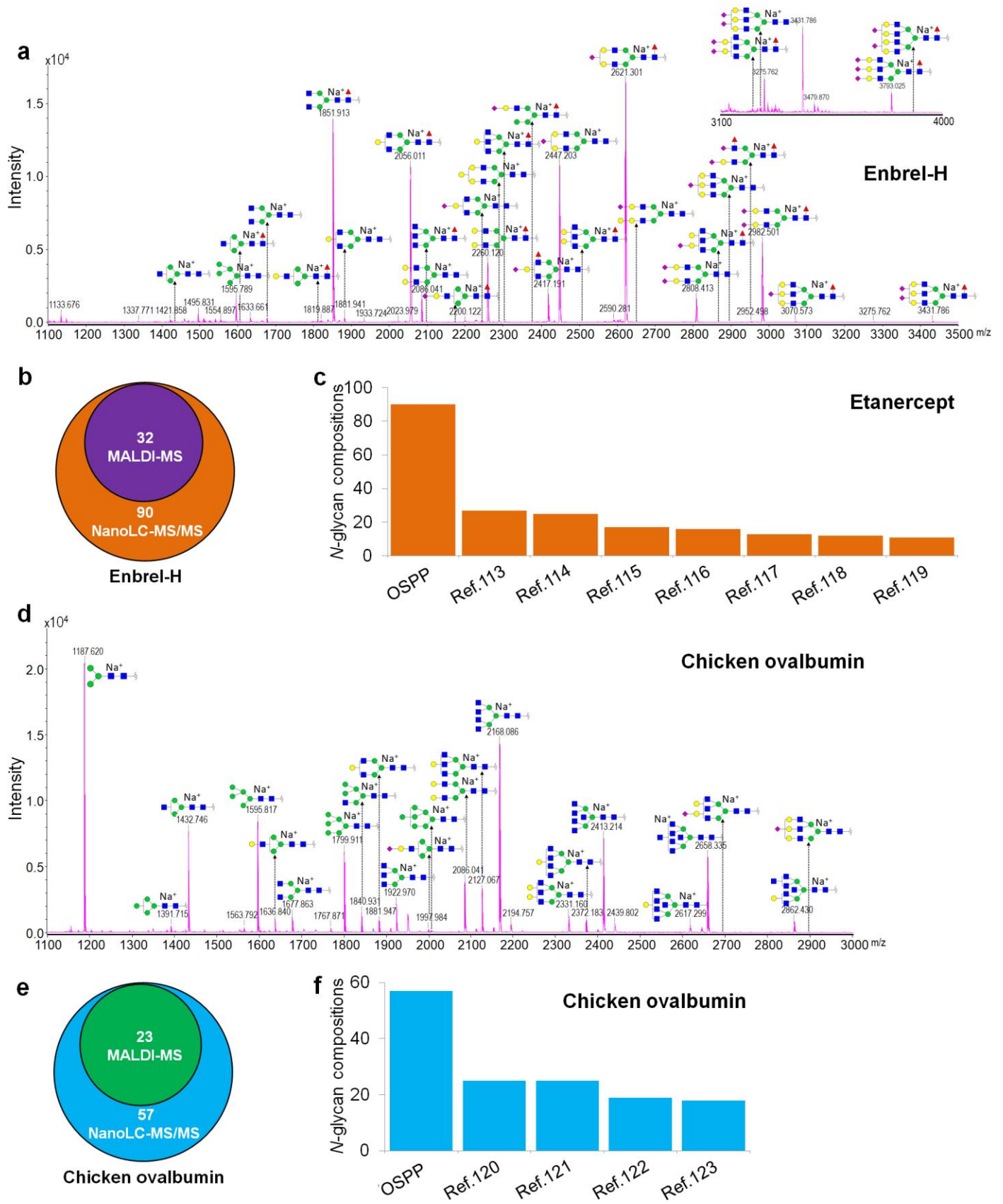


Figure 31. The identified *N*-glycans derived from purified glycoproteins by MALDI-MS and nanoLC-MS/MS. (a) Etanercept, using Enbrel-H product in this study, derived *N*-glycan analysis after OSPP preparation by MALDI-MS. (b) Venn diagram comparison of Enbrel-H derived *N*-glycan identification between MALDI-MS and nanoLC-MS/MS analysis. (c) The quantitative comparison of identified *N*-glycans from etanercept between this study and other recent studies. (d) Chicken ovalbumin derived *N*-glycan analysis after OSPP by MALDI-MS. (e) Venn diagram comparison of chicken ovalbumin derived *N*-glycan identification between MALDI-MS and nanoLC-MS/MS analysis. (f) The quantitative comparison of identified *N*-glycans from chicken ovalbumin between this study and other recent studies.

In addition to the monosaccharide compositions, nanoC₁₈-LC-MS/MS enabled the separation and identification of *N*-glycan isomers. For example, Neu5Ac₁HexNAc₃Hex₄Fuc₁Red-HexNAc₁ from Enbrel-H was measured at m/z 2,417.229 ($M+Na^+$) by MALDI-MS (Figure 32a), while four isomers were separated by nanoC₁₈-LC system (Figure 32b, c), showing more structural information of this *N*-glycan species.

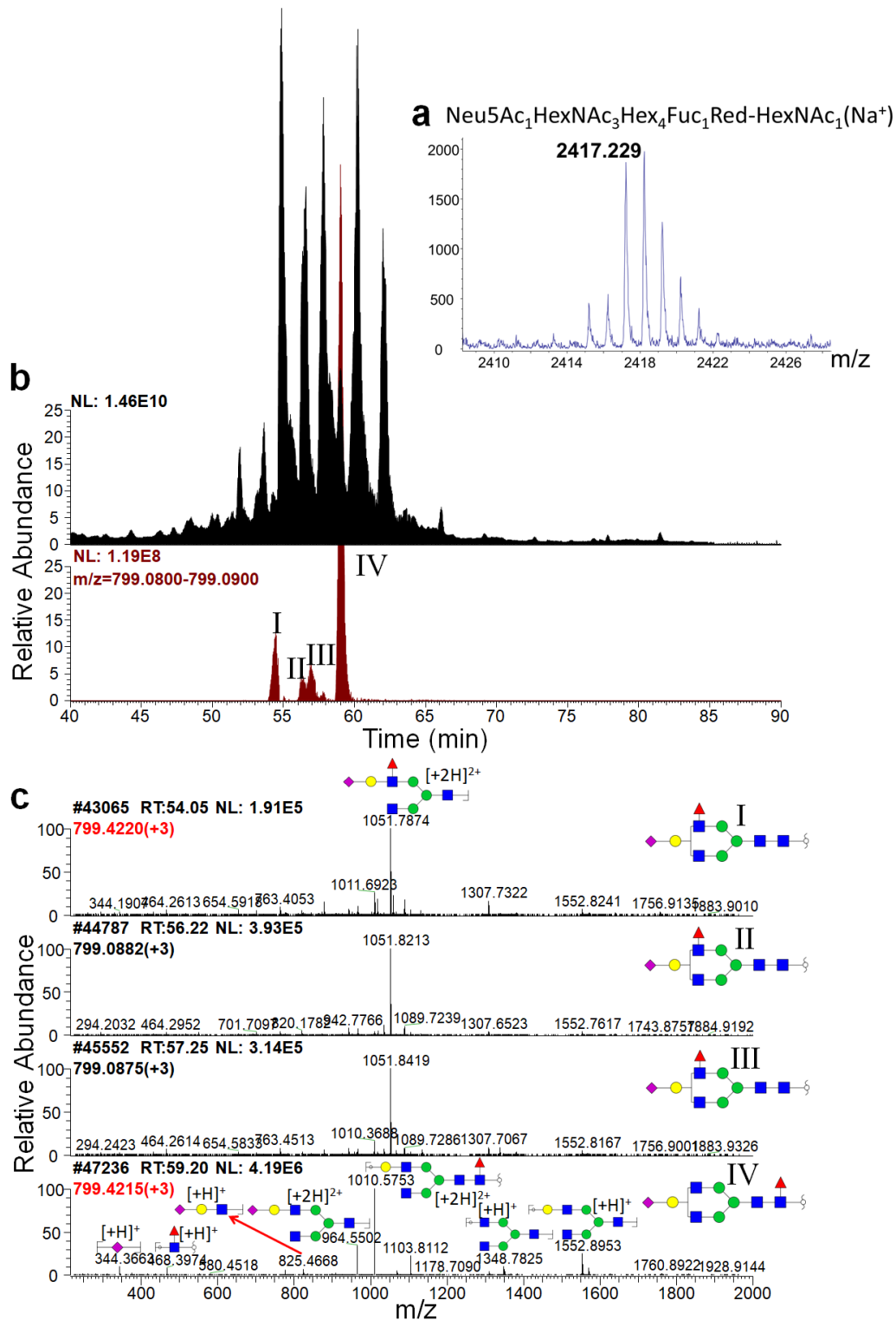


Figure 32. The comparison of MALDI-MS and nanoLC-MS/MS for glycan analysis. (a) MALDI-MS analysis of Neu5Ac₁HexNAc₃Hex₄Fuc₁Red-HexNAc₁, detected as sodium adduct. (b) EICs of Neu5Ac₁HexNAc₃Hex₄Fuc₁Red-HexNAc₁ isomers, separated as peak I, II, III and IV by nanoLC-MS/MS. (c) The MS2 spectra for structural elucidation of Neu5Ac₁HexNAc₃Hex₄Fuc₁Red-HexNAc₁ isomers. *The black m/z represented the monoisotopic precursor; red m/z represented the second peak of isotopic patterns.

4.10 The identification of *O*-glycans from etanercept

O-glycans of Enbrel-H were also analyzed by nanoLC-MS/MS after OSPP, while the peeling reaction from β -elimination happened. In this analysis, three Neu5Ac₁Hex₁Red-HexNAc₁ isomers and Neu5Ac₁Red-Hex₁ were identified (Figure 33). Neu5Ac₁Red-Hex₁ has been confirmed as the peeled *O*-glycan by β -elimination (115). Therefore, a method employing gentle release of intact *O*-glycans, similar to the PNGase F-mediated release of *N*-glycans, was critical for accurate *O*-glycome profiling.

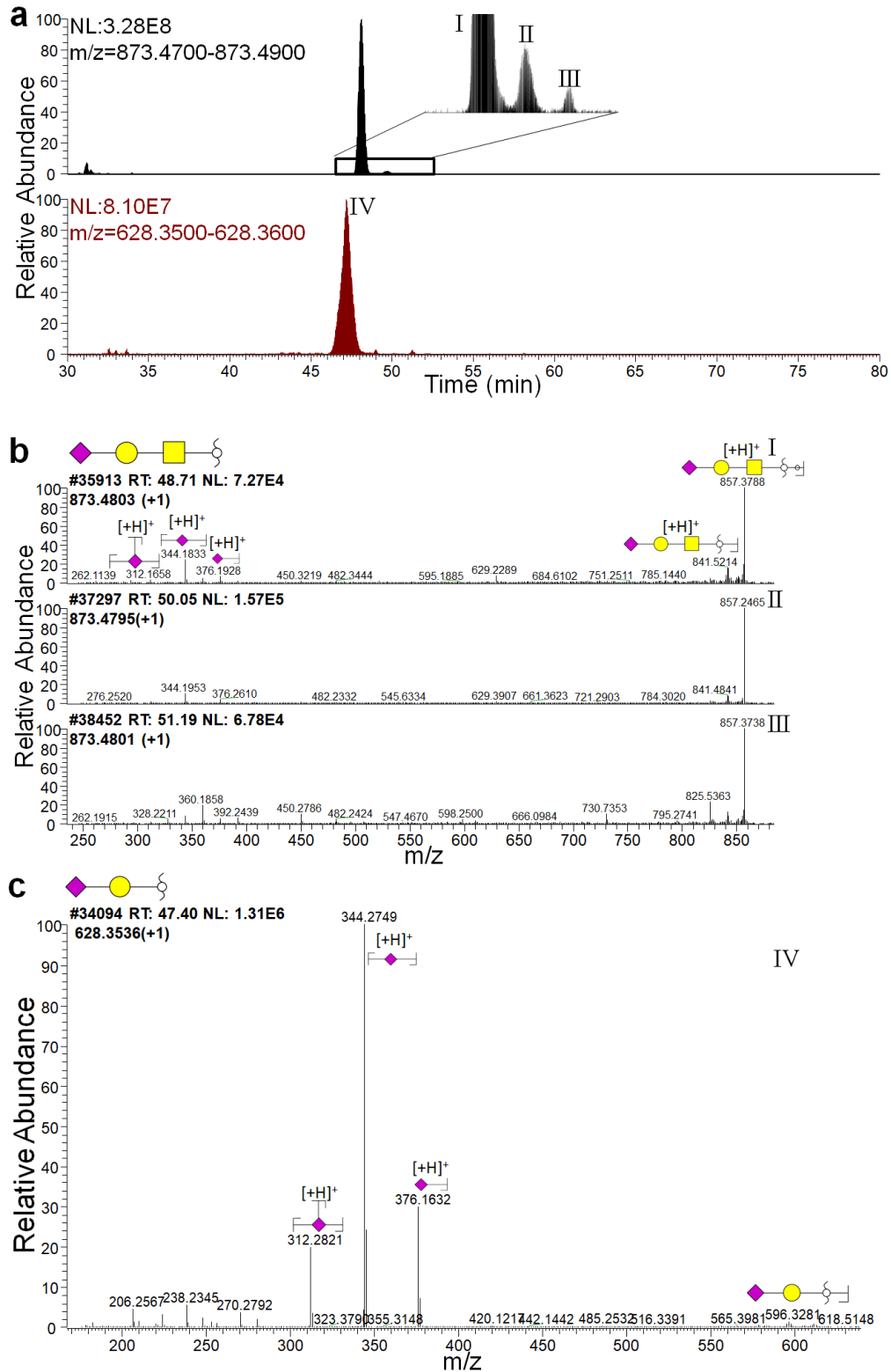


Figure 33. The identification of *O*-glycans from Enbrel-H after permethylation. (a) The EICs of Neu5Ac₁Hex₁Red-HexNAc₁ isomers, separated as peak I, II and III and the EIC of Neu5Ac₁Red-Hex₁, detected as peak IV. (b) The MS2 structural characterization for Neu5Ac₁Hex₁Red-HexNAc₁ from peak I, II and III. (c) The MS2 structural characterization for Neu5Ac₁Red-Hex₁ from peak IV.

4.11 The analysis of *N*-glycome from APL cells

The *N*-glycome samples, APL-H and APL-6-H, were analyzed using nanoLC-MS/MS by three replicates respectively. Recently, Xiao *et al.* have reported a large-scale characterization of 128 monosaccharide compositions (214 *N*-glycans) from human liver based on the database engine GlySeeker (124). In this study, 245 monosaccharide compositions (398 *N*-glycans) were identified from APL cells. The deviation for each *N*-glycan identification was analyzed, with an average of about 0.4 p.p.m. at MS1 level (Figure 34a). The identified *N*-glycans were classified based on fucosylation (Fuc) and sialylation (Neu5Ac) (Figure 34b). 47% of the species were fucosylated and sialylated. The ratio of fucosylation was 24% and sialylation 17%. Based on the *N*-glycan biosynthesis process (125, 126), identified *N*-glycans were further traced back into the cell biosynthesis map and separated into endoplasmic reticulum (ER) and Golgi apparatus (Figure 34c). In this biosynthesis pathway, all the *N*-glycans were identified at both MS1 and MS2 levels, summarized in Supplementary table 1 (Page 113).

