國立臺灣大學生命科學院生化科學研究所

博士論文

Graduate Institute of Biochemical Sciences

College of Life Science

National Taiwan University

Doctoral Dissertation

# 內皮細胞及 B 細胞多乳糖胺聚醣與雙唾液酸醣質體的有效鑑定

# Glycomic mapping of polylactosaminoglycans, terminal disialyl and sialyl sulfo N-acetyllactosamine motifs on mammalian cells

王穗華

Shui-hua Wang

指導教授：邱繼輝 博士

Advisor: Kay-Hooi Khoo, Ph.D.

中華民國 100 年 7 月

July, 2011

# 中文摘要

以質譜為基礎的醣質體及醣蛋白質體大多著重在鑑定醣鏈末端乙酰基乳糖胺上的唾液酸及岩藻醣化，對於 N 型醣鏈上是否為直鏈或是帶有支鏈的多乳糖胺聚醣則較少著墨，且大多只是憑藉著質譜所得的數據來推算其存在和可能的組成為何，因此，本篇論文主要是利用人類內皮細胞、老鼠及人類的 B 淋巴球為樣品，鑑定並挑戰分析多乳糖胺聚醣的結構特徵及其末端的唾液酸化和硫酸化修飾。

首先，以人類內皮細胞：EA.hy926 和 HUVEC 為起始原料，利用介質輔助雷射脫附法和電噴灑離子化方法在 MS 及 MS/MS 的階段來仔細地探討多乳糖胺聚醣結構，同時並搭配 endo-β-galactosidase 酵素及 Smith 降解反應來鑑定其是否為分支醣鏈及長鏈延伸的起始位置。和 HUVEC 相比，EA.hy926 細胞的 N 型醣鏈帶有較少的唾液酸化和岩藻醣化，但多乳糖胺聚醣的長度較長且具有分支醣鏈，其延伸點並不侷限在目前已被證實具有相當重要生物意義的甘露醣 6 號碳上。拓展至醣質體方面的研究上，則利用 *Lycopersicon esculentum* 凝集素分別在醣、醣蛋白及醣胜肽的階段做純化，較為專一的兩步驟純化方法讓我們得知了至少有 40 個以上的醣蛋白候補者可能帶有乳糖胺聚醣。

小鼠 B 細胞株—BCL1 上 N 型醣鏈的修飾主要為核心岩藻醣化，末端 α 鏈結半乳糖及唾液酸(Neu5Gc)化，且為非分支的多乳糖胺聚醣；相反的是，其 O 型醣鏈主要為簡單的 core 1 結構，帶有單唾液酸或雙唾液酸修飾。利用唾液酸酶、endo-β-galactosidase、MS/MS 及化學分析方法可知，雙唾液酸主要靠 α2-8 鏈結，並同時存在於具有多乳糖胺聚醣延伸或非延伸的 N 型醣鏈上。以螢光標記唾液酸並搭配高效液相層析儀顯示雙唾液酸結構的含量在 N 型醣鏈和 O 型醣鏈是相當的，且 CD45 為其攜帶者之一。透過小鼠 α2,8-sialyltransferase VI(ST8sia VI)基因抑制實驗可知，此酵素同時參與 N 型醣鏈和 O 型醣鏈上雙唾液酸的生合成，且 ST8Sia VI 和雙唾液酸結構的表現量皆會隨著 B 細胞分化而增加。

有趣的是，儘管其生理功能目前還不是很清楚，BCL1 的 N 型醣鏈末端的單唾液酸、雙唾液酸化乙酰基乳糖胺，及 N 型醣鏈核心結構皆可被硫酸化修飾。另外，在人類活化的 B 淋巴球上也鑑定到了 α2,6-sialyated 6-sulfo-LacNAc 結構，單和 α2-6 唾液酸化乙酰基乳糖胺相比，其為目前已知 CD22 更好的配體，這些 B 細胞上多乳糖胺聚醣鏈附加修飾的鑑定，使得 Galectin 和 Siglec 對 B 細胞分化的調控可以更為複雜精緻。總而言之，質譜分析技術的發展和進步，對於我們詳細地鑑定多乳糖胺聚醣結構來講，為一個相當重要的基礎，有助於我們對醣生物學及其它生理功能更近一步的了解。

關鍵字：質譜儀、多乳糖胺聚醣、雙唾液酸、硫酸化

## Abstract

Most mass spectrometry (MS)-based glycomic and glycoproteomic analyses focus on identifying changes in terminal glyco-epitopes represented by sialylation and fucosylation at specific positions of the terminal *N*-acetyllactosamine units. Much less attention was accorded to the underlying linear or branched poly-*N*-acetyllactosamine (polyLacNAc) extension from the N-glycan trimannosyl core other than a simple inference of its presence due to mass data and hence glycosyl compositional assignment. To advance the frontiers of glycomics, this thesis work aims primarily to address the analytical challenges in structural characterization of polylactosaminoglycans and associated terminal modifications such as sialylation and sulfation decorating the human endothelial cells, mouse and human B cells.

Using the human endothelial cells, EA.hy926 and HUVEC, as starting materials, we have systematically investigated the MALDI- and ESI-MS-based methodologies for probing the structural details of polyLacNAc at both MS and MS/MS levels in conjunction with the use of endo-β-galactosidase and Smith degradation to identify branching motifs and initiation sites. N-glycans in EA.hy926 were found to be less sialylated and fucosylated but more extended and branched than those of HUVEC, thus demonstrating a fundamental glycomic difference. For EA.hy926, its polyLacNAc chains were shown to be not restricted to extending from a specific antenna including the biologically important 6-arm position. Extending to glycoproteomics, the *Lycopersicon esculentum* lectin based enrichment strategy was optimized at glycan, glycoprotein, and glycopeptide levels, leading to identification of over 40 protein carriers utilizing a two-step enrichment workflow.

For mouse B cells, the N-glycans of a B lymphoma cell line, BCL1, were found to be mostly core-fucosylated, capped with α-Gal or Neu5Gc sialic acid, and carry non-branched polyLacNAcs. In contrast, its O-glycans were based on simple core 1 structures, mono- or disialylated on both arms. Sialidase digestion, in conjunction with further MS/MS and chemical analyses, established the identity of the terminal disialyl motif as Neu5Gcα2-8Neu5Gc-, which was shown by endo-β-galatosidase digestion to be additionally present on both polyLacNAc extended and non-extended N-glycans. Fluorescent-labeling of released sialic acids coupled with fluorometric high performance liquid chromatography analysis revealed that the amount of the disialyl motif was comparable for both N- and O-glycans, and CD45 is one of the protein carriers. Gene knockdown studies provided positive correlation indicative of mouse α2,8-sialyltransferase VI (ST8sia VI) being involved in the biosynthesis of disialic acid on both N- and O-glycans. Importantly, both the expression level of ST8sia VI and the total amount of disialic acids increase during B cell differentiation.

Interestingly, sulfation was additionally found on the terminal mono- and disialylated LacNAc of the polyLAcNAc chains, as well as on the LacNAc proximal to the trimannosyl core in BCL1 although its biological relevance is at present unclear. On the other hand, similar analysis led to identification of α2,6-sialylated 6-sulfo-LacNAc epitope on both the N- and O-glycans of activated human B cells, which is known to constitute a better ligand than the non-sulfated α2,6-sialylated

LacNAc for human CD22. These additional modifications of polyLacNAcs apparently complicate the simplistic interpretation of the modulating roles of galectins and Siglecs in the B cell differentiation model. The development of enabling analytical techniques sensitive enough to identify and characterize the fine structural details of the underlying polyLAcNAc is an important step towards a better understanding of the glycobiology of this and many other physiological processes.

Key words: Mass spectrometry, polylactosaminoglycans, disialic aicd, sulfation

## Abbreviation

CID: collision-induced dissociation
DABP: 3,4-diaminobenzophenone
DHB: 2, 5-dihydroxybenzoic acid
DMB: 1, 2-diamino-4,5-methyleneoxybenzene
DP: degree of polymerization
DSA: *Datura stramonium agglutinin*
EI: electron impact
ESI: electrospray ionization
FAB: fast-atom-bombardment
Fuc: fucose; Gal: Galactose; Glc: Glucose; Man: mannose
GC: gas chromatography
GlcNAc: N-acetylglucosamine; GalNAc: N-acetylgalactosamine
β3GnT: UDP-GlcNAc:βGal β-1,3-N-acetylglucosaminyltransferase
Hex: Hexose; HexNAc: N-acetylhexosamine
HPLC: high-performance liquid chromatography
HUVEC: human umbilical vein endothelial cells
IT: ion trap
LacNAc: N-acetyllactosmaine
LC: liquid chromatography
LEL: *Lycopersicon esculuntum* lectin
LPS: lipopolysaccharides
LTQ: linear ion trap
MALDI: matrix assisted laser desorption/ionization
MS: mass spectrometry
Neu5Ac: N-Acetylneuraminic acid; Neu5Gc: N-glycolylneuraminic acid
PA: 2-aminopyridine
PolyLacNAc: poly-*N*-acetyllactosamine
Q: quadrupole
Siglec: sialic acid binding Ig-like protein
ST8Sia: α2,8-sialyltransferase
TFA: trifluoracetic acid
TIM: total ion mapping
TOF: time-of-flight

## Table of contents

## List of figures and tables

# Chapter I Introduction

## 1.1 Glycosylation, Glycomics and Glycoproteomics

The cells of all organisms consist of four fundamental biomolecules: nucleic acids (DNA and RNA), proteins, lipids, and glycans. Unlike the biosynthesis of DNA, RNA and proteins, which use a nucleic acid sequence as a template, the fidelity of glycan biosynthesis relies on the specific activities of various glycosyltransferases acting in concert on their acceptor substrates. Inherent structural complexity is accentuated by the fact that individual subunits of glycans, namely the monosaccharides, can be linked together in more than one way, i.e., to one of three or four different positions of the neighboring monosaccharide units and in either α-glycosidic or β-glycosidic linkage. In vertebrates, the assembled glycan chains are mostly attached to secreted proteins, and the membrane proteins and lipids, thus forming a densely glycosylated cell surface (Fig 1.1) (Varki, 2008) populated by several main families of glycoconjugates:

a、 Glycoproteins (Fig 1-1a, b): In the glycoproteins of higher organisms, glycans are typically N-linked to asparagine (Asn) situated within the Asn-X-Thr/Ser sequon (X is any amino acid except Pro), or O-linked to the serine (Ser) or threonine (Thr) in the peptide chain;

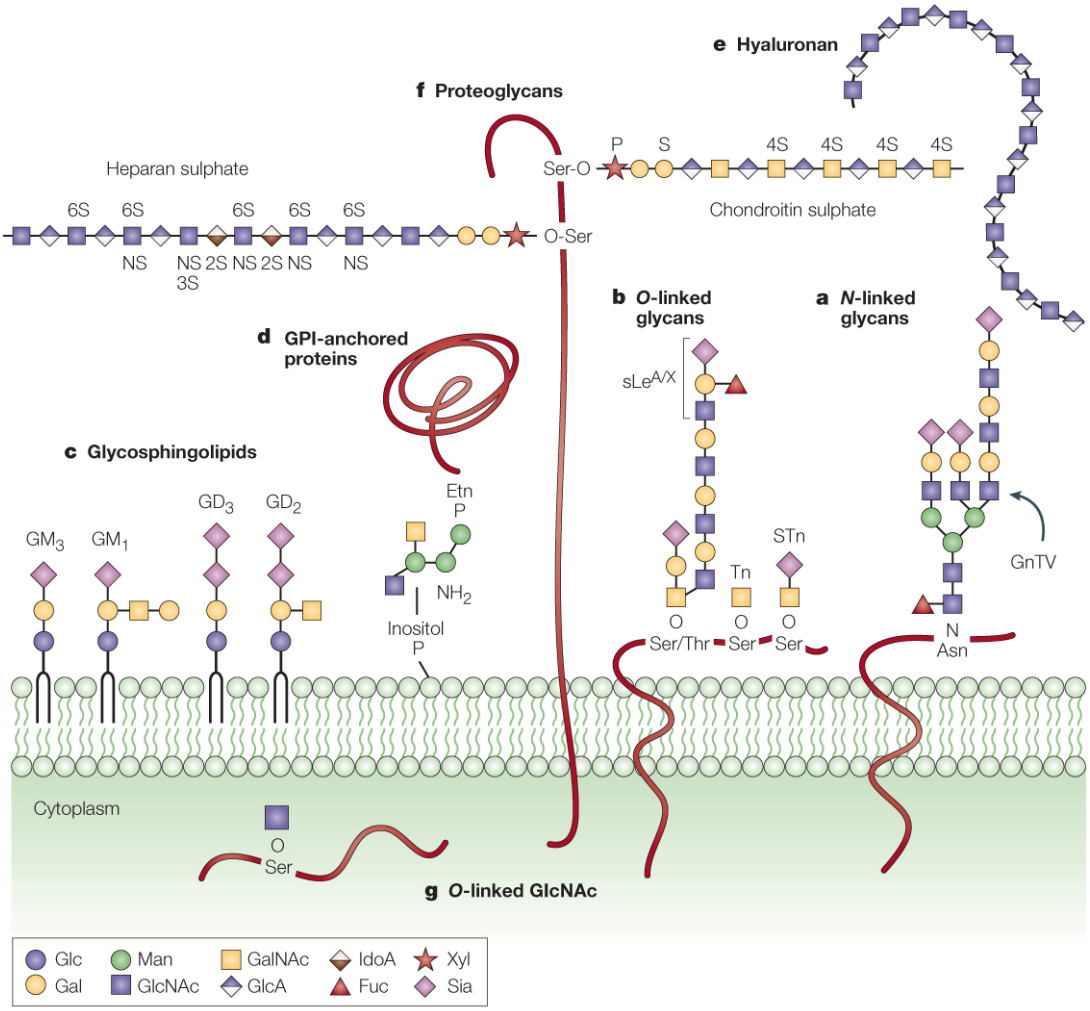
b、 Glycolipids (Fig 1-1c): glycolipids typically form less than 5% of the lipid content of most mammalian tissues, but represent more than 25% of the lipid content of myelin sheaths (Kanter et al., 2006). In mammals, the glycan chains of glycosphingolipids are mostly extended from a lactosylceramide core, as one of the three major structural series of the so-called ganglio-, (iso)globo- and (neo)lacto- sequences. They have been correlated with tumor growth and as antigens in various autoimmune neurological diseases such as Guillain-barré syndrome (GBS);

c、 GPI-anchored proteins (Fig 1-1d): GPI anchors are proteins that bear a glycan chain linked to a phosphatidylinositol lipid anchor, which confer specific properties on proteins, such as association with membrane microdomains. The structure of lipid and glycan moieties on GPI anchors are remodeled during biosynthesis and after attachment to proteins, which are critical for transport and microdomain-association of GPI-anchored proteins;

d、 Glycosaminoglycans (GAG) (Fig 1-1e, f): these are long, polymeric glycan chains consist of distinctive negatively charged disaccharide repeating units of hyaluronan, heparin sulfate, keratan sulfate, chondroitin sulfate, and dermatan sulfate. Except for hyaluronan, which occurs as free polysaccharides, other GAGs are normally attached to a secreted or membrane protein core forming the proteoglycans. GAGs play significant roles in the modulation of cellular signals

through interactions with proteins, including growth factors and growth factor receptors (Zhang, 2010);

e、 O-linked GlcNAc (Fig 1-1g): the Ser and Thr of various cytoplasmic and nuclear proteins can be reversibly modified by an O-linked N-acetylglucosamine (O-GlcNAc) (Torres and Hart, 1984), in place of phosphorylation. Cytosolic O-GlcNAcylation is important for the proper transduction of signaling cascades such as NF-κB pathway, whereas nuclear O-GlcNAc is crucial for regulating the activity of numerous transcription factors.




*Nature Reviews Cancer*, **5**, 526-542 (2005)

**Figure 1-1. Different glycoconjugates** including (a) N-linked glycans; (b) O-linked glycans; (c) glycosphingolipids; (d) glycosylphosphatidylinositol (GPI)-linked proteins; (e) glycosaminoglycans (such as hyaluronan); (f) proteoglycans; (g) various cytoplasmic and nuclear proteins contain O-linked N-acetylglucosamine (O-GlcNAc).

Glycobiology is the study of the biological role of glycans and glycoconjugates, including glycoproteins, glycolipids, proteoglycans, and of protein-glycans interactions. Glycan chains in glycoproteins and glycolipids play fundamental role in

many biological processes such as embryonic development, immune response and cell-to-cell interactions involving sugar-sugar- or sugar-protein-specific recognition. In fact, aberrant glycosylation has long been implicated in many diseases, including hereditary disorders, immune deficiencies, neurodegenerative diseases, cardiovascular diseases and cancer. Changes in glycosylation are often a hallmark of disease states and serve as antigenic markers revealing the presence of unnatural or infectious agents. Glycomics aims to define all the expressed glycans or the glycome at any particular patho-physiological stage, from which to establish the connections between glycan structures and their functions (functional glycomics). By monitoring glycosylation changes in disease diagnosis and prognosis, one may further elucidate the molecular mechanisms involved in pathogenesis (Dennis et al., 2009; Freeze and Aebi, 2005; Peracaula et al., 2008).

To be more precise, a glycomic investigation should endeavor to tackle the full repertoire of glycans attached to each of the major classes of glycoconjugates, including those on proteins and lipids, without necessarily considering the identity of their carriers. On the other hand, glycoproteomics aims more specifically to delineate the distribution of each of the assorted glycans on myriad protein carriers, including their attachment sites and to identify any site-specific glycosylation pattern that confers altered physico-chemical properties and functional activity on the nascent protein chain. It should be noted that many clinical biomarkers and therapeutic targets such as Her2/neu (breast cancer), prostate-specific antigen (PSA, prostate cancer) and CA125 (ovarian cancer)(Diamandis, 2004; Pan et al., 2006) are glycoproteins, and their site-specific glycosylation patterns bear important biological consequences.

**1.2 Strategies and approaches for glycosylation analysis**

Full characterization of protein glycosylation involves a number of different levels of analysis, to find the protein sites that are glycosylated, the level of glycan occupancy at each site, and the actual glycan structures at each site. Due to the effects of substoichiometric levels of glycan attachment with the highly heterogeneous structures, it's a quite difficult to perform such a complete analysis. There is no single technique capable of providing a full qualitative or quantitative structural analysis of glycoprotein or glycopeptides, let alone at the omics level as applied to whole cells or tissues. To achieve these goals, the development of precise, robust, and sensitive methodologies for glycan and glycopeptide analysis is critical.

Although many alternatives are possible, it is a common practice to first map the glycosylation profile of a protein or proteome at the released glycan level, especially if this information is not yet available. One would get many useful cues to formulate subsequent steps of analysis, be it a more targeted and in depth probing

into minor glycotopes not uncovered during such a first-screen glycomic mapping, or to devise enrichment strategies for capturing the targeted glycoproteins or glycopeptides bearing the glycotopes of interest thus revealed. A strategic workflow therefore starts from glycomic or glycan analysis, involving several modules corresponding to different degrees of precision, followed by knowledge-based driven glycoprotein/glycopeptides enrichment and glycan fractionation steps for more targeted glycoproteomics and glycomics, respectively (Fig.1-2).

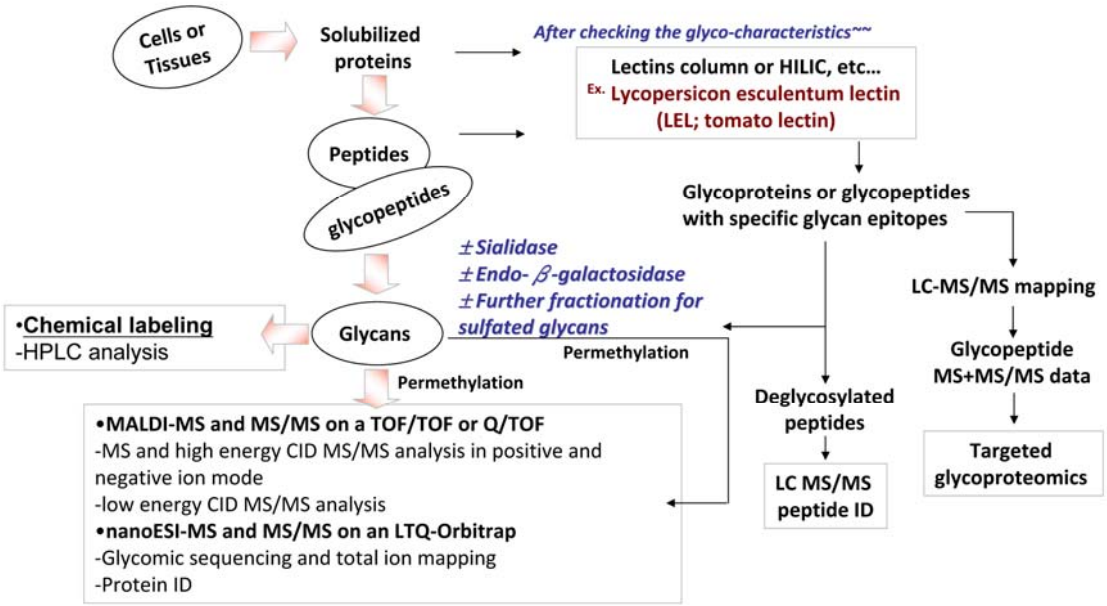



**Figure 1-2. The conceptual workflows from glycomics to glycoproteomics**

An initial compositional analysis can be performed on the released glycans by quantification of monosaccharides. All sugars apart from sialic acids—including fucose, hexosamines and uronic acids—can be confidently determined after optimized treatment with trifluoroacetic acid (TFA). Monosaccharides analysis and identification can be preformed by gas chromatography (GC), or by high-performance liquid chromatography (HPLC), for example, high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD)(Townsend and Hardy, 1991). However, the latter method can be limited by unstable baselines, loss of sensitivity, high pH and high salt concentration. Alternatively, the released glycans can be labeled with a fluorophore e.g. 2-aminobenzaminde (2-AB), 2-aminopyridine (PA), and 2-aminobenzoic acid (2-AA), to enable detection at the femtomole level in conventional HPLC analysis. For the issue of sialic acids, more than 20 different sialic acids are known, and understanding their key biological roles is becoming important. The fluorescence labeling of sialic acid with 1, 2-diamino-4, 5-methyleneoxybenzene (DMB)(Hara et al., 1987) is a useful tool, involving selective release of sialic acid and analysis of the quinoxaline derivatization product by using revered-phase HPLC.

The conventional approach for linkage analysis of glycans is GC-electron impact (EI)-mass spectrometry (MS) analysis of partially methylated alditol acetates (Lindberg and Lonngren, 1978) derived from permethylated glycan sample by acid hydrolysis, reduction and then peracetylation. The differential methylation versus acetylation pattern induces distinctive EI-MS fragmentation pattern, which allows the determination of the initial glycosyl substituted positions in the glycan sample. Although lagging behind in sensitivity, this methylation or linkage analysis method remains widely in use today to define the presence of critical linkage-specific glycosyl substitution or branching pattern within a glycome, following an initial MS glycomic mapping based on the released, permethylated glycans (see later section).

Nuclear magnetic resonance (NMR) spectroscopy is another powerful technique for obtaining important sequence information on glycans. Carbohydrates have two natural NMR-active nuclei: $^{13}C$ and $^{1}H$. The anomeric chemical shifts of the monosaccharides can be classified further based on the neighboring monosaccharides (at the reducing end), which would provide the abundance of the specific linkage between the two monosaccharides. The information is particularly important for terminal sialic acids, which can be α2-3 or α2-6 linked to the penultimate monosaccharide. However, a major limitation of this technique is that a relatively large amount, in the range of 10-100 ng, of purified glycans is required.

While such traditional approaches to glycan analysis, including specific exoglycosidase, derivatization, monosaccharide analysis, chromatography, and NMR analysis described above remain important in a modern glyco-analytical workflow, mass spectrometry is the single most rapidly evolving tool to meet the increasingly demanding need of glycobiology along with the recent emergence of microarrays, including antibody, glycosyltransferases, glycan and lectin microarrays.

The microarray platform is now widely used to characterize glycan structures based on their recognition by a panel of immobilized lectins, and to study glycosylation-related biological interactions by using the spotted or captured glycosylated molecules on the arrays as probes to interrogate a wide range of glycan binding biomolecules. These have proved to be a much more sensitive and higher throughput technique than other technologies available for glycan analysis, such as chromatography and mass spectrometry because they allow multiple analytes to be investigated simultaneously due to their physical attachment to unique regions on a single microarray slide that can be analyzed quickly with a reagent of choice (Tao et al., 2007; Uttamchandani and Yao, 2008; Voduc et al., 2008). However, these are not a substitute for precise structural determination. Instead, lectin arrays are commonly used as a high sensitivity first screen to gain an impression of probable glycosylation characteristics and for comparative analysis of a large sample size, particularly in biomarkers type of applications to map differential glycomic expression. It is highly

complementary to a more laborious following-up studies to try to validate the differentially expressed glycan structures.

Without doubt, recent advances in mass spectrometry techniques and instruments capable of detecting and sequencing glycans in the low picomole to high femtomole range with great accuracy has made MS-based glycomic analysis the method of choice for its unrivaled precision in structural determination of complex glycans derived from biological sample. It has in fact revolutionized glycomic analysis.

### 1.2.1 Mass spectrometry for glycosylation analysis

The history of glycan MS analysis commenced in the 80s and early 90s when fast-atom bombardment (FAB) or liquid secondary ionization MS (LSIMS) on sector type of instrument is the only viable MS ionization technique for efficient analysis of oligosaccharides (Dell, 1987; Heinz Egge, 1987). Molecular ion and fragment ion signals from intact glycans could be simultaneously afforded by FAB-MS, which are highly informative for a direct mapping of non-reducing terminal glycotope of glycan mixtures such as polyLacNAcs (Fukuda et al., 1984a; Fukuda et al., 1984b). The only drawback is that these fragment ions cannot be traced back to a particular parent ion unless the glycan of interest is first purified and analyzed in isolation. Another major limitation of FAB-MS is insufficient sensitivity especially when analyzing non-derivatized, native glycans. Thus from inception, efficient MS mapping of glycans has been relying mostly on analysis of permethyl derivatives, which not only enhances detection sensitivity, but also facilitates sequencing by giving more predictable fragmentation and reducing the predominant neutral loss of labile residues especially for fucosylated and sialylated glycans (Dell et al., 1994).

The process of permethylation generates methyl derivatives for each free OH and NH group and also esterifies the carboxyl group in the sialyl residues, increasing the overall sensitivity and making possible the simultaneous analysis of neutral and sialylated glycans in the same positive ion mode. The initial methylation method introduced by Hakomori (Hakomori, 1964) has now been virtually superseded by the less demanding NaOH/DMSO slurry method developed by Ciucanu and Kerek (Ciucanu and Kerek, 1984). Due to its popularity and proven utilities, several groups have since optimized it further (Wada et al., 2007), adapted to microscale, as well as coupled to solid phase (Kang et al., 2005). Importantly, our laboratory has recently shown that this standard NaOH permethylation procedure can be efficiently applied to sulfated glycans without losing the sulfates (Yu et al., 2009). A critical step in ensuring a good recovery of the permethylated sulfated glycans is the passing of the quenched and neutralized reaction mixtures through Sep-pak C18 instead of direct partitioning against organic solvent such as chloroform. Following a simple Sep-pak C18 clean-up, an amine-based stationary phase can be employed to further separate

the permethylated glycans into non-sulfated, mono-sulfated and di-sulfated fractions. This development essentially extended the versatility of MS-based glycan mapping, enabling both MS and MS/MS analyses of specifically enriched, sulfated, permethylated glycans, at high sensitivity.

On the MS technique itself, recent advances have almost supplanted the pioneering FAB ionization and sector instruments responsible for most of the fruitful glycan structural studies in the 80s and 90s. Matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) sources fitted on a range of mass analyzers were introduced in the mid 90s. Both MALDI and ESI are much softer ionization techniques compared with FAB and deemed not imparting sufficient energy to induce much in source fragmentation. Using the MALDI MS technique, sample is mixed with a chemical matrix that facilitates the production of intact gas-phase ion from large, non-volatile, and thermally labile compounds. The most commonly used matrix in glycan analysis include 2, 5-dihydroxybenzoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CHCA), and 6-aza-2-thiothymine (2-AA), 2,4,6-trihydroxyacetophenone (THAP), and 3,4-diaminobenzophenone (DABP) (Yu et al., 2009) for acidic glycans. Singly charged ions predominate using MALDI and usually desorb as $[M+Na]^+$ adducts. ESI entails spraying a solution containing the analyte through a needle, to which electrical potential is applied. Multiply charged ions result from ESI, and this is advantageous in that molecular ions can be detected by mass spectrometers having a limited mass-to-charge (m/z) range. However, the multiply-charged fashion also disperses one molecule into ions of different m/z, which is prone to complicate the MS profile. In addition, the combination of liquid chromatography (LC) with MS provides a powerful analytical tool for both qualitative and quantitative glycan analysis and dramatically improves the amount of obtainable information.

The mass resolution that is achievable by a mass spectrometer depends on both the type of analyzer and the experimental conditions. Several different types of mass analyzers exist, including the ion trap (IT), the time-of-flight (TOF), quadrupole (Q), and Fourier transform ion cyclotron resonance (FTICR). Each analyzer can be used alone or in tandem with another. Most of the current, high performance mass spectrometers comprise two different or identical analyzer modules to enable efficient MS/MS sequencing. MALDI-MS/MS is mostly performed on TOF/TOF (Harvey, 2000; Naven et al., 1997), but also on Q/TOF (Harvey et al., 2000; Loboda et al., 2000) and more recently also coupled to IT, which alone is capable of multistage MS ($MS^n$) (Peterman and Mulholland, 2006). On the other hand, ESI is mostly coupled to either IT or Q/TOF. Either offline (direct infusion) or online ESI-MS/MS coupled with micro or nanoLC is widely used. Because of the popularity of the LTQ-Orbitrap platform, glycans mapping is also increasingly performed on this instrument. Any of these platforms can allow simple glycan MS mapping and MS/MS sequencing although

each varies slightly in practical details and can be highly complementary in special cases.

### 1.2.2 MALDI-MS/MS sequencing of glycans

The predominant fragment ions produced at low energy MALDI CID MS/MS for monosodiated permethylated glycans are always the B/Y and C/Z ion pairs (Fig. 1-3(A)) according to the Domon and Costello nomenclature (Costello, 1988). Glycan sequencing by MALDI Q/TOF tends to be elegantly simple and most fragment ions can be readily assigned as either corresponding to B or Y ion or a combination of both, deriving from facile cleavages at HexNAc of permethylated glycans. The C/Z ions are less readily formed except at sialic acids and 3-linked saccharides which are sufficient to define a 3-linked glycosyl substituent, especially in distinguishing type I and type II LacNAc (Fan et al., 2008). In addition, multiple glycosidic cleavages is also a characteristic feature of low energy CID MS/MS of sodiated parent ions on the Q/TOF, which can be turned into advantages for distinguishing branching and multiple substitution by virtue of the O-Me tag introduced by permethylation in the first step.