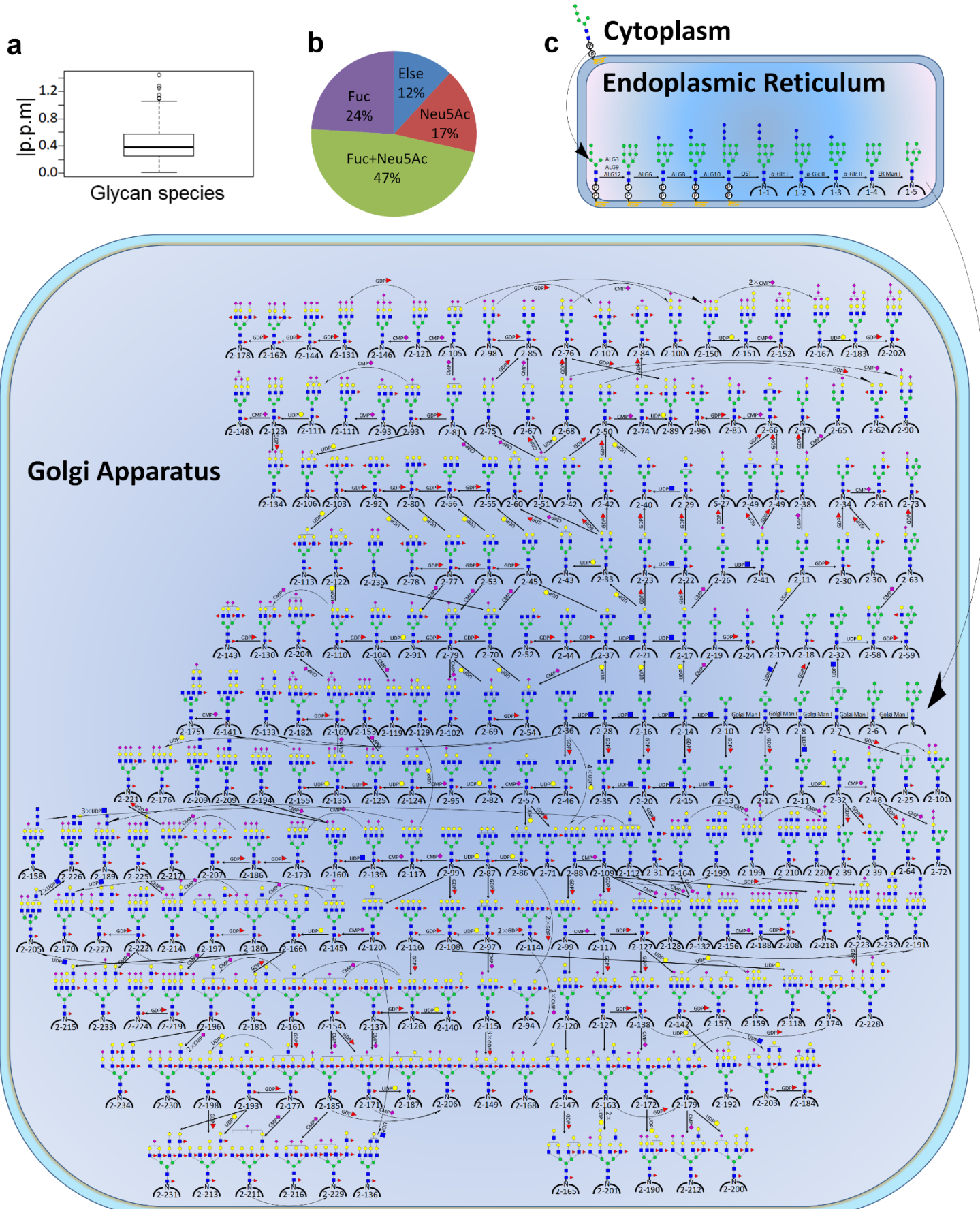


Figure 34. The deep analysis of *N*-glycome from APL cells by nanoLC-MS/MS. (a) Deviation analysis for all the identified *N*-glycans from APL cells. (b) The ratios of different *N*-glycan species derived from fucosylation (Fuc) and sialylation (Neu5Ac). (c) The biosynthesis map of identified *N*-glycans from APL cells.

4.12 The analysis of *N*-glycome from mouse corpus callosum

The *N*-glycans from mouse corpus callosum were also characterized at MS1 and MS2 levels. In total, 343 monosaccharide compositions (832 *N*-glycans) were identified. Of them, 64% was sialylated by only Neu5Ac, 5% by only Neu5Gc and 3% by both of them (Figure 35a). Based on MS2 diagnostic ions, 344.1704 from Neu5Ac and 374.1809 from Neu5Gc, the quantitative analysis of Neu5Ac and Neu5Gc were performed to show different sialylation levels on *N*-glycome based on their peak areas. In mouse corpus callosum, the amount of Neu5Gc was about 14% of Neu5Ac (Figure 35b).

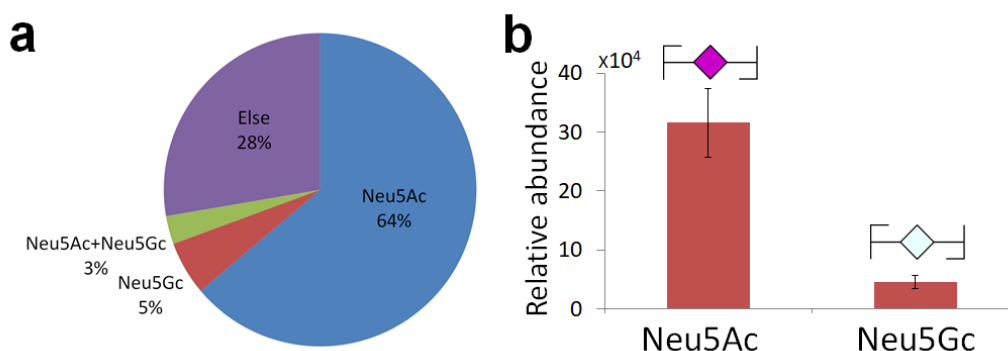


Figure 35. The sialylation comparison between Neu5Ac and Neu5Gc of *N*-glycome from mouse corpus callosum. (a) The ratios of different *N*-glycans with sialylation by Neu5Ac and Neu5Gc. (b) Quantitative analysis by MS2 ions of Neu5Ac and Neu5Gc.

4.13 Post-glycosylational modification analysis of *N*-glycans

Although post-glycosylational modifications were not considered in the above R-scripts, they were also identified in this study. Phosphate groups are commonly attached on C6 position of Man residues of high-mannose type glycans (19, 127). Phosphorylated Man (P-Hex) was searched from the MS raw data of erythropoietin derived *N*-glycans, in which the mass

increase of 93.98198 Da was identified from HexNAc₁Hex₅P-Hex₁Red-HexNAc₁ compared to non-phosphorylated species (Figure 36a). The diagnostic fragment ion at m/z 313.1720 was generated from terminal P-Hex of the *N*-glycan by CID fragmentation (Figure 36b).

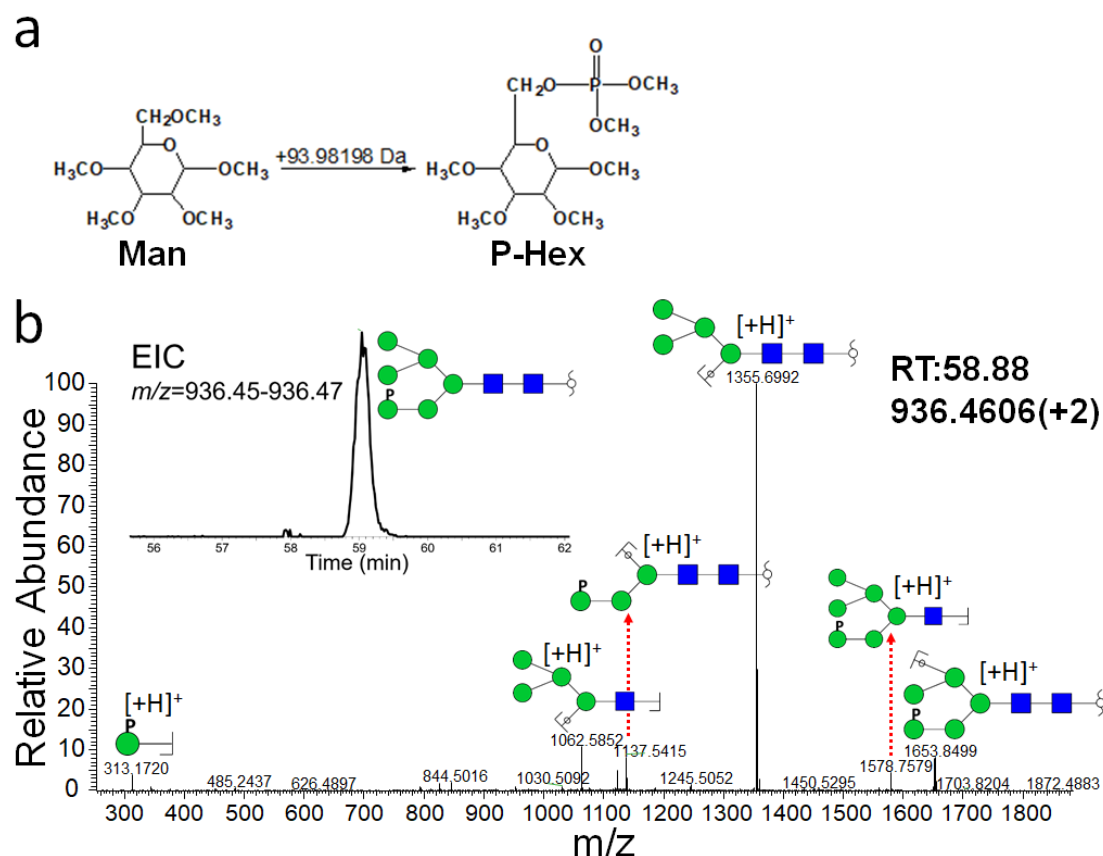


Figure 36. The MS-based characterization of phosphorylated *N*-glycans. (a) The structures of permethylated Man and phosphorylated Man (P-Hex). (b) The characterization of HexNAc₁Hex₅P-Hex₁Red-HexNAc₁ at MS1 and MS2 levels, by MS raw data of erythropoietin derived *N*-glycans.

O-acetylation is also a common post-glycosylational modification linked to either C4-, C7-, C8-, or C9-hydroxyl position of sialic acid residues (128), while it is removed during permethylation (129). However, pGlyco cannot identify these glycan modifications at glycopeptide level, which will promote next, necessarily required improvement of this software. Glycopeptides enable the identification of O-acetylated sialic acids

manually, with a mass increase of 42.01057 Da from non-O-acetylated species. As an example, the glycopeptide LVN(83)SSQPW(Neu5Ac₂HexNAc₆Hex₇Fuc₁) from erythropoietin after chymotrypsin digestion, was identified by using pGlyco with m/z 1338.5282 (Figure 37a). With an addition of one O-acetylated sialic acid, LVN(83)SSQPW(Neu5Ac+OAc₁Neu5Ac₁HexNAc₆Hex₇Fuc₁) was identified with m/z 1352.5311. Based on the diagnostic fragments, 316.1012, 334.1119 and 699.2453 could be identified at the MS2 level and provided evidences for the existence of Neu5Ac+OAc monosaccharide species on the glycopeptide (Figure 37b).

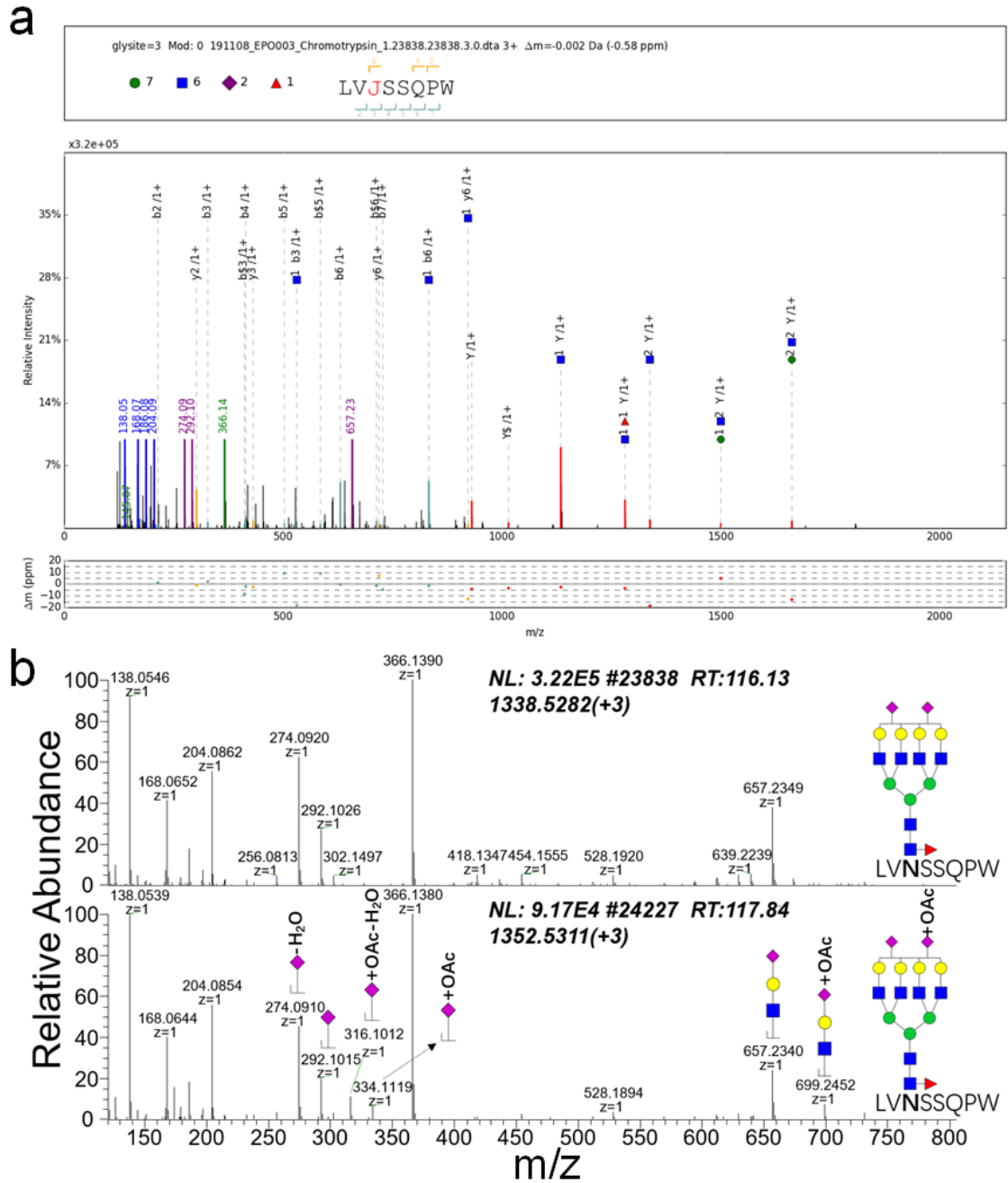


Figure 37. The identification of O-acetylated sialic acid at glycopeptide level. (a) The identified glycopeptide, LVN(83)SSQPW(Neu5Ac₂HexNAc₆Hex₇Fuc₁) by pGlyco. (b) The MS2 spectrum comparison between LVN(83)SSQPW(Neu5Ac₂HexNAc₆Hex₇Fuc₁) and LVN(83)SSQPW(Neu5Ac+OAc₁Neu5Ac₁HexNAc₆Hex₇Fuc₁).

4.14 The comparison between the OSPP-based *N*-glycan identification and glycopeptide analysis

pGlyco is an efficient software to analyze glycopeptides, providing the monosaccharide compositions, peptide sequences and glycosylation sites.

It was utilized to verify the monosaccharide compositions identified in my developed strategy. The samples were analyzed by technical triplicates and the identifications in at least 2 replicates were considered as valid data. For Enbrel-H, the monosaccharide compositions identified by pGlyco were compared with the *N*-glycan species identified based on OSPP preparation (Figure 38a). 33 monosaccharide compositions were identified by both approaches. 3 monosaccharide compositions were only identified by pGlyco, including EEQYN(317)STYR(HexNAc₃Hex₂Fuc₁) (Figure 39). Additionally, 57 monosaccharide compositions were exclusively identified after OSPP preparation. The *N*-glycome of APL cells was also compared between OSPP-based glycan analysis and pGlyco-based glycopeptide analysis (Figure 38b). 238 monosaccharide compositions were identified by pGlyco. 127 monosaccharide compositions were identified by both approaches; 111 species were exclusively identified by pGlyco and 118 by OSPP-based glycan analysis. A large overlap between two approaches was confirmed in *N*-glycan analysis, while the isomeric *N*-glycans were difficult to distinguish by pGlyco.

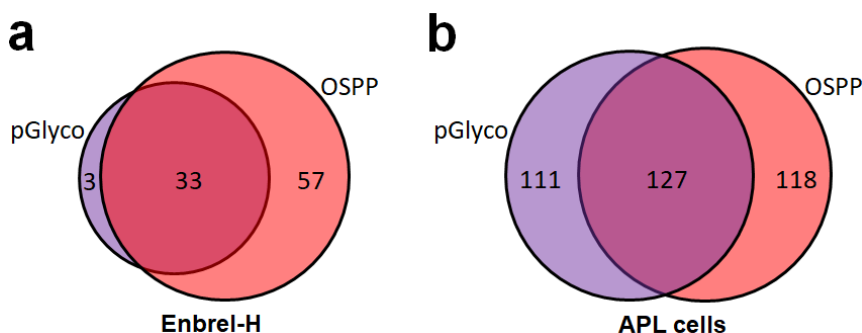


Figure 38. Comparative analysis of identified monosaccharide compositions between OSPP-based *N*-glycan analysis and glycopeptide identification with pGlyco. (a) The

responses of patients (130). During the biosynthesis, cytidine 5'-monophosphate (CMP)-Neu5Ac in the cytosol is the precursor to generate CMP-Neu5Gc, catalyzed by CMP-Neu5Ac hydroxylase (CMAH) (131). Then CMP-Neu5Gc is transferred into Golgi and assembles glycoconjugates by various sialyltransferases. The ratio between Neu5Ac and Neu5Gc, used as an index to quantify the generated Neu5Gc-containing *N*-glycans of erythropoietin, was measured based on the peak areas of the MS2 fragments: m/z 344.1704 from Neu5Ac and 374.1809 from Neu5Gc. The analysis revealed that the amount of Neu5Ac was about 18-fold larger compared to Neu5Gc in EPO-3 and 17-fold larger compared to Neu5Gc in EPO-4 (Figure 40b). After two-sample student's T-test analysis by Perseus software (EPO-3 vs EPO-4), the significantly differential abundant *N*-glycan species were visualized in a volcano plot. *N*-glycans that showed a at least 1.5- or 2-FC between the compared batches, were considered in further analysis (Figure 40c). Neu5Ac₁HexNAc₇Hex₈Fuc₁Red-HexNAc₁ (Green dot) showed a 1.7-FC higher abundance in EPO-4 compared to EPO-3. In contrast, eleven *N*-glycan species showed a 1.5-FC higher abundance in EPO-3 compared to EPO-4 (1.5-FC < Blue dots < 2-FC and Red dots > 2-FC). To further investigate the significantly different *N*-glycans, all candidates, showing an abundance difference higher than 2-FC, were compared in the heat map (Figure 40d). Most significantly, HexNAc₂Hex₃Red-HexNAc₁ was 4.86-

FC higher in EPO-3 compared to EPO-4. Besides, most *N*-glycans, 121 species (91%), fluctuated within 1.5-FC between two batches of erythropoietin samples. Therefore, the production of erythropoietin showed high stability and repeatability across different batches.

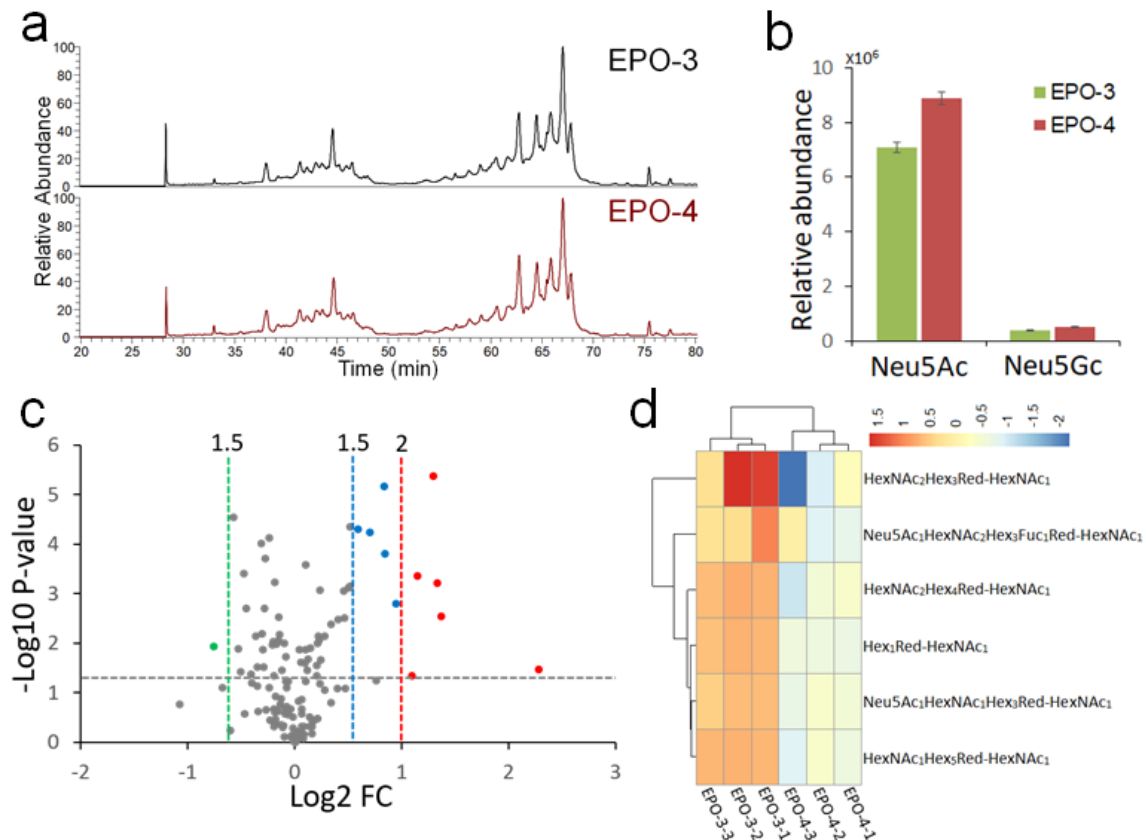


Figure 40. The quantitative comparison of identified monosaccharide compositions between EPO-3 and EPO-4. (a) The TIC comparison of permethylated *N*-glycans from EPO-3 and EPO-4. (b) The quantitative comparison of sialylation levels with Neu5Ac and Neu5Gc between EPO-3 and EPO-4. (c) The quantitative comparison of all identified monosaccharide compositions with volcano plot (EPO-3 vs EPO-4). (d) The heat map of identified monosaccharide compositions with more than 2-FC (EPO-3 vs EPO-4).

4.16 Stable isotopic labeling quantification of APL and APL-6 cells derived *N*-glycans

The stable isotopic labeling quantification has been an extensive approach to quantify the biomolecules by MS, especially for proteomics (132). The

N-glycome samples from APL cells, permethylated by iodomethane- d_3 and iodomethane- ^{13}C , were compared with APL-6-H respectively.

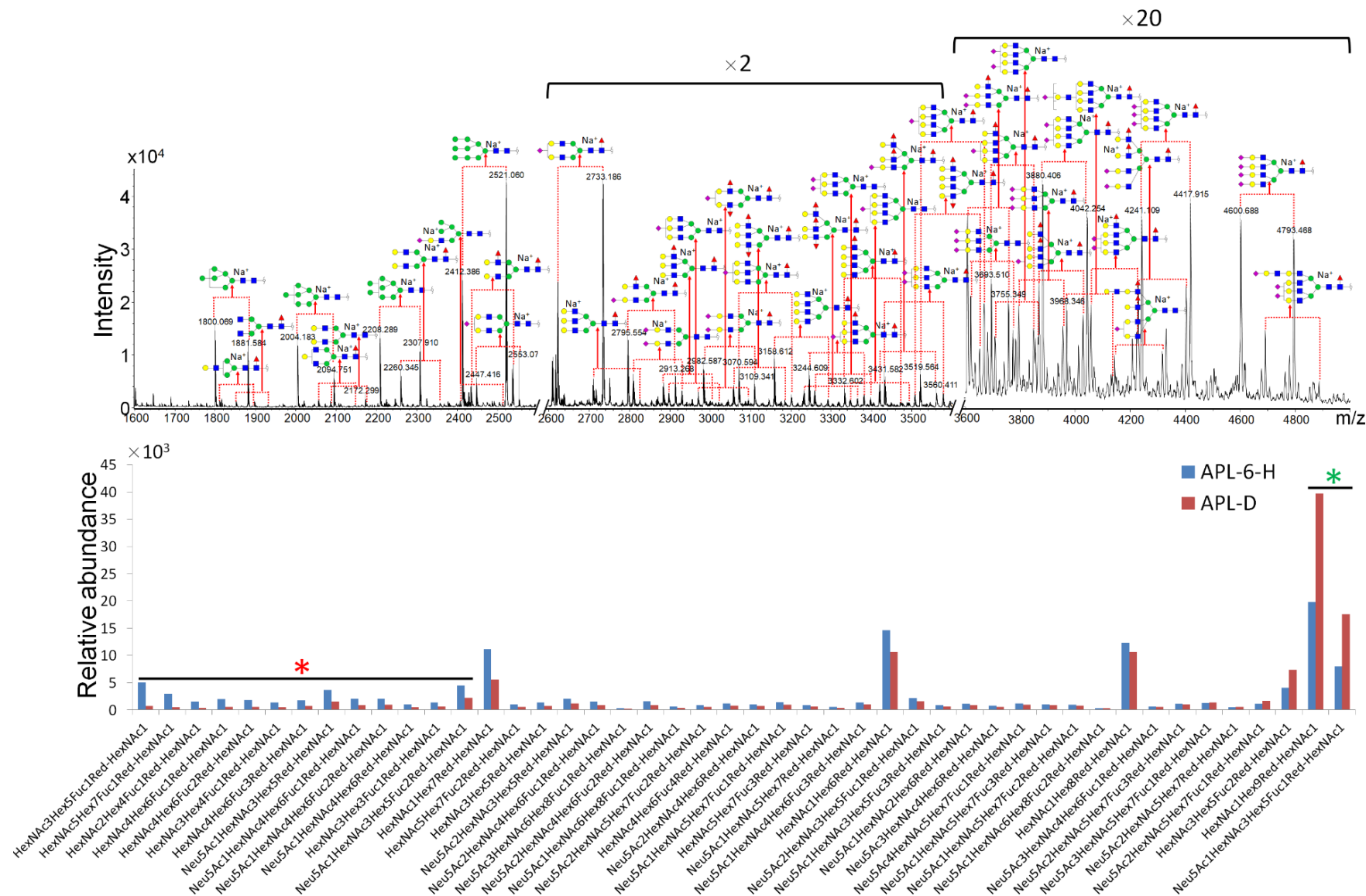


Figure 41. Isotopic labeling quantification for permethylated *N*-glycome from APL cells with CD₃I (APL-D) and APL-6 cells with CH₃I (APL-6-H) using MALDI-MS (The *N*-glycan labeled with asterisks meant more than 2-FC).

A total of 46 monosaccharide compositions, common to both APL-D and APL-6-H, were quantitatively compared (Figure 41). This comparison demonstrated that 13 monosaccharide compositions including HexNAc₃Hex₅Fuc₁Red-HexNAc₁ were at least 2-FC higher in APL-6-H than APL-D, while only HexNAc₁Hex₉Red-HexNAc₁ and Neu5Ac₁HexNAc₃Hex₅Fuc₁Red-HexNAc₁ were significantly reduced by more than 2-FC. APL-6-H was also compared with APL-13C, which showed similar data with the quantification between APL-6-H and APL-D (Figure 42). In addition, HexNAc₁Hex₅Fuc₁Red-HexNAc₁ was found to be elevated (2.47-FC) and HexNAc₁Hex₁₀Red-HexNAc₁ was reduced (4-FC) in APL-6-H compared to APL-13C. In total, 56 monosaccharide compositions in total were quantified with isotopic labeling by MALDI-MS (Figure 43). However, most *N*-glycans could not be quantitatively profiled using this detection method, due to the low signal strength and resolution in mass spectra of MAIDI-MS.

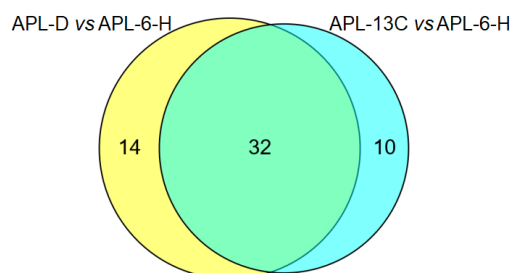


Figure 43. The comparison of different isotopic labeling quantification approaches, measured by MALDI-MS.

4.17 Relative quantification of APL and APL-6 cells derived *N*-

glycome by nanoLC-MS/MS

Relative quantification was performed based on the peak area of each *N*-glycan species of APL-6-H and APL-H, whose base peaks almost showed the same ingredients with slight difference in abundance (Figure 44a). By Pearson correlation analysis (Figure 44b), the R-squared was found to be above 0.96 in each group of technical replicates and below 0.86 between APL-6-H and APL-H. Compared to APL-H, the *N*-glycans of APL-6-H in up-regulation (Red dots) and down-regulation (Green dots), at more than 2-FC with p-value less than 0.05, were shown in the volcano plot (Figure 44c). To get more details for the quantification, above significantly different *N*-glycan species were exhibited in the heat map (Figure 44d), which provided more in-depth analysis than the quantification by MALDI-MS (See Figure 41, 42).