The advantages of MALDI TOF/TOF over Q/TOF for CID MS/MS is its ability to select parent ions of much higher m/z value and compensate for the shortage of linkage information by providing cross-ring and internal double-cleavage ions in addition to glycosidic ions. Characteristic ions in TOF/TOF are shown as reducing terminal fragment ions (the X, G, H, Y ions), non-reducing terminal ions (the A, B, C, E ions), and double cleavage ions (D and C”/Y). For practical purpose, the $^{1,5}$X ions and Y ions can provide general sequential information. For a 4-linked HexNAc, the G and H ions differ by 14 u and thus can be easily identified as a characteristic pairs of signals. A 3-linked GlcNAc would only give a G ion and not an H ion since both the 4- and 6-positions are not substituted. For the linkage of terminal sialic acids, only the 3-linked sialic acid will be eliminated in concert to generate the G ion. The D ion at the 3,6-branched β-Man coupled with elimination of the 3-arm helps us to define to 6-arm substituent. Its formation is not restricted at this position but also can be detected for each residue along the antenna but is more prominent, in terms of mass shift and abundance, when the cleavage site residue is substituted at the C3 position. In addition, since there are not many naturally occurring Hex residue that is 2-substituted apart from the Man of trimannosyl core, the C”/Y ions is mostly observed only with tri-and tetra-antennary structures. In these cases, the antenna 2-linked to Man will be lost, and thus would help to define the substituent on the antenna 4-linked or 6-linked to Man.

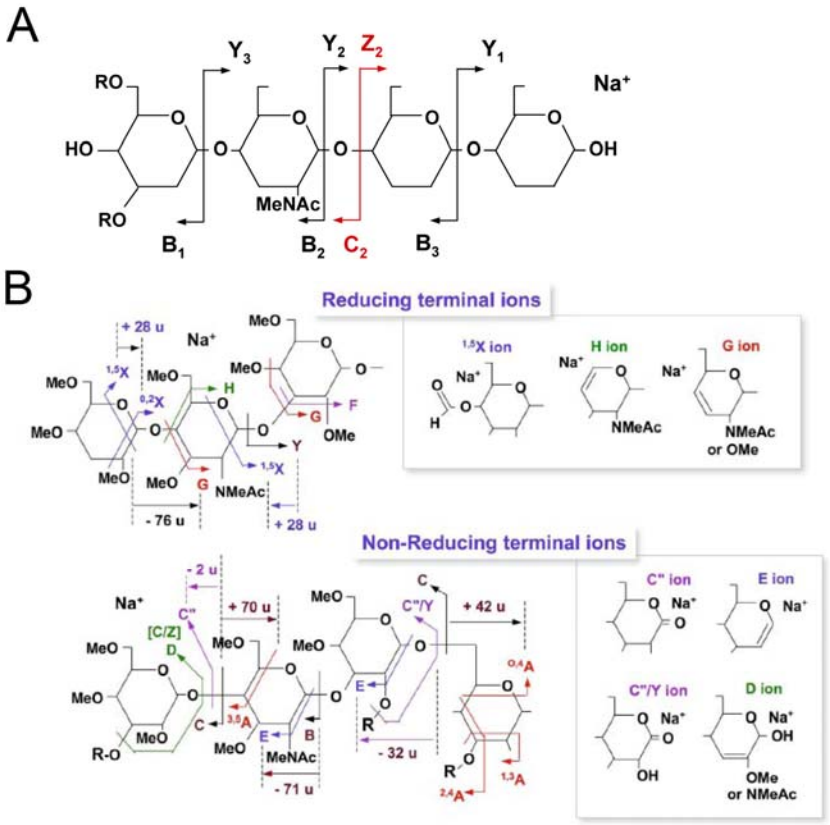



**Figure 1-3. MS/MS fragmentation pattern of low energy CID in Q/TOF (A) and high-energy CID in TOF/TOF (B).** Fragment ions in Q/TOF are mostly glycosidic B/Y or C/Z ions including multiple cleavages. Characteristic ions in TOF/TOF are shown as reducing terminal fragment ions (the X, G, H, Y ions), non-reducing terminal ions (the A, B, C, E ions), and double cleavage ions (D and C"/Y).

In addition to Q/TOF and TOF/TOF analyzers, ion traps, which alone are capable of multistage MS is also widely used for glycan sequencing. $MS^n$ experiments are especially useful for defining the linkages between carbohydrate residues via cross-ring fragmentation. Ion trap is robust and sensitive; however, they have relatively low mass resolutions and accuracy due to the limited number of ions that can accumulate in a trap at once before space-charging effects distort their measurements. The highly popular commercial hybrid LTQ Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA) consists of a hybrid mass spectrometer that combines a quadrupole linear ion trap (LTQ) with an Orbitrap FT mass analyzer (Olsen et al., 2005). The LTQ ion trap is capable of detecting MS and $MS^n$ spectra at high sensitivity, but has relatively low resolution and low mass accuracy. It accumulates and isolates ions, and fragments them in the mass range below m/z 2000. It is also very fast due to the high ion trap scan speed combined with the ability of the LTQ Orbitrap to measure precursors and fragments simultaneously. In contrast, the orbitrap MS is a high-resolution (up to 200,000), high-sensitivity, and high mass accuracy (1-2ppm) FT mass analyzer. The linear ion

trap Orbitrap hybrid instrument allows fragmentation either in the linear ion trap or the higher-energy collisional dissociation (HCD) cells. Resulting fragments can be measured either with the linear ion trap or with the Orbitrap detectors. Good fragment ion spectra in the Orbitrap analyzer require about an order of magnitude more ions than in the LTQ.

### 1.2.3 Glycopeptide enrichment for Glycoproteomics

Most glycoproteomic studies use MS for detection, typically combined with some type of glycan-specific enrichment or derivatization/tagging. Affinity purification has been achieved by using lectins or antibodies that are specific for certain glycan structures. Chemical methods such as hydrophilic interaction chromatography (HILIC) and $TiO_2$ affinity column have also been developed to accomplish selective isolation, identification, and quantification of glycoproteins or glycopeptides. For HILIC column, samples are loaded in organic solvents and recovered by increasing the aqueous content of the mobile phase. N-linked, O-linked, and GPI-anchored peptides can be purified in this way. More recently, glycopeptides containing sialic acid was shown to be selectively captured and recovered for MS analysis by using $TiO_2$ affinity enrichment (Larsen et al., 2007; Larsen et al., 2005). In addition, N-linked glycoproteins and glycopeptides can also be chemically derivatized and immobilized using periodate oxidation and hydrazide chemistries via their *cis* vicinal diols (Zhang et al., 2003). The immobilized proteins/peptides are subsequently released by treatment with N-glycosidase enzymes and identified by MS.

Lectins, a class of sugar-binding and cell-agglutinating proteins, are ubiquitous in nature, being found in all kinds of organisms, from virus to humans. They were originally obtained by classical biochemical techniques. The introduction in 1965 of affinity chromatography on immobilized carbohydrates for plant lectin isolation greatly facilitated the task and led to a fast increase in the number of these proteins purified (Agrawal and Goldstein, 1965). Another important discovery made somewhat later was that lectins, concanavalin A (ConA), can be employed for the isolation of glycoproteins (Donnelly and Goldstein, 1970). Many different lectins now are commercially available in an immobilized form suitable for glycoprotein purification. For example, ConA which binds to high mannose-type and hybrid-type oligosaccharides with high affinity (Kornfeld et al., 1981; Ohyama et al., 1985), wheat germ agglutinin A (WGA) which recognizes *N*-acetylglucosamine and sialic acid residues, *Arachis hypogaea* lectin (PNA, peanut agglutinin) specific to T-antigen found commonly in O-glycans (Neurohr et al., 1980), and *Aleuria aurantia lectin* (AAL) showing broad specificity for L-Fuc-containing glycans (Kochibe and Furukawa, 1980).

In the case of recognizing the basic building block of N- and O-glycans, galectins specific for LacNAc-containing N- and O-glycans have been used (Hirabayashi et al.,

1992; Hirabayashi et al., 1996), in addition to the *Lycopersicon esculentum* (tomato) lectin (LEL), which has the high affinity for N-linked glycopeptides containing three or more linear N-acetyllactosaminyl repeats (Lee et al., 1990; Merkle and Cummings, 1987), and *Datura stramonium agglutinin* (DSA), which recognizes two N-acetyllactosamine repeats arranged within the sugar chains in various manners. The ability of different lectins to recognize specific glycosylation motifs has been used to develop a multi-lectin affinity system that can achieve a comprehensive capture of glycoproteins as well as depleting non-glycosylated proteins. However, identification by lectins still may be misleading due to their different affinities of binding to similar structures.

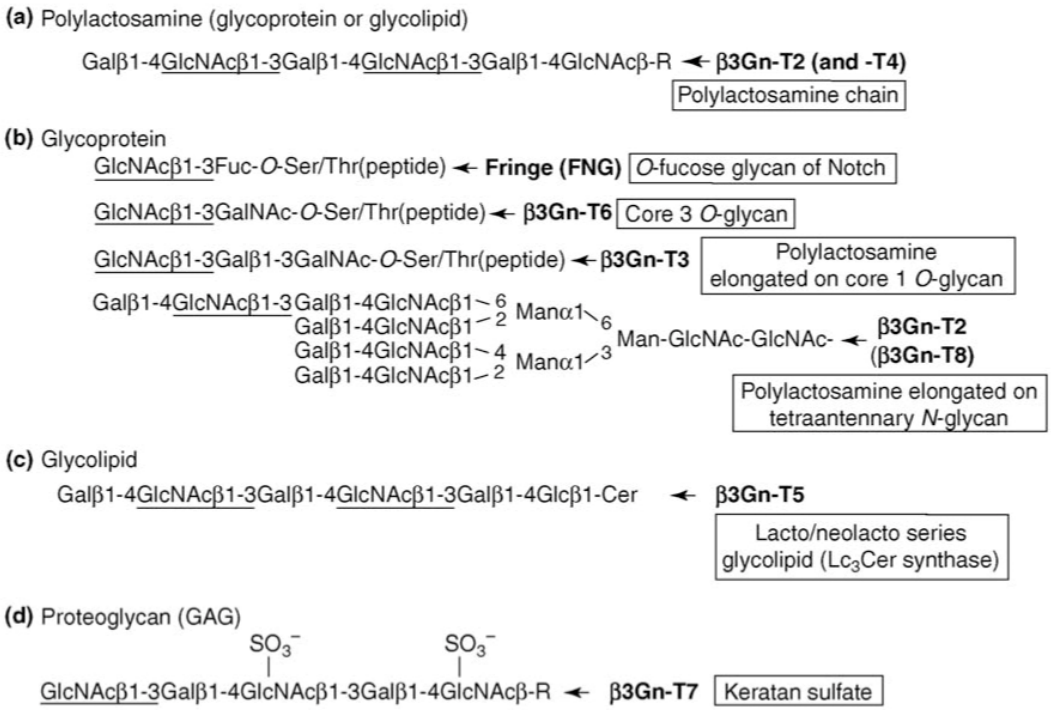
### 1.3 Structural and Functional Implications of poly-N-acetyllactosamine

Current glycomics, glycoproteomics and functional analyses of mammalian protein glycosylation are often founded on our knowledge of its general structural architecture governed by the well established biosynthetic pathways. It is understood that, structurally, the trimannosyl core of the N-glycans is commonly extended by an N-acetyllactosamine (Galβ1-4GlcNAcβ1-, or LacNAc) unit from specific positions of the 3- and 6-arm mannoses to give bi- to multiantennary complex type N-glycans. Likewise, either or both 3- and 6-arms of the common O-glycan cores can be extended by LacNAc. These LacNAc units can be further decorated by sialylation, fucosylation and/or sulfation at specific positions to form various terminal epitopes including the blood group antigens, ligands mediating myriad cell-cell, and pathogen-host recognition events (Fig. 1-6).

Often, for certain cell types or tissues and under onco-developmental regulation, the LacNAc units can be repeatedly extended to form linear chains of -[3Galβ1-4GlcNAcβ1-]$_n$, or polyLacNAc, which may be further branched at the 6-positions of the 3-linked Gal, to give the characteristic branched point structural motif of -GlcNAcβ1-3(-GlcNAcβ1-6)Galβ1-4GlcNAc- (Fukuda et al., 1984b; Ujita et al., 1999a; Ujita et al., 1999b). It is generally thought that branching of polyLacNAcs furnishes an extra dimension of multivalent scaffold that will better position the various glyco-epitopes as high avidity ligands and antagonists of cell adhesion proteins (Dennis et al., 1999; Leppanen et al., 1998; Muramatsu et al., 2008; Renkonen, 2000). Yet, despite being implicated as critical determinants of the basal levels of lymphocyte and macrophage activation (Togayachi et al., 2007), tumor cell metastasis (Saitoh et al., 1992; Srinivasan et al., 2009) and embryogenesis (Muramatsu et al., 2008; Muramatsu, 1988; Spillmann and Finne, 1994), the structural details of polyLacNAcs remain largely unclear and undefined in most cases. This, in turn, hindered much progress in our better understanding of the biosynthesis and functioning of polyLacNAcs.

### 1.3.1 Biosynthesis of polyLacNAc and its potential biological function

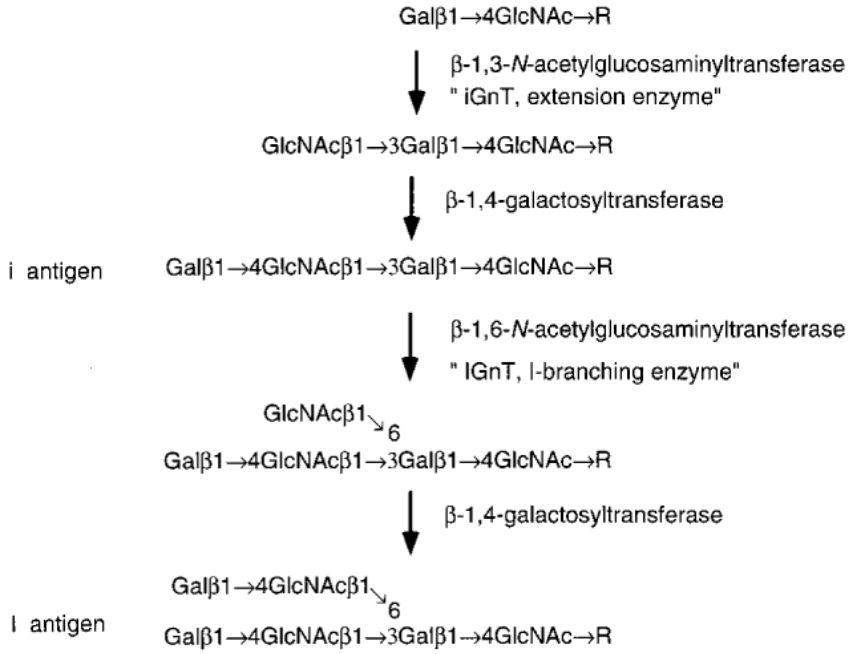
PolyLacNAc is biosynthesized by the alternating addition of GlcNAc by a UDP-GlcNAc:βGal β-1,3-N-acetylglucosaminyltransferase (β3GnT) and Gal by a UDP-Gal:βGlcNAc β-1,4-galactosyltransferase (β4GalT). Eight β3GnTs have been identified to date and five of them have been reported to be related with the biosynthesis of polyLacNAc in N-glycans: β3GnT1, 2, 4, and 8. For the biosynthesis of polyLacNAc, β-1,3-N-acetylglucosaminyltransferase (iGnT, β3GnT1) essential for the formation of the i antigen (Fig 1-4) was isolated by the expression cloning method (Sasaki et al., 1997). Three additional β3GnTs, β3GnT2 to β3GnT4, were subsequently isolated based on structural similarity with the β3Gal T family. β3GnT2 was found to have strong activity *in vitro* against oligosaccharide substrates with polylactosamine structures, suggesting that it is the main polylactosamine synthase (Shiraishi et al., 2001; Togayachi et al., 2001). β3GnT3 is the enzyme which elongates polylactosamine chains on core 1 O-glycans. β3GnT4 has weak polylactosamine synthase activity. β3GnT8 was the most recently identified enzyme among the β3GnTs. It has been reported that the complex of β3GnT2 and 8 exhibits potent polyLacNAc synthesis activity against tetraantennary N-glycans (Ishida et al., 2005; Seko and Yamashita, 2005). However, the differential role played by these enzymes in the body is still unclear.




*Current Opinion in Structural Biology*, **16**, 567-575 (2006)

**Figure 1-4. Schematic representation of the implied substrate specificities of each β3GnT.**

In erythrocyte, the fetal i antigen is converted to the adult I antigen by I branching β-1,6-N-acetylglucosaminyltransferase (IGnT) during development (Fukuda et al., 1984a; Fukuda et al., 1984b). When IGnT is also present with iGnT, the I antigen is synthesized instead of the i antigen. The cDNA encoding IGnT1 (cIGnT) was first cloned from human embryonal carcinoma cells, PA-1 (Bierhuizen et al., 1993). Later, IGnT2 (also called C2GnT-2) (Schwientek et al., 1999; Yeh et al., 1999) and IGnT3 (Inaba et al., 2003), were isolated by analyzing the genome database with sequence of IGnT1. It has been known that IGnT1 has the strong centrally acting I-branching activity, whereas IGnT2 exhibits only weak peridistal I-branching activity (Fig 1-4).




*Proc. Natl. Acad. Sci.*, **94**, 14294-14299 (1997)

**Figure 1-5. Structure and biosynthesis of i and I antigens.** The i antigen is synthesized by iGnT followed by β-1,4-galactosyltransferase. The i antigen is converted to the I antigen by the stepwise addition of a GlcNAc β1→6 and a Galβ1→4 residue (Gu et al., 1992). As an alternative pathway, another IGnT adds β-1,6-*N*-acetylglucosamine to GlcNAcβ1→3Galβ1→4GlcNAc→R precursor and this product, GlcNAcβ1→3(GlcNAcβ1→6)Galβ1→4GlcNAc→R, is converted to Galβ1→4GlcNAcβ1→3(Galβ1→4GlcNAcβ1→6)Galβ1→4GlcNAc→R, forming I antigen (Bierhuizen et al., 1993; Fukuda et al., 1979; Gu et al., 1992; Koenderman et al., 1987; Piller and Cartron, 1983; Piller et al., 1984).

### 1.3.2 Mapping the occurrence of polyLacNAc

It has been shown that tumor cells express more polyLacNAc than normal counterparts (Dennis et al., 1987; Pierce and Arango, 1986; Saitoh et al., 1992; Yamashita et al., 1984). Since the functional roles of polyLacNAc on N-glycans are

most often correlated with the expression of N-acetylglucosaminyltransferase V (mannosylglycoprotein N-acetyl glucosaminyltransferase 5; GnT-V; MGAT5), which specifically add a GlcNAc to the C6 position of the 6-arm Man in tri- and tetra-antennary structures (Cummings et al., 1982; Saito et al., 1994), it is widely believed that polyLacNAc extension from this 6-arm position matters most biologically functions (Chakraborty and Pawelek, 2003; Demetriou et al., 2001; Guo et al., 2007; Lagana et al., 2006; Lau et al., 2007; Togayachi et al., 2007). In addition to GnTV, GnT-IVa and –IVb were also required for the formation of multiantennary compex-type N-glycans. Elevated expression of GnT-IVa, -IVb, and –V was found in many types of carcinoma cells(Ide et al., 2006; Kobata, 1988; Nan et al., 1998; Takamatsu et al., 1999). In GnT-IVa/-IVb double knockout mice, the loss of tetra-antennary structures is frequently accompanied by concomitant increases in polylactosamine-containing bi- and tri-antennary glycans of the same overall compositions (Takamatsu et al., 2010).

In general, from the perspectives of MS analysis, the presence of polyLacNAcs on N-glycans can be assumed whenever a molecular composition carrying more than 4 [$Hex_1HexNAc_1$] units in addition to the trimannosyl core, $Hex_3HexNAc_2$, is detected by MS-based glycomic mapping. For the O-glycans, any composition in excess of more than 2 [$Hex_1HexNAc_1$] units in addition to the residues attributed to the core, would implicate the presence of polyLacNAc. However, most such MS analyses did not probe into possible occurrence of branching or the initiation sites of polyLacNAcs (Babu et al., 2009; Denecke et al., 2008; Kremser et al., 2008). Instead, it is generally assumed that polyLacNAc preferentially extends from the 6-arm position of both N- and O-glycans, without much supporting biochemical evidence. (Fig. 1-6) In actuality, it is difficult to ascertain the presence and contribution of polyLacNAc chains extending from other antennary positions by any analytical means. For the N-glycans, identification of the -Galβ1-4GlcNAcβ1-6(-GlcNAcβ1-2)Manα1-6(-Manα1-3)Manβ1- outer arm branching motif is facilitated by the use of *Phaseolus vulgaris* leucoagglutinin (PHA-L) lectin but positive staining or binding does not strictly imply the presence of polyLacNAc (Cummings and Kornfeld, 1982). Likewise, of the more commonly used lectins that bind polyLacNAc and will retain polyLacNAc containing glycopeptides on column, neither the DSA nor the LEL has an absolute requirement for polyLacNAc (Cummings and Kornfeld, 1984; Lee et al., 1990; Merkle and Cummings, 1987; Yamashita et al., 1987).

It thus appear that, with further development and in conjunction with the use of selective cleavages, MS analysis remains the analytical method of choice that will allow unambiguous identification of polyLacNAc and determination of its attachment point, branching and additional sialylation, fucosylation and sulfation that may affect lectin binding. Such additional manifold decorations of polyLacNAc are thought to occur both at the primary chains and at the branches, and are extremely difficult to

map definitively. Most of them are expressed only distally, but some are present all along the backbones. These decorations include α3- and α6-linked sialic acids, 3/6-O-sulfate on Gal, 6-O-sulfate on GlcNAc, α2-linked Fuc on Gal, α3/4-Fuc on GlcNAc, β3-linked Gal on GlcNAc, or β3-/β4-/α3-linked GalNAc and α3-linked Gal on Gal (Fig. 1-6). Intriguingly, there are reports showing that terminal substitution such as the preferred sialylation linkages on the polyLacNAc may differ from that carried on non-extended LacNAc (Fukuda et al., 1984a; Nemansky et al., 1995). This and other observations underscore an important need to precisely map the occurrence and fine structural details of polylacNAcs, along with its terminal and/or peripheral substitution and branching, for a better functional delineation.

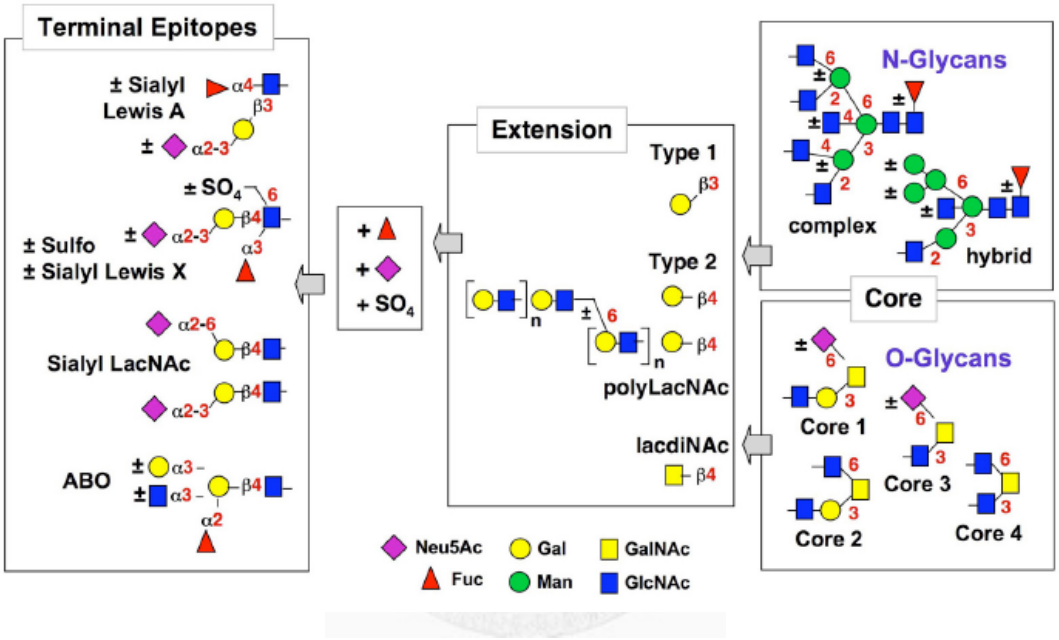



**Figure 1-6. The structural motifs of glycans.** The right panels are common core structures of the mammalian N-glycans and O-glycans. Glycans extended and branched from the core structures are capped with terminal glycosylation modifications.

### 1.4 Occurrence of terminal disialylation

Sialic acids are common constituents of glycoproteins and glycolipids of cell membranes. The sialic acids are a family of 9-carbon carboxylated sugars, containing nearly 50 members that are derivatives of N-acetylneuraminic acid (Neu5Ac), N-glycolylneuraminic acid (Neu5Gc), and deaminoneuraminic acid (2-keto-3-deoxy-D-glycero-D-galacto-nononic acid; KDN) (Angata and Varki, 2002). It is an important non-reducing terminal residue in glycoconjugates and is involved in various biological activities. Sialic acid residues can also form linear homopolymers of α2,8-linked sialic acids that are named according to their DP (degree of polymerization), disialic acid (diSia) for DP=2, oligosialic acid (oligoSia) for 3≤DP≤7 and polysialic acid (PSA) for DP≥8 (Troy, 1996).

Di- and oligo-sialic acid chains are commonly found on gangliosides of b and c series that are known to play important roles in differentiation, signal transduction and cell adhesion (Cheresh et al., 1986; Kasahara et al., 1997; Liu et al., 1997). [More recently, these shorter sialic acid chains were also described on both N- and O-glycoproteins. The first suggestion for expression of α2,8-linked diNeu5Ac residues in rat brain and other tissues was made by Finne *et al* (Finne et al., 1977a; Finne et al., 1977b).] Other than the glycosphingolipids, detailed structural characterization of its carrier glycans has been generally lacking despite an inferred wide occurrence of diSia on mammalian glycoproteins (Sato, 2004; Sato et al., 2000; Yasukawa et al., 2005a). Evidence for a direct attachment of diSia to core 1 *O*-glycans was first obtained on bovine chromogranins (Kiang et al., 1982) and human erythrocyte glycophorins (Fukuda et al., 1987) by MS analysis in the 80s. This motif has since been additionally found on the N-glycans of fetuin and α2-macroglobulin of calf serum, and on the O-glycans of bovine adipo-Q, CD166 of Neuro2A cells (Sato et al., 2002; Sato et al., 2001). Only the disialic acid on CD166 was further shown to be involved in neural differentiation and regulatory pathway of T cell activation.

### 1.4.1 Sialyltransferases responsible for the biosynthesis of disialic acid

The biosynthesis of sialosides is carried out by a family of 20 sialyltransferases that are high conserved from mouse to human (Paulson and Rademacher, 2009). These enzymes transfer sialic acid from a donor substrate, CMP-sialic acid, to terminal positions of the glycan chains. Sialyltransferases can be grouped into four subfamilies based on the linkage formed in the adduct: β-galactoside α2,6-sialyltransferase (ST6Gal), N-acetyllacosaminide α2,6 sialyltransferase (ST6GalNAc), β-galactoside α2,3-sialyltransferase (ST3Gal), and α 2,8-sialyltransferases (ST8Sia). Six members of the ST8Sia family have been described and cloned from various animal species, and also in mouse. However, the role of each enzyme in the biosynthesis of diSia, oligoSia, and polysialic acids *in vivo* remain unclear.

Among the six known α2,8-sialyltransferases (ST8Sia), ST8Sia II and IV are polysialyltransferases catalyzing the polymerization of α2-8 sialic acid residues on N-glycans (Angata and Fukuda, 2003; Angata et al., 1998; Kojima et al., 1995), while the related ST8Sia III is generally considered as a Siaα2-3Galβ1-4GlcNAc α2,8-sialyltransferase acting primarily on N-glycans (Lee et al., 1998; Yoshida et al., 1995). ST8Sia I and V are both involved in the α2-8 sialylation of gangliosides (Kono et al., 1996; Sasaki et al., 1994) leaving ST8Sia VI as the only ST8Sia reported to prefer O-glycans as acceptor substrates (Takashima et al., 2002; Teintenier-Lelievre et al., 2005). Functional implications and correlation between the expression of diSia-glycans and ST8Sia are, however, not well established, not least due to lack of suitable cell lines and model systems. In this context, although the detection

sensitivity of MS-based glycomic analysis is still lower than that of biological probes, it does offer a more global view of the carrier glycans at an unrivaled structural precision.

### 1.4.2 Detection of disialic acids

From the MS perspectives, the occurrence of a terminal disialyl motif on the O-glycans from a variety of cell lines and tissues are increasingly being identified through recent MS-based glycomic profiling (Avril et al., 2006; Canis et al., 2010). In most cases, these are carried on simple core 1 structure in which both the Gal and GalNAc can be disialylated. Occurrence on the N-glycans is, however, less readily identified by MS mapping since it cannot be read off directly from MS mapping alone. Instead, it requires additional MS/MS analysis to determine the distribution of the various sialic acids on typically multi-sialylated complex type N-glycans. Such low level of occurrence renders more confident identification difficult, which leaves several reported cases of incidence by mAb staining unsubstantiated. Among these are components of mouse serum (Yasukawa et al., 2005b) identified as immunoglobulin (Ig) light chain, vitronectin, plasminogen (Yasukawa et al., 2006), and carbonic anhydrase II (Yasukawa et al., 2007), which suggested that putative disialyl modification of serum glycoproteins is biologically important for immunologic events and fibrinolysis.

As noted above, the limiting sample amounts often prevented conventional analysis and identification of di- and oligosialic acids by methods such as mild acid hydrolysis-thin-layer chromatography (TLC), methylation, NMR, or even advanced MS. Instead, several highly sensitive, chemical and biochemical methods have been successfully developed and widely adopted as standard means to confirm the occurrence of tiny amount (femtomole level) of di/oligosialic acid in glycoproteins.

a. Chemical analysis

DMB was originally developed as a fluorogenic reagent for α-keto acids (ulosonic acids) and applied to micro-determination of sialic acids (C9-ulosonic acid, nonulosonic acid) (Hara et al., 1987). Since the reaction occurs only in relatively strong acid, partial hydrolytic cleavage of inter-residue sialyl linkages is unavoidable although the α2,3- and/or α2,6-sialyl linkages to the proximal sugar residues of core glycans are cleaved about 2.5-fold faster than α2-8 linkages (Inoue et al., 2000; Inoue et al., 2001). Under the optimized experimental conditions, DMB labeling of released free oligo/polysialic can be used to detect the presence in glycoconjugates various types of sialyl di/oligo/polymer, which may differ in sialic acid species, inter-residue linkages and DP. For example, fluorescent labeled C7 and C9-compounds after periodate oxidation products with DMB would strongly suggest the presence of

internal sialyl residues or oligomeric structure of α2,8-linked N-acylneuraminic acid therein. To detect the DP, a series of di/oligo/polymers were produced by mild acid hydrolysis, labeled with DMB and separated on anion exchange HPLC for flurometric detection. The detection limit of the disialic acids was 13 fmol (Lin et al., 1999). In addition, these analyses can also be applied to glycoproteins on PVDF membrane.

b. Biochemical analysis

In addition to chemical analyses, antibodies that specifically recognize di/oligo/polysialic acids and endo-/exo-sialidases were also powerful tools for structure and functional analyses. Anti-oligoSia antibodies, 2-4B (Sato et al., 1998b), S2-566, 1E6 (Sato et al., 2000), and AC1, recognize oligosialic acids with DP 2-4. For the recognition of disialic acids, S2-566 which was activated by human GD3 and 1E6 which was activated by $(Neu5Ac)_2$-bearing artificial glycopolymer both could recognize Neu5Ac with DP=2. 2-4B recognizes Neu5Gc with DP ≥ 2.