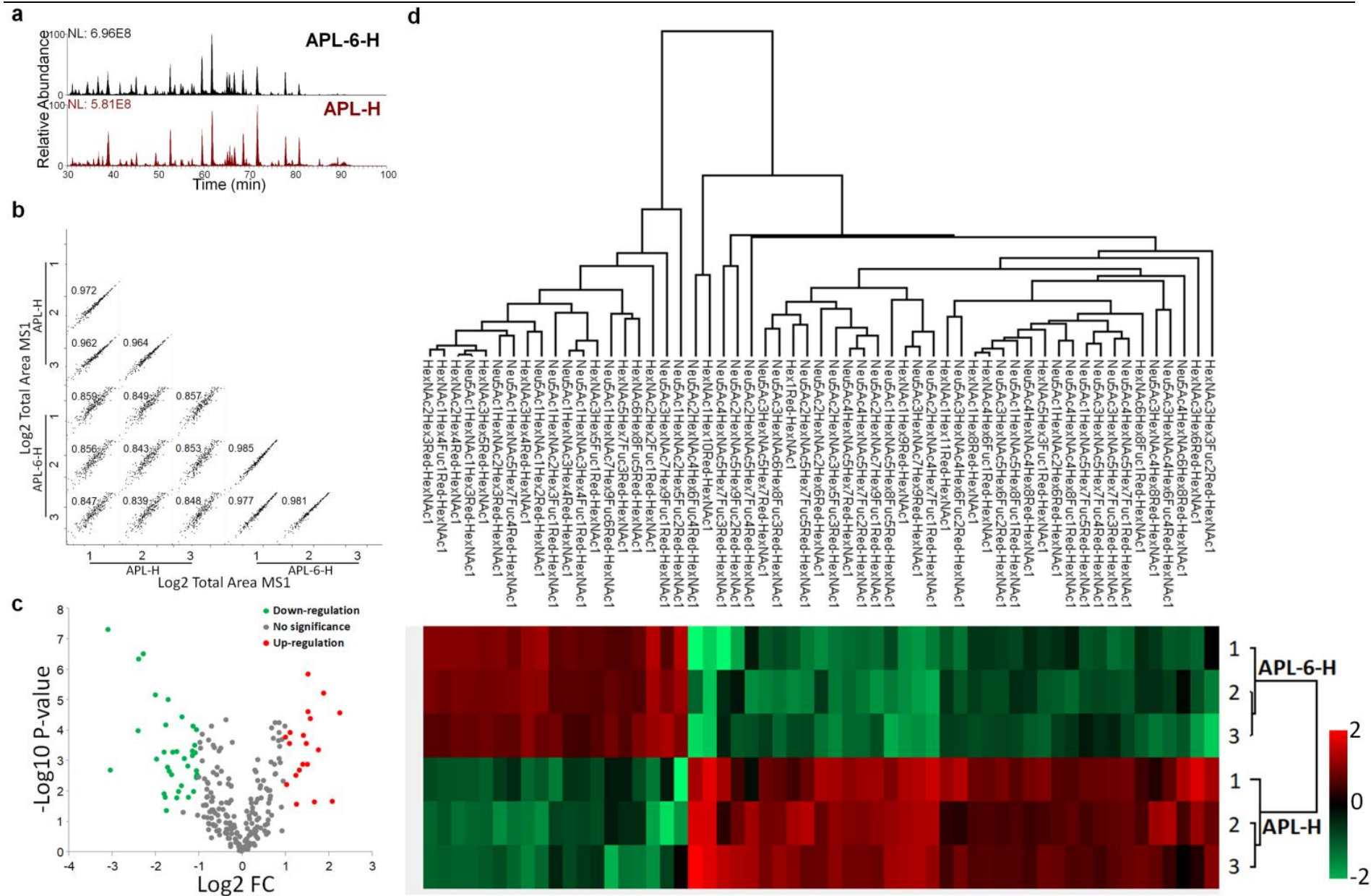


Figure 44. Relative quantification of *N*-glycome from APL-6-H and APL-H. (a) The comparison of base peaks between APL-6-H and APL-H. (b) The scatter plots of Log2 Total Area MS1 values of identified *N*-glycans and the Pearson correlation analysis with R-squared between APL-6-H and APL-H. (c) The volcano plot for relative quantification of *N*-glycome (APL-6-H vs APL-H). (d) Heat map profile for *N*-glycan species in significantly different abundance (APL-6-H vs APL-H).

5. Conclusions

A series of improved sample preparation approaches, including high (glyco)protein recovery, high *N*-glycan purification and efficient OSPP, were developed and enabled to acquire more accurate information of *N*-glycans. In MS-based deep *N*-glycomics, the database-free glycoinformatics solution was efficient to match the experimental precursors to the theoretical monosaccharide compositions using the newly designed R-scripts. In addition, an isotope-based FDR analysis was significant to control the quality of data obtained by R-scripts at MS1 level. Furthermore, the summarized fragmentation of permethylated reducing *N*-glycans promoted the application of a novel algorithm, named as bundled sequencing, and simplified the identification of *N*-glycan structures and isomers at MS2 level. In comparison to previous studies, this improved strategy can identify more *N*-glycans at low abundance from purified glycoproteins or even from glycoproteome. A large overlap of identified monosaccharide compositions was found between this improved strategy and glycopeptide analysis by pGlyco. Besides identification, relative quantification based on LC-MS chromatography data showed more comprehensive quantification than stable isotopic labeling approach using MALDI-MS. This analytical approach was also applicable to *O*-glycomics. Samples from humans and mice studied as model organisms can be

extended to other species having *e.g.* no Neu5Gc and Neu5Gc sugars using this strategy.

This improved strategy for in-depth and quantitative *N*-glycomics will be beneficial to identify glycan biomarkers and improve our understanding of the structure, biosynthesis and function of the glycoproteome.

6. References

1. Kuska B. 1998. Beer, Bethesda, and biology: how "genomics" came into being. *J Natl Cancer Inst* 90(2):93.
2. Fortina P, Al Khaja N, Al Ali MT, Hamzeh AR, Nair P, Innocenti F, Patrinos GP, Kricka LJ. 2014. Genomics into Healthcare: the 5th Pan Arab Human Genetics Conference and 2013 Golden Helix Symposium. *Hum Mutat* 35(5):637-640.
3. Omenn GS. 2017. Advances of the HUPO Human Proteome Project with broad applications for life sciences research. *Expert Rev Proteomics* 14(2):109-111.
4. Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57-63.
5. Miura Y, Endo T. 2016. Glycomics and glycoproteomics focused on aging and age-related diseases-Glycans as a potential biomarker for physiological alterations. *Biochim Biophys Acta* 1860(8):1608-1614.
6. Yang L, Li M, Shan Y, Shen S, Bai Y, Liu H. 2016. Recent advances in lipidomics for disease research. *J Sep Sci* 39(1):38-50.
7. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, Ge Y, Gunawardena J, Hendrickson RC, Hergenrother PJ, Huber CG, Ivanov AR, Jensen ON, Jewett MC, Kelleher NL, Kiessling LL, Krogan NJ, Larsen MR, Loo JA,

Ogorzalek Loo RR, Lundberg E, MacCoss MJ, Mallick P, Mootha VK, Mrksich M, Muir TW, Patrie SM, Pesavento JJ, Pitteri SJ, Rodriguez H, Saghatelian A, Sandoval W, Schlüter H, Sechi S, Slavoff SA, Smith LM, Snyder MP, Thomas PM, Uhlén M, Van Eyk JE, Vidal M, Walt DR, White FM, Williams ER, Wohlschlager T, Wysocki VH, Yates NA, Young NL, Zhang B. 2018. How many human proteoforms are there? *Nat Chem Biol* 14(3): 206-214.

8. Apweiler R, Hermjakob H, Sharon N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473(1):4-8.

9. Johansen PG, Marshall RD, Neuberger A. 1961. Carbohydrates in protein. 3 The preparation and some of the properties of a glycopeptide from hen's-egg albumin. *Biochem J* 78(3):518-527.

10. Rademaker GJ, Pergantis SA, Blok-Tip L, Langridge JI, Kleen A, Thomas-Oates JE. 1998. Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal Biochem* 257(2):149-160.

11. Qi W, Fong C, Lamport DT. 1991. Gum arabic glycoprotein is a twisted hairy rope: a new model based on o-galactosylhydroxyproline as the polysaccharide attachment site. *Plant Physiol* 96(3):848-855.

12. Colley KJ, Baenziger JU. 1987. Identification of the post-translational modifications of the core-specific lectin. The core-specific lectin contains

hydroxyproline, hydroxylysine, and glucosylgalactosylhydroxylysine residues. *J Biol Chem* 262(21):10290-10295.

13. Smythe C, Cohen P. 1991. The discovery of glycogenin and the priming mechanism for glycogen biogenesis. *Eur J Biochem* 200(3):625-631.

14. Wopereis S, Lefeber DJ, Morava E, Wevers RA. 2006. Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin Chem* 52(4):574-600.

15. Zanetta JP, Pons A, Richet C, Huet G, Timmerman P, Leroy Y, Bohin A, Bohin JP, Trinel PA, Poulain D, Hofsteenge J. 2004. Quantitative gas chromatography/mass spectrometry determination of C-mannosylation of tryptophan residues in glycoproteins. *Anal Biochem* 329(2):199-206.

16. Lote CJ, Weiss JB. 1971. Identification in urine of a low-molecular-weight highly polar glycopeptide containing cysteinyl-galactose. *Biochem J* 123(4):25P.

17. Haynes PA. 1998. Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology* 8(1):1-5.

18. Ferguson MA. 1999. The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. *J Cell Sci* 112 (Pt 17):2799-2809.

19. Muthana SM, Campbell CT, Gildersleeve JC. 2012. Modifications of glycans: biological significance and therapeutic opportunities. *ACS Chem Biol* 7(1):31-43.

20. Lowe JB, Marth JD. 2003. A genetic approach to mammalian glycan function. *Annu Rev Biochem* 72:643-691.
21. Potapenko IO, Haakensen VD, Lüders T, Helland A, Bukholm I, Sørlie T, Kristensen VN, Lingjaerde OC, Børresen-Dale AL. 2010. Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Mol Oncol* 4(2):98-118.
22. Varki A. 2017. Biological roles of glycans. *Glycobiology* 27(1):3-49.
23. Blomme B, Van Steenkiste C, Callewaert N, Van Vlierberghe H. 2009. Alteration of protein glycosylation in liver diseases. *J Hepatol* 50(3):592-603.
24. Kanninen K, Goldsteins G, Auriola S, Alafuzoff I, Koistinaho J. 2004. Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach. *Neurosci Lett* 367(2):235-240.
25. Kyselova Z, Mechref Y, Kang P, Goetz JA, Dobrolecki LE, Sledge GW, Schnaper L, Hickey RJ, Malkas LH, Novotny MV. 2008. Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem* 54(7):1166-1175.
26. Kladney RD, Cui X, Bulla GA, Brunt EM, Fimmel CJ. 2002. Expression of GP73, a resident Golgi membrane protein, in viral and nonviral liver disease. *Hepatology* 35(6):1431-1440.
27. Tang X, Guo N, Xu L, Gou X, Mi M. 2012. CD147/EMMPRIN: an effective therapeutic target for hepatocellular carcinoma. *J Drug Target*

- 21(3):224-231.
28. Siegel RL, Miller KD, Jemal A. 2016. Cancer statistics, 2016. *CA Cancer J Clin* 66(1):7-30.
29. Wu J, Xie X, Liu Y, He J, Benitez R, Buckanovich RJ, Lubman DM. 2012. Identification and confirmation of differentially expressed fucosylated glycoproteins in the serum of ovarian cancer patients using a lectin array and LC-MS/MS. *J Proteome Res* 11(9):4541-4552.
30. Nicastrì A, Gaspari M, Sacco R, Elia L, Gabriele C, Romano R, Rizzuto A, Cuda G. 2014. N-glycoprotein analysis discovers new up-regulated glycoproteins in colorectal cancer tissue. *J Proteome Res* 13(11):4932-4941.
31. Lebrilla CB, An HJ. 2009. The prospects of glycan biomarkers for the diagnosis of diseases. *Mol Biosyst* 5(1):17-20.
32. Rifai N, Gillette MA, Carr SA. 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 24(8):971-983.
33. Anderson NL, Anderson NG. 2002. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1(11):845-867.
34. Sethi MK, Fanayan S. 2015. Mass spectrometry-based N-glycomics of colorectal cancer. *Int J Mol Sci* 16(12): 29278-29304.
35. Stavenhagen K, Kolarich D, Wührer M. 2015. Clinical glycomics

employing graphitized carbon liquid chromatography-mass spectrometry. *Chromatographia* 78(5-6): 307-320.

36. Thaysen-Andersen M, Packer NH. 2014. Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim Biophys Acta* 1844(9):1437-1452.

37. Barrabés S, Sarrats A, Fort E, De Llorens R, Rudd PM, Peracaula R. 2010. Effect of sialic acid content on glycoprotein pI analyzed by two-dimensional electrophoresis. *Electrophoresis* 31(17):2903-2912.

38. Stumpe M, Miller C, Morton NS, Bell G, Watson DG. 2006. High-performance liquid chromatography determination of α 1-acid glycoprotein in small volumes of plasma from neonates. *J Chromatogr B Analyt Technol Biomed Life Sci* 831(1-2):81-84.

39. Puerta A, Díez-Masa JC, Martín-Álvarez PJ, Martín-Ventura JL, Barbas C, Tuñón J, Egido J, de Frutos M. 2011. Study of the capillary electrophoresis profile of intact α -1-acid glycoprotein isoforms as a biomarker of atherothrombosis. *Analyst* 136(4):816-822.

40. Chandramouli KH, Zhang Y, Wong YH, Qian PY. 2012. Comparative glycoproteome analysis: dynamics of protein glycosylation during metamorphic transition from Pelagic to Benthic life stages in three invertebrates. *J Proteome Res* 11(2):1330-1340.

41. Thaysen-Andersen M, Packer NH, Schulz BL. 2016. Maturing

glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Mol Cell Proteomics* 15(6):1773-1790.

42. Novotny MV, Mechref Y. 2005. New hyphenated methodologies in high-sensitivity glycoprotein analysis. *J Sep Sci* 28(15):1956-1968.

43. Corfield AP, Berry M. 2015. Glycan variation and evolution in the eukaryotes. *Trends Biochem Sci* 40(7):351-359.

44. Dove A. 2001. The bittersweet promise of glycobiology. *Nat Biotechnol* 19(10):913-917.

45. O'Neill RA. 1996. Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis. *J Chromatogr A* 720(1-2):201-215.

46. Patel T, Bruce J, Merry A, Bigge C, Wormald M, Jaques A, Parekh R. 1993. Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. *Biochemistry* 32(2):679-693.

47. Huang Y, Konse T, Mechref Y, Novotny MV. 2002. Matrix-assisted laser desorption/ionization mass spectrometry compatible beta-elimination of O-linked oligosaccharides. *Rapid Commun Mass Spectrom* 16(12):1199-1204.