Endo-sialidase can serve as a specific probe to modify α2,8-linked polysialic acid chains. Endo-N from bacteriophage K1F, endo-NE, and endosialidase requires a minimum chain length of DP>5, DP≥11, and DP≥3, respectively. Exo-sialidase derived from *Arthrobacter ureafaciens*, *Vibrio cholerae*, and *Clostridium perfringens* can cleaves α2,8-linkeages of Neu5Ac and Neu5Gc (Rosenberg, 1995), but not KDN (Kitajima et al., 1994). In conjunction with endo- and exo-sialidase treatments, it can help us to confirm the chain length of given di-, oligo-, and polysialic acid chains although they were not applicable in disialic acids.

**1.5 Specific aims**

Despite recent advances in MS-based glycomics, mapping of the occurrence and fine structural details of polyLacNAc, along with its terminal substituents, remain difficult. Other glycotopes occurring at low level such as the disialyl motif, or the sulfated glycotopes that were refractory to positive ion mode MS screening, are often mis- or under-represented and thus hitherto not reported. This thesis work set out to address some of the outstanding technical issues in advanced glycomic mapping, with an emphasis on optimizing a full range of analytical methods in characterizing the polyLacNAcs and associated glycotopes thus identified on the 2 biological systems used, namely the endothelial cells and B cells.

Specifically, this thesis work aims to:

1) Use the human endothelial cells to develop an efficient and comprehensive analytical workflow that will better characterize the polyLacNAcs based on advanced MS analysis in conjunction with chemical and enzymatic manipulations, as well as lectin based enrichment strategies for glycoproteomics

applications;

For these purposes, EA.hy926 cells, derived from fusion of primary human umbilical vein endothelial cells (HUVEC) (Edgell et al., 1983) with human pulmonary adenocarcinoma A549, were used as the primary source material. This EA.hy926 cell line is a permanent, easily maintained, widely used endothelial cell line, which still maintains many of the characteristics of primary cultured endothelial cells (Edgell et al., 1983; Emeis and Edgell, 1988; Saijonmaa et al., 1991; Suggs et al., 1986). Occurrence of polyLacNAcs has been implicated but direct evidence for EA.hy926 or any other human primary endothelial cells has been generally lacking. Extensive glycomics profiling were undertaken in parallel with a more restricted mapping performed on HUVEC for a direct comparison.

2) Apply the methodologies developed here and elsewhere by others to

i) map the occurrence of polyLacNAcs on mouse and human B cells, along with any other unique glycotopes thus uncovered, including the disialyl motifs and sulfated sialyl LacNAc on the termini of both polyLacNAcs and non-extended LacNAcs; and

ii) in the case of disialyl motifs, to attempt correlating their expression level with the activities of ST8Sia in the process of B cell differentiation

For these investigations, the B cell lymphoma used, BCL1, is the first reported spontaneous B tumor line that arose in a 2-year-old BALB/c mouse (Knapp et al., 1979a; Knapp et al., 1979b; Slavin and Strober, 1978; Strober et al., 1979; Vitetta et al., 1979). The mouse B cells can be divided into two categories : conventional (or B-2) cells and B-1 cells, the latter are distinguished by their different ontogeny, which can be further classified into B-1a and B-1b cells, differ from conventional B cells in that they are generated predominantly during fetal and neonatal development. BCL1 expresses CD5, surface IgM, Mac-1, CD43 and low level of B220, and is likely to have B-1a cell origin. For further mapping of the occurrence of disialyl motifs, primary mouse B cells were isolated from spleen and induced to differentiate by lipopolysaccharides (LPS) or interleukins (IL). In the case of human B cells, resting and activated B cells were isolated from human peripheral blood of a healthy donor.

# Chapter II Material and methods

## 2.1 Cells and culture conditions

HUVEC were purchased from Bioresource Collection & Research Center (Hsinchu, Taiwan; H-UV001). Cell cultures were maintained with 5% $CO_2$, at 37˚C, in M199 medium (GIBCO, Invitrogen Corporation, NY) containing 20% (vol/vol) FBS, 2 mM L-glutamin, 100 U/ml penicillin, 100 μg/ml streptomycin, 50 μg/ml heparin, and 50 μg/ml endothelial cell growth factor (Sigma, St. Louis, MO). Every culture dish was coated with 2% gelatin before use. HUVEC at 3–5 passages were used for all experiments. The EA.hy926 cell line (Edgell et al., 1983), a hybrid of HUVEC and a human lung carcinoma cell line A549, was provided by Dr. Nin Wang from the Institute of Biomedical Sciences, Academia Sinica, Taiwan (Chen et al., 2008). Cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% FBS. BCL1 cells were maintained in RPMI 1640 (Invitrogen) containing 10% FBS (Invitrogen), 50 μM 2-mercaptoethanol. All the above mentioned cell lines were grown in medium containing 100 U/ml penicillin, and 100 μg/ml streptomycin. The flasks were gassed with 5% $CO_2$ at 37°C, and the media were changed every 2-3 days.

Mouse splenic B cells, including all pre-treatment, for all MS and chemical analyses reported in this thesis work were provided by Chih-Ming Tsai from the laboratory of Dr. Kuo-I Lin at Genomics Research center, Academia Sinica, through collaborative arrangement. Briefly, splenic B cells were isolated from 6-8 week-old C57BL/6 mice (Purchased from National Laboratory Animal Center, Taiwan), and purified using B220 microbeads (Miltenyi Biotec). The purity of B cell preparations (purity >95%) was evaluated by FACS analysis on FACSCanto flow cytometer (BD Biosciences) and cultured as described (Tsai et al., 2008). Purified splenic B cells were stimulated with LPS (2ug/ml; Sigma-Aldrich) or IL4 (10ng/ml) + IL5 (20ng/ml)+ CD40L (10ng/ml; Pepro Tech).

Human primary B cells used for all MS analyses reported in this thesis work were provided by Dr. Norihito Kawasaki from the laboratory of Dr. James C. Paulson at The Scripps Research Institute, US, arranged through the Consortium for Functional Glycomics, in lyophilized form. Briefly, human B cells were isolated from human peripheral blood of a healthy donor by negative selection using B cell isolation kit (Miltenyi). 3.0 x $10^7$ of B cells were washed with PBS three times and suspended in 0.1 M ammonium bicarbonate, boiled 10 min, and lyophilized. 1.0 x $10^6$ of Isolated B cells were seeded in 6-well plate in 2 mL of RPMI medium (Invitrogen) supplemented with 10% heat-inactivated FCS, 2 mM glutamine, 100 U/mL penicillin, 100 mg/mL streptomycin, 1 mM non-essential amino acid, 1 mM sodium pyruvate, and 50 mM 2-melcaptoethanol and stimulated with 10 mg/mL of anti-human IgG, A, M (Jackson immunoresearch) and 50 nM CpG2006 (Invitrogen) for 3 days. Cells were harvested

(3.4 x $10^7$ cells recovered) and washed with PBS three times and suspended in 0.1 M ammonium bicarbonate, boiled 10 min, and lyophilized.

**2.2 Release of N-glycans and O-glycans from whole cell lysate**

Harvested EA.hy926 (1x$10^7$), HUVEC (3x$10^6$), and BCL1(1x$10^7$) cells were first delipidated by 1.5 ml of chloroform and methanol in the ratio of 2:1 (v:v) in a sonicator for 30 min. Cell pellet was then spun down by centrifugation at 8000 rpm for 20 min and dried under a stream of nitrogen or by SpeedVac. Delipidated cell pellet was redissolved and incubated overnight in 1 ml of extraction buffer (6M Guanidine-HCl in 0.1 M Tris-HCl, pH 8.6, with 1 mM $CaCl_2$) at 4˚C. The supernatant containing the crude extracts was subjected to reduction by 10 mM dithiothreitol (DTT; Sigma) at 37˚C for 1 h, alkylation by 50 mM iodoacetamide (IAA, Sigma) at 37˚C for 1h, and then dialysed in 6-8 kDa cut-off dialysis bag (Spectrum Laboratories Inc., Rancho Dominguez, CA) against double distilled $H_2O$, with several replacements, before drying down by Speed Vac. The dried sample was digested stepwise with trypsin (Sigma) and chymotrypsin (Sigma) in an approximate ratio of 20:1 (w/w) in 50 mM ammonium bicarbonate buffer, pH 8.5 at 37˚C for 4 h and then boiled for 5 min.

The resulting peptide and glycopeptide mixtures were digested with 3 units of N-glycosidase F (PNGase F, Roche Diagnostics, Mannheim, Germany) at 37˚C for 48 h, followed by passing through a Sep-Pak classic C18 cartridge (Waters, Milford, MA) to separate the released N-glycans from de-N-glycosylated peptides. The Sep-Pak C18 cartridge was conditioned successively by 3 ml washes of 100% acetonitrile (ACN), 100% 1-propanol and 5 ml of 5% acetic acid. Sample was loaded onto the Sep-Pak C18 cartridge and the non-binding N-glycans were collected in the flow through and the first 5 ml of 5% acetic acid wash, whereas the bound peptides were subsequently eluted with 2 ml each of 20, 40, and 60% 1-propanol in 5% acetic acid, pooled and dried by Speed Vac.

Peptide mixtures were washed via milli-Q water and dried by SpeedVac for at least three times. O-glycans was released from these de-N-glycosylated peptide mixtures by reductive elimination in 400 µl of 1M $NaBH_4$ in 0.05M NaOH at 45˚C for 16 h. The reaction was terminated by dropwise addition of 30% acetic acid on ice, followed by passing through a Dowex 50W-X8 (Bio-Rad) column in 5% acetic acid to desalt. After drying down in SpeedVac, borate was removed from the released O-glycan sample by repeated co-evaporation with 10% acetic acid in methanol.

**2.3 Enrichment of polyLacNAc-carrying glycopeptides for glycoproteomic studies**

**2.3.1 First step glycoprotein enrichment from total lysates**

For the two-step enrichment method, polyLacNAc-containing glycoproteins

were first enriched from total cell lysates, followed by tryptic digestion and then subjected to a second step enrichment (Section 2.3.3) at the glycopeptide level. The whole cell lysates of EA.hy926 ($1.5x10^7$) were sonicated and solubilized with buffer A containing 10 mM HEPES, pH 7.3, 150 mM NaCl, 0.1 mM $CaCl_2$, 1.0% Triton X-100, and complete protease inhibitor cocktail (Roche Diagnostics, EDTA-free), at 4˚C for 1 h. Insoluble residue was removed by centrifugation at 13,000 rpm for 20 min. The supernatant was loaded onto a 500 µl LEL-agarose (AL-1173; Vector Laboratories, Burlingame, CA) gel pre-equilibrated with 3 ml of buffer A. The mixture was left to rotate gently for 3 h at 4˚C, and then packed into a column (Bio-Rad, Hercules, CA, 0.8 x 4 cm). The unbound material was removed by washing with 3 ml each of buffer A and B (buffer A containing 1:3000 fold dilution of a commercial preparation of chitin hydrolysates (SP-0090; Vector; available as a 100 mM GlcNAc equivalent stock solution in a nearly saturated NaCl solution). Thereafter, bound glycoproteins were eluted slowly by 3 ml of buffer C (buffer A containing 1:250 fold dilution of chitin hydrolysates), followed by additional steps of elution with higher concentration of the chitin hydrolysates. Each of the collected fractions was reduced with 10 mM DTT at 37˚C for 1h, alkylated with 50 mM IAA at 37˚C for 1 h, followed by addition of 1 volume of trichloroacetic acid stock (T0699; Sigma) to 9 volumes of protein sample to precipitate and concentrate the proteins. Supernatants were discarded and the pellets washed with cold acetone for five times. The dried proteins were then re-suspended with 50 mM ammonium bicarbonate and digested by sequence grade modified porcine trypsin (V511A; Promega) for 48 h at 37˚C with a 1:20 trypsin:protein (w/w) ratio. Initially, the tryptic digest from each of the collected fractions was de-N-glycosylated by PNGase F and the released N-glycans purified by Sep-Pak C18 for rapid MS screening to check for the presence of polyLacNAc-carrying N-glycans, while their respective peptides were subjected to proteomic identification by LC-MS/MS. Only the tryptic digest from the fraction containing most of the polyLacNAc-carrying glycoprotein (eluted with 1:250 fold dilution of chitin hydrolysates) was then subjected to a second round of tomato lectin enrichment at the peptide/glycopeptide level.

### 2.3.2 Isolation of membrane fraction for one step glycopeptide enrichment

For the one-step enrichment method, membrane fraction was first isolated and then digested directly into peptides without the enrichment step at the glycoprotein level. Alternatively, harvested EA.hy926 cells were resuspended in the homogenization buffer (20 mM HEPES, pH 7.3, 1 mM EDTA in 0.25 M sucrose, and complete protease inhibitor cocktail (Roche Diagnostics, EDTA-free) and sonicated on ice for 10 s for 4 times. The lysate was centrifuged at 2000 g for 20 min at 4˚C, and the supernatant was subjected to ultra-centrifugation at 105,000 g for 3 h at 4˚C with a small cushion of 0.3 M buffered sucrose (20 mM HEPES, pH 7.3, 1 mM EDTA in 0.3 M sucrose). After ultra-centrifugation, the pellet containing the membrane fraction

was diluted with 0.1 M $Na_2CO_3$ and maintained on ice for 30 min to strip extrinsic proteins off membranes without affecting the disposition of integral components, and then re-pelleted down again by ultracentrifugation. The membrane proteins were extracted from the resulting pellet using 1 ml of 6 M Guanidine-HCl in 0.1 M Tris-HCl, pH 8.6, with 1 mM $CaCl_2$, at 4˚C for overnight. The solubilized extracts were reduced and alkylated with DTT and IAA, and then digested by sequence grade modified porcine trypsin (Promega) for 48 h at 37˚C with a 1:20 trypsin:protein (w/w) ratio, as described above. The resulting tryptic peptides/glycopeptides were subsequently subjected to one-step enrichment by tomato lectin.

### 2.3.3 Glycopeptide enrichment

The tryptic digests of the lectin-enriched glycoprotein fraction from the total lysates (the sencond round of two-step enrichment) or the total membrane fraction (Secton 2.3.2, for one-step enrichment) were dissolved in buffer D (10 mM HEPES, pH 7.3, 150 mM NaCl, 0.1 mM $CaCl_2$, and complete protease inhibitor cocktail) and co-incubated with 500 µl LEL-agarose gel pre-equilibrated with 3 ml of buffer D at 4˚C for 3 h. The peptides and unbound glycopeptides were washed away with 3 ml of buffer D and buffer E (buffer D containing 1:3000 diluted chitin hydrolysates), while the polyLacNAc-carrying glycopeptides were eluted with 3 ml of buffer F (buffer D containing 1:250 diluted chitin hydrolysates). Each fraction was desalted and cleared of a majority of the chitin hydrolysates by Sep-Pak C18, preconditioned successively with 3 ml of methanol, 100% ACN, 50% ACN with 0.1% formic acid (FA), and 5 ml of 0.1% FA. Each fraction was then loaded onto the Sep-Pak C18 cartridge and washed with 5 ml 0.1% FA. The peptides/glycopeptides fraction was eluted by 2 ml of 25% ACN with 0.1% FA, 50% ACN with 0.1% FA, and 100% ACN, and dried by Speed Vac. Each fraction was then dissolved in 200 µl of ammonium bicarbonate (50 mM, pH 8.5) and incubated with 3 unit of PNGase F for 48 h at 37˚C to release the N-glycans. The de-N-glycosylated peptides recovered from further round of Sep-Pak C18 extraction were taken for protein identification.

## 2.4 Immunoprecipitation of CD45 from BCL1 for chemical analysis

CD45 was immunoprecipitated from the membrane glycoprotein extracts of BCL1 using the Abcam anti-CD45 antibody ab82407. Crude membrane fraction was first isolated from 1 x $10^8$ BCL1 cells by ultracentrifugation and the glycoproteins were extracted with 1 ml of 20 mM Tris-HCl buffer, pH 7.5, containing 137 mM NaCl, 10% glycerol, 1% NP-40, and 2 mM EDTA. Aliquots of the extracts were pre-cleared with protein A/G ultralink resin (55132, Thermo Scientific) for 30 min at 4°C, before adding 2 µg of the anti-CD45 antibody per 200 µl of the membrane glycoprotein extracts and incubated at 4˚C for 2 h. Fresh Protein A/G ultralink resin was then added and the reaction mixtures further incubated for 3 h at 4˚C, after which the

supernatant was removed and the resin washed 5 times with 20 mM Tris-HCl buffer, pH 7.5, containing 137 mM NaCl, 0.1% NP-40, 2mM EDTA. The immunoprecipitates collected were boiled for 5 min in SDS–PAGE sample buffer containing β-mercaptoethanol (2-ME) and run into 8% SDS–PAGE gels. Recovery and purity of immunoprecipitated CD45 was evaluated by protein silver staining and Western blot analysis using a 1:1000 diluted anti-CD45 antibody (1 μg/ml) with 1:3000 diluted goat anti-rabbit secondary antibody (Abcam ab6517, 1 μg/3ml). The sample transferred from gel to PVDF membrane was also used directly for chemical analysis. In separate run, the gel was stained with Sypro-Ruby (Molecular Probes, Eugene, OR, USA) and the three major bands corresponding to CD45 were excised for in-gel trypsin digestion and N-glycan release, followed by proteomic analysis of the de-N-glycosylated peptides using LTQ Orbitrap Velos.

**2.5 Glycosidase digestions and subsequent clean-up or fractionation**

Desialylation was performed by treating the N-glycans with 10 mU of neuraminidase from Arthrobacter ureafaciens (Roche) or 50mU from *Streptococcus pneumoniae* (sialidase S; recombinant, E. coli, Prozyme, San Leandro, CA) in 100 μl of 50 mM sodium acetate buffer, pH 5.5, at 37˚C for 24 h. For endo-β-galactosidase digestion, the N-glycans were digested with 5 mU of endo-β-galactosidase from *E. freundii* (Seikagaku, Tokyo, Japan) in 100 μl of 50 mM sodium acetate buffer, pH 5.8 at 37 °C for 48 h. Higher dosages of enzyme (up to 50 mU) were also investigated initially to ascertain if the internal branched structures released by 5 mU enzyme remained resistant to digestion and that similar MS profiles for the digestion products were reproduced. Endo-β-galactosidase was also applied directly to tryptic peptide digests of BCL1 before releasing the N-glycans. Digested sample was passed through a Sep-pak C18 in 5% acetic acid and the endo-β-galactosidase released glycans were collected in the flow through fraction.

In general, digested sample was desalted by Superclean ENVI-Carb (SUPELCO, Bellefonte, PA), which was pre-conditioned stepwise with 2 ml each of 100% ACN, 75% ACN with 0.1% TFA, 50% ACN with 0.1% TFA, 25% ACN with 0.1% TFA, and $H_2O$. Sample was loaded and washed with 3 ml of $H_2O$, after which N-glycans were eluted with 2 ml each of 25% ACN with 0.1% TFA and 50% ACN with 0.1% TFA, pooled and dried down for analysis. However, the sialidase or endo-β-galactosidase treated sample also could be permethylated directly although without the clean-up steps, samples were more prone to under-methylation.

In the case of endo-β-galactosidase digestion on BCL1, the negatively charged mono- and disialylated terminal fragments were purified away from the neutral, non-sialylated ones by fractionation on an Oasis MAX extraction cartridge (Waters, Milford, MA, USA). The column was preconditioned stepwise with 1 ml of 95% ACN,

2ml of 100 mM NaOAc, 6 ml of $H_2O$, and 6 ml of 95% ACN. Dried glycan sample was redissolved in 500 µl of 95% ACN (50 µl of milli-Q water was added to first dissolve the glycans and then 450 µl of 100% ACN was added slowly to raise the concentration of ACN. The final concentration of ACN should be not lower than 90%.) Sample was loaded in 3 ml of 95% ACN, washed for three times, with the first 2 wash fractions reloaded. Neutral glycans were eluted with 6 ml of 50% ACN, and the negatively charged sialylated fragments by 6 ml of 100 mM sodium acetate fraction. Both fractions were further desalted by ENVI-Carb column.

In addition, Bio-Gel P2 (Fine, particle size rang 45-90 µM; typical fractionation range: 100-1800 Daltons) and P4 (Fine, particle size range 45-90µM; typical fractionation range: 800-4000 Daltons) were also used in this work to size fractionate the endo-β-galactosidase digested N-glycans from BCL1 and desialylated oligosaccharides from EA.hy926, respectively. Both gels were packed in Bio-Rad column (1x 47 cm) and balanced with 100 mM $NH_4HCO_3$. Dried glycan samples were dissolved in 200 µl of 100 mM $NH_4HCO_3$ and loaded into column. Fetuin N-glycans was used as standard to calibrate the approximate eluting time corresponding to N-glycans with specific range of molecular weight or shape. Eluents were collected via fraction collector (20 drops per tube) and the glycans collected in every fraction was validated by permethylation and MS analysis.

**2.6 Mild periodate oxidation and Smith degradation**

Mild periodate oxidation of O-glycans from BCL1 was performed by adding 100 µl of 10mM $NaIO_4$ in 50 mM sodium acetate, pH 5.5 to dried O-glycans and incubated on ice for 30 min in dark. The reaction was stopped by adding 1 µl of 1M ethylene glycol, incubated on ice for a further 40 min, and then dried by Speed Vac. Smith degradation on the N-glycans from EA.hy926 was performed by incubating the sample at more concentrated $NaIO_4$, for longer time at higher temperature. added 100 µl of 0.08M $NaIO_4$ in 50mM sodium acetate, pH 5.5 was added to dried glycans and the reaction mixture was incubated at room temperature in the dark. After 4 h, 5 µl of ethylene glycol was added to stop the reaction. The sample was dried down by Speed Vac after a further incubation of 1 h.

After the periodate oxidation step, the created aldehyde groups were reduced back to hydroxyl groups by adding 200 µl of 10 mg $NaBH_4$ in 1 ml of 2 M $NH_3$ and incubated at room temperature for 2 h before stopping the reaction by adding 30% acetic acid. The reaction products were applied to a column of Dowex 50W-X8, eluted with 5% acetic acid, and dried by Speed Vac. Additional borate was removed by repeated co-evaporation with 10% acetic acid in methanol.

**2.7 Permethylation and microscale fractionation**

Glycan sample dried down in the glass tube was redissolved in a slurry of finely ground NaOH pellets in dimethyl sulfoxide (~200µl), followed by the addition of methyl iodide in equal amounts. For non-sulfated glycans, the reaction mixture was incubated at room temperature for 30 min and extracted by water/chloroform method. Reaction was terminated by adding ~1 ml of milli-Q water drop-wise and then an equal volume of chloroform was added immediately. Permethylated glycans were extracted into the bottom layer (chloroform) and excess reagents as well as other hydrophilic contaminants were removed by repeated extraction with milli-Q water for 3-5 times. The permethylated glycans in chloroform were then dried under a stream of nitrogen gas.

For sulfated glycans, the reaction mixture was gently vortexed for 3 h at 4˚C and then quenched on ice with one drop of cold 10% acetic acid. The neutralized reaction mixtures were applied directly to a pre-washed and equilibrated Sep-Pak C18 cartridge (Waters). Hydrophilic salts and contaminants were step-wise washed off with 5 ml of milli-Q water, 10ml of 5% ACN. Subsequently, permethylated sulfated N-linked glycans were eluted with 2 ml of 25% ACN, 2ml of 50% ACN, and 75% ACN. Under similar conditions, both the sulfated and nonsulfated smaller O-glycans were eluted in earlier fractions.

To further separate the non-sulfated glycans from sulfated glycans, all Sep-Pak C18 fractions screened positive for sulfated glycans by MALDI-MS analysis in negative ion mode were pooled and subjected to fractionation by amine beads (712200.10, NUCLEOSIL 100-5 $NH_2$, 5 µm particle size, 100 Å pore size; MACHEREY-NAGEL GmbH & Co. KG, Duren, Germany) self-packed into pipette tip with the tapered end plugged by filter paper. Depending on the sample quantity, the volume of packed beads can range from as little as 0.5 µl to about 5 µl. If the beads were packed in 20 µl tips, the wash/elution volume would be 20 µl for three times for each step, or 100 to 150 µl for one time for each step if packed in 200 µl tips. The packed amine column was conditioned and washed with 100% MeOH, 100%ACN, 95%ACN/0.1%FA, 50% ACN/0.1% FA, and 95% ACN/0.1% FA sequentially. Permethylated sample was dissolved in 100% ACN for loading. Non-sulfated permethylated glycans such as neutral and sialylated ones were collected in unbound fraction (100% ACN) and an additional wash by 95% ACN. Mono- and di-sulfated fractions were eluted with 50% ACN/2.5mM ammonium acetate, 50% ACN/10mM ammonium acetate, respectively.

Alternatively, permethylated sample in the NaOH/DMSO reaction mixtures was directly loaded into an Oasis MAX extraction cartridge (Waters, Milford, MA, USA), to be desalted and fractionated at the same time. The cartridge was pre-equilibrated with 3 ml of 95% ACN and 3 ml of 100mM $NH_4OAc$, before loading the reaction mixtures (without idomethane) in 1 ml of 100 mM $NH_4OAc$. (The $NH_4OAc$ was loaded to column first, then loaded reaction mixture to column.) The remaining excess

iodomethane was dried under $N_2$ and the glass tube was washed for three times using 2 ml of 10% ACN each time. The washes were pooled and additionally loaded into the cartridge, which was then washed by 1 ml of $NH_4OAc$, followed by 6 ml of milli-Q water to remove as much salts as possible. Subsequently, the permethylated neutral glycans were eluted by 6 ml of 95% ACN, mono-sulfated glycans by 6ml of 1 mM $NH_4OAc$ in 80% ACN, and glycans with more than one sulfate by 3ml of 100mM $NH_4OAc$ in 60% ACN, 20% methanol.

Both permethylated sulfated and non-sulfated glycans can be further cleaned-up or concentrated by $ZipTip_{C18}$ (Millipore), when necessary. The permethylated derivatives were dissolved in 20 µl of 10% ACN (adding 2ul of 100% ACN first, and then 18 µl of milli-Q water to dilute the sample into 10% ACN). $ZipTip_{C18}$ was first conditioned by 50% ACN by pipetting for several times, followed by 10 µl of 0.1% TFA (or 0.1% FA or 1% acetic acid) for three times. The sample was repeatedly taken up in the $ZipTip_{C18}$ for up to 50 times to allow binding. After washing several times with 10 µl of 0.1% TFA, the permethylated glycans were eluted by 10 µl of 75% ACN / 0.1% TFA, or less volume, for collection into microtubes or direct spotting onto the MALDI target plate.

**2.8 Analysis of DMB-sialic acid derivatives**

Conditions for mild acid release of intact di- and oligosialic acids in conjunction with DMB-derivatization and fluorometric HPLC detection were adapted from protocols developed by Sato and Inoue et al (Lin et al., 1999; Sato et al., 1999). The lyophilized cell lysates, released N- and O-glycan samples, or wet PVDF membranes carrying blotted glycoproteins were incubated with 50 µl of 0.02 N TFA and 50 µl of 7 mM DMB solution in 0.02 N TFA containing 1 M 2-mercaptoethanol and 18 mM sodium hydrosulfite, at 50˚C for 2 h, after which the reaction was terminated by adding 20 µl of 1 N NaOH. The resulting DMB derivatives were applied to a Mono Q HR5/5 (0.5 x 5 cm) anion-exchange column fitted on an Agilent HP 1100 HPLC system and eluted by a NaCl gradient in 20 mM Tris-HCl, pH 8.0, starting at 10 min increasing linearly from 0 to 0.3 M in 20 min and then 0.3 to 1 M in the next 20 min. The flow rate was 0.5 ml/min and the wavelengths for the Agilent HP1100 fluorescence detector were set at 373 nm and 448 nm for excitation and emission, respectively.

For the quantification of monosialic acids in cells, O-glycans or N-glycans, sample was first acid hydrolyzed in 100 µl of 0.1 N TFA at 80˚C for 1 h and then subjected to DMB labeling. The resulting monomeric DMB-sialic acid derivatives were separated isocratically on a C18 reverse phase HPLC column (250 × 4.6 mm, 5 µm, Vydac) by a solvent mixture of methanol/ACN/water (7:9:84), at 1.0 ml/min at a column temperature of 26˚C (Sato et al., 1998a). The elution time of Neu5Ac or Neu5Gc can be inferred from the elution positions of standard Neu5Ac and Neu5Gc derivatives.

**2.9 GC-MS methylation analysis**

For GC-EI-MS linkage analysis, permethylated glycans were hydrolyzed in 200 µl of 2N TFA and incubated at 120˚C for 2 h, after which the hydrolyzed sample was dried under a stream of nitrogen gas. 200 µl of NaBD4 (10-20 mg per ml of equal mix of 2N $NH_3$ and pure ethanol) was then added and the reaction mixtures incubated at room temperature for 2 h to reduce the monosaccharides to alditols. The reaction was stopped by addding 2 to 3 drops of acetic acid, followed by drying down under nitrogen and removal of excess borate by evaporation with 10% acetic acid in methanol. Peracetylation was performed by addition of 200 µl of acetic anhydride and incubate at 100˚C for 1 h, followed by drying down under nitrogen. The partially methylated alditol acetates were dissolved in chlorofrom and washed with water for 4 times, and the chlorofrom fraction was dried down under nitrogen. GC-EI-MS was carried out using a Hewlett-Packard Gas Chromotograph 6890 conncected to HP 5973 Mass Selective Detector. Smple was dissolved in hexane before spolitless injection into an HP-5MS fused silica capillary column (30 m x 0.25 mm internal diameter, Hewlett-Packard). The column head pressure was maintained at ~56.6 kPa to give a constant flow rate of 1 ml/min using helium as carrier gas. The initial oven temperature was held at 60˚C for 1 min, increased to 90˚C in 1 min, and then to 290˚C in 25 min at a rate of 8˚C /min.

**2.10 MALDI-MS and MS/MS analysis**

The permethylated sampled in 100% ACN were mixed 1:1 with DHB (10 mg/ml in 50% ACN) or 1:1 with DABP matrix (10 mg/ml in 50% ACN without acids) (Acros Organics, NJ) for spotting onto the target plate for positive- and negative-ion mode analyses, respectively. MALDI-TOF/TOF MS of permethylated glycans were performed on a 4700 Proteomic Analyzer (Applied Biosystems, Framingham, MA), operated in the reflectron mode. Each MS data acquired normally comprised a total of 16 sub-spectra of 125 laser shots, and laser energy was set 4500 for both matrices.

MALDI MS/MS sequencing of the permethylated glycans was performed on either Q-TOF Ultima$^{TM}$ MALDI (Micromass, Manchester, UK) in positive ion mode or 4700 Proteomics Analyzer in reflectron positive and negative ion mode. In MALDI TOF/TOF, the potential difference between the source acceleration voltage and the collision cell was set at 3 kV to obtain the desirable high-energy CID fragmentation in positive ion mode, and at 1 kV in negative ion mode for permethylated sulfated glycans. The number of shots was accumulated until satisfactory signal to noise ratio was achieved but usually not exceeding 4000 shots. Two or more spectra can be combined in post-acquisition processing with mass tolerance set at 10 ppm to improve S/N ratio. In Q-TOF, argon was used as the collision gas with collision energy manually adjusted (between 50-200 V) to achieve optimum degree of fragmentation for the parent ions

under investigation.

**2.11 NanoESI-MS and Total ion mapping analysis**

For additional nanoESI-MS and MS/MS analyses on an LTQ-Orbitrap XL hybrid FT mass spectrometer (Thermo Scientific, San Jose, CA), the permethylated glycans in 50% ACN /1 mM NaOH were infused directly by static nanospray for data acquisition using either the linear ion trap or Orbitrap mass analyzer. Using the total ion mapping (TIM) functionality provided by the XCalibur instrument control software package (version 2.0.7), automated MS/MS spectra (at 35% collision energy) were obtained across the entire mass range m/z 500–2000 with a parent ion stepping of 2 mass units, in collection windows that were 2.8 mass units in isolation width (±1.4 amu). Two microscans, each 1000 ms, were averaged for each collection window. The 0.8 mass unit overlap from step to step was set to allow minor parent ion signals occurring at the edge of an individual window to be sampled in a more representative fashion, as introduced and described initially by Aoki et al (Aoki et al., 2007). For data analysis, the Qual Browser (Thermo Scientific) was used in ion-map mode to filter out the parent ions containing specific terminal epitope, either by product ion or neutral loss sorting, to create pseudo-precursor ion or neutral loss scan profiles. These low resolution reconstructed TIM mass profiles were then matched against the high resolution, high accuracy MS profile obtained with the Orbitrap to facilitate manual peak assignment. False positives were manually eliminated by further examination of the respective ion trap MS2 data.

**2.12 LC-MS/MS shotgun proteomics analysis**

For the LEL-enriched de-N-glycosylated peptide samples from EA.hy926, nano-LC-nano-ESI-MS/MS analysis performed on a Paradigm MS4 nanoflow system (Michrom BioResources, Auburn, CA), connected to an LTQ-Orbitrap XL hybrid FT mass spectrometer (Thermo Scientific) equipped with a PicoView nanospray interface (New Objective, Woburn, MA). Peptide mixtures were loaded onto a 75 µm × 250 mm fused silica capillary column packed in house with C18 resin (5 µm, Nucleosil 120-5 C18, Macherey-Nagel GmbH & Co. KG., Dueren, Germany) and were separated using a 45 min linear gradient from 5 to 65% solvent B (95% ACN with 0.1% FA) at a flow rate of 200 nl/min. Solvent A was 0.1% FA in $H_2O$. The 10 most intense ions were sequentially isolated for CID MS/MS fragmentation and detection in the linear ion trap (automatic gain control target value at 7000), with previously selected ions dynamically excluded for 90 s. Ions with single and unrecognized charge states were also excluded. All the measurements in the Orbitrap were performed with the lock mass option for internal calibration.

For the immunoprecipitated CD45 de-N-glycosylated peptides mixtures, nano-LC-nano-ESI-MS/MS analysis was performed on a nanoACQUITY system (Waters,

Milford, MA) coupled to an LTQ-Orbitrap Velos hybrid mass spectrometer (Thermo Scientific) equipped with a PicoView nanospray interface (New Objective). Peptide mixtures were loaded onto a 75-µm × 250-mm nanoACQUITY UPLC BEH130 column packed with C18 resin (Waters, Milford USA) and were separated at a flow rate of 300 nl/min using a linear gradient of 5 to 40% solvent B (95% acetonitrile with 0.1% formic acid) in 30 min, followed by a sharp increase to 85% B in 1 min and held at 85% B for another 10 min. Solvent A was 0.1% formic acid in water. The 20 most intense peptide ions with charge states ≥2 were sequentially isolated to a target value of 5,000 and fragmented in the high-pressure linear ion trap by low-energy CID with normalized collision energy of 35%. Ion selection threshold was 500 counts for MS/MS, and the maximum allowed ion accumulation times were 500 ms for full scans and 100 ms for CID-MS/MS measurements in the LTQ. An activation q = 0.25 and activation time of 10 ms were used. Standard mass spectrometric conditions for all experiments were: spray voltage, 1.8 kV; no sheath and auxiliary gas flow; heated capillary temperature, 200 °C; predictive automatic gain control (AGC) enabled, and an S-lens RF level of 50%. Data processing and search criteria were similar to those described above for the LEL-enriched glycopeptide samples.

The mass spectrometer was operated in the data-dependent mode. Briefly, survey full-scan MS spectra were acquired in the Orbitrap (m/z 350–1600) with the resolution set to 60,000 at m/z 400 and automatic gain control target value at 106. The MS and MS/MS raw data were processed using the Extract_Msn module in Bioworks 3.3.1 (molecular weight range, 600–6000; grouping tolerance set to 0; precursor charge set to auto; minimum scans per group; and intermediate scans set to 1) and searched against an in-house generated SwissProt database comprising human ORFs in their forward and reversed orientations (sprot_56.8 (410518 sequences; 148080998 residues)) or mouse ORFs (Sprot 57.14 (514,789 sequences; 181,163,771 residues) for CD45 using an in house Mascot Daemon 2.2 server. Search criteria used were trypsin digestion, variable modifications set as carbamidomethyl (Cys), oxidation (Met), and deamindated (NQ), allowance of up to one missed cleavage, and mass accuracy of 10 ppm on the parent ion and 0.60 Da on the fragment ions.

In the identification results in lectin enrichment, all significant protein hits from Mascot ($p < 0.05$, Peptide ion score > 19 for sample from one-step enrichment of the total membrane fraction, or >27 for sample from two-step enrichment of the total lysates) thus obtained contained no false positive hit from the reverse database. Gene ontology categories for the identified proteins were assigned by using the GOMINER bioinformatics resources and used to classify the proteins according to their localization.

**2.13 RNA isolation, RT-Quantitative (Q)PCR, and knockdown of ST8sia VI**

Total RNA isolation, cDNA synthesis and subsequent RT-QPCR analysis of the cDNA in an ABI/Prism 7300 sequence detection system (Applied Biosystems) was performed as described previously (Tsai et al., 2008). The SYBR Green primers used here were:

*ST8sia I*, 5’-GTGGTCCTCTGTTGGCTC-3’ and 5’-GAGATGGGCAGGGTCA- C-3’;
*ST8sia II*, 5’-ACTGGATCAGAATCAGAACCC -3' and 5’-GCCGACAGTC- AGTTTCAATG-3’;
*ST8sia III*, 5’-GTGGAATCTTGACAGGGAGT-3’ and 5’-GTT- GTTACGGTCCTGAATGG-3’;
*ST8sia IV*, 5’-CAAGTGCGAACTGCCTAT-3’ and 5’-CTTTTCCATTCAGATCCTTGGG-3’;
*ST8sia V*, 5’-TCATAATTGACAGGTTCCACAAGTT-3’ and
5’-TCGTCCAGCATGTATTTGACTC-3’;
*ST8sia VI*, 5’-TCA- CAAACATACAGAGATGCCC-3’ and 5'-TCTTGCTTTCCACCTCGTAG-3’.

The RNA interference (RNAi) lentiviral vector pLKO.1 expressing short hairpin RNAs (shRNAs) targeting ST8sia VI (TRCN0000110385), and luciferase for control (targeting sequence, 5’-CCTAAGGTTAAGTCGCCCTCG-3’) were obtained from the National RNAi Core Facility, Taiwan.

# Chapter III Results

## Part I. PolyLacNAcs on Endothelial cells

Human endothelial cell culture systems are indispensable for a range of biological studies. Due to many inherent problems associated with the use of primary cells including time-consuming isolation, short-term viability and differences in genetic background of the isolates, immortalized cell lines were widely used instead as experimental models. One of the most commonly used primary endothelial cells, the HUVEC, was implicated to display polyLacNAcs through positive binding with galectins (Nangia-Makker et al., 2000; Yang et al., 2007) and the presence of a series of N-glycans containing many $Hex_1HexNAc_1$ units in our preliminary MS-based glycomic screen. Moreover, TNFα was shown to induce profound changes in the expression of a set of glycosyltransferases including β4GalT1/5 and β3GnT2 involved in polyLacNAc biosynthesis, along with increased DSA staining pattern (Garcia-Vallejo et al., 2006). EA.hy926 cells were derived from HUVEC by fusion with a human pulmonary adenocarcinoma A549 and have been shown to best retain the characteristics of endothelial cells among other HUVEC derivatives (Emeis and Edgell, 1988; Saijonmaa et al., 1991; Thomas et al., 1997; Unger et al., 2002). It is therefore highly relevant to ascertain if this cell line indeed affords a glycomic expression of polyLacNAcs similar to that of HUVEC. It also represents a more readily obtained biological source material for developing and optimizing the enabling analytical techniques.

### 3.1 Direct glycomic profiling of glycans with polyLacNAc

#### 3.1.1 MALDI-based MS analysis of HUVEC and EAhy926

MALDI-MS profiling of permethylated glycans is a well established and widely adopted analytical method for glycomics (North et al., 2009; Zaia, 2008). Peak assignment is straightforward since the molecular ions detected are normally monosodiated and thus singly charged. Occurrence of polyLacNAc-carrying N-glycans can be easily recognized from the regularity of multiple LacNAc unit increments at higher mass range. At first glance, the MS profile afforded by the N-glycans of EA.hy926 resembled that of HUVEC in carrying a full range of heterogeneity but clearly distinguishable from the latter by a lower degree of additional fucosylation and sialylation (Fig. 3-1A and B). When mono-fucosylated, the single Fuc was primarily located on the GlcNAc at the reducing end, which could be readily established by MS/MS analysis (data not shown). Thus, for the N-glycans of EA.hy926, the prominent $[M+Na]^+$ molecular ions at m/z 2244, 2693, 3142, 3591, 4041 and 4490 corresponded to fucosylated trimannosyl core with 2-7 LacNAcs, which were not detected in the spectrum afforded by the N-glycans of HUVEC. It should be noted

that substitution of 2 LacNAc units (898 u) with 2 Neu5Ac and 1 Fuc (896 u) differs by 2 mass units, which can be easily resolved at lower mass range with high resolution data acquired in the reflectron mode but difficult to discriminate at higher mass range for the low resolution data acquired in the linear mode. It was nonetheless apparent that the higher degree of sialylation and fucosylation shown by the HUVEC N-glycans was equally reflected at higher mass range (Fig. 3-1C and D), in which the most abundant peaks for HUVEC were those sialylated with 3-4 Neu5Ac ± additional Fuc, whereas for EA.hy926, the more abundant ones were those with 0-2 Neu5Ac. Molecular species carrying non-sialylated $LacNAc_{5-10}$ were only detected in the EA.hy926 sample and not HUVEC, which afforded instead peaks at 2 u lower.

Even when sample amount was not severely limiting such as that for the EA.hy296 cells, decent molecular ion signals extended only as far up to core fucosylated N-glycans with 18 LacNAcs or so. In comparison, using 3x less cells, the largest N-glycan detected for HUVEC corresponded to one with 8 LacNAcs and 4 sialylation (Fig. 3-1C). Most MS analyses of biological samples reported to date similarly detected up to about 20 LacNAcs or so for permethylated N-glycans without sialic acids or additional peripheral Fuc, approaching an m/z value around 10,000 where the signal intensities tailed off (Bateman et al., 2010; North et al., 2010). It is therefore not practical to infer the maximum number of LacNAc units carried by such analysis, which is highly dependent on the state of sample preparation, amount loaded, and sensitivity of the MS instrument. Perhaps more importantly, despite an informative MALDI-MS profile, definitive structural assignment particularly in resolving ambiguities associated with isobaric and isomeric peaks requires extensive MS/MS analysis, often further aided by desialylation to reduce heterogeneity. A common problem though is insufficient sensitivity and precursor ion resolution for productive MALDI-MS/MS on each of the observed peaks, especially those of low intensity at higher mass range. In this context, a readily automated, more comprehensive glycomic mapping at MS2 level is highly desirable.

### 3.1.2 nanoESI-based Total ion mapping (TIM) at MS2 level

Instead of relying on the stochastic nature of data dependent acquisition during an LC-MS/MS analysis, a data independent process of TIM appeared to be better suited to survey the full complexity of a glycome that is often populated by similar glycan structures poorly resolved by conventional LC system. The rapid scan rate of a linear ion trap system and the facile neutral loss of terminal sialic acids have been successfully utilized to afford a high sensitivity nanoESI-based TIM aiming at identifying minor sialylated components otherwise not obvious in any MS profiling (Aoki et al., 2007). Briefly, MS2 spectra were acquired across the entire useful mass range in overlapping steps of every 2 mass units. Pseudo precursor ion or neutral loss scan profiles can then be constructed based on production of specific fragment ions.

In the case of demonstrating the presence of polyLacNAc, previous work has utilized the TIM functionality to obtain mass profiles filtered for the presence of a singly charged non-reducing terminal diLacNAc fragment ion (Abbott et al., 2008a). However, the intensity of such terminal fragment ion is often low in comparison with ions resulting from neutral loss, which leads to many false positives when the sample amount is limiting. Furthermore, the charge state of the precursors yielding this fragment ion cannot be directly inferred from the low resolution ion trap MS data.

As an alternative, we have investigated the possibilities of obtaining a full range profiling of polyLacNAc-carrying N-glycans based on the ubiquitous neutral loss of terminal LacNAc units. As shown in Fig. 3-2, a permethylated, desialylated N-glycan sample from EA.hy926 afforded a much simpler MALDI-MS profile when being rid of heterogeneity contributed by sialylation. Small degree of additional peripheral fucosylation can also be readily identified. In contrast, a typical offline nanoESI-MS profile of non-LC-separated sample only afforded clear doubly charged signals for the N-glycans with 2-5 LacNAc units in a spectrum dominated by the high mannose structures. The triply charged N-glycans with 6-7 LacNAc units were less apparent and it was not possible to confidently identify molecular ion signals for larger N-glycans. However, if TIM data was acquired instead and the resulting MS profiles plotted for neutral loss of LacNAc unit at different m/z values, the triply and quadruply charged N-glycans with 4-10 and 7-14 LacNAc units could be clearly filtered out, respectively, each accompanied by species carrying an additional Fuc. It was observed that, in general, the neutral loss of a LacNAc did not alter the charge state of the parents, which facilitated their direct identification. The nanoESI-MS TIM data is therefore superior not only to a direct nanoESI-MS profile, but also to a MALDI-MS profile as it offers better signal-to-noise identification of larger components, each already supported by characteristic MS2 neutral loss.

The specificity of the pseudo neutral loss and precursor ion scan profiles was demonstrated in examples filtering for the presence of peripheral fucosylation. As above, plotting for neutral loss of a LacNAc unit from doubly charged parent ions led to clear signals of core fucosylated N-glycans with 2-5 LacNAc units, each accompanied by signals carrying additional Fuc (Fig. 3-3A). When plotting instead for parents that either gave a singly charged terminal $Fuc_1LacNAc_1$ product ion or afforded a neutral loss of a terminal $Fuc_1LacNAc_1$, the signals at m/z 1134, 1360, 1584, 1808 were no longer observed, confirming that when monofucosylated, the Fuc was preferentially at the core. On the other hand, signals were detected for N-glycans with 2 and 3 LacNAcs carrying up to 2 and 3 additional peripheral Fuc, respectively. The true MS2 spectra for the parent ions can be further examined as these were already acquired during TIM, albeit with a wider than usual precursor ion isolation window width. A virtual superimposition of all neutral loss and product ion profiles clearly afforded a much more enriched mapping data set than that possible by a

direct, single dimensional MALDI- or nanoESI-MS mapping. In addition, any peak of interest can be further subjected to $MS^n$ analysis during the long lasting offline nanoESI-MS acquisition time. Unfortunately, despite being a viable alternative to MALDI-MS mapping, information on polyLacNAc branching and initiation site on these large N-glycans is still equally difficult to obtain.

### 3.2 Mapping the fine structural details of polyLacNAc

Except for very few cases when sample amount were less limiting (Sutton-Smith et al., 2007), most productive MS/MS sequencing to date have been applied to permethylated N-glycans with fewer LacNAcs (< 5000 Da) (Babu et al., 2009; North et al., 2010; Pang et al., 2009; Takamatsu et al., 2010). For larger glycans such as those derived from cell wall polysaccharides, selective enzymatic or chemical cleavage to smaller fragments of a size more amenable to detailed structural analyses has always been the common practice.

#### 3.2.1 Applications of endo-β-galactosidase digestion

Endo-β-galactosidase from *E. freundii* has been in use since the early days when the structural details of several archetypal linear and branched polyLacNAc-carrying N-glycans were first determined (Carlsson and Fukuda, 1990; Fukuda et al., 1985; Fukuda et al., 1984a; Fukuda et al., 1984b; Fukuda et al., 1988; Fukuda et al., 1984c; Fukuda and Matsumura, 1976; Spooncer et al., 1984). With an established specificity for cleaving the β4-galactosyl bond of an internal -3Galβ1-4GlcNAc-, where the non-reducing end must carry at least another LacNAc unit and the 4-linked GlcNAc not further fucosylated at position 3, three distinct sets of endo-β-galactosidase digestion products are expected from polyLacNAc-carrying N-glycans:

1. The unsubstituted linear polyLacNAc chain itself will be degraded to the expected disaccharide unit of GlcNAcβ1-3Gal, whereas larger internal oligosaccharide units carrying additional fucosylation will be produced, if present. In this context, branched polyLacNAc chains will be resistant to digestion at the branched point 3,6-linked Gal unless subjected to a much higher than usual dosage of enzyme (Fukuda et al., 1984b).
2. Fragments corresponding to the non-reducing terminal epitopes of a polyLacNAc chain in the general formula of R-GlcNAcβ1-3Gal, where R can be a further substituted or non-extended Gal, will be obtained.
3. The trimannosyl core with one or more single GlcNAc residues retained on each of the polyLacNAc-antennary initiation points of the 3- and 6-arm α-Man, will remain, along with any non-extended LacNAc and endo-β-galactosidase resistant chains either due to branching or further substitution at the LacNAc proximal to the core.

Applied at glycomic level, simultaneous detection and additional detailed sequencing of each of these products will collectively inform about the chemical nature of polyLacNAc initiation, branching and termination. Specifically, any polyLacNAc extended N-glycans would have been clipped to smaller sizes to allow more effective MS/MS analyses. By locating the distribution of exposed terminal GlcNAc versus intact LacNAc unit, one can infer if the polyLacNAcs are attached to one or more positions.

As demonstrated with application to the EA.hy926 N-glycans, a 48 h digestion with 5 mU of endo-β-galactosidase yielded a profile virtually identical to that produced by using 50 mU of enzyme under the same experimental conditions (Fig. 3-4). Apart from the expected non-reducing terminal products of Gal-GlcNAc-Gal at m/z 722 and its fucosylated and sialylated counterparts at m/z 896 and 1083, respectively, three products detected could be attributed to internal branch point structures, ±Galβ1-4GlcNAcβ1-3(±Galβ1-4GlcNAcβ1-6)Galβ1-4GlcNAcβ1-3Gal, at m/z 1212, 1416 and 1620 (Fig. 3-4A). For each of these branched structures, MALDI-MS/MS by high energy CID unambiguously showed a lack of linear structural isomers that could be attributed to incomplete digestion (Fig. 3-5A). As exemplified by the $Hex_4HexNAc_3$ structure at m/z 1620, reducing terminal $^{1,5}X$ ions due to loss of non-reducing terminal Hex and $Hex_1HexNAc_1$ were observed at m/z 1430 and 1185, respectively, but the next $^{1,5}X$ ion in series was found at m/z 532, corresponding to loss of $Hex_3HexNAc_2$ and not 3 or 4 glycosyl residues, as would otherwise be expected for a linear structure. Likewise, prominent non-reducing terminal E ion series was detected at m/z 211 ($Hex_1$) and 415 ($Hex_1HexNAc_1$), then leaped to m/z 1109 ($Hex_3HexNAc_2$) and 1313 ($Hex_3HexNAc_3$), without the corresponding ions in between. Importantly, the D ion at m/z 676, in conjunction with the $^{3,5}A$ ion at m/z 574, firmly established the branch point motif. These and the corresponding fragment ion assignments for the structure at m/z 1416 lacking one non-reducing terminal Gal were schematically illustrated in the drawings accompanying Fig. 3-5A. Collectively, the results from MS/MS analyses, including that for the structure at m/z 1212 lacking both terminal Gal (data not shown), indicated that further polyLacNAc extension could occur at either or both arms of a distal branch point Gal. Curiously, these branched products were not recovered from similar digestion and analysis of the HUVEC N-glycans, which indicated the lack of polyLacNAc branching following an internal linear chain in the latter sample.

At the mass range of ~ m/z 1700 to 4000, the MALDI-MS profile of the endo-β-galactosidase digested EA.hy926 (Fig. 3-4B) or HUVEC N-glycans closely resembled those of the non-treated samples with respect to the major peaks, which indicated that most of the N-glycans carrying 2-4 LacNAcs were respectively bi, tri and tetra-antennary with single non-extended LacNAc at each antenna. However, closer examination revealed that each of these peaks was additionally accompanied

by counterparts lacking one or more terminal Gal (Fig. 3-4B), not previously present in the original spectrum (Fig. 3-1B). For example, the non-sialylated biantennary structures lacking either 1 or both terminal Gal were detected at m/z 2040 and 1835, respectively. Likewise, additional non-sialylated tetraantennary structures lacking 1-4 terminal Gal were observed at m/z 2938, 2734, 2530 and 2326, respectively. These digestion products clearly showed that polyLacNAc extension could occur on as many antennae available. Further MS/MS analyses (Fig. 3-5B) demonstrated that, in fact, when only 1 or 2 antennae were extended out of the tetraantennary structures, the locations of polyLacNAc chains were not preferentially restricted to the 6-arm either, as indicated by the presence of multiple D ions at m/z 921, 1125 and 1329.

All structures carrying 5 or more LacNAc units were no longer observed or greatly reduced in signal intensities, which was consistent with a rather complete digestion. In view of the prohibitively costly enzyme, the use of excessive dosage (250 mU/200 µl as employed by Fukuda et al (Fukuda et al., 1984b) was deemed unnecessary. Instead, we have shown here that up to 50 mU of active enzyme in 100 µl reaction solution could efficiently digest the polyLacNAc chains, leaving the branch point intact for it to be detected and thus identified. A caveat though is that if the branch point is at the first LacNAc attached to the trimannosyl core, it will not be duly digested. A small amount of such structures were indeed detected at higher mass range (Fig. 3-4C), where the MS profile was significantly different from that of the non-treated sample. Unlike the original EA.hy926 sample, only up to 12 LacNAcs extension could be detected, which corresponded in number to having a minimal branched tri-LacNAcs on each of the tetraantennae although many other permutations were possible. In the case of HUVEC, only trace amount of N-glycans with up to 6 LacNAcs extension were barely detected. Unfortunately, the extremely low amount of these structures prevented further MS/MS to more accurately assign the structures. The usefulness of the endo-β-galactosidase digestion in conjunction with MS analyses to probe the fine structures of polylactosaminoglycans was nevertheless amply demonstrated here.

### 3.2.2 Applications of Smith degradation after further size fractionation

Although analyzing the remaining trimannosyl core structures after endo-β-galactosidase digestions allows one to determine if one or more of the antennae are originally extended by polyLacNAcs, several limitations remain. First, fucosylation at GlcNAc of the inner LacNAc units will make it resistant to digestion. Secondly, all linear polyLacNAcs, irrespective of their initial attachment sites, will be digested into disaccharide units and therefore cannot be detected intact to allow determination of its length. The method also cannot distinguish which of the released LacNAc units were originally assembled on which antennae. Ideally, one would like to be able to chemically release the polyLacNAc chains intact for

subsequent characterization.

Smith degradation selectively degrade a polysaccharide to smaller oligosaccharides, from which structure information can be deduced. The procedure involves three steps, oxidation with periodate ($IO_4^-$), reduction to a polyalcohol with borohydride ($BH_4^-$) followed by hydrolysis with dilute acid under mild conditions, normally at room temperature. Vicinyl hydroxyl groups in sugars can be oxidized by periodate to aldehydes with cleavage of the carbon-carbon bond between them. The beauty of this method when applied to polyLacNAc containing N-glycans is that the entire polyLacNAc chain, being of -3Gal-4GlcNAc- linkages, is resistant to periodate oxidative cleavage. On the other hand, all the terminal glycosyl residues will be trimmed away. At the trimmanosyl core, the respective cleavages will be dependent on whether the 3-arm or 6-arm Mannose is singly (2-linked), doubly (2,6- or 2,4-linked), or triply (2,4,6-linked) substituted. Thus, oligosaccharides corresponding to GlcNAc-[Gal-GlcNAc]$_n$-X will be the expected Smith degradation products, in which X can be a remnant fragment of the cleaved Man residue, or the intact Man attached to the β-Man-GlcNAc-GlcNAc in cases of it being 2,4- or 2,4,6-substituted (Fig. 3-6).

To evaluate if MS can be applied to efficiently identify the Smith degradation products, the whole N-glycan mixtures from EA.hy926 were first reacted. However, it was found that since the smaller size N-glycans, such as the high mannose and complex type N-glycans with less than 4 LacNAcs, typically are much more abundant than the larger N-glycans with polyLacNAc, the resultant MS profile is highly complicated (data not shown). As expected (Fig. 3-6), the high mannose type N-glycans gave a Manα1-6Manβ1-4GlcNAcβ1-4GlcNAc-itol (m/z=983) (Fig 3-7A), whereas complex type bi- and triantennary N-glycans in which single LacNAc were extended from C2 of α1-3 Man and C2 and C6 of α1-6 Man gave a Manβ1-4GlcNAcβ1-4GlcNAc-itol (m/z=779), and tetraantennary complex type N-glycans gave a GlcNAcβ1-2 (GlcNAcβ1-4)Manα1-3Manβ1-4GlcNAcβ1-4GlcNAc-itol (m/z=1473) (Fig 3-7B). The signal intensity of these components was very high and may have largely suppressed other minor but important signals derived from the released polyLacNAc fragments.

To reduce such interference and to better control the chemistry, the N-glycans were first de-sialylated and then fractionated by P4 gel filtration into several parts (Fig. 3-8). F1 contained mostly N-glycans with more than 3 LacNAcs and had no high mannose structures. F2 and F3 contained mostly N-glycans with 2-4 LacNAcs and high mannose, respectively. Taking the F1 fraction through the Smith degradation procedure, a much simplified MS spectrum was obtained (Fig. 3-8B), with most of the signals below m/z 1500 further analyzed by MS/MS. It was found that signals at m/z 851, 1300 and 1749 corresponded to the expected polyLacNAc chain cleaved off from the C6 of Man, namely GlcNAc-[Gal-GlcNAc]n-X, with n=1-3, respectively. The linear

stretch of the LacNAc units carried within was confirmed by the series of X ions at m/z 416, 620, 865, and 1069 afforded by MS/MS analysis, as shown in Fig. 3-9. It therefore indicated that at least 4 LacNAcs could be extended linearly from C6 Man. In agreement with the results from endo-β-galactosidase digestion described above (Section 3.2.1), which identified the presence of branched polyLacNAc, signals corresponding to structures carrying an extra HexNAc above m/z 851 and 1300 were also detected (Fig. 3-7B). Finally, LacNAc extensions from the GlcNAcβ1-2 (GlcNAcβ1-4)Manα1-3Manβ1-4GlcNAcβ1-4GlcNAc-itol (m/z=1473) were present at m/z 1677 and 1923. However, the expected fragment released from C2 of α1-6 Man was not identified and there were still many peaks that could not be assigned (therefore not labeled).

### 3.3 Enrichment of polyLacNAc-carrying glycopeptides for glycoproteomics

An efficient way to enrich for polyLacNAc-carrying glycopeptides is a prerequisite to identify the protein carriers by means of MS-based glycoproteomic analysis. Lack of specific monoclonal antibodies implies that the use of lectin is the only option available for affinity capture. Current literature suggests that the more commonly used DSA can also recognize N-glycans with only 2 LacNAc units arranged in a specific manner including the non-polyLacNAc-extended tri- and tetraantennary structures (Cummings and Kornfeld, 1984; Yamashita et al., 1987), which are highly abundant in most glycomic samples. In contrast, the LEL was reported to interact with high affinity only with N-glycopeptides containing 3 or more linear LacNAc (Lee et al., 1990; Merkle and Cummings, 1987) and thus appears to be more specific and better suited for enrichment of polyLacNAc-carrying N-glycopeptides. Previous studies have relied mostly on using defined -4GlcNAcβ- oligomers such as tri-N-acetylchitotrioses and tetra-N-acetylchitotetraoses, for elution (Atrih et al., 2005; Callaghan et al., 1990; Callaghan et al., 1992; Fukuda et al., 1989; Merkle and Cummings, 1987) while a recent work has used instead 0.1 M glycine/2% acetic acid-HCl buffer for elution after affinity capture at the protein level (Jung et al., 2009). The specificity of the retained glycoproteins and the degree of non-specific bindings were, however, not reported.

#### 3.3.1 Specificity of tomato lectin at the glycan level

To investigate the binding specificity for polyLacNAcs of LEL, the released N-glycans from EA.hy926 were first taken through the lectin affinity capture step to see if the expected specificity can be obtained. Instead of eluting by chitotriose, chitotetraose, or chitin hydrolysates as suggested by the vendor (Vector Laboratories), 0.1M Gly-HCl, pH 2.5 was used to elute the bound glycans to avoid the difficulty associated with removing these massive amounts of glycans such as chitin hydrolysates from the targeted glycans. At the glycan level, only N-glycans containing

more than 5 LacNAcs (Fig. 3-10B) and up to 10 LacNAcs could be captured by tomato lectin although they were also present in the unbound fraction (Fig. 3-10A). In addition, it seems that LEL has the higher affinity toward non-siaylated polyLacNAc with one or more Fuc. Most glycans in the bound fraction could only have more than 1 Neu5Ac in addition to LacNAc and Fuc.

### 3.3.2 One-step and two-step enrichment of glycopeptides with polyLacNAc

Taking into consideration both the cost and highly desirable enrichment specificity, the use of a commercially available preparation of chitin hydrolysates at different dilution levels was systematically investigated as eluents for LEL affinity captured glycoproteins prepared from the total lysates of EA.hy926. Using the 1:20 fold dilution (or even higher concentration) of the chitin hydrolysates recommended by the vendor resulted in identification of mostly non-specific, high affinity binding proteins such as interleukin enhancer-binding factor, ribosomal proteins and histones (data not shown). In contrast, glycoproteins containing polyLacNAcs could be eluted by more diluted (1:250 fold) chitin hydrolysates, as monitored by MALDI-MS screening of the released N-glycans from the respective fractions eluted by increasing concentration of the chitin hydrolysates. This concentration of the chitin hydrolysates was thus adopted for eluting the enriched glycoproteins and subsequently found to be also suitable for eluting the enriched glycopeptides.

The inability to directly identify the glycopeptide itself by direct LC-MS/MS sequencing has long been one of the major limitations in glycoproteomics, particularly so for N-glycopeptides carrying large polyLacNAcs with a full complexity and heterogeneity of sialylation. While enrichment at protein level followed by direct identification of the enriched glycoproteins based on LC-MS/MS analysis of their proteolytic digests has been commonly in use, this method suffers from inability to distinguish non-specific bindings since identification is invariably based on the majority non-glycosylated peptides. Even if de-N-glycosylated peptides can be reliably identified, this pool would comprise glycopeptides not carrying polyLacNAcs, which are derived from other glycosylated sites of the captured, polyLacNAc-containing glycoproteins. It therefore cannot furnish site-specific glycosylation information and mere identification of the proteins does not reliably provide true positives. Two possible workflows were therefore evaluated. The first was to start from a total tryptic digest pool of a purified membrane fraction, which was itself enriched with glycoproteins, and to effect a one-step LEL-enrichment at the glycopeptide level. The alternative approach was to start from a crude total lysates and to perform two-step enrichment at first the protein level, followed by tryptic digestion and then a second enrichment at the peptide level. The resulting glycopeptide pools from both methods were then de-N-glycosylated and subjected

to a simple one dimensional LC-MS/MS analysis to evaluate the enrichment efficiency based on identified proteins.