48. Huang Y, Mechref Y, Novotny MV. 2001. Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass

- spectrometry and capillary electrophoresis. *Anal Chem* 73(24):6063-6069.
49. Michael C, Rizzi AM. 2015. Tandem mass spectrometry of isomeric aniline-labeled N-glycans separated on porous graphitic carbon: Revealing the attachment position of terminal sialic acids and structures of neutral glycans. *Rapid Commun Mass Spectrom* 29(13):1268-1278.
50. Zhao J, Li S, Li C, Wu SL, Xu W, Chen Y, Shameem M, Richardson D, Li H. 2016. Identification of low abundant isomeric N-glycan structures in biological therapeutics by LC/MS. *Anal Chem* 88(14):7049-7059.
51. Houel S, Hilliard M, Yu YQ, McLoughlin N, Martin SM, Rudd PM, Williams JP, Chen W. 2014. N- and O-glycosylation analysis of etanercept using liquid chromatography and quadrupole time-of-flight mass spectrometry equipped with electron-transfer dissociation functionality. *Anal Chem* 86(1):576-584.
52. Lareau NM, May JC, McLean JA. 2015. Non-derivatized glycan analysis by reverse phase liquid chromatography and ion mobility-mass spectrometry. *Analyst* 140(10):3335-3338.
53. Szabo Z, Guttman A, Rejtar T, Karger BL. 2010. Improved sample preparation method for glycan analysis of glycoproteins by CE-LIF and CE-MS. *Electrophoresis* 31(8):1389-1395.
54. Mørtz E, Sareneva T, Julkunen I, Roepstorff P. 1996. Does matrix-assisted laser desorption/ionization mass spectrometry allow analysis of carbohydrate heterogeneity in glycoproteins? A study of natural human

- interferon-gamma. *J Mass Spectrom* 31(10):1109-1918.
55. Everest-Dass AV, Abrahams JL, Kolarich D, Packer NH, Campbell MP. 2013. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom* 24(6):895-906.
56. Harvey DJ, Scarff CA, Crispin M, Scanlan CN, Bonomelli C, Scrivens JH. 2012. MALDI-MS/MS with traveling wave ion mobility for the structural analysis of N-linked glycans. *J Am Soc Mass Spectrom* 23(11):1955-1966.
57. Hua S, An HJ, Ozcan S, Ro GS, Soares S, DeVere-White R, Lebrilla CB. 2011. Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. *Analyst* 136(18):3663-3671.
58. Isailovic D, Kurulugama RT, Plasencia MD, Stokes ST, Kyselova Z, Goldman R, Mechref Y, Novotny MV, Clemmer DE. 2008. Profiling of human serum glycans associated with liver cancer and cirrhosis by IMS-MS. *J Proteome Res* 7(3):1109-1117.
59. Egorova KS, Toukach PV. 2018. Glycoinformatics: bridging isolated islands in the sea of data. *Angew Chem Int Ed* 57(46):14986-14990.
60. Damerell D, Ceroni A, Maass K, Ranzinger R, Dell A, Haslam SM. 2012. The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol Chem* 393(11):1357-1362.
61. Hayes CA, Karlsson NG, Struwe WB, Lisacek F, Rudd PM, Packer NH,

- Campbell MP. 2011. UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics* 27(9):1343-1344.
62. Agravat SB, Saltz JH, Cummings RD, Smith DF. 2014. GlycoPattern: a web platform for glycan array mining. *Bioinformatics* 30(23):3417-3418.
63. Barnett CB, Aoki-Kinoshita KF, Naidoo KJ. 2016. The glycome analytics platform: an integrative framework for glycobioinformatics. *Bioinformatics* 32(19):3005-3011.
64. Ranzinger R, Herget S, Wetter T, von der Lieth CW. 2008. GlycomeDB-integration of open-access carbohydrate structure databases. *BMC Bioinformatics* 9:384.
65. Kang P, Mechref Y, Kyselova Z, Goetz JA, Novotny MV. 2007. Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal Chem* 79(16):6064-6073.
66. Cao W, Zhang W, Huang J, Jiang B, Zhang L, Yang P. 2015. Glycan reducing end dual isotopic labeling (GREDIL) for mass spectrometry-based quantitative N-glycomics. *Chem Commun (Camb)* 51(71):13603-13606.
67. Aoki K, Perlman M, Lim JM, Cantu R, Wells L, Tiemeyer M. 2007. Dynamic developmental elaboration of N-linked glycan complexity in the *Drosophila melanogaster* embryo. *J Biol Chem* 282(12):9127-9142.
68. Prien JM, Prater BD, Qin Q, Cockrill SL. 2010. Mass spectrometric-based stable isotopic 2-aminobenzoic acid glycan mapping for rapid glycan

- screening of biotherapeutics. *Anal Chem* 82(4):1498-1508.
69. Xia B, Feasley CL, Sachdev GP, Smith DF, Cummings RD. 2009. Glycan reductive isotope labeling for quantitative glycomics. *Anal Biochem* 387(2):162-170.
70. Desaire H. 2013. Glycopeptide analysis, recent developments and applications. *Mol Cell Proteomics* 12(4): 893-901.
71. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. 2004. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14(2):103-114.
72. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K, Takahashi N, Isobe T. 2003. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* 21(6):667-672.
73. Stavenhagen K, Hinneburg H, Thaysen-Andersen M, Hartmann L, Varón Silva D, Fuchser J, Kaspar S, Rapp E, Seeberger PH, Kolarich D. 2013. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom* 48(6):627-639.
74. Sun B, Ranish JA, Utleg AG, White JT, Yan X, Lin B, Hood L. 2007. Shotgun glycopeptide capture approach coupled with mass spectrometry

- for comprehensive glycoproteomics. *Mol Cell Proteomics* 6(1):141-149.
75. Wang Y, Wu SL, Hancock WS. 2006. Monitoring of glycoprotein products in cell culture lysates using lectin affinity chromatography and capillary HPLC coupled to electrospray linear ion trap-fourier transform mass spectrometry (LTQ/FTMS). *Biotechnol Prog* 22(3):873-880.
76. Shao W, Liu J, Yang K, Liang Y, Weng Y, Li S, Liang Z, Zhang L, Zhang Y. 2016. Hydrogen-bond interaction assisted branched copolymer HILIC material for separation and N-glycopeptides enrichment. *Talanta* 158:361-367.
77. Alvarez-Manilla G, Atwood JA, Guo Y, Warren NL, Orlando R, Pierce M. 2006. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *J Proteome Res* 5(3):701-708.
78. Stavenhagen K, Plomp R, Wuhler M. 2015. Site-specific protein N- and O-glycosylation analysis by a C18-porous graphitized carbon liquid chromatography-electrospray ionization mass spectrometry approach using pronase treated glycopeptides. *Anal Chem* 87(23):11691-11699.
79. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 101(26):9528-9533.
80. Alley WR Jr, Mechref Y, Novotny MV. 2009. Characterization of glycopeptides by combining collision-induced dissociation and electron-

transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* 23(1):161-170.

81. Seipert RR, Dodds ED, Clowers BH, Beecroft SM, German JB, Lebrilla CB. 2008. Factors that influence fragmentation behavior of N-linked glycopeptide ions. *Anal Chem* 80(10):3684-3692.

82. Hart-Smith G, Raftery MJ. 2012. Detection and characterization of low abundance glycopeptides via higher-energy C-trap dissociation and orbitrap mass analysis. *J Am Soc Mass Spectrom* 23(1):124-140.

83. Liu MQ, Zeng WF, Fang P, Cao WQ, Liu C, Yan GQ, Zhang Y, Peng C, Wu JQ, Zhang XJ, Tu HJ, Chi H, Sun RX, Cao Y, Dong MQ, Jiang BY, Huang JM, Shen HL, Wong CCL, He SM, Yang PY. 2017. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat Commun* 8(1):438.

84. Hinneburg H, Stavenhagen K, Schweiger-Hufnagel U, Pengelley S, Jabs W, Seeberger PH, Silva DV, Wührer M, Kolarich D. 2016. The art of destruction: Optimizing collision energies in quadrupole-time of flight (Q-TOF) instruments for glycopeptide-based glycoproteomics. *J Am Soc Mass Spectrom* 27(3):507-519.

85. Taniguchi N, Paulson JC. 2007. Frontiers in glycomics; bioinformatics and biomarkers in disease. September 11-13, 2006 Natcher Conference Center, NIH Campus, Bethesda, MD, USA. *Proteomics* 7(9):1360-1363.

86. Taniguchi N. 2008. Human disease glycomics/proteome initiative (HGPI). *Mol Cell Proteomics* 7(3):626-627.
87. Yang F, Guan Y, Feng X, Rolfs A, Schlüter H, Luo J. 2019. Proteomics of the corpus callosum to identify novel factors involved in hypomyelinated Niemann-Pick Type C disease mice. *Mol Brain* 12(1):17.
88. Villén J, Gygi SP. 2008. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* 3(10):1630-1638.
89. Harada Y, Hirayama H, Suzuki T. 2015. Generation and degradation of free asparagine-linked glycans. *Cell Mol Life Sci* 72(13):2509-2533.
90. Wessel D, Flügge UI. 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 138(1):141-143.
91. Fic E, Kedracka-Krok S, Jankowska U, Pirog A, Dziedzicka-Wasylewska M. 2010. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis* 31(21):3573-3579.
92. Zhou S, Dong X, Veillon L, Huang Y, Mechref Y. 2017. LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Anal Bioanal Chem* 409(2):453-466.
93. Zhou H, Froehlich JW, Briscoe AC, Lee RS. 2013. The GlycoFilter: a simple and comprehensive sample preparation platform for proteomics, N-

- glycomics and glycosylation site assignment. *Mol Cell Proteomics* 12(10):2981-2991.
94. Morelle W, Michalski JC. 2007. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* 2(7):1585-1602.
95. Mechref Y, Kang P, Novotny MV. 2009. Solid-phase permethylation for glycomic analysis. *Methods Mol Biol* 534:53-64.
96. Kang P, Mechref Y, Klouckova I, Novotny MV. 2005. Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun Mass Spectrom* 19(23):3421-3428.
97. Song X, Ju H, Lasanajak Y, Kudelka MR, Smith DF, Cummings RD. 2016. Oxidative release of natural glycans for functional glycomics. *Nat Methods* 13(6):528-534.
98. Ciucanu I, Costello CE. 2003. Elimination of oxidative degradation during the per-O-methylation of carbohydrates. *J Am Chem Soc* 125(52):16213-16219.
99. Szabo Z, Thayer JR, Reusch D, Agroskin Y, Viner R, Rohrer J, Patil SP, Krawitzky M, Huhmer A, Avdalovic N, Khan SH, Liu Y, Pohl C. 2018. High performance anion exchange and hydrophilic interaction liquid chromatography approaches for comprehensive mass spectrometry-based characterization of the N-glycome of a recombinant human erythropoietin. *J Proteome Res* 17(4):1559-1574.
100. Jiang K, Zhu H, Xiao C, Liu D, Edmunds G, Wen L, Ma C, Li J, Wang

- PG. 2017. Solid-phase reductive amination for glycomic analysis. *Anal Chim Acta* 962:32-40.
101. Ruhaak LR, Steenvoorden E, Koeleman CA, Deelder AM, Wuhrer M. 2010. 2-picoline-borane: a non-toxic reducing agent for oligosaccharide labeling by reductive amination. *Proteomics* 10(12):2330-2336.
102. Gaunitz S, Nagy G, Pohl NL, Novotny MV. 2017. Recent advances in the analysis of complex glycoproteins. *Anal Chem* 89(1):389-413.
103. Varki, A. 2009. Multiple changes in sialic acid biology during human evolution. *Glycoconj J* 26(3):231-245.
104. Domon B, Costello CE. 1988. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj J* 5(4):397-409.
105. Huang LJ, Chiang CW, Chen SL, Wei SY, Chen SH. 2019. Complete mapping of disulfide linkages for etanercept products by multi-enzyme digestion coupled with LC-MS/MS using multi-fragmentations including CID and ETD. *J Food Drug Anal* 27 (2):531-541.
106. Morelle W, Donadio S, Ronin C, Michalski JC. 2006. Characterization of N-glycans of recombinant human thyrotropin using mass spectrometry. *Rapid Commun Mass Spectrom* 20(3): 331-345.
107. Ciucanu I. 2006. Per-O-methylation reaction for structural analysis of carbohydrates by mass spectrometry. *Anal Chim Acta* 576(2):147-155.
108. Murase T, Kajihara Y. 2010. Unique cleavage of 2-acetamido-2-

deoxy-D-glucose from the reducing end of biantennary complex type oligosaccharides. *Carbohydr Res* 345(12):1702-1207.

109. Toghiani KE, Compton RG. 2010. Electrochemical non-enzymatic glucose sensors: A perspective and an evaluation. *Int J Electrochem Sci* 5(9):1246-1301.

110. Largeaud F, Kokoh KB, Beden B, Lamy C. 1995. On the electrochemical reactivity of anomers: electrocatalytic oxidation of α - and β -D-glucose on platinum electrodes in acid and basic media. *J Electroanal Chem* 397(1-2): 261-269.

111. Harvey DJ. 2011. Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci.* 879(17-18):1196-1225.

112. Ashwood C, Lin CH, Thaysen-Andersen M, Packer NH. 2018. Discrimination of isomers of released N- and O-glycans using diagnostic product ions in negative ion PGC-LC-ESI-MS/MS. *J Am Soc Mass Spectrom* 29(6):1194-1209.

113. Fazel R, Guan Y, Vaziri B, Krisp C, Heikaus L, Saadati A, Hidayah SN, Gaikwad M, Schlüter H. 2019. Structural and in vitro functional comparability analysis of altebrel™, a proposed etanercept biosimilar: focus on primary sequence and glycosylation. *Pharmaceuticals (Basel)* 12(1):14.

114. Jeong YR, Kim SY, Park YS, Lee GM. 2018. Simple and robust N-

glycan analysis based on improved 2-aminobenzoic acid labeling for recombinant therapeutic glycoproteins. *J Pharm Sci* 107(7):1831-1841.

115. Cho IH, Lee N, Song D, Jung SY, Bou-Assaf G, Sosic Z, Zhang W, Lyubarskaya Y. 2016. Evaluation of the structural, physicochemical, and biological characteristics of SB4, a biosimilar of etanercept. *MAbs* 8(6):1136-1155.

116. Borza B, Szigeti M, Szekrenyes A, Hajba L, Guttman A. 2018. Glycosimilarity assessment of biotherapeutics 1: Quantitative comparison of the N-glycosylation of the innovator and a biosimilar version of etanercept. *J Pharm Biomed Anal* 153:182-185.

117. Szigeti M, Guttman A. 2017. Automated N-glycosylation sequencing of biopharmaceuticals by capillary electrophoresis. *Sci Rep* 7(1):11663.

118. D'Atri V, Nováková L, Fekete S, Stoll D, Lauber M, Beck A, Guilleme D. 2019. Orthogonal middle-up approaches for the characterization of the glycan heterogeneity of etanercept by hydrophilic interaction chromatography coupled to high resolution mass spectrometry. *Anal Chem* 91(1):873-880.

119. Wohlschlager T, Scheffler K, Forstenlehner IC, Skala W, Senn S, Damoc E, Holzmann J, Huber CG. 2018. Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat Commun* 9(1):1713.

120. Wang C, Qiang S, Jin W, Song X, Zhang Y, Huang L, Wang Z. 2018.

Reductive chemical release of N-glycans as 1-amino-alditols and subsequent 9-fluorenylmethyloxycarbonyl labeling for MS and LC/MS analysis. *J Proteomics* 187:47-58.

121. Wu Y, Wang C, Luo J, Liu Y, Zhang L, Xia Y, Feng X, Liu BF, Lin Y, Liu X. 2017. Microwave-assisted deglycosylation for rapid and sensitive analysis of N-glycans via glycosylamine derivatization. *Anal Bioanal Chem* 409(16):4027-4036.

122. Liu Y, Fu D, Yu L, Xiao Y, Peng X, Liang X. 2016. Oxidized dextran facilitated synthesis of a silica-based concanavalin a material for lectin affinity enrichment of glycoproteins/glycopeptides. *J Chromatogr A* 1455:147-155.

123. Abrahams JL, Campbell MP, Packer NH. 2018. Building a PGC-LC-MS N-glycan retention library and elution mapping resource. *Glycoconj J* 35(1):15-29.

124. Xiao K, Han Y, Tian Z. 2018. Large-scale identification and visualization of human liver N-glycome enriched from LO2 cells. *Anal Bioanal Chem* 410(17):4195-4202.

125. Kornfeld R, Kornfeld S. 1985. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 54:631-664.

126. Chung CY, Majewska NI, Wang Q, Paul JT, Betenbaugh MJ. 2017. SnapShot: N-glycosylation processing pathways across kingdoms. *Cell* 171(1):258-258.e1.