Proteomic analysis of the de-N-glycosylated tryptic peptides produced from the one-step enriched fraction resulted in only 61% of the total identified proteins being membrane proteins and only 26% of all identified peptides carried the consensus N-glycosylation sites with the required Asn to Asp conversion (Fig. 3-11A). In contrast, using the same LC-MS/MS conditions for the de-N-glycosylated tryptic peptides from two-step enriched fraction, 83% of the identified proteins were membrane proteins by GOMINER classification, with 77% of all confidently identified peptides carried the consensus N-glycosylation sites. Considered from a slightly different perspective, there were more membrane proteins being identified by one-step enrichment of the total membrane fraction in terms of absolute number, as compared with that resulting from two-step enrichment (Fig. 3-11B). However, only 26 out of the 91 identified membrane proteins (28%) by one-step enrichment actually contained de-N-glycosylated sites to give it some level of confidence that these truly represent proteins with glycosylation site(s) carrying N-glycans extended with polyLacNAc. In contrast, two rounds of LEL-enrichment of the total lysates resulted in less number of membrane proteins identified but 46 out of the 52 (88%) identified carried the de-N-glycosylated sites (Table 3-1). In other words, when both occurrence of de-N-glycosylated sites and classification as membrane proteins were taken as the minimal criteria for considering true positives, the two-step enrichment process actually returned more candidate glycoproteins with sites carrying polyLacNAc-extended N-glycans.

To further evaluate the enrichment specificity, the released N-glycans from the LEL-non-binding and bound glycopeptide fractions were taken through permethylation and profiled by MALDI-MS (Fig. 3-12). As expected, the non-binding fraction from the second step of the two-step enrichment process was indeed largely devoid of polyLacNAc-carrying N-glycans. The major N-glycans detected were restricted to those containing 4 or less LacNAc units, with no obvious larger N-glycans being detected at higher mass range in the linear mode (Fig. 3-12A). N-glycans with 5 or more LacNAc units were only found in the MALDI-MS profile of the sample derived from the bound fraction (Fig. 3-12B), and the difference was particularly striking in linear mode (Fig. 3-12C). It could be concluded that LEL did enrich for N-glycans with polyLacNAcs although it did not completely exclude all non-carriers especially those with 4 LacNAcs, which comprised both tri- and tetra- antennary structures with mostly non-extended LacNAc, as well as biantennary ones with extended polyLacNAcs. A comparative analysis for one step enrichment at peptide/glycopeptide level for the total membrane extracts indicated that it did not fare better than the two-step enrichment and returned more non-targeted glycopeptides such as those carrying high mannose structures by the same glycan

level screening (data not shown).

We have thus developed a low cost, highly efficient LEL-based affinity enrichment method for rapid identification of candidate membrane glycoproteins carrying polyLacNAcs. From EA.hy926, well known polyLacNAc carriers such as LAMP-1 and LAMP-2 were confidently identified, along with many other candidates (Table 3-1), including basigin/CD147/EMMPIN, which has previously been shown to contain MGAT5-catalyzed, β1,6-branched, polyLacNAc-type N-glycans in HUVEC (Tang et al., 2004). There is currently no report on occurrence of polyLacNAc on all other membrane glycoproteins identified here. However, CD98 (4F2) heavy chain was previously shown to interact with basigin (Xu and Hemler, 2005), whereas the epithelial membrane protein-1 (EMP-1) was suggested to cross-talk with the EGFR signaling pathway (Jain et al., 2005). Furthermore, ICAM-2 is involved in endothelial survival and migration through hemophilic interactions and angiogenesis (Huang et al., 2005). All these surface molecules may thus collectively present a multivalent polyLacNAc lattice engaging molecules such as Galectin-3 (Brewer et al., 2002; Garner and Baum, 2008; Nangia-Makker et al., 2000), which can induce endothelial cell morphogenesis and angiogenesis.

**Part II. PolyLacNAcs on B cells**

In addition to endothelial cells, polyLacNAc extension is also widely implicated in the metastasis of cancer cells and immune cells. It has been reported that galectin-1 binds stronger to mature B cells than plasma B cells (Tsai et al., 2008). Since galectin-1 is known to have specificity toward polyLacNAc, we have attempted to apply the developed analytical strategies to map the occurrence of polyLacNAcs on B cells, in collaboration with Dr. Kuo-I Lin, starting with a murine B lymphoma—BCL1, which is more readily available than either mouse or human primary B cells. This work has since led to unexpected identification of not only polyLacNAcs, but also sulfated glycans and disialyl epitopes in BCL1, the characterization of which is reported in this Result section. It also allowed us to use BCL1 as a model to study the functional relationship between the sialyltransferase responsible for the expression of disialic acids and B cell differentiation, in addition to polyLacNAcs.

**3.4 Identification of polyLacNAcs and terminal disialyl motif on BCL1**

An initial screening by MALDI-MS showed that both N- and O-glycans of BCL1 were sialylated with Neu5Gc and not Neu5Ac. It was obvious that the N-glycans carried polyLacNAcs, were mostly core-fucosylated, and capped with α-Gal or sialic acid. Subsequently, the N-glycans was digested with endo-β-galactosidase to better define the structural details of polyLacNAc. In contrast, the O-glycans were relatively simple, corresponding to core 1 structures without polyLacNAc extension, but unusually decorated with up to four Neu5Gc.

### 3.4.1 High abundance of terminal disialyl motif on the O-glycans

As introduced in Section 1.4.2, the occurrence of terminal disialyl motifs on simple core 1 O-glycans are readily revealed by MS-based glycomic analysis. In principle, a disialylated core 1 may correspond to one with a disialyl motif on either arm, or one sialic acid each on both arms. This requires further MS/MS to confirm. However, in the case of a Hex-HexNAc-itol composition with 3 and 4 sialic acids, the presence of terminal disialyl motif is self evident. For BCL1, a simple Core 1 decorated with up to 4 Neu5Gc is highly abundant (Fig 3-13A). The presence of terminal disialyl is established by normal CID MS/MS on the permethylated glycans, since a consecutive loss of 2 sialic acids is apparent without creating an extra OH group, as would be the case if the 2 sialic acids are distributed on 2 different sites. For the $Neu5Gc_4Hex_1HexNAc$-itol, the y ions at m/z 1694, and 1302, coupled with B ion at m/z 819, are among the critical fragment ions (Fig 3-13B and C). Switching to high energy CID, the usual array of additional cleavage ions (Yu et al., 2006) are present (Fig 3-13C). Among these are the highly abundant, characteristic ion at m/z 777, confirming the presence of disialyl, and the G ion at m/z 1254, due to loss of the disialyl substituent from the 3-arm. The results indicated that one of the two disialic acids was extended from the C3 of Gal in O-glycans with core 1 structure. Similar range of ions (not shown) established that in the case of m/z 1707 ($Neu5Gc_3Hex_1HexNAc$-itol), disialyl on either arm is present and thus represent a mix of structural isomers. On the other hand, for m/z 1316 ($Neu5Gc_2Hex_1HexNAc$-itol), there is only a minor signal of fragment ion 777 in the MS/MS spectrum. It suggested that the structure represented by m/z 1316 was mostly a common core 1 structure with monosialylation on both arms although a small amount of structural isomer containing a disialyl motif is also evidently present.

In order to determine whether the disialyl motif was overall preferably located on C6 of GalNAc or C3 of GalNAc, as well as the linkage between Neu5Gc, the native O-glycans in BCL1 were subjected to mild periodate oxidation. It was expected that the O-glycans in the form of oligo-glycosyl alditol after reductive elimination would be cleaved into two series of products, distinguishable by retaining either the 2 or 4 carbon remnants from the cleaved GalNAc-itol (Fig. 3-14A). The original glycan chains attached to the C3 (3-arm) and C6 (6-arm) of GalNAc-itol are thus designated accordingly as R-C4 and R-C2, carrying a mass increment of 204 and 75 u for the C4 and C2 end, respectively, after a further reduction followed by permethylation. At the same time, the adjacent OH groups on C7, C8 and C9 on the side chain of Neu5Gc would also be oxidized to aldehyde group, resulting in the loss of the C8-C9 fragment after the chemical conversion (Fig. 3-14A). If the disialyl motif was linked via α2-9 linkage, the C7-C8 bond would be oxidatively cleaved, thereby releasing the terminal NeuGc. There would be no remaining disialyl substituted fragments to be detected in the MS spectrum. On the other hand, if the disialyl motif were linked via α2-8 linkage,

only the terminal Neu5Gc would be oxidized and clipped at its side chain.

Taking into consideration the changes in mass as a result of oxidative cleavages, the products afforded 3 major molecular ions at m/z 749, 807 and 1140, which could be assigned as NeuGc-Gal-C4, $NeuGc_2$-C2, and $NeuGc_2$-Gal-C4, respectively (Fig. 3-14B). Interestingly, a minor molecular ion was detected at m/z 1531, which indicated that a trisialyl extension was probably also present at the 3-arm. More importantly, the data clearly supported the presence of NeuGcα2-8NeuGc disialyl motif that can be attached to both C6 of the reducing end GalNAc and the Gal on C3 of the GalNAc, with the 6-arm substituent marginally more abundant based on the respective peak intensity. It does not, however, rule out the presence of α2-9 linked disialyl since the expected product would be indistinguishable from NeuGc-Gal-C4 or NeuGc-C2 that could originate from monosialylated arms.

### 3.4.2 DMB-sialic acid analysis on the total O- and N-glycan pools and CD45

Although a direct MS profiling of the O-glycans revealed only the presence of disialyl motif, with possibly a very small amount of trisialylated motif, it is possible that their low abundance, heterogeneity, large size and poor ionization prevent the presence of polysialyl chain from being detected by MS. A more sensitive method was therefore employed to additionally confirm the findings. Chemical release of the poly- and oligo-sialyl chain followed by fluorescent labeling with DMB for quantitative mapping by fluorometric HPLC is a well established technique commonly employed for sialylation analysis (Lin et al., 1999; Sato et al., 1999). When applied to the O-glycans, it was found that extension beyond disialyl unit is negligible on a comparative basis, relative to the abundance of disialyl conjugates (Fig 3-15A). Unexpectedly, such analysis, when extended to N-glycans, clearly showed that disialic acids were not only carried on the O-glycans of BCL1, but also on the N-glycans. On a semi-quantitative basis, the relative amount of detectable disialyl unit on N and O-glycans, as revealed by DMB analysis, are in fact comparable despite its being non-apparent from the initial MALDI-MS mapping profile of the N-glycans. This prompted us to investigate the N-glycans in more details later (see Section 3.4.3).

Since the O-glycans on CD45 has been widely implicated in the glycobiology of lymphocytes in mouse and human (Amano et al., 2003; Dang et al., 1996; Earl et al., 2010; Garcia et al., 2005; Ma et al., 2009), the DMB-sialic acid analysis was further extended to immunoprecipitated CD45 from BCL1. Silver staining and western blotting (Fig. 3-15B) suggested that there are at least three obvious isoforms of CD45 around 220kDa. In-gel digestion followed by proteomic analysis of the extracted, de-N-glycosylated peptide mixtures suggested that all three of these bands are indeed mouse CD45. All the CD45 immunoprecipitated from 1 x $10^8$ cell was then

loaded in one 8% SDS-PAGE and the resulting protein bands transferred to PVDF membrane. DMB labeling was performed directly on excised PVDF membrane pieces corresponding to the location of CD45 by protein staining. The HPLC data clearly indicated that disialyl motif was present in CD45 (Fig. 3-15B) even after de-N-glycosylation (data not shown), without any significant level of further extension. In contrast, no DMB-disialyl unit was detected from the released N-glycan pools (data not shown), suggesting that the disialyl motif was mostly located on the O-glycans of CD45.

### 3.4.3 Terminal Disialyl motif on both polyLacNAc extended and non-extended N-glycans

As noted from the onset, the initial N-glycan profiling showed the usual presence of high mannose and complex type N-glycans (Fig. 3-16A), with obvious polyLacNAc extension and terminal Gal-Gal epitope in addition to terminal NeuGc-sialylation. Due to the multiantennary nature of the complex type N-glycans, and the possible presence of disialylated antenna in the form of internal α2-6 monosialylation, the presence of a terminal disialyl motif similar to the one observed for the O-glycans could not be readily inferred from the MS profile. Nonetheless, molecular ions at m/z 3230 and 3417 were the smallest N-glycans with a glycosyl composition that apparently carried an excess of Neu5Gc for what appeared to be biantennary structures. MS/MS analysis on these 2 potential candidates that may carry disialyl termini indeed afforded the characteristic fragment ion at m/z 777 (Fig. 3-16B), although the overall spectra quality is not high due to their low abundance.

Seeking further evidence for their presence on larger N-glycans, endo-β-galactosidase was used to digest the N-glycans since the overall glycomic mapping revealed a strong presence of polyLacNAc (Fig 3-16A). Interestingly, endo-β-galactosidase treatment gave rise to the terminal disialyl-LacNAc-Gal, which was identified along with non-sialylated, monosialylated and α-Gal capped terminal units, corresponding to the signals at m/z 1504, 722, 1113, and 926, respectively (Fig. 3-17A). This indicated that the disialyl termini could be found on non-extended LacNAc unit, as well as on polyLacNAc chains. The abundance of ion at m/z 518, which corresponded to the released internal HexNAc-Hex unit from polyLacNAc repeats, along with the absence of the branched internal units similar to those derived from EA.hy926 (see Section 3.2.1), indicated that the polyLacNAcs of BCL1 were mostly linear. For the remaining N-glycan core structures, signals at m/z 1835, 2081, and 2326, which corresponded to fucosylated trimannosyl core with additional 2-4 HexNAcs, indicated that polyLacNAc could be extended from all available antennae in bi-, tri-, and tetra-antennary N-glycans. Other N-glycan signals assigned to core structures with various degree of LacNAc extension and terminal sialylation as annotated in Fig. 3-17A, such as m/z 2530, 2676, 2734, 2921, 3125, 3271, 3331,

and 3518, suggested that polyLacNAc could also be extended from only one, two, or three antennae. Overall, for these multiantennary N-glycans with one or more polyLacNAc chain extension, there was a slight preference for 6 arm extension although not exclusively since as many as 4 extension on all 4 antennae could be detected. It also appeared that digestion by endo-β-galactosidase was rather complete compared to the results from EA.hy926, suggesting that most N-glycans contain polyLacNAcs and supported the conclusion that there was no significant degree of polyLacNAc branching.

Further MS/MS analysis on (Fig. 3-17B) the released terminal disialyl LacNAc-Gal unit (m/z 1504) not only confirmed its identity but also indicated that the disialyl epitope on the terminus of polyLacNAc was extended from NeuGc α2-3 linked to Gal, based on the G ion at m/z 660. Similar MS/MS analysis of the Neu5Gc-LacNAc-Gal (m/z 1113) also showed that, in the case of monosialylated termini, the Neu5Gc was likewise attached to Gal via α2-3 linkage (data not shown). Additional confirmation was obtained by treating the endo-β-galactosidase released, sialylated terminal fragments with α2,3-specific sialidase S. To distinguish the desialylated products from non-sialylated starting materials, the sialylated LacNAc-Gal fragments released directly from total tryptic peptide pool of BCL1 were first isolated via 2 step purification, namely a reverse phase Sep-Pak C18 to get rid of all remaining peptides/glycopeptides, followed by an ion exchange step on Oasis-Max cartridge to remove the non-sialylated LacNAc-Gal fragments. Unfortunately after an addition step of reduction, only the monosialylated and not the disialylated LacNAc-Gal-itol was recovered and detected by MS at m/z 1129 (Fig. 3-18). This was almost completely digested by sialidase S, shifting the peak to m/z 738, corresponding to non-sialylated LacNAc-Gal-itol. It therefore supported the MS/MS data in establishing that the majority, if not all, of the NeuGc monosialylation on the termini of polyLacNAc was via α2-3 linkage. The trace remaining of the non-digested peak at m/z 1129 may be attributed to incomplete digestion by sialidase S, which was known to be more efficient against NeuAc (Chokhawala et al., 2007), although it remains possible that a tiny fraction of α2-6 sialylation existed.

However, if all of the permethylated N-glycan pool was taken through GC-EI-MS linkage analysis, not only 3-Gal (17.213 min) but also 6-Gal (17.712 min) was identifed (Fig. 3-19). 3-linked Gal could be produced from internal Gal in polyLacNAc, α3-Gal capped Gal, α2,3-Neu5Gc capped Gal, whereas 6-linked Gal could only be derived from Neu5Gc α2-6 Gal. The results therefore suggested that Neu5Gc at the termini of polyLacNAc were almost exclusively α2-3 linked, from which an additional α2,8-linked NeuGc can be attached to give the NeuGcα2-8NeuGc disialyl motif. However, on non-extended LacNAc termini, both α2-3 and α2-6 linked Neu5Gc were present. It is likely that the disialyl motif on such non-extended LacNAc termini was also preferentially α2,3-linked to the Gal but evidence is currently lacking due to

limited amount of sample to perform a more thorough analysis.

### 3.5 Sialyltransferases responsible for the biosynthesis of disialic acids

This Section of the thesis work was performed in close collaboration with Dr. Kuo-I Lin and Chih-Ming Tsai from Genomic Research Center, Academia Sinica. The generation of knockdown clones, induced B cell differentiation, and data related to ST8Sia expression were provided by Chih-Ming Tsai.

#### 3.5.1 ST8Sia in BCL1

The ready identification of disialyl motif on both N- and O-glycans of BCL1 provided a good experimental model system to investigate the ST8Sia responsible for its biosynthesis. As introduced in Section 1.4.1, six members of the ST8Sia have been cloned from mouse and characterized. Among them, ST8Sia III and VI are widely accepted as the ST8Sia responsible for disialylation on N- and O-glycan respectively. However, when the expression of the six ST8Sia were examined by RT-PCR, only ST8Sia II, IV, and VI were found to be significantly expressed in BCL1 (Fig. 3-20A). This made ST8Sia VI as the most likely candidate responsible for the expression of disialyl on both N- and O-glycans. To ascertain this functional attribution, short hairpin (sh) RNA specifically against mouse ST8sia VI was generated to further examine whether depletion of ST8sia VI decreases the expression of disialyl motifs in BCL1.

By knocking down the expression level of ST8Sia VI to approx 10% level of its original level by siRNA silencing (Fig. 3-20B), there was no significant difference in the amount of disialyl O-glycans that can be detected by MALDI-MS analysis between the control (Ctrli) and knockdown (ST86i) samples (data not shown). For the N-glycans, the 2 molecular ions already shown to carry disialyl termini, namely signals at m/z 3230 and 3417, both decreased in signal intensity relative to the monosialylated counterparts. For a better quantitative assessment, DMB-sialic acid analysis by fluorometric HPLC was additionally performed. Taking the amount of detectable disialic acids in Ctrli as 100%, the amount of disialyl motifs in ST86i was found to decrease to 54% in O-glycans and to 26% in N-glycans (Fig. 3-20B). Contrary to expectation based on known specificity and activity of ST8Sia VI *in vitro*, disialyl expression on the N-glycans thus appeared to be more affected than that on the O-glycans.

#### 3.5.2 ST8Sia VI expression and disialic acid content during B cell differentiation

Finally, we investigated whether the elevated expression of ST8Sia VI and disialyl content may be correlated with differentiation of primary B cells and therefore may play an important role in B cell differentiation. Primary B cells was stimulated by LPS-induced T-independent (TI) or IL4/IL5/CD40L-induced T-dependent (TD) pathway

to examine the expression of ST8sia II, ST8sia IV and ST8sia VI in $B220^{hi}CD138^{-}$ (activated mature B cells) or $B220^{low}CD138^{+}$ (plasma cells) cell population. It was found that the expression of ST8sia VI was highly increased in both $B220^{low}CD138^{+}$ plasma cell populations by TI and TD stimulation. The expression of ST8sia II and ST8sia IV were also slightly elevated by TI stimulation but registered no significant difference by TD stimulation (Fig. 3-21A, B). Consistent with the increased expression of ST8sia VI, the disialyl content was found to also increase by DMB-sialic acid analysis (Fig. 3-21C, D). A comparative analysis of the crude disialyl content of primary B cells showed that its amount increased from almost negligible level in naive B cells, to a significantly higher level in mature and then plasma B cells.

### 3.6. Sulfated glycans in B cells

As noted in the Introduction, a unique advantage of analyzing permethylated glycans as opposed to native, underivatized iones, provided proper subsequent clean up and fractionation steps were taken, is that the presence of sulfated glycans can be readily screened in negative ion mode, as an integrated part of all MS-based glycomic analysis. This additional step was duly performed in the analysis of BCL1, which led to identification of sulfated N-glycans. Since α2,6-sialylated 6-sulfo-LacNAc has recently been implicated to be present on human B cells and constitute a relevant, preferred ligand of human CD22 (Kimura et al., 2007), we have thus extended the work to activated human B cells in an attempt to specifically identify its presence.

#### 3.6.1 Sulfated N-glycans in BCL1

Initial screening in negative ion mode showed that relevant signals that could be assigned to sulfated glycans were given by the N-glycans (Fig. 3-22) but not O-glycans. This was extended to the endo-β-galactosidase digested N-glycan sample, which additionally identified the presence of sulfated glycotopes on the termini of polyLacNAcs (Fig. 3-23A). Signals assigned as terminal mono- and disialyl sulfo-LacNAc-Gal epitopes at m/z 1155 and 1546, respectively, were further confirmed by MS/MS analyses (Fig. 3-23B). The X ion at m/z 574 and G ion at m/z 498 indicated that sulfate was located at the C6 of GlcNAc. Additional G ion at m/z 702 was indicative of the Neu5Gc being linked via α2-3 linkage. This is the first time a sulfated disialyl LacNAc was reported. In addition, since a high proportion of the remaining trimannosyl core, with only additional HexNAc but not Hex, also carried sulfate, the data suggested that polyLacNAc can also be extended from a sulfated GlcNAc proximal to the core.

#### 3.6.2 Sulfated N- and O-glycans in activated human B cells

Unlike BCL1, the activated human cell glycome comprises a significant amount of complex O-glycans in addition to N-glycans. In positive ion mode, its glycomic

profiles on non-sulfated glycans were found to be similar to those published by others (cf CFG website, url: http://www.functionalglycomics.org/glycomics/common/jsp/samples/searchSample.jsp?5=Human&15=&10=&9=B-Cells&operation=refine&templateKey=2&12=CellType&submit=submit ) and hence not further pursued. As expected, all sialylation detected was restricted to NeuAc instead of the predominant NeuGc found in mouse BCL1. Notably, in the O-glycans, NeuAc-NeuAc disialylated species could be detected at m/z 1617 although clearly, unlike BCL1, these were very minor components.

Focusing on mapping the sulfated glycans in negative ion mode, the major $[M-H]^-$ molecular ions detected could be assigned to the usual biantennary, triantennary, and tetraantennary N-glycans with or without sialylation and fucosylation, as annotated in Fig. 3-24. Apart from the expected change of Neu5Gc to NeuAc and the lack of extra α3-Gal capping, another difference from the corresponding sulfated N-glycan profile of BCL1 was the more obvious presence of extra Fuc in additional to the one assigned as core fucosylation. After digestion with sialidase S, which is specific to the α2,3-linked sialic acids, the intensity of peaks assigned as N-glycans with higher degree of sialylation (eg. 3 sialic acids for those with 4 or more lacNAc units) was significantly reduced relative to those having less sialylation (eg. 2 sialic acids). This indicated that a certain proportion of the terminal sialylation was of α2,3-linkage. However, overall, there was a substantial amount of those sialylated components that retained one or more of the sialic acids after sialidase S treatment, which must be attributed to α2,6-sialylation, consistent with the presence of the sought after α2,6-sialylated 6-sulfo-LacNAc epitope.

Similarly for the O-glycans, which could be assigned as sulfated Core 1 and 2 structures with sialylation and/or fucosylation (Fig 3-25A), sialidase S digestion shifted most of the signals assigned as di-sialylated sulfated O-glycans to O-glycans with one sialic acid or no sialylation (Fig 3-25B). The data was thus in agreement with the NeuAc directly attached to the 3-arm Gal-GalNAc-itol being mostly α2,3-sialylated as reported before (Gillespie et al., 1992). On the other hand, the second NeuAc on the 6 arm was mostly attached by α2,6-linkage to the extended polyLacNAc sequence (up to 6 LacNAcs detected). In summary, both N- and O-glycans appeared to carry the α2,6-sialylated 6-sulfo-LacNAc epitope. For the N-glycans, it may be largely restricted to non-extended LacNAc, whereas the sulfated LacNAc epitope on the termini of polyLacNAc may be mostly α2-3 sialylated. On the O-glycans, the α2,6-sialylated 6-sulfo-LacNAc epitope may be carried mostly on the 6-arm of Core 2 structures with and without polyLacNAc extension, whereas those on the non-extended 3-arm appeared to be α2-3 sialylated. These tentative conclusions obviously need to be further substantiated by MS/MS and other chemical analysis, which are, however, beyond the scope of this thesis work and will require substantially more sample amount.

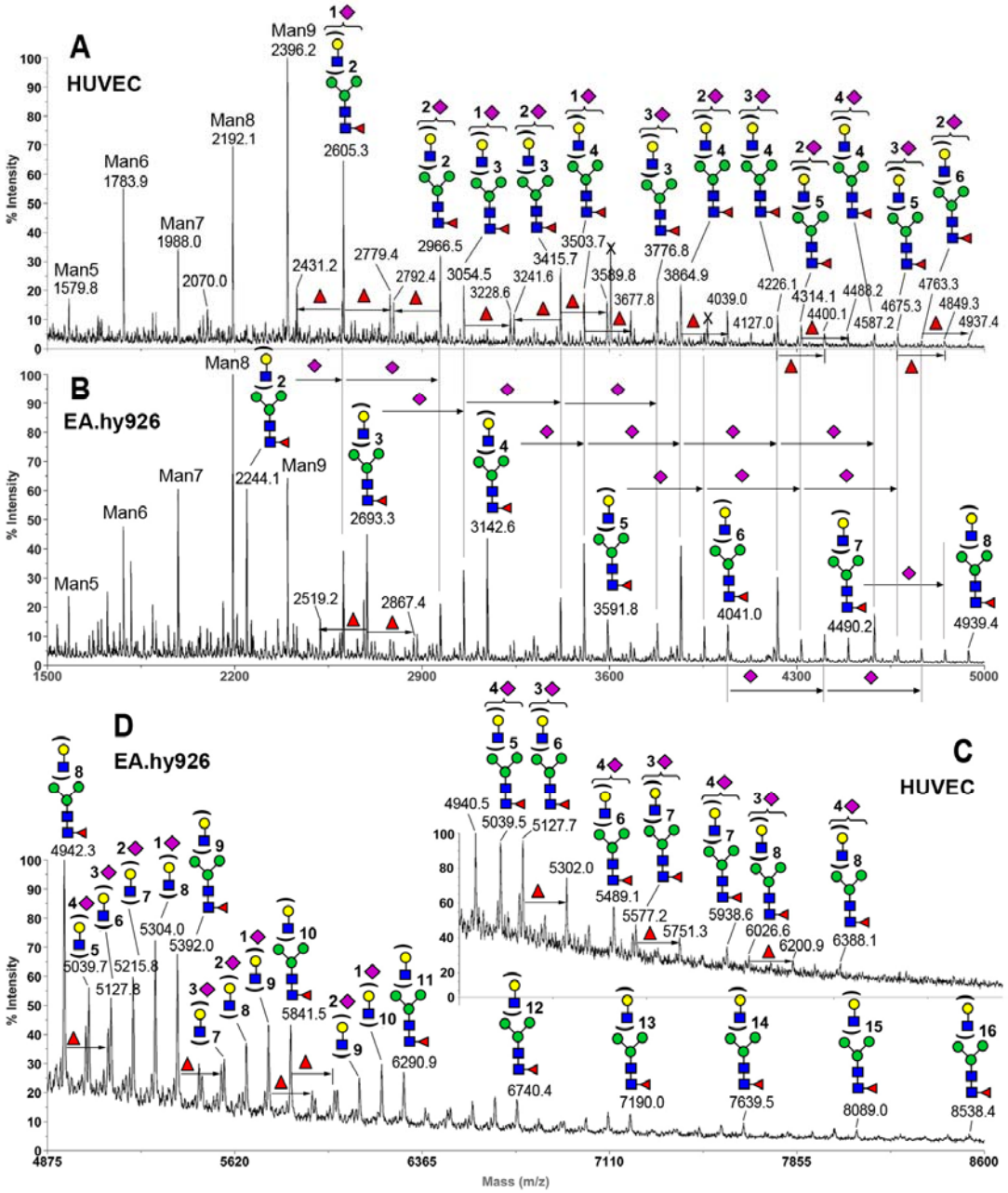



**Figure 3-1. MALDI-MS profiles of permethylated N-glycans released from HUVEC (A, C) and EA.hy926 (B, D), in reflectron (A, B) and linear modes (C, D) for lower and higher mass ranges, respectively.** The glycosyl compositions of the major $[M+Na]^+$ molecular ion signals could be assigned as annotated, assuming all sharing a common fucosylated trimannosyl core, except for the high mannose structures, labeled as Man5-9. Non-fucosylated counterparts and those with additional fucosylation could also be identified and selected few of the more abundant ones were annotated as Fuc increments. Likewise for the varying degrees of sialylation observed. For (D), peaks in between the non-sialylated glycans that were not annotated showed the same mass difference relationship as those annotated in between the species with $LacNAc_9$, $LacNAc_{10}$ and $LacNAc_{11}$, namely several degrees of a Neu5Ac substituting for a LacNAc (-88 u), making up for a full range of differential sialylation. The m/z values

of the molecular ion signals were labeled as monoisotopic mass for spectra acquired in reflectron mode, whereas for the higher mass regions acquired in linear mode, the annotated m/z values correspond to non-resolved peak top, often the average masses. The symbols used for cartoonish annotation of select signals conform to the standard representation recommended by the Consortium for Functional Glycomics. In MS terms, circle represents Hex, square for HexNAc, triangle for Fuc and diamond filled in purple for Neu5Ac. This representation was also used for subsequent figures. Peaks marked with X in (A) are well known contaminating glycan peaks derived from culture media.