127. Nimtz M, Wray V, Rüdiger A, Conradt HS. 1995. Identification and structural characterization of a mannose-6-phosphate containing oligomannosidic N-glycan from human erythropoietin secreted by recombinant BHK-21 cells. *FEBS Lett* 365(2-3):203-208.
128. Klein A, Roussel P. 1998. O-acetylation of sialic acids. *Biochimie* 80(1), 49-57.
129. Wu Z, Li H, Zhang Q, Liu X, Zheng Q, Li J. 2017. Characterization of O-acetylation in sialoglycans by MALDI-MS using a combination of methylamidation and permethylation. *Sci Rep* 7:46206.
130. Ghaderi D, Taylor RE, Padler-Karavani V, Diaz S, Varki A. 2010. Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins. *Nat Biotechnol* 28(8):863-867.
131. Kooner AS, Yu H, Chen X. 2019. Synthesis of N-glycolylneuraminic acid (Neu5Gc) and its glycosides. *Front Immunol* 10:2004.
132. Chahrour O, Cobice D, Malone J. 2015. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J Pharm Biomed Anal* 113:2-20.
133. <https://www.sigmaaldrich.com/safety-center/hazard-and-precautionary-statements.html>

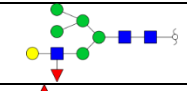
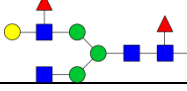
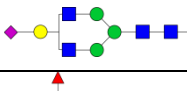
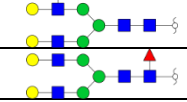
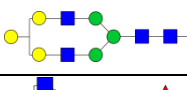
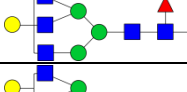
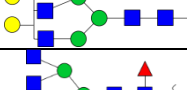
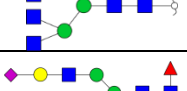
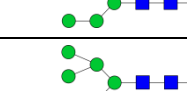

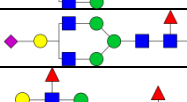
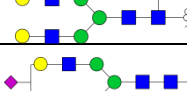

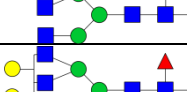

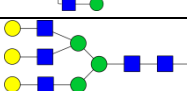
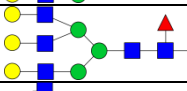
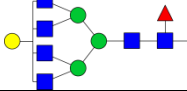
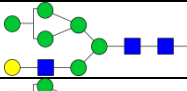
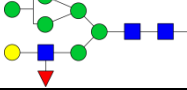

Supplementary table 1. The identified 245 monosaccharide compositions (398 *N*-glycans) from APL cells at MS1 and MS2 levels.

Monosaccharide composition	Code in Figure 29	Preferred <i>N</i> -glycan structures	Permethylated monoisotopic MW		Deviation (p.p.m.)
			Experiment	Theory	
HexNAc ₁ Hex ₁₂ Red-HexNAc ₁	1-1		3001.5258	3001.5231	0.92
HexNAc ₁ Hex ₁₁ Red-HexNAc ₁	1-2		2797.4239	2797.4233	0.22
HexNAc ₁ Hex ₁₀ Red-HexNAc ₁	1-3 (a, b)		2593.3234	2593.3235	0.04
HexNAc ₁ Hex ₉ Red-HexNAc ₁	1-4 (a, b, c)		2389.2234	2389.2238	0.14
HexNAc ₁ Hex ₈ Red-HexNAc ₁	1-5 (a, b, c)		2185.1232	2185.1240	0.35
HexNAc ₁ Hex ₇ Red-HexNAc ₁	2-6 (a, b, c)		1981.0248	1981.0242	0.30
HexNAc ₁ Hex ₆ Red-HexNAc ₁	2-7 (a, b)		1776.9252	1776.9245	0.43
HexNAc ₁ Hex ₅ Red-HexNAc ₁	2-8 (a, b)		1572.8261	1572.8247	0.92
HexNAc ₁ Hex ₄ Red-HexNAc ₁	2-9 (a, b, c)		1368.7254	1368.7249	0.37
HexNAc ₁ Hex ₃ Red-HexNAc ₁	2-10 (a, b, c)		1164.6258	1164.6251	0.58
HexNAc ₂ Hex ₅ Red-HexNAc ₁	2-11 (a, b)		1817.9503	1817.9510	0.37
HexNAc ₁ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-12 (a, b)		1542.8144	1542.8141	0.19
HexNAc ₁ Hex ₃ Fuc ₁ Red-HexNAc ₁	2-13 (a, b)		1338.7137	1338.7143	0.47
HexNAc ₂ Hex ₃ Red-HexNAc ₁	2-14 (a, b, c)		1409.7519	1409.7515	0.33
HexNAc ₂ Hex ₃ Fuc ₁ Red-HexNAc ₁	2-15 (a, b)		1583.8417	1583.8407	0.66
HexNAc ₃ Hex ₃ Red-HexNAc ₁	2-16 (a, b)		1654.8784	1654.8778	0.39
HexNAc ₂ Hex ₄ Red-HexNAc ₁	2-17 (a, b, c)		1613.8505	1613.8512	0.44
HexNAc ₁ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-18		1746.9151	1746.9139	0.70

Supplementary table 1

Neu5Ac ₁ HexNAc ₂ Hex ₃ Red-HexNAc ₁	2-19 (a, b)		1770.9254	1770.9251	0.17
HexNAc ₃ Hex ₃ Fuc ₁ Red-HexNAc ₁	2-20 (a, b, c)		1828.9683	1828.9670	0.73
HexNAc ₃ Hex ₄ Red-HexNAc ₁	2-21 (a, b, c)		1858.9781	1858.9776	0.31
HexNAc ₂ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-22 (a, b, c)		1787.9408	1787.9404	0.21
HexNAc ₃ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-23 (a, b, c)		2033.0658	2033.0668	0.46
Neu5Ac ₁ HexNAc ₂ Hex ₃ Fuc ₁ Red-HexNAc ₁	2-24 (a, b)		1945.0154	1945.0143	0.56
HexNAc ₁ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-25		1951.0129	1951.0137	0.38
Neu5Ac ₁ HexNAc ₂ Hex ₄ Red-HexNAc ₁	2-26 (a, b)		1975.0230	1975.0249	0.95
Neu5Ac ₁ HexNAc ₂ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-27 (a, b)		2149.1133	2149.1141	0.36
HexNAc ₄ Hex ₃ Red-HexNAc ₁	2-28 (a, b)		1900.0045	1900.0041	0.22
HexNAc ₂ Hex ₄ Fuc ₂ Red-HexNAc ₁	2-29		1962.0305	1962.0296	0.44
HexNAc ₂ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-30 (a, b)		1992.0412	1992.0402	0.51
HexNAc ₃ Hex ₃ Fuc ₂ Red-HexNAc ₁	2-31		2003.0567	2003.0562	0.26
HexNAc ₂ Hex ₆ Red-HexNAc ₁	2-32 (a, b)		2022.0501	2022.0508	0.32
HexNAc ₃ Hex ₅ Red-HexNAc ₁	2-33 (a, b, c)		2063.0782	2063.0773	0.44
HexNAc ₂ Hex ₅ Fuc ₂ Red-HexNAc ₁	2-34		2166.1289	2166.1294	0.23
HexNAc ₄ Hex ₃ Fuc ₁ Red-HexNAc ₁	2-35 (a, b, c)		2074.0924	2074.0933	0.43
HexNAc ₅ Hex ₃ Red-HexNAc ₁	2-36		2145.1313	2145.1304	0.42
HexNAc ₄ Hex ₄ Red-HexNAc ₁	2-37		2104.1042	2104.1039	0.17
Neu5Ac ₁ HexNAc ₂ Hex ₅ Red-HexNAc ₁	2-38		2179.1253	2179.1247	0.30
HexNAc ₂ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-39 (a, b)		2196.1390	2196.1400	0.44

Supplementary table 1

					
HexNAc ₃ Hex ₄ Fuc ₂ Red-HexNAc ₁	2-40		2207.1543	2207.1560	0.75
Neu5Ac ₁ HexNAc ₃ Hex ₄ Red-HexNAc ₁	2-41		2220.1513	2220.1512	0.05
HexNAc ₃ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-42 (a, b, c)		2237.1673	2237.1665	0.35
HexNAc ₃ Hex ₆ Red-HexNAc ₁	2-43		2267.1760	2267.1771	0.47
HexNAc ₄ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-44		2278.1928	2278.1931	0.11
HexNAc ₄ Hex ₅ Red-HexNAc ₁	2-45 (a, b)		2308.2033	2308.2036	0.14
HexNAc ₅ Hex ₃ Fuc ₁ Red-HexNAc ₁	2-46		2319.2201	2319.2196	0.21
Neu5Ac ₁ HexNAc ₂ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-47		2353.2145	2353.2139	0.28
Neu5Ac ₁ HexNAc ₂ Hex ₆ Red-HexNAc ₁	2-48		2383.2252	2383.2244	0.33
Neu5Ac ₁ HexNAc ₃ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-49 (a, b, c, d)		2394.2399	2394.2404	0.21
HexNAc ₃ Hex ₅ Fuc ₂ Red-HexNAc ₁	2-50		2411.2547	2411.2557	0.42
Neu5Ac ₁ HexNAc ₃ Hex ₅ Red-HexNAc ₁	2-51 (a, b)		2424.2503	2424.2510	0.27
HexNAc ₄ Hex ₄ Fuc ₂ Red-HexNAc ₁	2-52		2452.2815	2452.2823	0.31
HexNAc ₄ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-53 (a, b)		2482.2927	2482.2929	0.05
Neu5Ac ₁ HexNAc ₄ Hex ₄ Red-HexNAc ₁	2-54		2465.2797	2465.2775	0.89
HexNAc ₄ Hex ₆ Red-HexNAc ₁	2-55 (a, b)		2512.3041	2512.3034	0.28
HexNAc ₄ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-56 (a, b)		2686.3929	2686.3926	0.11
HexNAc ₅ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-57		2523.3184	2523.3194	0.39
HexNAc ₂ Hex ₇ Red-HexNAc ₁	2-58		2226.1495	2226.1505	0.46
HexNAc ₂ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-59		2400.2397	2400.2398	0.01

Supplementary table 1

HexNAc ₃ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-60		2441.2672	2441.2663	0.38
Neu5Ac ₁ HexNAc ₂ Hex ₅ Fuc ₂ Red-HexNAc ₁	2-61		2527.3033	2527.3031	0.09
HexNAc ₃ Hex ₅ Fuc ₃ Red-HexNAc ₁	2-62		2585.3429	2585.3450	0.79
Neu5Ac ₁ HexNAc ₂ Hex ₇ Red-HexNAc ₁	2-63		2587.3239	2587.3242	0.11
Neu5Ac ₁ HexNAc ₂ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-64		2557.3134	2557.3137	0.09
Neu5Ac ₂ HexNAc ₂ Hex ₅ Red-HexNAc ₁	2-65		2540.2987	2540.2983	0.16
Neu5Ac ₁ HexNAc ₃ Hex ₄ Fuc ₂ Red-HexNAc ₁	2-66 (a, b, c, d)		2568.3283	2568.3296	0.51
Neu5Ac ₁ HexNAc ₃ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-67 (a, b)		2598.3400	2598.3402	0.07
Neu5Ac ₁ HexNAc ₃ Hex ₆ Red-HexNAc ₁	2-68 (a, b)		2628.3504	2628.3508	0.13
Neu5Ac ₁ HexNAc ₄ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-69 (a, b, c)		2639.3655	2639.3667	0.46
Neu5Ac ₁ HexNAc ₄ Hex ₅ Red-HexNAc ₁	2-70 (a, b)		2669.3762	2669.3773	0.40
HexNAc ₅ Hex ₄ Fuc ₂ Red-HexNAc ₁	2-71		2697.4077	2697.4086	0.33
Neu5Ac ₂ HexNAc ₂ Hex ₆ Red-HexNAc ₁	2-72		2744.4001	2744.3981	0.74
Neu5Ac ₁ HexNAc ₂ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-73		2761.4141	2761.4134	0.26
Neu5Ac ₁ HexNAc ₃ Hex ₅ Fuc ₂ Red-HexNAc ₁	2-74		2772.4287	2772.4294	0.25
Neu5Ac ₂ HexNAc ₃ Hex ₅ Red-HexNAc ₁	2-75		2785.4254	2785.4247	0.28
Neu5Ac ₁ HexNAc ₃ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-76 (a, b)		2802.4395	2802.4400	0.16
HexNAc ₄ Hex ₅ Fuc ₂ Red-HexNAc ₁	2-77		2656.3840	2652.3821	0.74
HexNAc ₄ Hex ₅ Fuc ₃ Red-HexNAc ₁	2-78		2830.4716	2830.4713	0.12
Neu5Ac ₁ HexNAc ₄ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-79 (a, b, c)		2843.4654	2843.4665	0.38
HexNAc ₄ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-80 (a, b)		2860.4808	2860.4818	0.35
Neu5Ac ₁ HexNAc ₄ Hex ₆ Red-HexNAc ₁	2-81 (a, b, c)		2873.4763	2873.4771	0.28

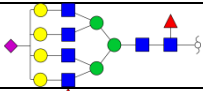
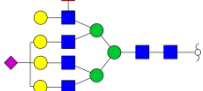
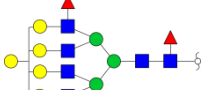
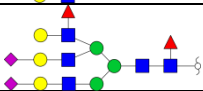
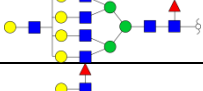
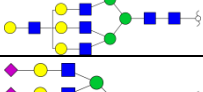
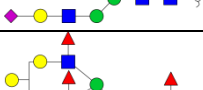
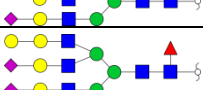
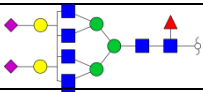
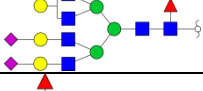
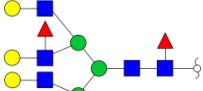
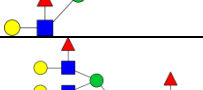
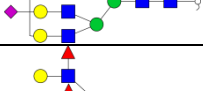
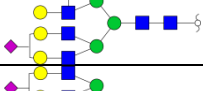
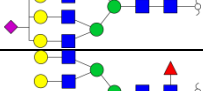
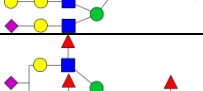
Supplementary table 1

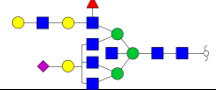
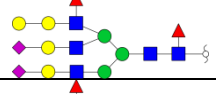
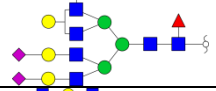
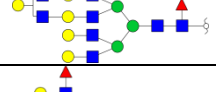
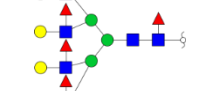

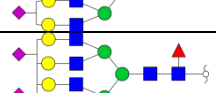
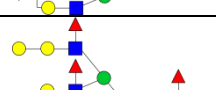
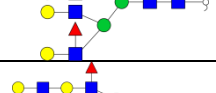
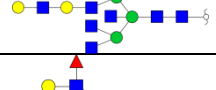
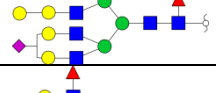
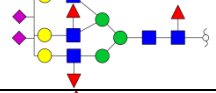
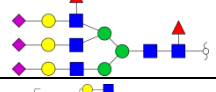
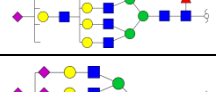
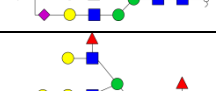
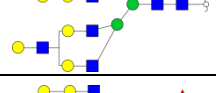
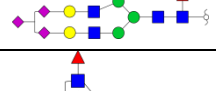
Neu5Ac ₁ HexNAc ₅ Hex ₄ Fuc ₁ Red-HexNAc ₁	2-82		2884.4904	2884.4931	0.92
Neu5Ac ₂ HexNAc ₃ Hex ₄ Fuc ₂ Red-HexNAc ₁	2-83 (a, b)		2929.5025	2929.5033	0.26
Neu5Ac ₁ HexNAc ₃ Hex ₅ Fuc ₃ Red-HexNAc ₁	2-84		2946.5174	2946.5186	0.40
Neu5Ac ₂ HexNAc ₃ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-85		2959.5134	2959.5139	0.15
HexNAc ₅ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-86		2727.4173	2727.4192	0.68
HexNAc ₅ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-87		2931.5190	2931.5190	0.03
HexNAc ₅ Hex ₇ Red-HexNAc ₁	2-88		2961.5291	2961.5295	0.13
Neu5Ac ₁ HexNAc ₃ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-89		2976.5307	2976.5292	0.52
Neu5Ac ₂ HexNAc ₃ Hex ₆ Red-HexNAc ₁	2-90		2989.5234	2989.5244	0.33
Neu5Ac ₁ HexNAc ₄ Hex ₅ Fuc ₂ Red-HexNAc ₁	2-91 (a, b)		3017.5525	3017.5557	1.06
HexNAc ₄ Hex ₆ Fuc ₃ Red-HexNAc ₁	2-92		3034.5704	3034.5710	0.20
Neu5Ac ₁ HexNAc ₄ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-93 (a, b, c, d)		3047.5652	3047.5663	0.35
HexNAc ₅ Hex ₅ Fuc ₃ Red-HexNAc ₁	2-94		3075.5960	3075.5976	0.51
Neu5Ac ₁ HexNAc ₅ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-95		3088.5935	3088.5928	0.22
Neu5Ac ₂ HexNAc ₃ Hex ₄ Fuc ₃ Red-HexNAc ₁	2-96		3103.5935	3103.5929	0.33
HexNAc ₅ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-97		3105.6069	3105.6082	0.40
Neu5Ac ₂ HexNAc ₃ Hex ₅ Fuc ₂ Red-HexNAc ₁	2-98		3133.6028	3133.6031	0.08
HexNAc ₅ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-99 (a, b, c)		3135.6178	3135.6187	0.28