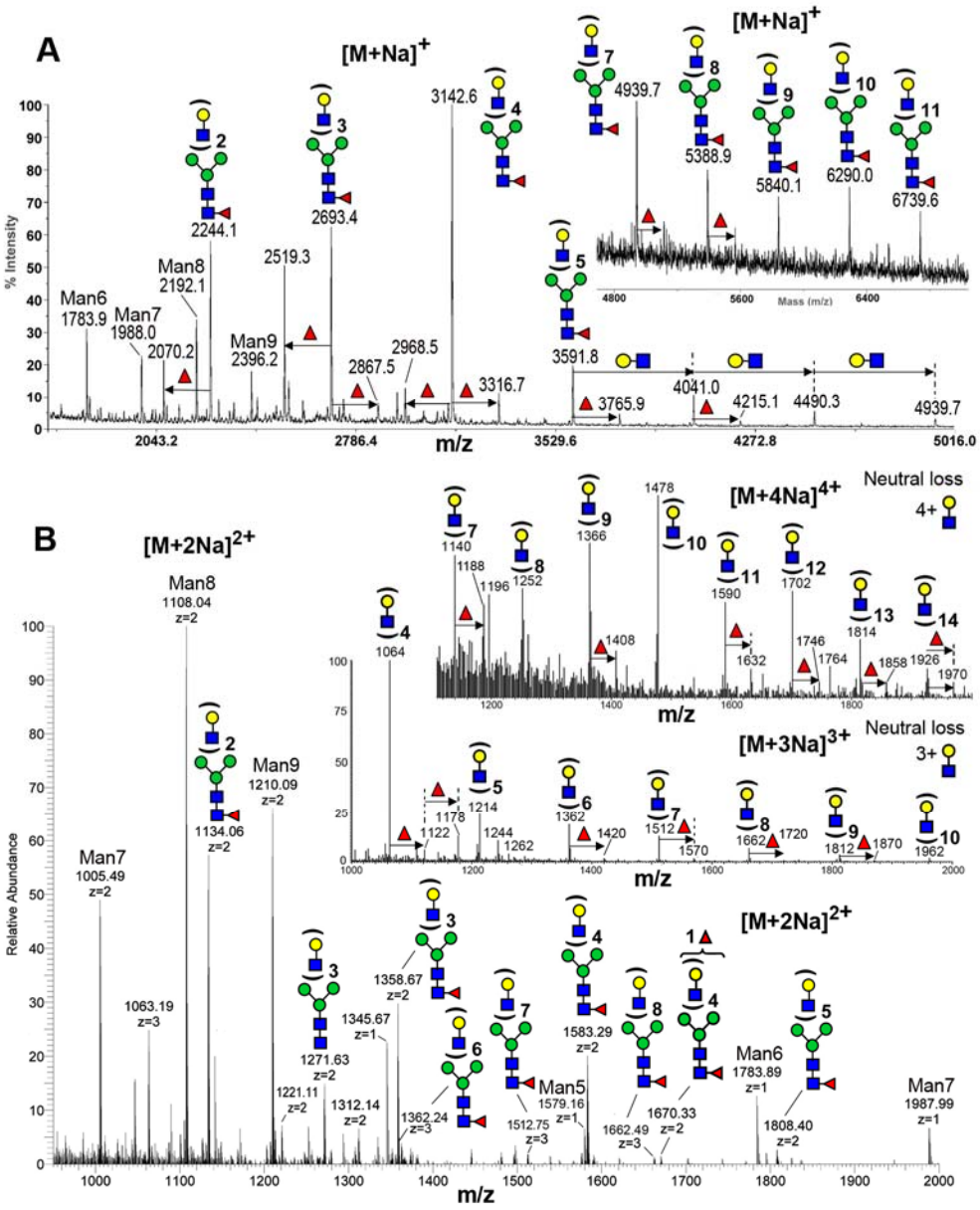



**Figure 3-2. MALDI-MS profile of desialylated, permethylated N-glycans from EA.hy926 (A), in comparison with its corresponding offline nanoESI-MS profile (B) and those reconstructed from nanoESI-MS2 total ion mapping (TIM) filtered for neutral loss of LacNAc (B, insets).** Singly charged $[M+Na]^+$ molecular ions were afforded by MALDI-MS, whereas the nanoESI gave doubly, triply, and quadruply sodiated species for the 2+, 3+, and 4+ molecular ions, respectively. For neutral loss filtering, the triply charged N-glycans with 4-10 LacNAcs were identified by the presence of a primary Y ion with an m/z value of 154.4 (463.2/3) lower than the parent, corresponding to loss of a LacNAc unit without altering its charge state. The quadruply charged N-glycans with 7-14 LacNAcs were similarly identified by a neutral loss corresponding to an *m/z* value of 115.8 (463.2/4). It is obvious that each of the N-glycans afforded more than one molecular ion signals differing in charge states but detection was limited to within m/z 1000-2000 in this case study. For example, the core fucosylated N-glycans with 4 LacNAcs gave an $[M+2Na]^{2+}$ ion at *m/z* 1583.29 in

the direct nanoESI-MS profile, as well as an $[M+3Na]^{3+}$ ion at *m/z* 1064 in the TIM profile filtered for triply charged parent ions. Likewise, the N-glycans with 9 LacNAcs gave both $[M+3Na]^{3+}$ ion at *m/z* 1812 and $[M+4Na]^{4+}$ ion at *m/z* 1366 in the respective TIM profiles. The direct nanoESI-MS profile was acquired at high resolution using the Orbitrap, which therefore allowed auto-identification of charge state, as annotated (B). The TIM MS2 data were acquired at steps of every 2 mass units across the entire mass range and therefore the reconstructed profiles afforded only peaks identified with integer number, corresponding to the *m/z* value set for parent ion isolation. Due to lack of space, all peaks in the insets in (B) were simply annotated as carrying how many LacNAc units on the fucosylated trimannosyl core.

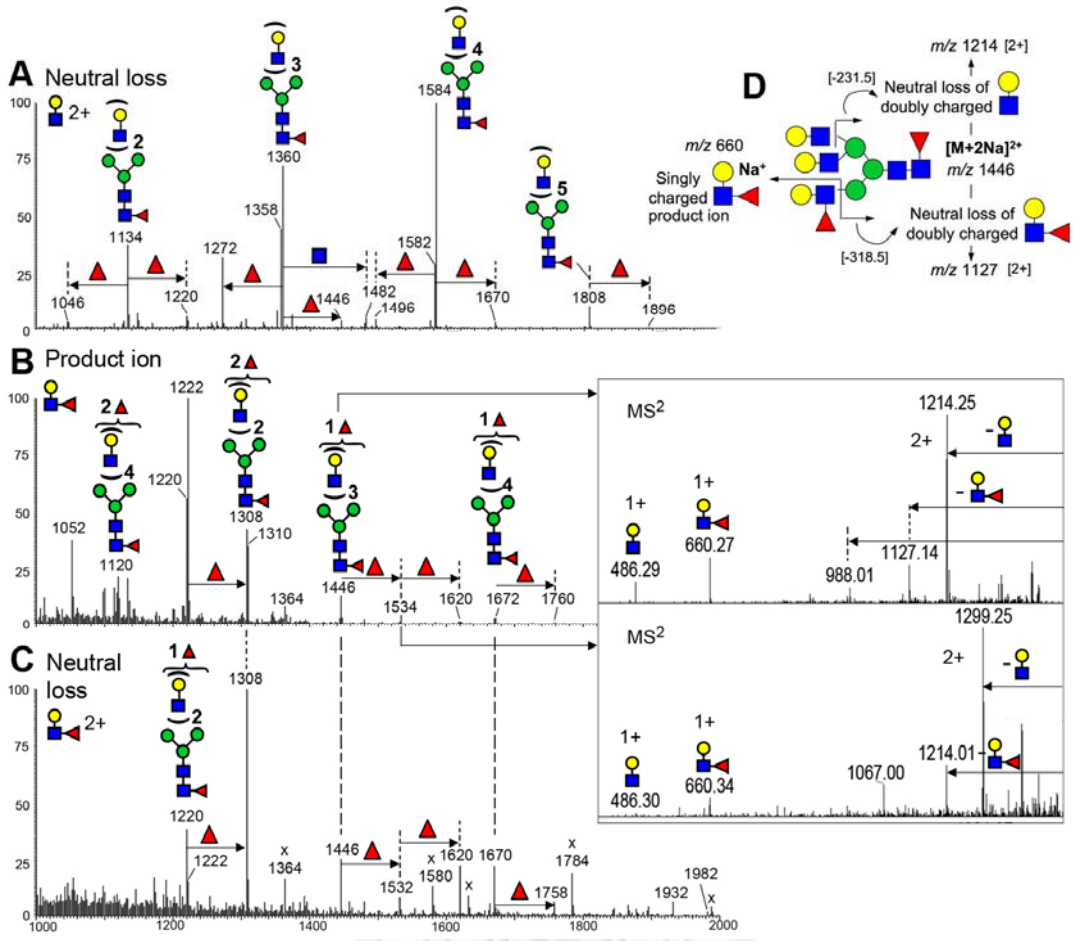



**Figure 3-3. Reconstructed nanoESI-MS2 TIM profiles of desialylated, permethylated N-glycans from EA.hy926, based on occurrence of neutral loss for a terminal LacNAc (A) or $Fuc_1LacNAc_1$ (C) from doubly charged parent ions, or the presence of a singly charged $Fuc_1LacNAc_1$ product ion at m/z 660.3 (B).** Neutral losses from doubly charged parent ions were filtered using *m/z* values of 231.5 and 318.7 for $LacNAc_1$ and $Fuc_1LacNAc_1$, respectively. A schematic drawing is provided in (D) to illustrate the generation of neutral loss and product ions for the use in TIM filtering. The exemplary core fucosylated N-glycan with 3 LacNAcs and 1 additional Fuc afforded a doubly sodiated, doubly charged molecular ion at *m/z* 1446, which was relatively minor compared with the one without the additional Fuc at *m/z* 1360. This relative ratio was roughly maintained when the TIM data was filtered for those parent ions that could lose a terminal LacNAc (A). However, if the data was filtered for parent ions that could lose a terminal $Fuc_1LacNAc_1$ (B), or that could give a product ion of $Fuc_1LacNAc_1$ (C), only the peak at m/z 1446 but not 1360 was retained. The full MS2 spectra for each of the implicated parent ions, e.g. m/z 1446 and 1670 shown here as insets, could be manually examined to support the identification by TIM filtering.

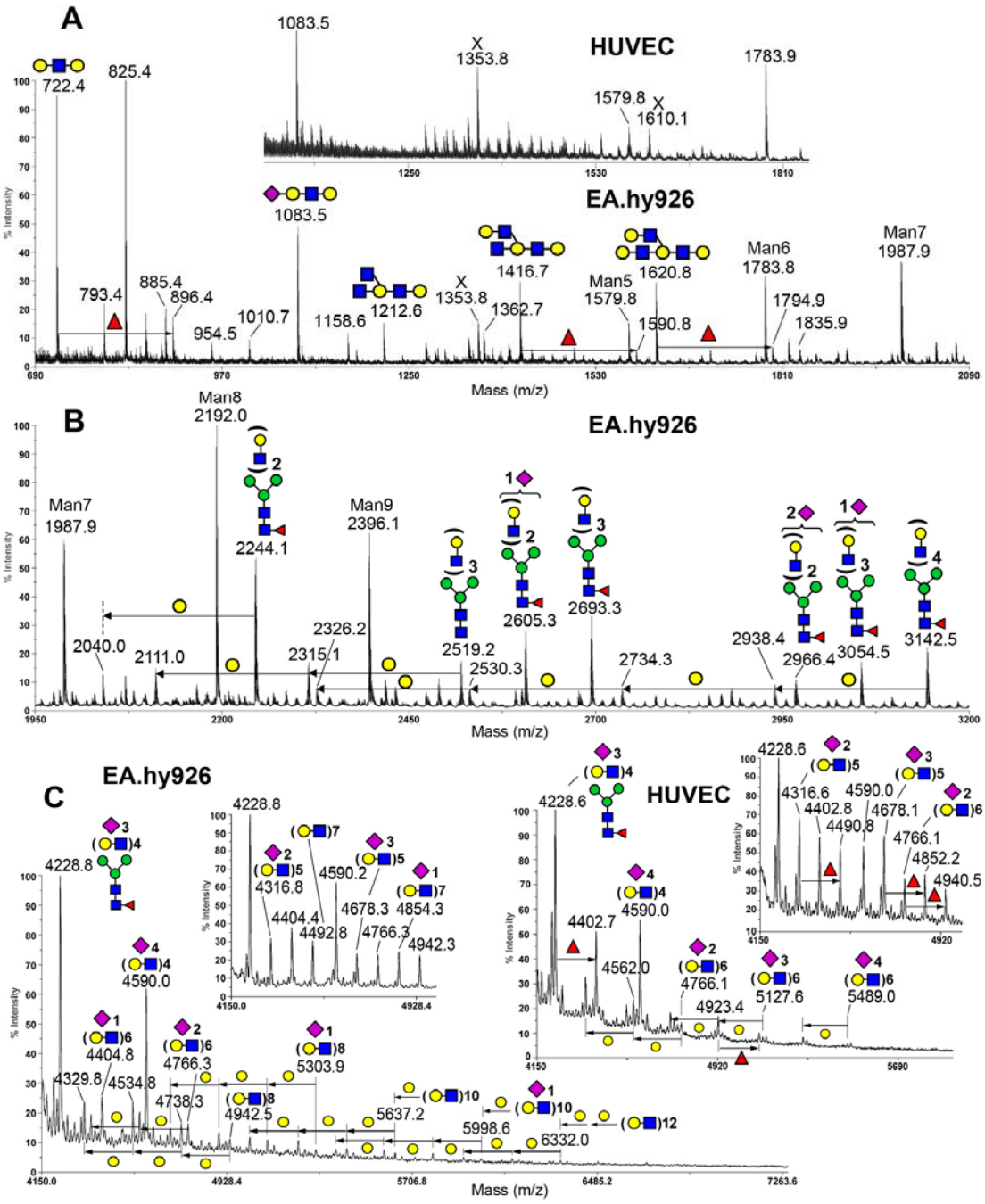



**Figure 3-4. MALDI-MS profiles of permethylated, endo-β-galactosidase digested N-glycans from EA.hy926 at different mass regions (A-C), in comparison with those similarly produced from HUVEC.** (A) The low mass region showed the prominent presence of the three branched structures at m/z 1212, 1416 and 1620, along with the two fucosylated counterparts at m/z 1590 and 1794, all of which were absent from the corresponding mass region of the spectrum for HUVEC (top inset). Oxonium ions derived from in source prompt fragmentation were detected at m/z 825, 954, 1158, and 1362, corresponding in composition to those at m/z 1083, 1212, 1416 and 1620 but lacking the reducing terminal Gal. (B) The intermediate mass region showed that most of the N-glycans with 2-4 LacNAcs remained intact after digestion but each was additionally accompanied by peaks lacking one or more terminal Gal, as

indicated for a representative few. The profile for HUVEC in this region (not shown) was likewise similar to that before digestion but carried additional minor peaks lacking terminal Gal. (C) At higher mass region in linear mode, the tri- and tetrasialylated N-glycans with 4 LacNAcs were clearly not much affected by digestion and remained as the only major signals, but most other peaks with higher degree of LacNAc extension were no longer observed or significantly reduced in intensities, in reference to their relative amounts in the original spectra (insets). Curiously, those with even number of LacNAcs (6, 8, 10, 12 for EA.hy926; 6 for HUVEC) extension better survived the digestion, consistent with them carrying branched structures proximal to the trimannosyl core.

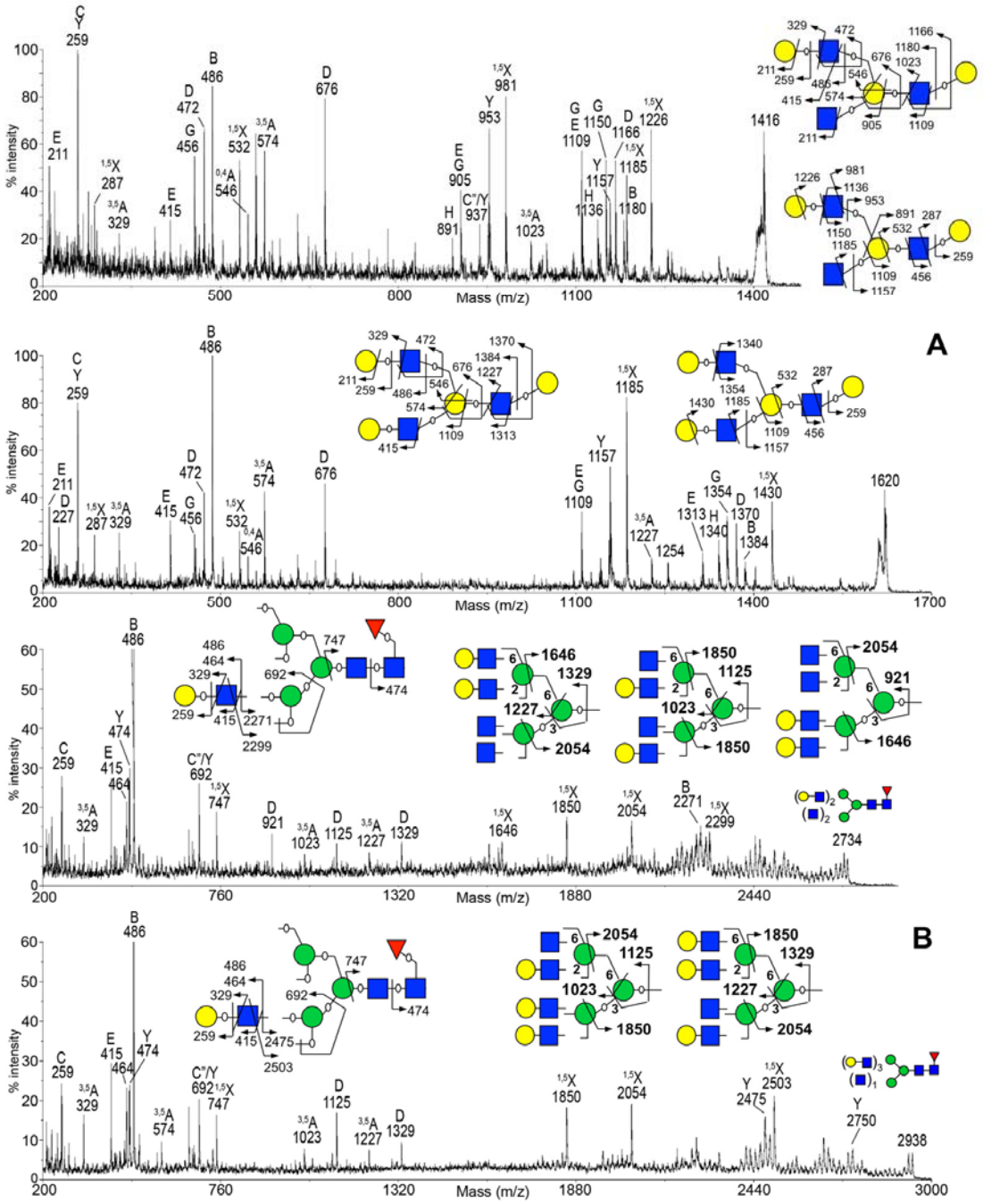



**Figure 3-5. High energy CID MALDI-MS/MS analyses of the permethylated, endo-β-galactosidase digested products derived from the N-glycans of EA.hy926.** (A) The branched nature of the released products were established by complementary sets of reducing ($^{1,5}$X, Y, G ions) and non-reducing (B, $^{3,5}$A, E ions) terminal ions, and the characteristic D ions formed at the branched Gal, as schematically illustrated. The structure assigned for m/z 1416 (top panel) can be additionally accompanied by an isomer with non-extended, intact LacNAc unit on the 3- and not 6-arm, which could neither be established nor ruled out by current set of fragment ions. (B) Two of the remaining trimannosyl core structures for tetraantennary N-glycans at m/z 2734 and 2938 were shown by the characteristic D and $^{3,5}$A ions formed at the branched

3,6-Man to carry the exposed GlcNAc termini non-exclusively on either arm, as schematically illustrated. All ion types and nomenclature were as described previously (Fan et al., 2008; Yu et al., 2006).

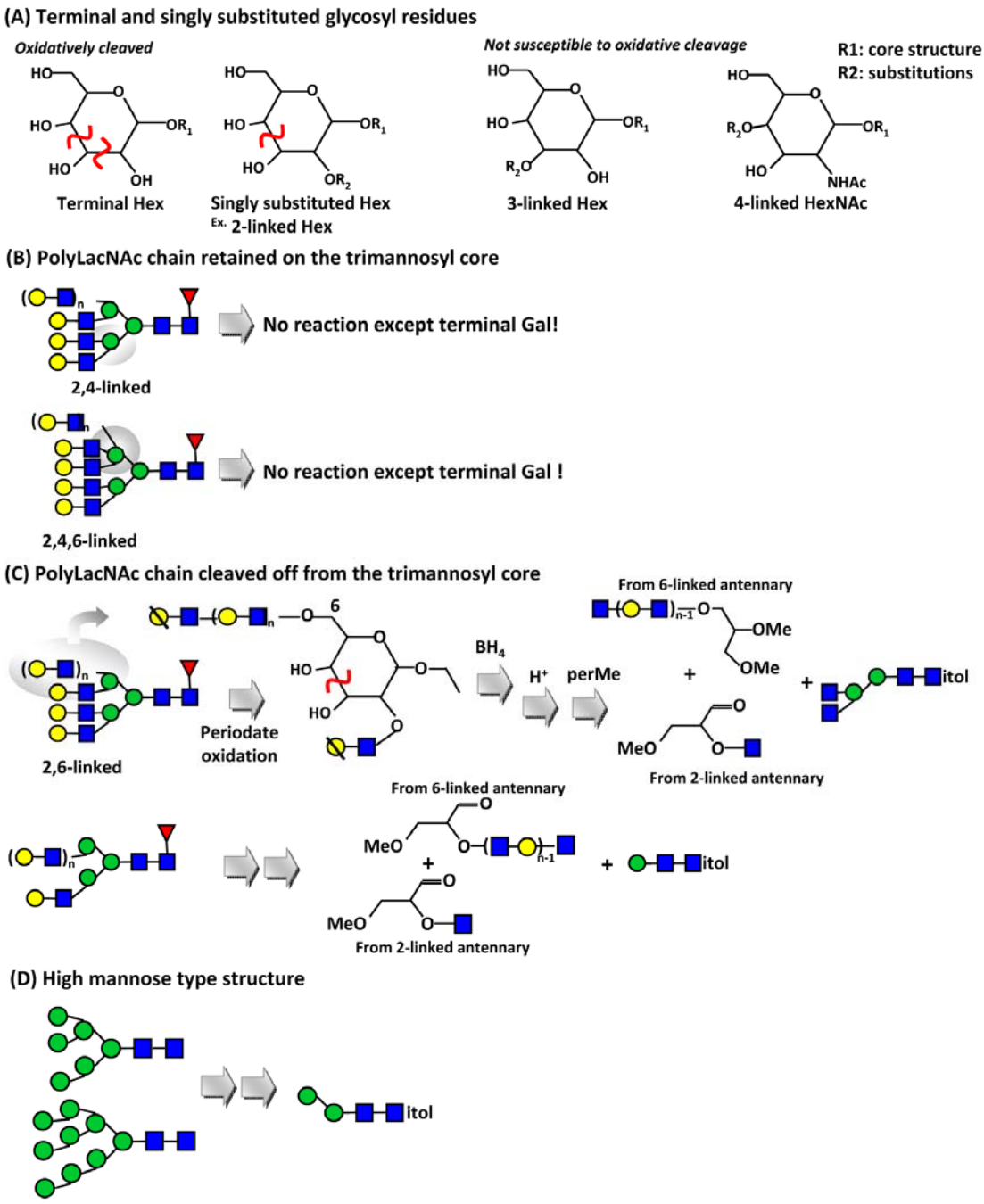



**Figure 3-6. Schematic illustration of the expected Smith degradation products.** Oxidative cleavage of vicinyl OH group will occur at all terminal and singly substituted glycosyl residue except a 3-linked residue and 4-linked GlcNAc (A). Consequently, all the α1-6 linked fucoses will be cleaved and removed from the core structure of N-glycans. In contrast, the polyLacNAc backbone, -3Gal-4GlcNAc-, is resistant to cleavage and remains intact. Depending on its attachment site on the trimannosyl core, the entire polyLacNAc chain may be retained on the core (B) or released from it along with remnant fragments derived from the oxidatively cleaved Man residue (C). All high mannose structures will be converted to a Manα1-6Manβ1-4GlcNAcβ1-4GlcNAc-itol, as shown in (D).

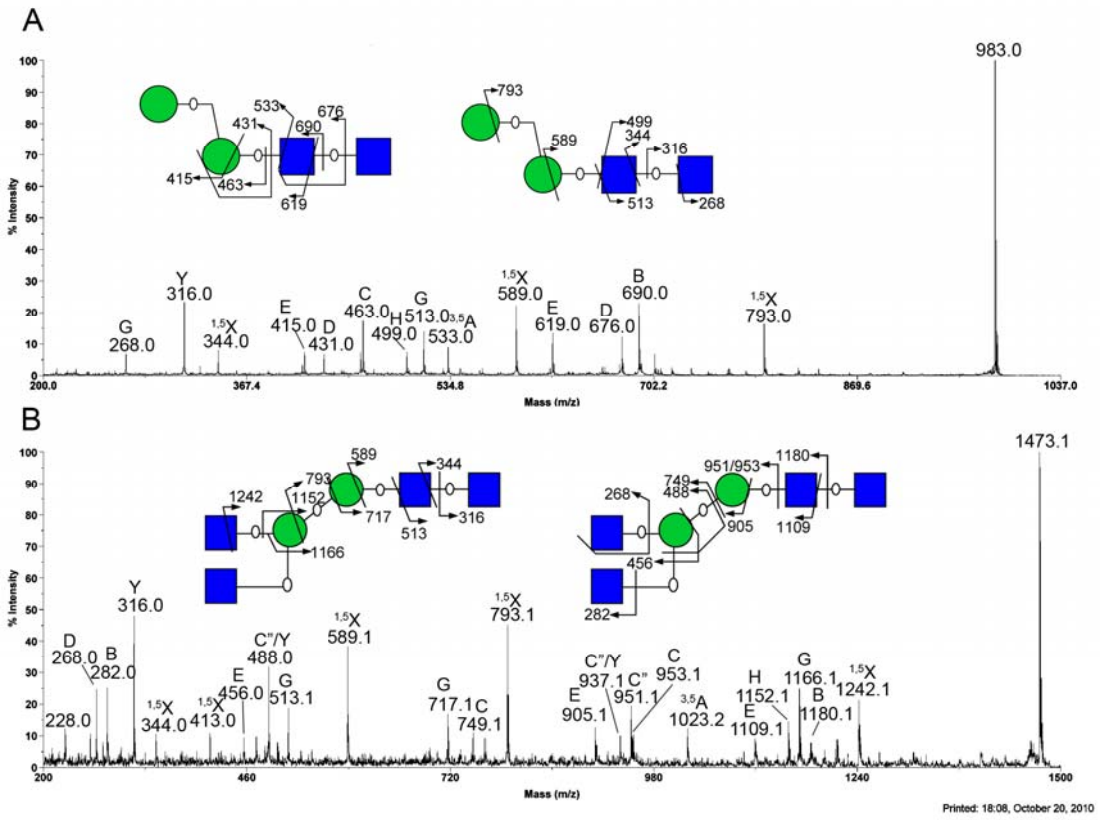



**Figure 3-7. MALDI MS/MS analysis of the major Smith degradation products of desialylated N-glycans of EA.hy926.** High mannose structures were converted to Manα1-6Manβ1-4GlcNAcβ1-4GlcNAc-itol (m/z=983) (A), whereas complex type N-glycans were converted to either Manβ1-4GlcNAcβ1-4GlcNAc-itol (m/n=779) (not shown) or GlcNAcβ1-2(GlcNAcβ1-4)Manα1-3Manβ1-4GlcNAcβ1-4GlcNAc-itol (m/z=1473) (B), depending on their antennary branching pattern, as described in the text.

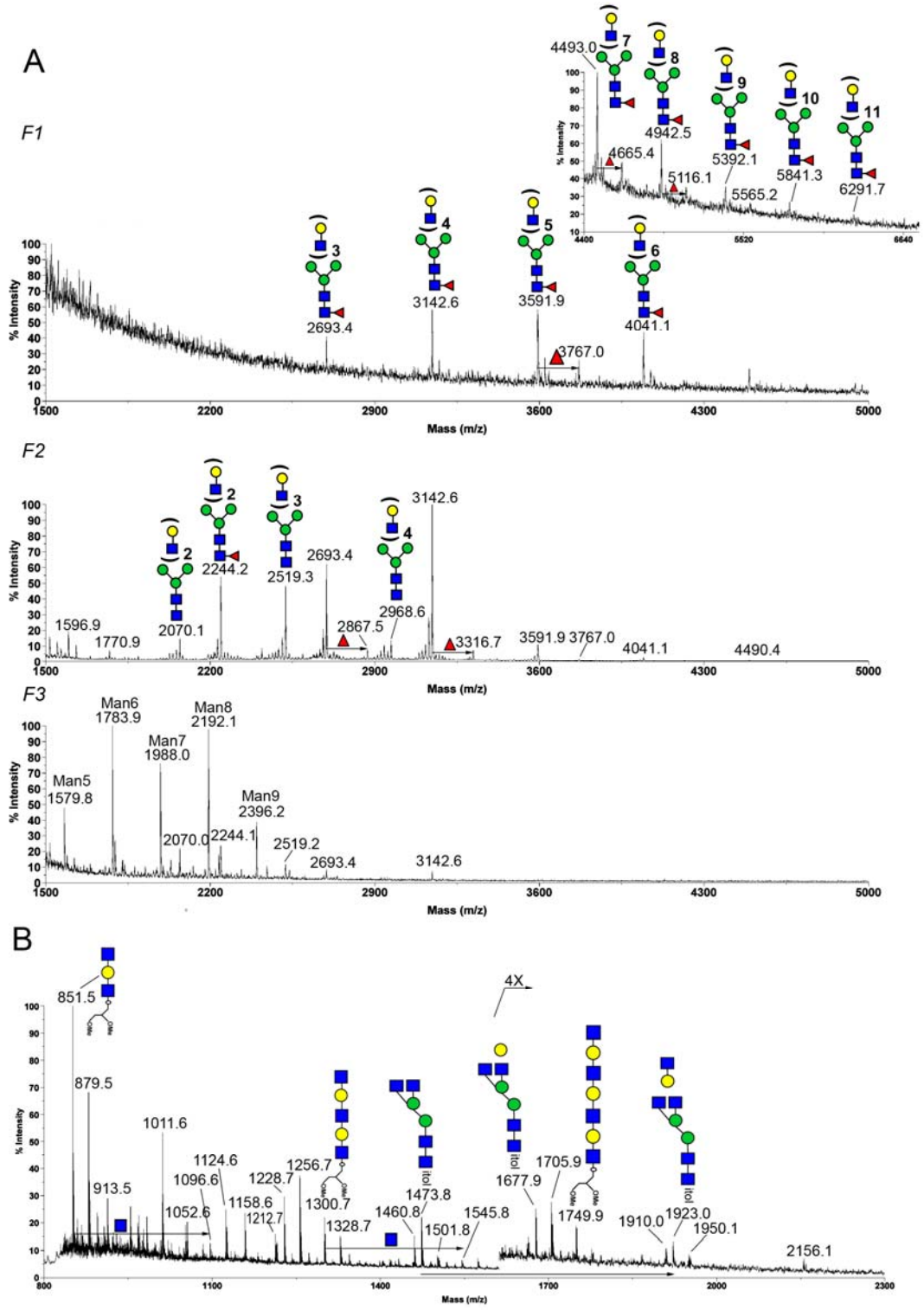



**Figure 3-8. Size fractionation of the de-sialylated N-glycans of EA.hy926 by P4 gel filtration (A) and subsequent Smith degradation on the isolated larger N-glycans with polyLacNAc.** The N-glycans fractionated into each of the 3 pooled fraction, F1-F3, were profiled by MALDI MS analysis after permthylation, and only F1 was taken through the Smith degradation. Most of the labeled structures in (B) at m/z less than 1500 were analyzed further by MS/MS.

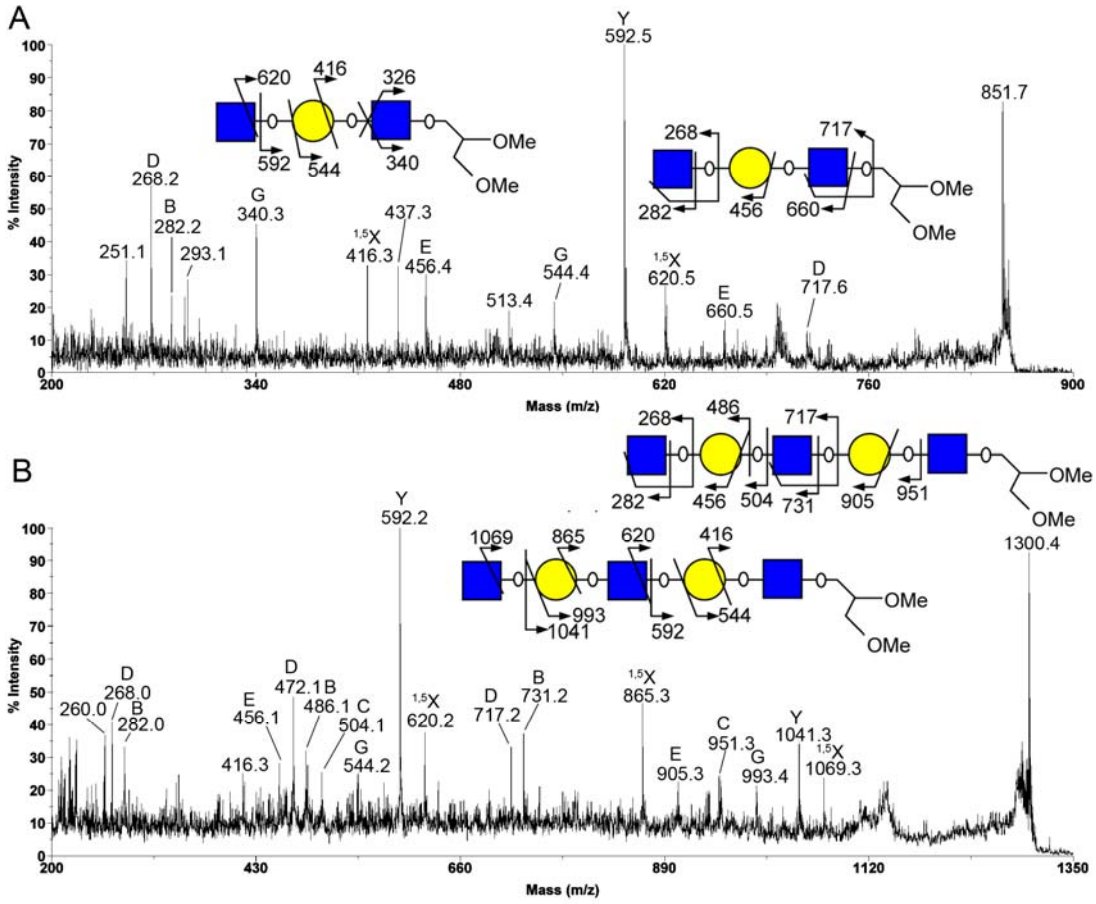



**Figure 3-9. MALDI MS/MS analysis of the permethylated polyLacNAc released by Smith degradation.** As illustrated in Fig. 3-6, the GlcNAc-(Gal-GlcNAc)$_{n-1}$-X fragments were most likely derived from the polyLacNAc chain originally attatched to C6 of the Man on 6-arm.

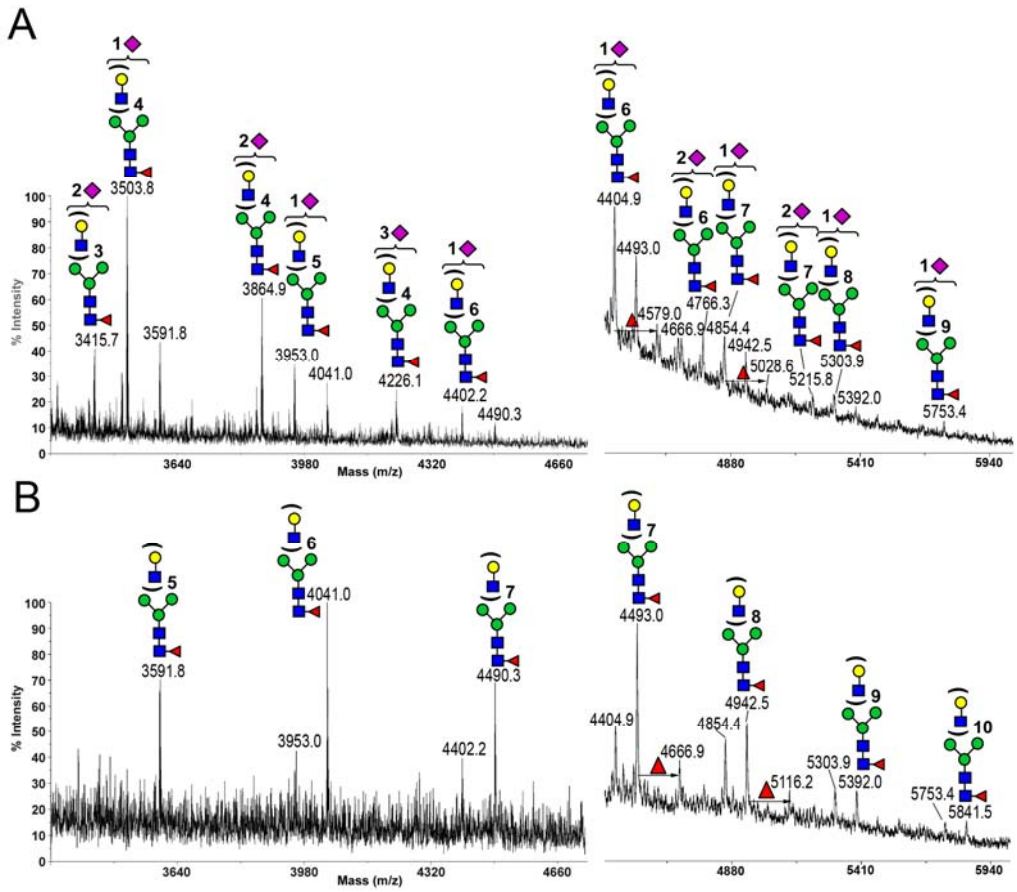



**Figure 3-10. MALDI MS profiles of permethylated N-glycans of EA.hy926 fractionated by LEL affinity chromatography.** Comparing the unbound (A) and bound (B) fractions, N-glycans with more than 5 LacNAcs were preferentially captured by LEL although a certain portion of these N-glycans were also found in the flow through fraction. It seems that LEL preferred non-sialylated N-glycans at the glycan level.

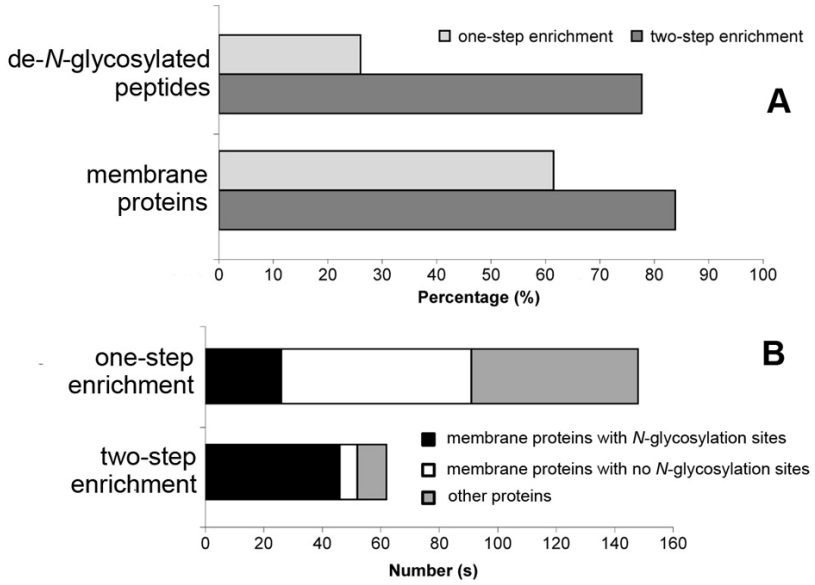



**Figure 3-11. Evaluation of one-step versus two-step LEL enrichment by means of proteomic analysis of the de-N-glycosylated tryptic peptides.** Experimental conditions for obtaining the glycopeptides via the one-step and two-step enrichment workflows are described in Sections 2.3.2 and 2.3.1, respectively. The two-step process involved an initial LEL enrichment at the glycoprotein level not undertaken for the one-step process. Both shared the subsequent step of LEL enrichment at the glycopeptide level as described in Section 2.3.3. Membrane protein classification was based on GOMINER, whereas identification of a de-N-glycosylated peptide containing the consensus Asn-Xxx-Ser/Thr motif with a requisite Asn to Asp conversion was taken to implicate an originally N-glycosylated peptide. The two-step enrichment workflow returned a higher percentage of both the de-N-glycosylated peptides and membrane proteins (A). In terms of the total number of identified proteins, the two-step enrichment workflow returned a higher number of membrane proteins represented by de-N-glycosylated peptides although the one-step enrichment workflow registered a higher number of all confidently identified proteins and membrane proteins irrespective of whether they also carried the de-N-glycosylation sites (B).

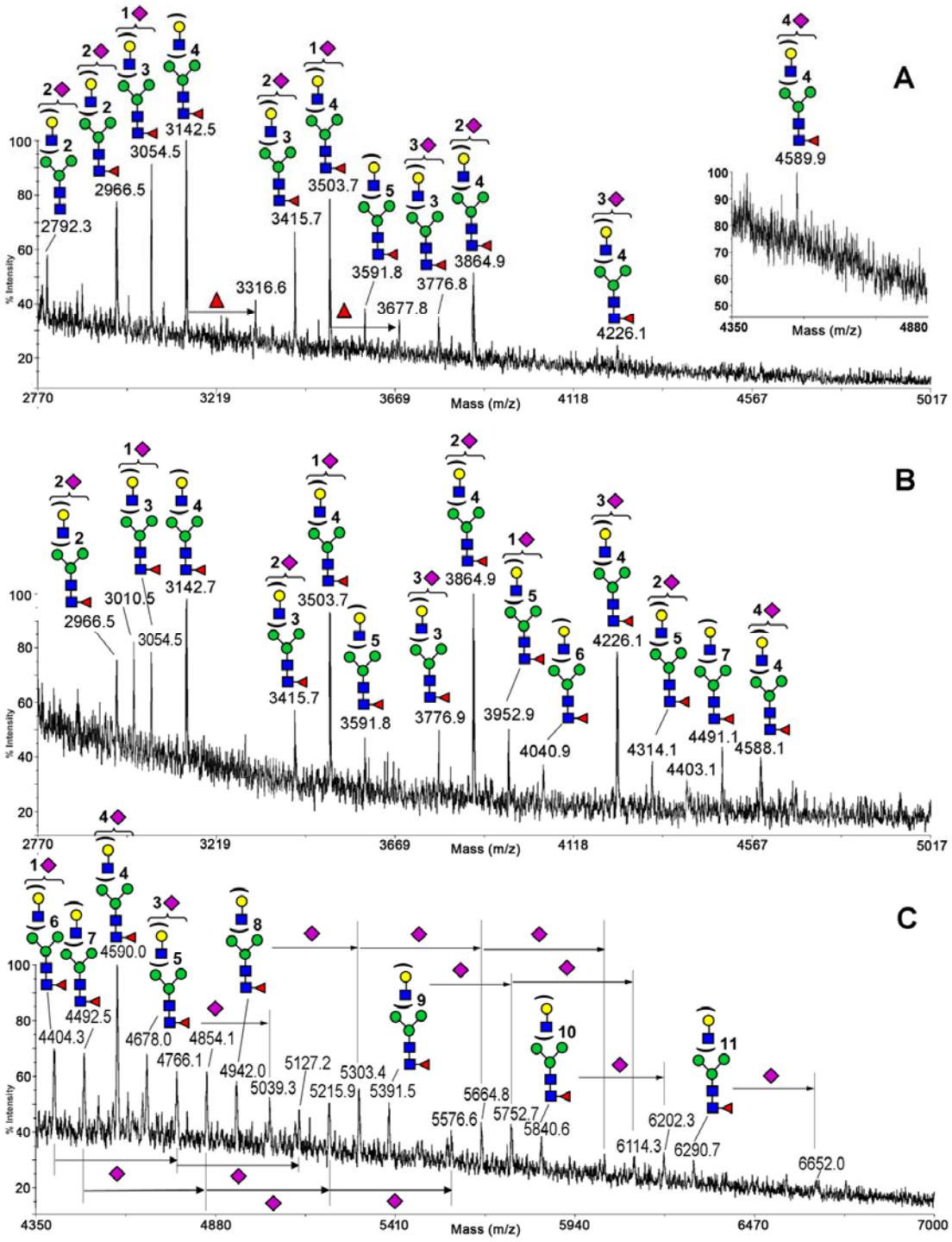



**Figure 3-12. MALDI-MS profiles of permethylated N-glycans released from the tryptic digests of the LEL-non-binding (A) and bound (B, C) fractions from the second step of the two-step LEL-enrichment process.** The non-binding fraction did not afford any significant signal corresponding to larger N-glycan containing 5 or more LacNAc units, as shown in the inset for the higher mass range acquired in the linear mode. In reflectron mode (A), the major signals detected corresponded to N-glycans with 2-4 LacNAc units, as annotated. In contrast, the bound fraction afforded major signals in both reflectron mode (B) and linear mode (C) for the lower

and higher mass range, respectively. The polyLacNAc extension detected was not as extensive as that shown in Fig. 1 due to much less amount of starting material used to obtain the N-glycans.

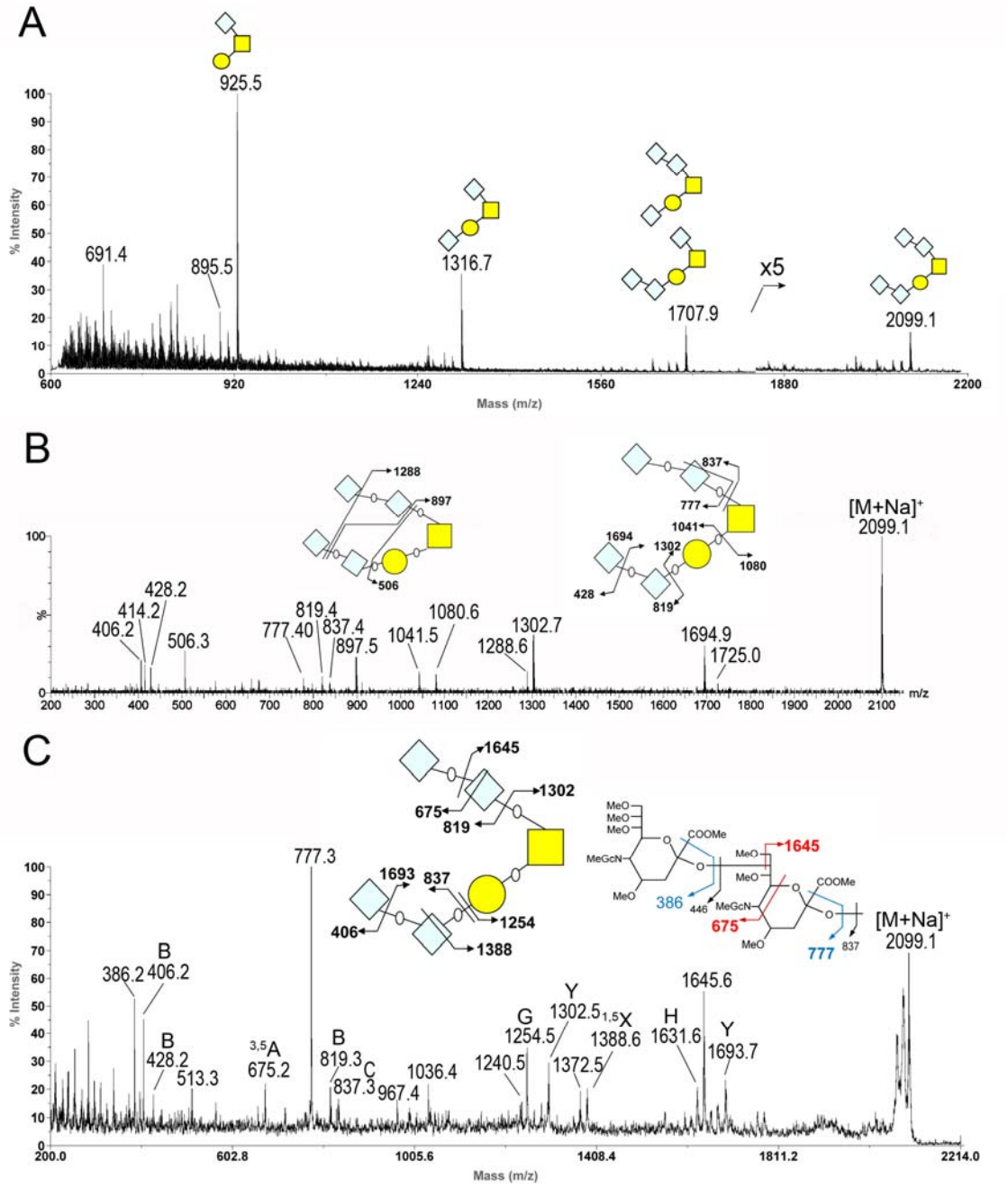



**Figure 3-13. MALDI-MS profile of the permethylated O-glycans of BCL1. (A) and the MALDI Q/TOF (B) and TOF/TOF (C) MS/MS spectra for the tetrasialylated core 1 structure, with schematic drawings illustrating the cleavage patterns.** The cleavage pattern and assignment of the fragment ions were illustrated by schematic drawings. Only the MS/MS spectrum for Neu5GcHex$_1$HexNAc$_1$-itol at m/z 2099 was shown here. MS/MS on the mono- (m/z 925), di- (m/z 1316), and tri-sialylated (m/z 1707) core 1 were also performed.

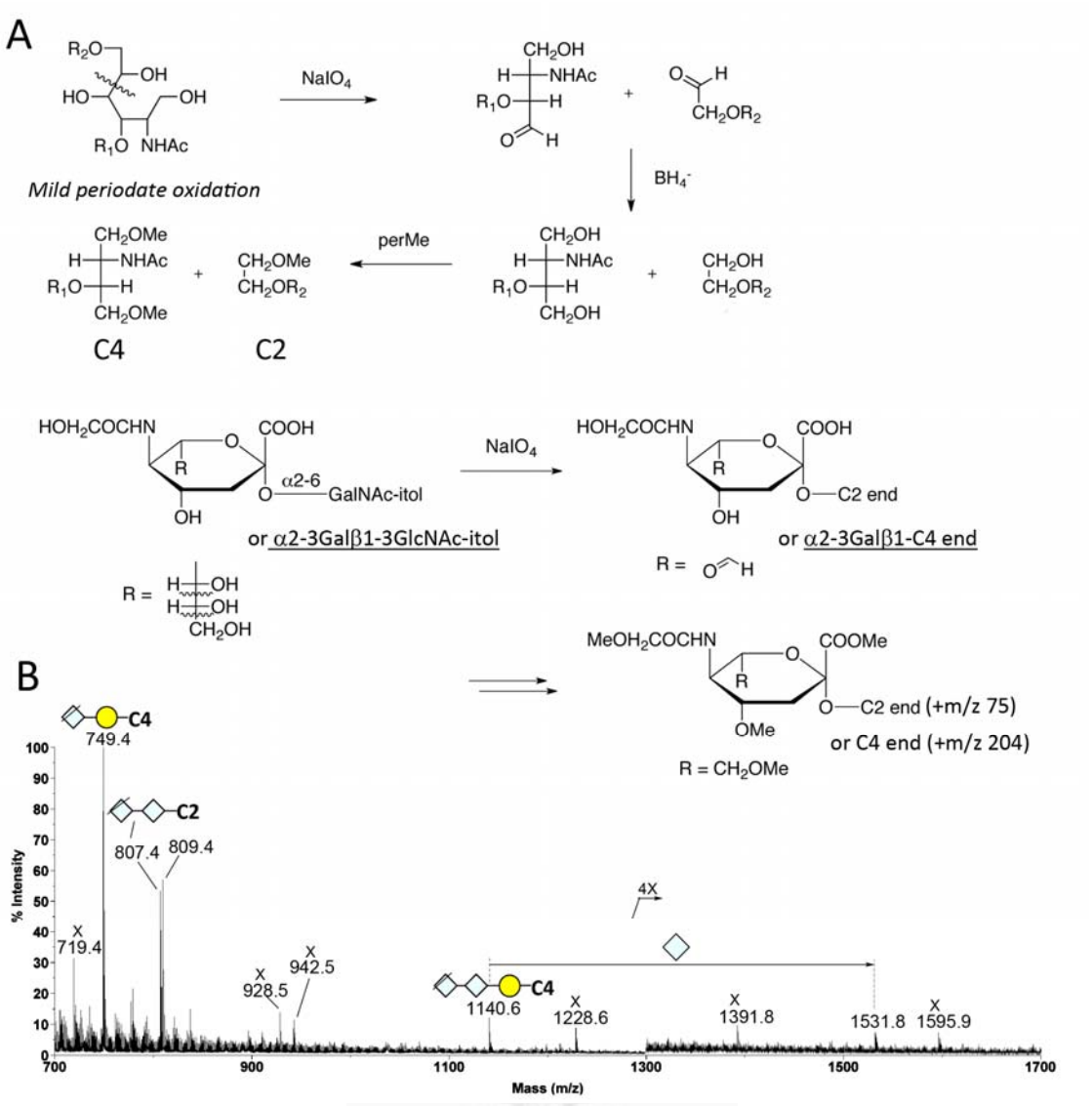



**Figure 3-14. Mild periodate oxidized O-glycans profiles of BCL1.** Through mild periodate oxidation, the O-glycans could be cleaved into two series of products, as illustrated in (A). The terminal sialic acid itself would lose the C8-C9 side chains corresponding to a reduction in mass of 88u, which was represented by a slash across the Neu5Gc symbol in the cartoonist peak labels.

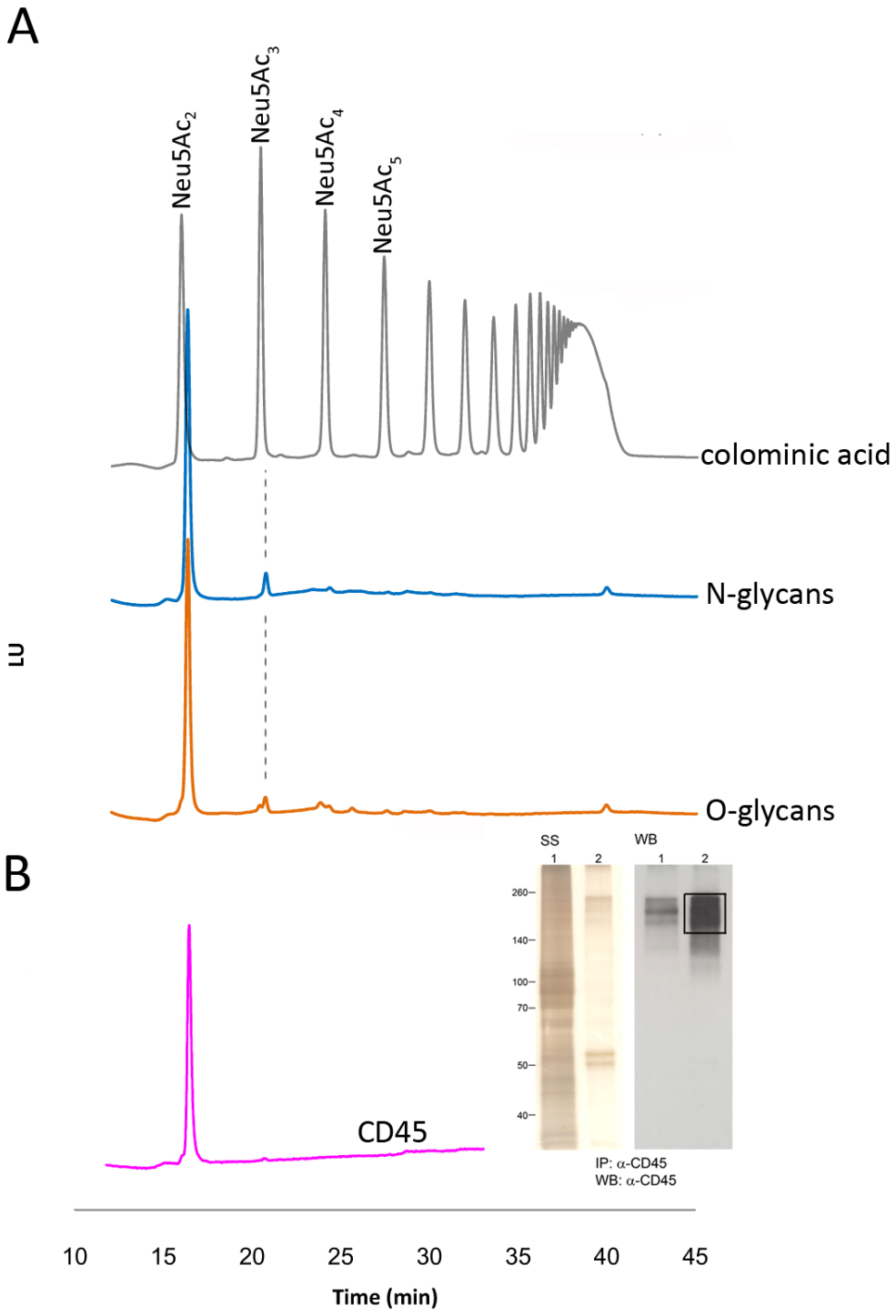



**Figure 3-15. DMB-sialic acid analysis of the glycans from BCL1 (A) and CD45 isolated from BCL1 (B).** DMB-labeled sialic acid oligomers were separated on a mono Q column, detected and quantified by fluorescence. Elution positions were calibrated against hydrolysates of colominic acid. Di-Neu5GC was known to be eluted a bit later than di-Neu5Ac from the mono Q column in anion exchange chromatography (Sato et al., 1999). The Western blot (WB) and protein silver stain (SS) of immuno-purified CD45 was shown as insets on (B). Lane 1, extracted total membrane proteins; Lane 2, sample immunoprecipitated by anti-CD45.

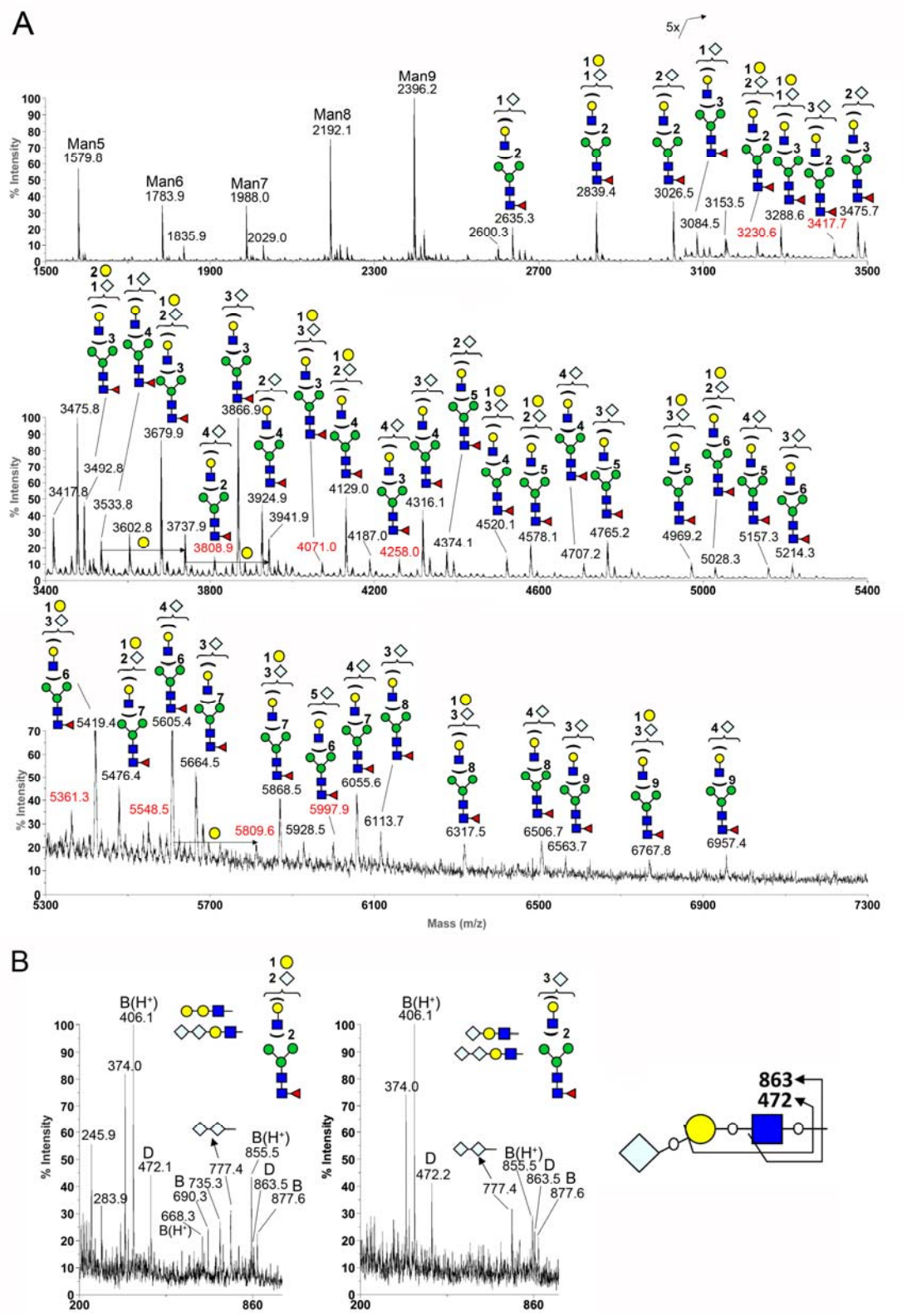



**Figure 3-16. MALDI MS (A) and MS/MS (B) analysis of N-glycans from BCL1 in positive ion mode.** The N-glycan profile showed the usual presence of high mannose and complex type N-glycans while the peaks labeled in red were potentially N-glycans with disialyl motif. Two of these at m/z 3230 and 3417 in (A) were further subjected to MS/MS analyses (B). MALDI MS/MS analysis indeed afforded characteristic fragment ion at m/z 777 (B). The most prominent fragment ins at m/z 406, 374, and 855 corresponded to oxonium ions of Neu5Gc$_+$, Neu5Gc-MeOH, and

Neu5Gc-Hex-HexNAc$_+$, respectively. The characteristic ion at m/z 777 was indicative of the presence of Neu5Gc-Neu5Gc (of Fig. 3-13C), whereas other ions were assigned as illustrated.