Supplementary table 1

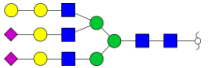
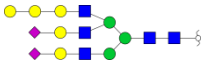
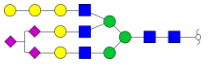
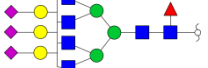
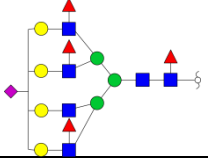
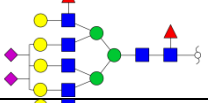

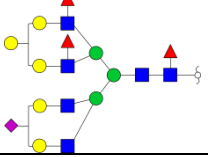
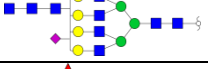
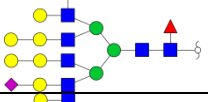
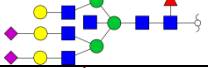
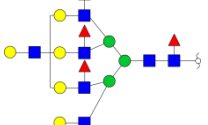
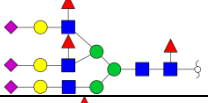
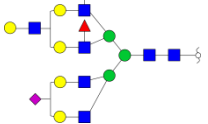

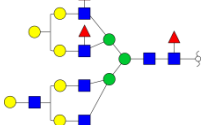
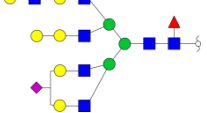
Neu5Ac ₂ HexNAc ₃ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-100		3163.6164	3163.6136	0.88
Neu5Ac ₂ HexNAc ₃ Hex ₇ Red-HexNAc ₁	2-101		3193.6247	3193.6242	0.17
Neu5Ac ₂ HexNAc ₄ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-102 (a, b, c)		3204.6388	3204.6402	0.42
HexNAc ₄ Hex ₆ Fuc ₄ Red-HexNAc ₁	2-103 (a, b)		3208.6590	3208.6603	0.38
Neu5Ac ₁ HexNAc ₄ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-104 (a, b)		3221.6543	3221.6555	0.36
Neu5Ac ₂ HexNAc ₄ Hex ₆ Red-HexNAc ₁	2-105 (a, b, c, d)		3234.6500	3234.6508	0.22
Neu5Ac ₁ HexNAc ₄ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-106 (a, b)		3251.6653	3251.6661	0.22
Neu5Ac ₂ HexNAc ₃ Hex ₅ Fuc ₃ Red-HexNAc ₁	2-107		3307.6915	3307.6923	0.23
HexNAc ₅ Hex ₇ Fuc ₂ Red-HexNAc ₁	2-108		3309.7043	3309.7079	1.09
Neu5Ac ₁ HexNAc ₅ Hex ₇ Red-HexNAc ₁	2-109 (a, b)		3322.7008	3322.7032	0.70
Neu5Ac ₁ HexNAc ₄ Hex ₆ Fuc ₃ Red-HexNAc ₁	2-110 (a, b)		3395.7435	3395.7447	0.35
Neu5Ac ₂ HexNAc ₄ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-111 (a, b, c)		3408.7389	3408.7400	0.30
HexNAc ₆ Hex ₈ Red-HexNAc ₁	2-112		3410.7535	3410.7556	0.61
HexNAc ₄ Hex ₇ Fuc ₄ Red-HexNAc ₁	2-113 (a, b)		3412.7601	3412.7600	0.03
HexNAc ₅ Hex ₆ Fuc ₄ Red-HexNAc ₁	2-114		3453.7838	3453.7866	0.80
Neu5Ac ₁ HexNAc ₅ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-115 (a, b)		3466.7847	3466.7818	0.84
HexNAc ₅ Hex ₇ Fuc ₃ Red-HexNAc ₁	2-116		3483.7961	3483.7971	0.29

Supplementary table 1

Neu5Ac ₁ HexNAc ₅ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-117 (a, b)		3496.7912	3496.7924	0.33
HexNAc ₅ Hex ₈ Fuc ₂ Red-HexNAc ₁	2-118 (a, b)		3513.8108	3513.8077	0.89
Neu5Ac ₂ HexNAc ₄ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-119 (a, b)		3582.8275	3582.8292	0.45
HexNAc ₆ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-120 (a, b)		3584.8413	3584.8448	0.97
Neu5Ac ₃ HexNAc ₄ Hex ₆ Red-HexNAc ₁	2-121 (a, b)		3595.8230	3595.8244	0.38
Neu5Ac ₁ HexNAc ₄ Hex ₇ Fuc ₃ Red-HexNAc ₁	2-122 (a, b, c)		3599.8435	3599.8445	0.26
Neu5Ac ₂ HexNAc ₄ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-123		3612.8434	3612.8397	1.03
Neu5Ac ₂ HexNAc ₅ Hex ₅ Fuc ₁ Red-HexNAc ₁	2-124 (a, b)		3449.7615	3449.7665	1.44
Neu5Ac ₂ HexNAc ₅ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-125		3653.8627	3653.8663	0.97
HexNAc ₅ Hex ₇ Fuc ₄ Red-HexNAc ₁	2-126		3657.8875	3657.8864	0.32
Neu5Ac ₁ HexNAc ₅ Hex ₇ Fuc ₂ Red-HexNAc ₁	2-127 (a, b)		3670.8799	3670.8816	0.45
Neu5Ac ₂ HexNAc ₅ Hex ₇ Red-HexNAc ₁	2-128		3683.8753	3683.8769	0.41
Neu5Ac ₁ HexNAc ₅ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-129		3700.8891	3700.8922	0.82
Neu5Ac ₂ HexNAc ₄ Hex ₆ Fuc ₃ Red-HexNAc ₁	2-130		3756.9164	3756.9184	0.52
Neu5Ac ₃ HexNAc ₄ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-131 (a, b, c, d)		3769.9129	3769.9136	0.18
Neu5Ac ₁ HexNAc ₆ Hex ₈ Red-HexNAc ₁	2-132		3771.9271	3771.9293	0.56

Neu5Ac ₁ HexNAc ₇ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-133 (a, b)		3782.9471	3782.9453	0.50
Neu5Ac ₂ HexNAc ₄ Hex ₇ Fuc ₂ Red-HexNAc ₁	2-134 (a, b, c)		3786.9329	3786.9289	1.06
Neu5Ac ₂ HexNAc ₅ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-135 (a, b)		3827.9532	3827.9555	0.59
HexNAc ₇ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-136		3829.9760	3829.9711	1.28
HexNAc ₅ Hex ₇ Fuc ₅ Red-HexNAc ₁	2-137		3831.9768	3831.9756	0.33
Neu5Ac ₁ HexNAc ₅ Hex ₇ Fuc ₃ Red-HexNAc ₁	2-138 (a, b)		3844.9701	3844.9708	0.17
Neu5Ac ₂ HexNAc ₅ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-139 (a, b)		3857.9642	3857.9661	0.47
HexNAc ₅ Hex ₈ Fuc ₄ Red-HexNAc ₁	2-140		3861.9867	3861.9861	0.16
HexNAc ₈ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-141 (a, b)		3870.9983	3870.9977	0.17
Neu5Ac ₁ HexNAc ₅ Hex ₈ Fuc ₂ Red-HexNAc ₁	2-142 (a, b)		3874.9833	3874.9814	0.51
Neu5Ac ₂ HexNAc ₄ Hex ₆ Fuc ₄ Red-HexNAc ₁	2-143		3931.0083	3931.0076	0.19
Neu5Ac ₃ HexNAc ₄ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-144		3944.0020	3944.0028	0.20
Neu5Ac ₁ HexNAc ₆ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-145		3946.0150	3946.0185	0.87
Neu5Ac ₄ HexNAc ₄ Hex ₆ Red-HexNAc ₁	2-146 (a, b)		3956.9962	3956.9981	0.46
HexNAc ₆ Hex ₉ Fuc ₂ Red-HexNAc ₁	2-147		3963.0314	3963.0338	0.60
Neu5Ac ₃ HexNAc ₄ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-148		3974.0153	3974.0134	0.49
Neu5Ac ₁ HexNAc ₅ Hex ₆ Fuc ₅ Red-HexNAc ₁	2-149		3989.0508	3989.0495	0.35

Supplementary table 1

Neu5Ac ₂ HexNAc ₄ Hex ₇ Red-HexNAc ₁	2-150		3438.7498	3438.7505	0.20
Neu5Ac ₂ HexNAc ₄ Hex ₈ Red-HexNAc ₁	2-151		3642.8496	3642.8503	0.18
Neu5Ac ₃ HexNAc ₄ Hex ₈ Red-HexNAc ₁	2-152		4004.0255	4004.0240	0.40
Neu5Ac ₃ HexNAc ₅ Hex ₆ Fuc ₁ Red-HexNAc ₁	2-153		4015.0349	4015.0400	1.25
Neu5Ac ₁ HexNAc ₅ Hex ₇ Fuc ₄ Red-HexNAc ₁	2-154		4019.0588	4019.0600	0.29
Neu5Ac ₂ HexNAc ₅ Hex ₇ Fuc ₂ Red-HexNAc ₁	2-155 (a, b)		4032.0527	4032.0553	0.62
Neu5Ac ₃ HexNAc ₅ Hex ₇ Red-HexNAc ₁	2-156		4045.0488	4045.0505	0.41
Neu5Ac ₁ HexNAc ₅ Hex ₈ Fuc ₃ Red-HexNAc ₁	2-157		4049.0736	4049.0706	0.76
Neu5Ac ₁ HexNAc ₈ Hex ₇ Red-HexNAc ₁	2-158		4058.0837	4058.0822	0.39
Neu5Ac ₁ HexNAc ₅ Hex ₉ Fuc ₂ Red-HexNAc ₁	2-159		4079.0800	4079.0811	0.27
Neu5Ac ₂ HexNAc ₆ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-160		4103.0888	4103.0924	0.86
HexNAc ₆ Hex ₈ Fuc ₄ Red-HexNAc ₁	2-161 (a, b)		4107.1128	4107.1124	0.09
Neu5Ac ₃ HexNAc ₄ Hex ₆ Fuc ₃ Red-HexNAc ₁	2-162		4118.0893	4118.0920	0.66
Neu5Ac ₁ HexNAc ₆ Hex ₈ Fuc ₂ Red-HexNAc ₁	2-163		4120.1055	4120.1077	0.52
Neu5Ac ₂ HexNAc ₆ Hex ₈ Red-HexNAc ₁	2-164		4133.1025	4133.1029	0.09
HexNAc ₆ Hex ₉ Fuc ₃ Red-HexNAc ₁	2-165		4137.1211	4137.1230	0.45
Neu5Ac ₁ HexNAc ₆ Hex ₉ Fuc ₁ Red-HexNAc ₁	2-166		4150.1209	4150.1183	0.65

Supplementary table 1

Neu5Ac ₄ HexNAc ₄ Hex ₇ Red-HexNAc ₁	2-167 (a, b)		4161.0947	4161.0979	0.74
Neu5Ac ₂ HexNAc ₅ Hex ₆ Fuc ₄ Red-HexNAc ₁	2-168		4176.1324	4176.1339	0.35
Neu5Ac ₃ HexNAc ₅ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-169		4189.1270	4189.1292	0.50
Neu5Ac ₁ HexNAc ₇ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-170		4191.1468	4191.1448	0.49
Neu5Ac ₁ HexNAc ₅ Hex ₇ Fuc ₅ Red-HexNAc ₁	2-171		4193.1482	4193.1492	0.24
Neu5Ac ₂ HexNAc ₅ Hex ₇ Fuc ₃ Red-HexNAc ₁	2-172		4206.1437	4206.1445	0.17
Neu5Ac ₃ HexNAc ₅ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-173 (a, b)		4219.1385	4219.1397	0.28
Neu5Ac ₁ HexNAc ₅ Hex ₈ Fuc ₄ Red-HexNAc ₁	2-174		4223.1579	4223.1598	0.44
Neu5Ac ₁ HexNAc ₈ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-175 (a, b)		4232.1696	4232.1714	0.40
Neu5Ac ₂ HexNAc ₅ Hex ₈ Fuc ₂ Red-HexNAc ₁	2-176		4236.1561	4236.1550	0.26
HexNAc ₆ Hex ₈ Fuc ₅ Red-HexNAc ₁	2-177		4281.2031	4281.2017	0.35
Neu5Ac ₃ HexNAc ₄ Hex ₆ Fuc ₄ Red-HexNAc ₁	2-178		4292.1798	4292.1813	0.33
Neu5Ac ₁ HexNAc ₆ Hex ₈ Fuc ₃ Red-HexNAc ₁	2-179		4294.1948	4294.1969	0.48
Neu5Ac ₂ HexNAc ₆ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-180 (a, b)		4307.1893	4307.1922	0.65
Neu5Ac ₁ HexNAc ₆ Hex ₉ Fuc ₂ Red-HexNAc ₁	2-181		4324.2051	4324.2075	0.54

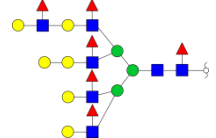
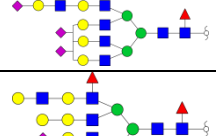
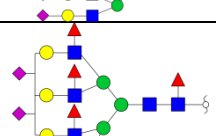
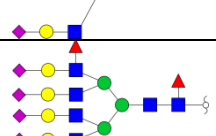
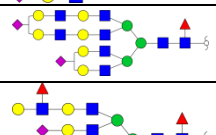
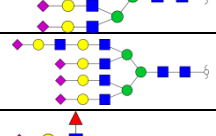
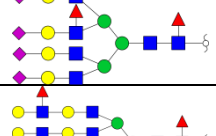
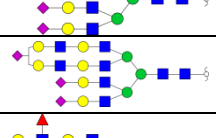
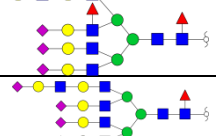
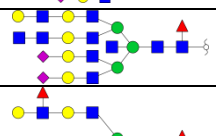
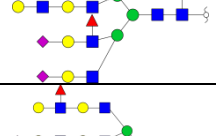
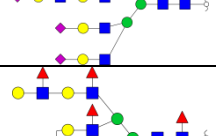
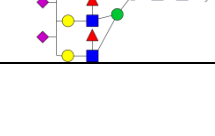


Supplementary table 1

Neu5Ac ₃ HexNAc ₅ Hex ₆ Fuc ₃ Red-HexNAc ₁	2-182		4363.2186	4363.2184	0.06
Neu5Ac ₄ HexNAc ₄ Hex ₈ Red-HexNAc ₁	2-183 (a, b, c)		4365.1992	4365.1976	0.37
Neu5Ac ₁ HexNAc ₇ Hex ₈ Fuc ₂ Red-HexNAc ₁	2-184		4365.2353	4365.2340	0.30
Neu5Ac ₂ HexNAc ₅ Hex ₇ Fuc ₄ Red-HexNAc ₁	2-185		4380.2357	4380.2337	0.47
Neu5Ac ₃ HexNAc ₅ Hex ₇ Fuc ₂ Red-HexNAc ₁	2-186 (a, b)		4393.2283	4393.2289	0.13
Neu5Ac ₁ HexNAc ₅ Hex ₈ Fuc ₃ Red-HexNAc ₁	2-187		4397.2455	4397.2490	0.79
Neu5Ac ₄ HexNAc ₅ Hex ₇ Red-HexNAc ₁	2-188		4406.2225	4406.2242	0.37
Neu5Ac ₁ HexNAc ₈ Hex ₇ Fuc ₂ Red-HexNAc ₁	2-189		4406.2601	4406.2606	0.10
Neu5Ac ₂ HexNAc ₅ Hex ₈ Fuc ₃ Red-HexNAc ₁	2-190		4410.2424	4410.2443	0.41
Neu5Ac ₂ HexNAc ₈ Hex ₇ Red-HexNAc ₁	2-191		4419.2580	4419.2558	0.51
Neu5Ac ₂ HexNAc ₅ Hex ₉ Fuc ₂ Red-HexNAc ₁	2-192		4440.2575	4440.2548	0.62
HexNAc ₆ Hex ₈ Fuc ₆ Red-HexNAc ₁	2-193 (a, b)		4455.2933	4455.2909	0.55
Neu5Ac ₃ HexNAc ₆ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-194 (a, b)		4464.2663	4464.2660	0.07
Neu5Ac ₁ HexNAc ₈ Hex ₉ Red-HexNAc ₁	2-195		4466.2802	4466.2817	0.32
Neu5Ac ₁ HexNAc ₆ Hex ₈ Fuc ₄ Red-HexNAc ₁	2-196 (a, b)		4468.2888	4468.2861	0.61
Neu5Ac ₂ HexNAc ₆ Hex ₈ Fuc ₂ Red-HexNAc ₁	2-197		4481.2803	4481.2814	0.23

Supplementary table 1

HexNAc ₆ Hex ₉ Fuc ₅ Red-HexNAc ₁	2-198		4485.2974	4485.3014	0.89
Neu5Ac ₃ HexNAc ₆ Hex ₈ Red-HexNAc ₁	2-199		4494.2725	4494.2766	0.90
Neu5Ac ₁ HexNAc ₆ Hex ₉ Fuc ₃ Red-HexNAc ₁	2-200		4498.2942	4498.2967	0.54
Neu5Ac ₁ HexNAc ₆ Hex ₁₀ Fuc ₂ Red-HexNAc ₁	2-201		4528.3077	4528.3072	0.11
Neu5Ac ₄ HexNAc ₄ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-202		4539.2886	4539.2868	0.40
Neu5Ac ₁ HexNAc ₇ Hex ₈ Fuc ₃ Red-HexNAc ₁	2-203		4539.3182	4539.3232	1.10
Neu5Ac ₄ HexNAc ₅ Hex ₆ Fuc ₂ Red-HexNAc ₁	2-204		4550.3013	4550.3028	0.32
Neu5Ac ₂ HexNAc ₇ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-205		4552.3181	4552.3185	0.07
Neu5Ac ₂ HexNAc ₅ Hex ₇ Fuc ₅ Red-HexNAc ₁	2-206		4554.3269	4554.3229	0.89
Neu5Ac ₃ HexNAc ₅ Hex ₇ Fuc ₃ Red-HexNAc ₁	2-207		4567.3166	4567.3181	0.33
Neu5Ac ₄ HexNAc ₅ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-208 (a, b)		4580.3118	4580.3134	0.33
Neu5Ac ₂ HexNAc ₈ Hex ₇ Fuc ₁ Red-HexNAc ₁	2-209 (a, b, c, d)		4593.3419	4593.3450	0.67
Neu5Ac ₁ HexNAc ₈ Hex ₉ Fuc ₁ Red-HexNAc ₁	2-210		4640.3736	4640.3709	0.59
Neu5Ac ₁ HexNAc ₆ Hex ₈ Fuc ₅ Red-HexNAc ₁	2-211 (a, b)		4642.3739	4642.3753	0.30
Neu5Ac ₂ HexNAc ₆ Hex ₈ Fuc ₃ Red-HexNAc ₁	2-212 (a, b, c)		4655.3691	4655.3706	0.31

Supplementary table 1






HexNAc ₆ Hex ₉ Fuc ₆ Red-HexNAc ₁	2-213 (a, b)		4659.3893	4659.3906	0.28
Neu5Ac ₃ HexNAc ₆ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-214 (a, b)		4668.3635	4668.3658	0.48
Neu5Ac ₂ HexNAc ₆ Hex ₉ Fuc ₂ Red-HexNAc ₁	2-215		4685.3767	4685.3811	0.93
Neu5Ac ₃ HexNAc ₅ Hex ₇ Fuc ₄ Red-HexNAc ₁	2-216 (a, b, c)		4741.4086	4741.4074	0.27
Neu5Ac ₄ HexNAc ₅ Hex ₇ Fuc ₂ Red-HexNAc ₁	2-217 (a, b)		4754.4038	4754.4026	0.27
Neu5Ac ₂ HexNAc ₇ Hex ₉ Fuc ₁ Red-HexNAc ₁	2-218 (a, b)		4756.4177	4756.4183	0.10
Neu5Ac ₃ HexNAc ₆ Hex ₈ Fuc ₂ Red-HexNAc ₁	2-219 (a, b, c)		4842.4538	4842.4550	0.24
Neu5Ac ₄ HexNAc ₆ Hex ₈ Red-HexNAc ₁	2-220		4855.4474	4855.4503	0.58
Neu5Ac ₄ HexNAc ₅ Hex ₇ Fuc ₃ Red-HexNAc ₁	2-221		4928.4899	4928.4918	0.38
Neu5Ac ₂ HexNAc ₇ Hex ₉ Fuc ₂ Red-HexNAc ₁	2-222		4930.5069	4930.5075	0.10
Neu5Ac ₃ HexNAc ₇ Hex ₉ Red-HexNAc ₁	2-223		4943.4999	4943.5027	0.55
Neu5Ac ₃ HexNAc ₆ Hex ₈ Fuc ₃ Red-HexNAc ₁	2-224 (a, b)		5016.5427	5016.5442	0.30
Neu5Ac ₄ HexNAc ₆ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-225 (a, b, c)		5029.5391	5029.5395	0.06
Neu5Ac ₂ HexNAc ₉ Hex ₈ Fuc ₁ Red-HexNAc ₁	2-226 (a, b)		5042.5693	5042.5711	0.35
Neu5Ac ₂ HexNAc ₇ Hex ₉ Fuc ₃ Red-HexNAc ₁	2-227		5104.5955	5104.5967	0.22
Neu5Ac ₃ HexNAc ₇ Hex ₉ Fuc ₁ Red-HexNAc ₁	2-228 (a, b)		5117.5904	5117.5919	0.28
Neu5Ac ₂ HexNAc ₆ Hex ₈ Fuc ₆ Red-HexNAc ₁	2-229 (a, b)		5177.6322	5177.6382	1.15

Supplementary table 1












Neu5Ac ₃ HexNAc ₆ Hex ₈ Fuc ₄ Red-HexNAc ₁	2-230		5190.6289	5190.6334	0.87
Neu5Ac ₁ HexNAc ₇ Hex ₉ Fuc ₆ Red-HexNAc ₁	2-231		5265.6879	5265.6906	0.51
Neu5Ac ₂ HexNAc ₉ Hex ₁₀ Red-HexNAc ₁	2-232		5276.6778	5276.6815	0.68
Neu5Ac ₃ HexNAc ₇ Hex ₉ Fuc ₂ Red-HexNAc ₁	2-233 (a, b)		5291.6788	5291.6811	0.43
Neu5Ac ₁ HexNAc ₈ Hex ₁₀ Fuc ₅ Red-HexNAc ₁	2-234 (a, b)		5540.8244	5540.8275	0.55
Neu5Ac ₂ HexNAc ₄ Hex ₅ Red-HexNAc ₁	2-235 (a, b)		3030.5481	3030.5510	0.94
Fuc ₁ Red-HexNAc ₁	2-236		481.2890	481.2887	0.62
Hex ₁ Red-HexNAc ₁	2-237 (a, b)		511.2994	511.2993	0.27
HexNAc ₁ Hex ₁ Red-HexNAc ₁	2-238		756.4263	756.4256	0.97
Neu5Ac ₁ Hex ₁ Red-HexNAc ₁	2-239 (a, b)		872.4721	872.4729	0.94
HexNAc ₁ Hex ₁ Fuc ₁ Red-HexNAc ₁	2-240		930.5150	930.5148	0.22
HexNAc ₁ Hex ₂ Red-HexNAc ₁	2-241		960.5247	960.5254	0.66
HexNAc ₁ Hex ₂ Fuc ₁ Red-HexNAc ₁	2-242		1134.6139	1134.6146	0.58
Neu5Ac ₁ HexNAc ₁ Hex ₂ Red-HexNAc ₁	2-243 (a, b)		1321.6983	1321.6990	0.54
HexNAc ₂ Hex ₂ Fuc ₁ Red-HexNAc ₁	2-244 (a, b)		1379.7400	1379.7409	0.64
Neu5Ac ₁ HexNAc ₁ Hex ₃ Red-HexNAc ₁	2-245 (a, b, c)		1525.7978	1525.7988	0.64

7. Risks and safety statements

According to Globally Harmonized System of Classification and Labeling of Chemicals (GHS), a list of potentially hazardous chemicals with the respective hazard and precautionary statements is given as follows (133):

Compound	GHS symbol	GHS hazard	Hazard statements	Precautionary statements
Methanol (LiChrosolv®)		GHS02 GHS06 GHS08	H225-H301+H311 +H331-H370	P210-P240-P280- P302+P352 -P304+P340- P308+P310- P403+P233
Ammonium bicarbonate		GHS07	H302	P301+P312+P330
Dithiothreitol		GHS07	H302-H315-H319- H335	P261- P305+P351+P338
Iodoacetamide		GHS06 GHS08	H301-H317-H334	P261-P280- P301+P310- P342+P311
Formic acid		GHS02 GHS06 GHS05	H226-H302-H314- H331-EUH071	P210-P280- P301+P330+P331- P304+P340- P305+P351+P338- P308+P310

Risks and safety statements

Trifluoroacetic acid	 	GHS05 GHS07	H314-H332-H412	P273-P280- P305+P351+P338- P310
Acetonitrile (LiChrosolv®)	 	GHS02 GHS07	H225- H302+H312+H332- -H319	P210-P240- P302+P352- P305+P351+P338- P403+P233
Trypsin	 	GHS07 GHS08	H315-H319-H334- H317-H335	P264-P272-P280- P302+P352- P305+P351+P338- P312
Iodomethane	 	GHS06 GHS08	H301-H312-H315- H331-H335-H351	P261-P280- P301+P310-P311
Sodium deoxycholate		GHS07	H302	P301+P312+P330
Tetraethylamm onium borohydride	 	GHS02 GHS07	H261-H315-H319- H335	P231+P232-P261- P305+P351+P338- P422

GHS hazard statements

H225	Highly flammable liquid and vapor
H226	Flammable liquid and vapor
H261	In contact with water releases flammable gas
H301	Toxic if swallowed
H302	Harmful if swallowed
H304	May be fatal if swallowed and enters airways
H311	May be fatal if swallowed and enters airways
H312	Harmful in contact with skin
H314	May be fatal if swallowed and enters airways
H315	Causes skin irritation
H317	May be fatal if swallowed and enters airways
H319	Causes serious eye irritation
H331	May be fatal if swallowed and enters airways
H332	Harmful if inhaled
H334	May be fatal if swallowed and enters airways
H335	May cause respiratory irritation
H336	May cause drowsiness or dizziness
H351	Suspected of causing cancer
H370	Causes damage to organs
H373	Causes damage to organs through prolonged or repeated exposure
H412	Harmful to aquatic life with long lasting effects
EUH071	Corrosive to the respiratory tract
H360FD	May damage fertility. May damage the unborn child

GHS precautionary statements

P201	Obtain special instructions before use
P210	Keep away from heat/sparks/open flames/hot surfaces. -No smoking
P240	Ground/bond container and receiving equipment
P260	Do not breathe dust/fume/gas/mist/vapours/spray
P261	Avoid breathing dust/fume/gas/mist/vapours/spray
P264	Wash hands thoroughly after handling
P272	Contaminated work clothing should not be allowed out of the workplace
P273	Avoid release to the environment
P280	Wear protective gloves/protective clothing/eye protection/face protection
P310	Immediately call a POISON CENTER or doctor/physician
P311	Call a POISON CENTER or doctor/physician
P312	Call a POISON CENTER or doctor/physician if you feel unwell
P330	Rinse mouth
P405	Store locked up P301 + P310
P422	Store contents under ...
P231+P232	Handle under inert gas. Protect from moisture.
P302+P352	IF ON SKIN: wash with plenty of soap and water
P304+P340	IF INHALED: Remove victim to fresh air and Keep at rest in a position comfortable for breathing
P301+P310	IF SWALLOWED: Immediately call a POISON CENTER or doctor/physician
P301+P312	IF SWALLOWED: call a POISON CENTER or doctor/physician IF you feel unwell
P304+P340	IF INHALED: Remove victim to fresh air and Keep at rest in a position comfortable for breathing
P308+P310	IF exposed or concerned: Immediately call a POISON CENTER or doctor/physician
P308+P313	IF exposed or concerned: Get medical advice/attention
P332+P313	IF SKIN irritation occurs: Get medical advice/attention

P337+P313	IF eye irritation persists: Get medical advice/attention
P342+P311	IF experiencing respiratory symptoms: call a POISON CENTER or doctor/physician
P362+P364	Take off contaminated clothing and wash it before reuse
P370+P378	In case of fire: Use ... for extinction
P403+P233	Store in a well-ventilated place. Keep container tightly closed
P301+P312+P330	IF SWALLOWED: call a POISON CENTER or doctor/physician IF you feel unwell. Rinse mouth.
P303+P361+P353	IF ON SKIN (or hair): Remove/Take off Immediately all contaminated clothing. Rinse SKIN with water/shower
P305+P351+P338	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses, if present and easy to do. Continue rinsing.
P301+P330+P331	IF SWALLOWED: Rinse mouth. Do NOT induce vomiting

8. Acknowledgements

First of all, I would like to sincerely thank my supervisor Prof. Hartmut Schlüter for supervising my work in his group at Universitätsklinikum Hamburg-Eppendorf (UKE) and teaching me how to be a scientist.

I also would like to thank Prof. Christian Betzel, who served as my co-supervisor and provided suggestions for my project.

Financial support by China Scholarship Council is gratefully acknowledged (CSC No. 201606220045).

Many thanks to Prof. Pengyuan Yang (Fudan University), Dr. Morten Thaysen-Andersen and Prof. Nicki Packer (Macquarie University), who improved the research a lot.

Also, I am grateful to Jiaxiang Hu for his contribution in R-script writing, Wencong Cui, Fan Yang and Weiqian Cao for their help in sample preparation. I would like to thank all of our group members, Fritz, Sönke, Christoph, Maryam, Ramin, Pascal, Marcus, Manka, Parnian, Laura, Dennis, Benjamin, Siti, Manasi, Min and Hannah, for the nice atmosphere in the lab.

Finally, I would like to thank my beloved family, my lovely mom, my great dad and my sisters, Sanju, Heli and Yuxia, for their support and endless love in my life.

9. Declarations

I hereby declare on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids. The submitted written version corresponds to the version on the electronic storage medium. I hereby declare that I have not previously applied or pursued for a doctorate (Ph.D. studies).

City and date:

Hamburg, 22.03.2020

Signature:

Yudong Guan