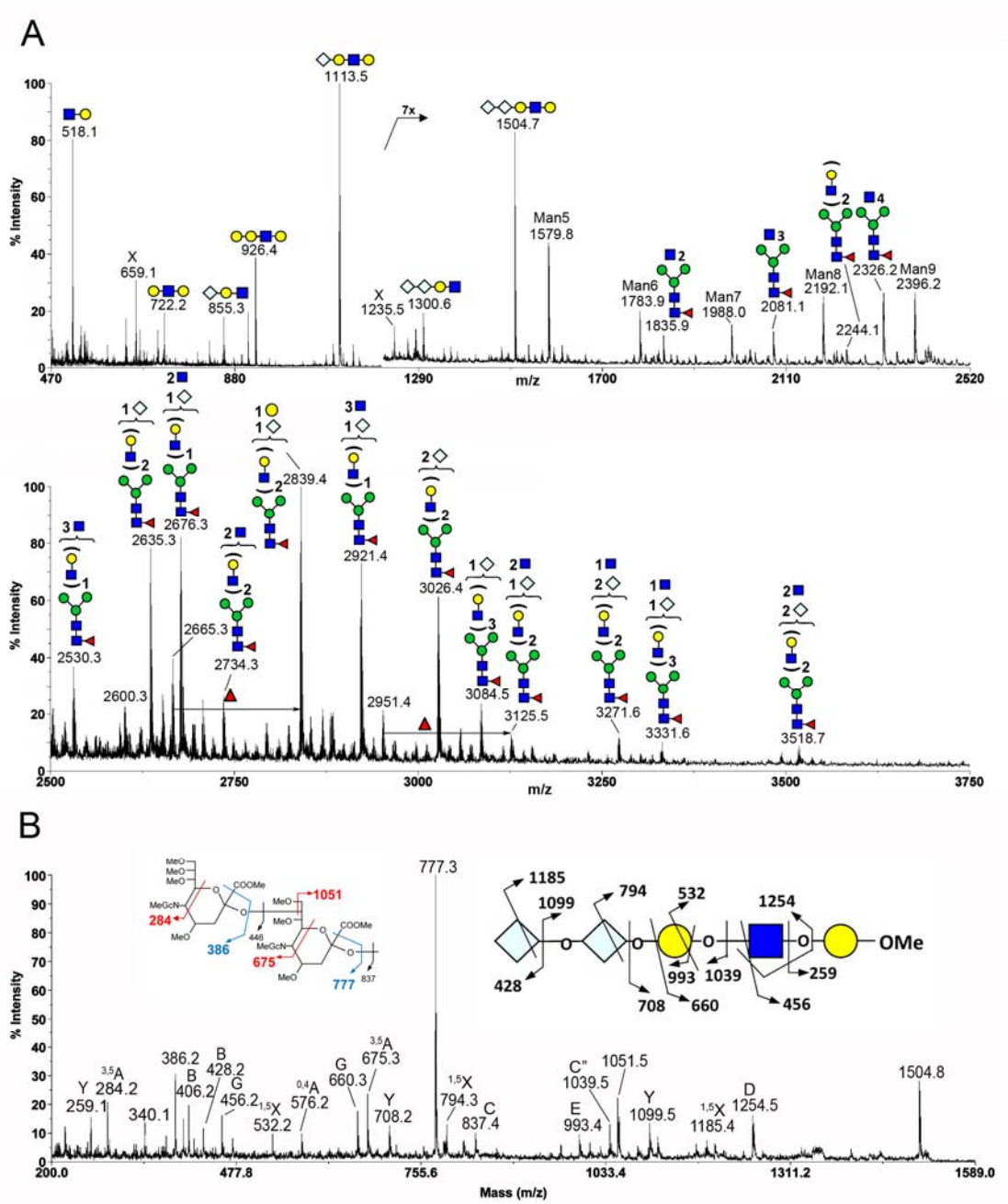



**Figure 3-17. MALDI MS (A) and MS/MS (B) analysis of endo-β-galactosidase digested N-glycans from BCL1.** Endo-β-galactosidase treatment give rise to the terminal disialyl-LacNAc-Gal (m/z 1504), which was identified along with non-sialylated, monosialylated and α-Gal capped terminal units, corresponding to the signals at m/z 722, 1113, and 926, respectively. All these ions were further selected for MS/MS analyses to confirm their assigned structures but only the MS/MS spectra for m/z 1504 was shown in (B). Other molecular ions detected corresponded to core structures with one or more of their polyLacNAc chains being removed, as annotated.

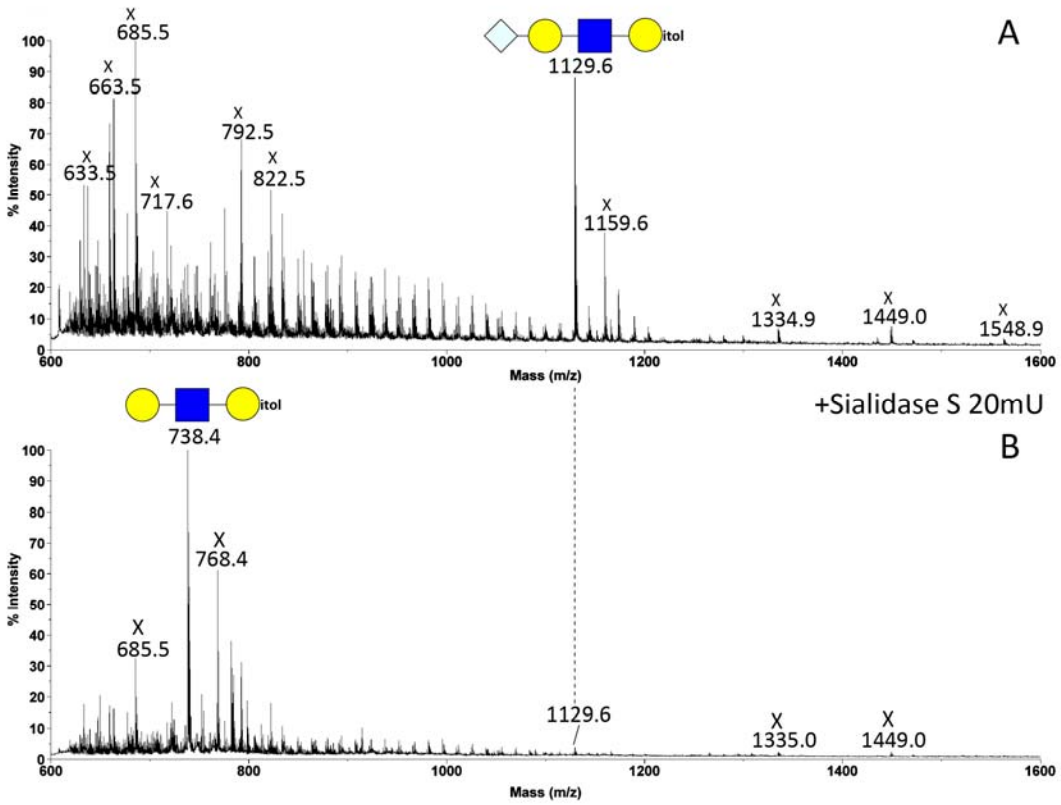



**Figure 3-18. The MS profile of isolated endo-β-galactosidase released terminal fragment before (A) and after (B) sialidase S digestion.** For this experiment, endo-β-galactosidase digestion was performed directly at the tryptic peptides level and the released fragments were isolated from the remaining glycopeptides/peptides via Sep-pak C18. Non-sialylated fragments were further removed by anion exchange Oasis MAX and the recovered sialylated glycans fragments-desalted, permethylated, and analyzed by MS, with and without additional sialidase S digestion. The resulting MS profile of the reduced, permethylated glycans showed that most of the Neu5Gc-LacNAc-Gal-itol (m/z 1129) was digested to give the desialylated product, LacNAc-Gal-itol at m/z 738.

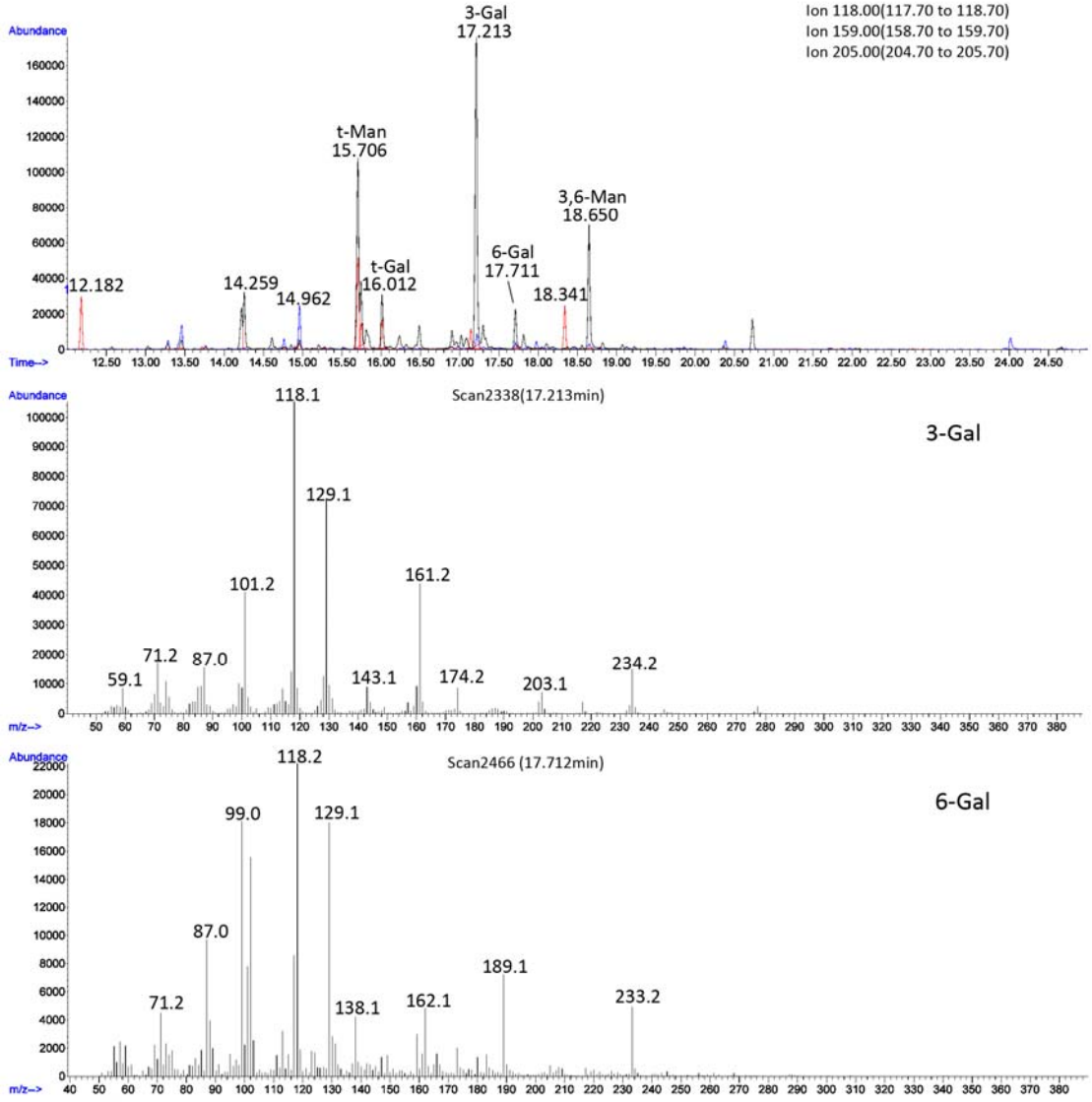



**Figure 3-19. GC-EI-MS linkage analysis of the N-glycans from BCL1.** The partially methylated acetates prepared from permethylated N-glycans were identified based on retention time in reference to authentic standards and their well established, characteristic EI-MS fragmentation pattern. the GC-MS profile shown in (A) was reconstructed from superimposed extracted ion chromatograms of m/z 118, 159, 205, which collectively filtered for all possible substituted glycosyl residues, as annotated. The EI-MS spectra for peak at 17.213 min and 17.712 min as shown in (B) is indicative of not only the presence of 3-linked Gal, but also 6-linked Gal in BCL1, and therefore the presence of both α2,3- and α2,6-sialylation.

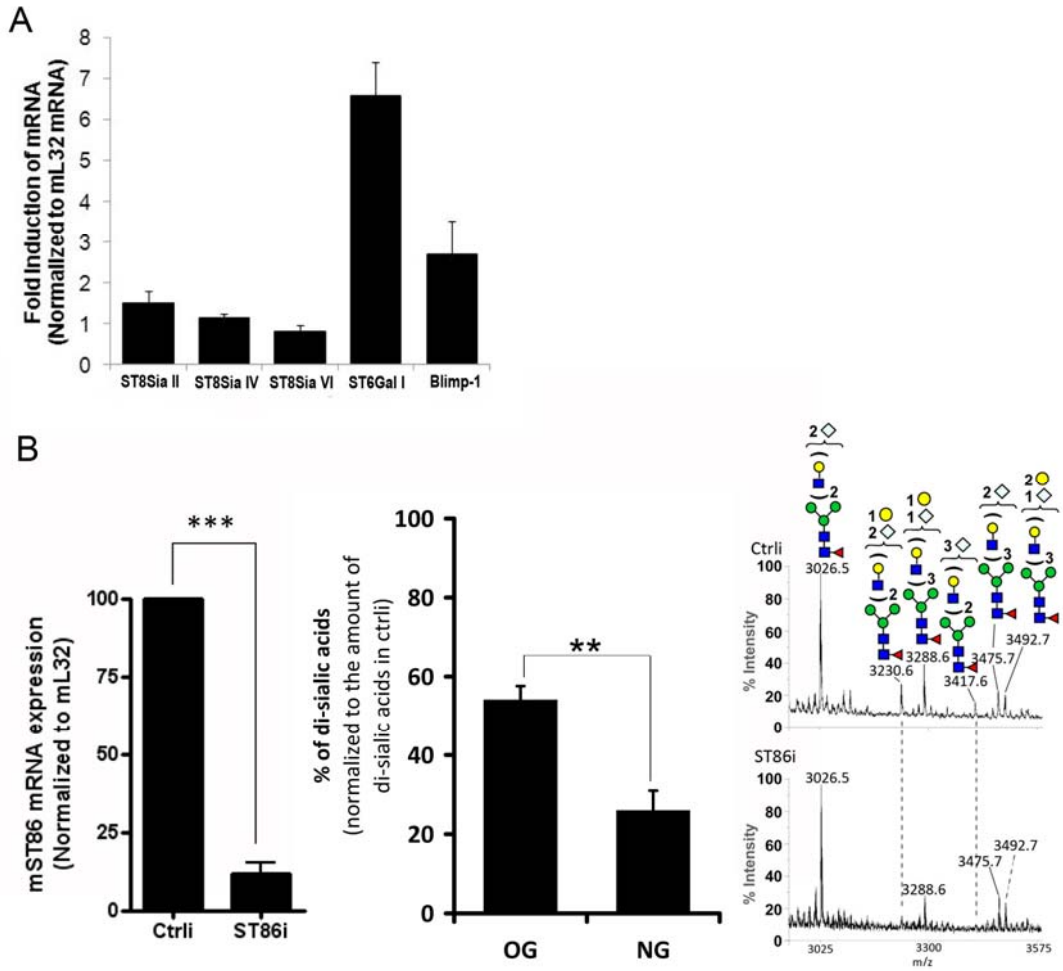



**Figure 3-20. The expression level of mouse α 2,8-sialyltransferase (mST8) II, IV, and VI in BCL1 by RT-Quantitative (Q)PCR (A) and the effect of siRNA knockdown of ST8Sia VI (B).** Only mST8Sia II, IV, and VI were expressed in BCL. Although the expression of mST8Sia II and IV were higher than mST8Sia VI, mST8Sia VI showed the highest fold of induction during plasma cell differentiation. After siRNA knockdown of ST8Sia VI to 10% of its original level (B, left panel), DMB-sialic acid analysis showed that the disialic acid content of the O- and N-glycans were reduced to 54% and 26%, respectively, relative to Ctrli (B, middle panel). This experiment was repeated for three times. *, $p<0.05$; **,$p<0.01$;***,$p<0.005$. The corresponding change in MALDI MS profile for O-glycans was not obvious (not shown), but the N-glycan peak assigned as containing the disialyl motif at m/z 3230 and 3417 clearly showed a reduction in peak intensity (B, right panel).

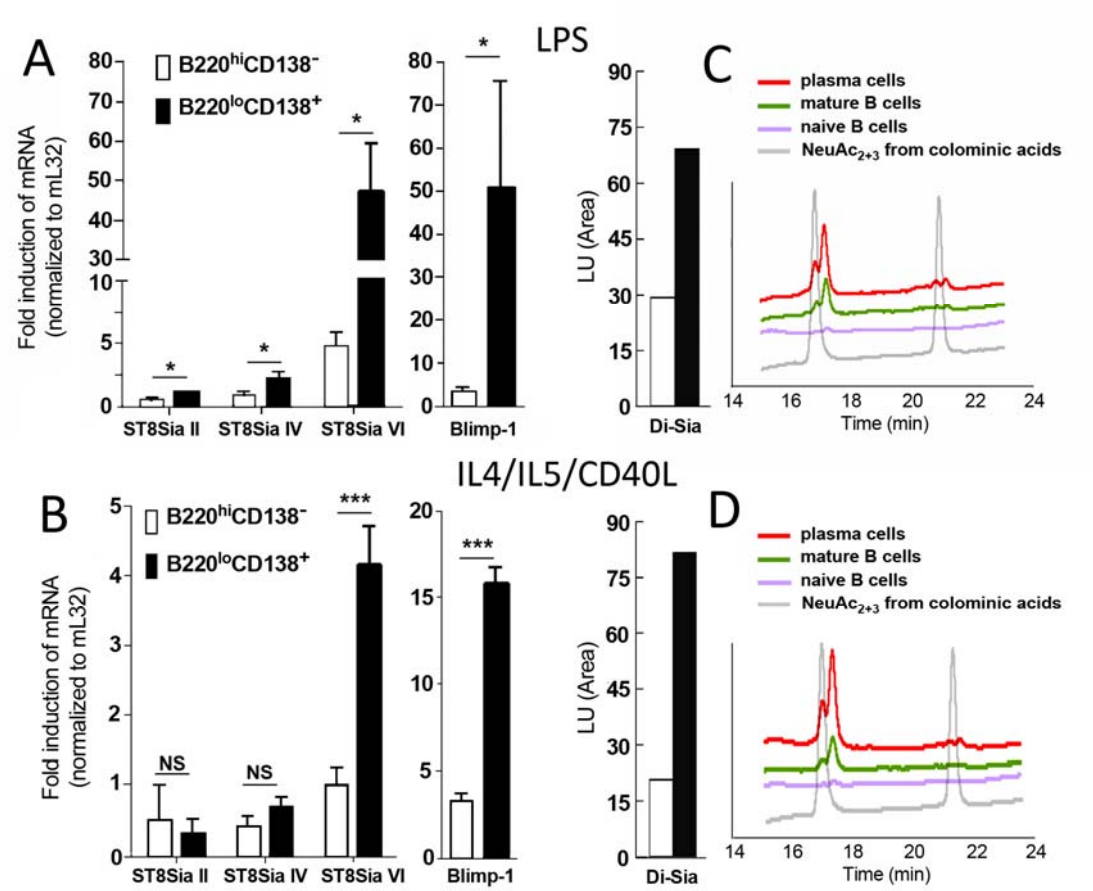



**Figure 3-21. The disialyl content of primary mouse B cells increased during plasma cell differentiation.** RT-QPCR showed an increased level of ST8sia II, ST8sia IV and ST8sia VI to internal control mL32 in B220$^{hi}$CD138$^{-}$ (activated mature B cells) and B220$^{low}$CD138$^{+}$ (plasma cells) by sorting from LPS (A) or IL4/IL5/CD40L (B) stimulated splenic B cells on day 3. DMB analysis of disialic acids in LPS (C) or IL4/IL5/CD40L (D) activated cell populations, showed a corresponding increase in the disialyl content. Results in A are means ±SD (n=3). N.S. represents no significant levels. *, $p$<0.05; **,$p$<0.01;***,$p$<0.005.

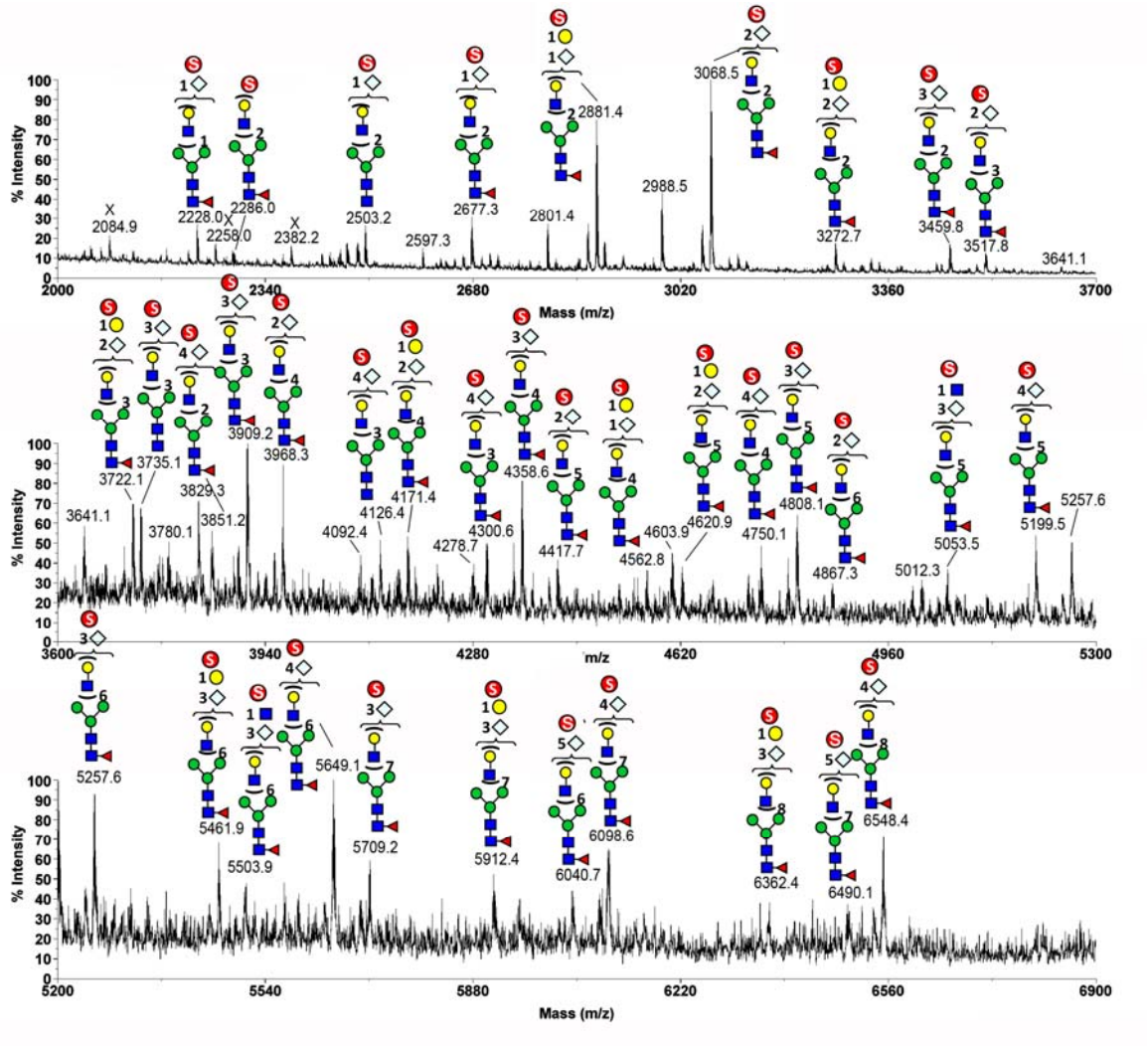



**Figure 3-22. MALDI MS analysis of permethylated sulfated N-glycans from BCL1 in negative ion mode after $NH_2$-microtip fractionation.** Permethylated sample was cleaned up by Sep-pak C18 cartridge and further fractionated by $NH_2$-mircrotip. Neutral glycans were in unbound and 95% ACN fraction. Mono-sulfated glycans were found to be mostly eluted by 2.5 mM $NH_4HCO_3$ in 50% ACN and detected in negative ion mode. Most of the complex type N-glycans detected by MALDI MS in positive ion mode (Fig. 3-16) afforded the corresponding sulfated signals in negative ion mode.

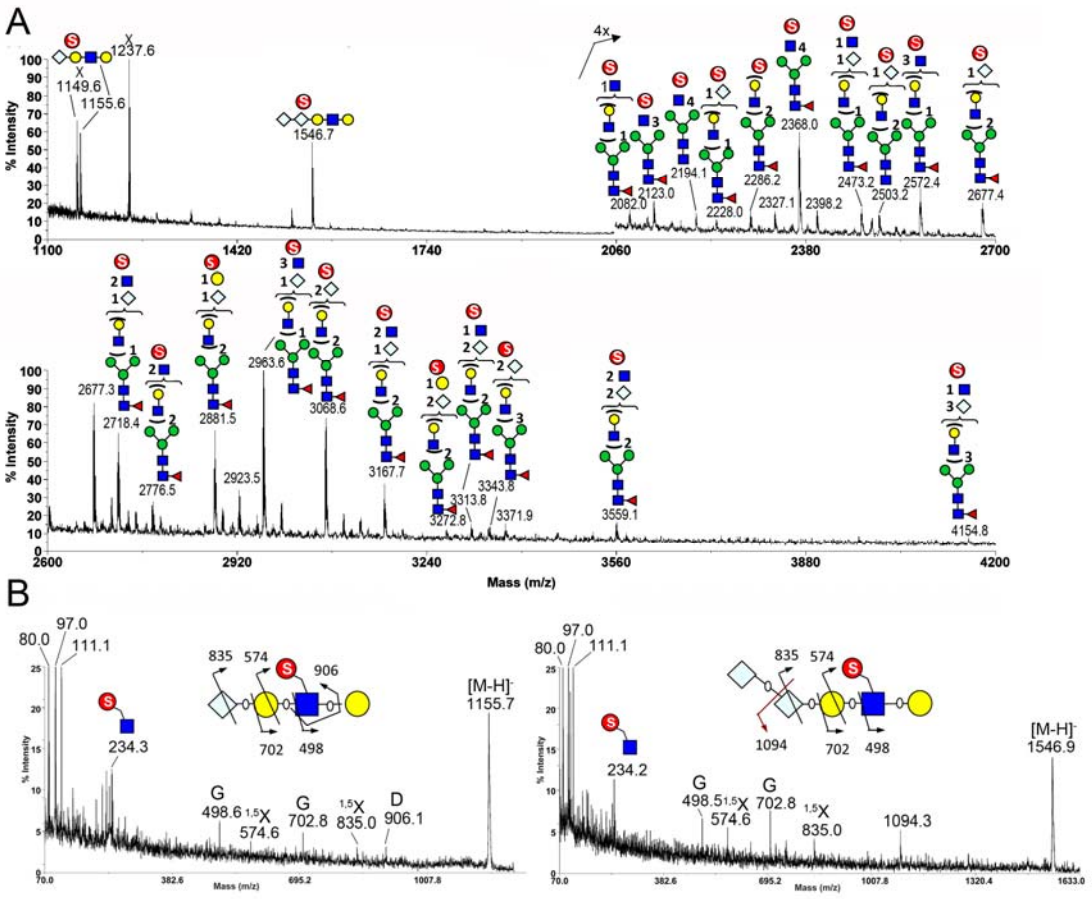



**Figure 3-23. MALDI MS (A) and MS/MS (B) analyses of permethylated N-glycans from BCL1 in negative ion mode after endo-β-galactosidase digestion.** All detected sulfated N-glycan core structure and released fragment were assigned as annotated **(**A). The identity of the released terminal mono- and di-sialyl sulfo LacNAc-Gal fragments were further validated by MS/MS sequencing (B). The major fragment ions were assigned as illustrated schematically.

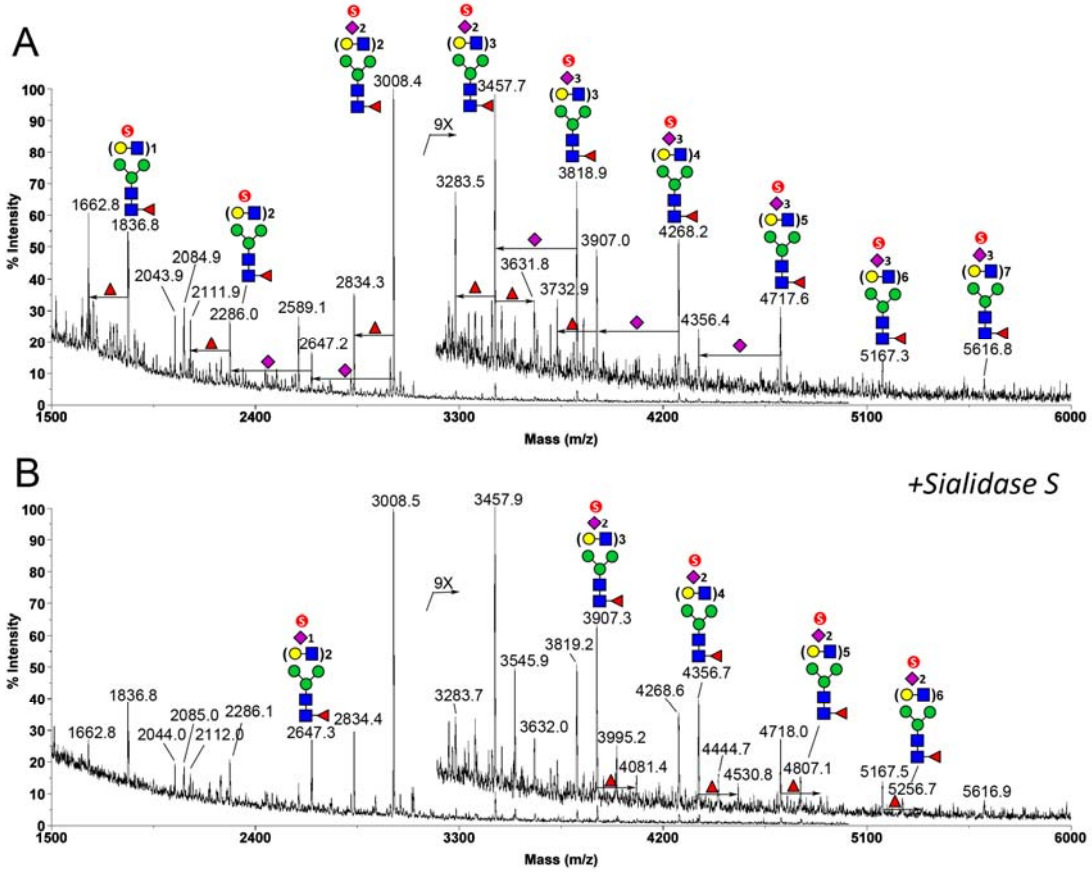



**Figure 3-24. Negative ion mode MALDI MS profiles of permethylated sulfated N-glycans from activated human B cells before (A) and after (B) sialidase S digestion.** The permethylated sample was cleaned stepwise by Spe-pak C18, $NH_2$ microtip and $ZipTip_{C18}$ before MALDI MS analysis in the negative ion mode, using DABP as matrix. All major peaks were assigned as annotated in (A), while additional more prominent peak produced after sialidase S digestion were additionally annotated in (B).

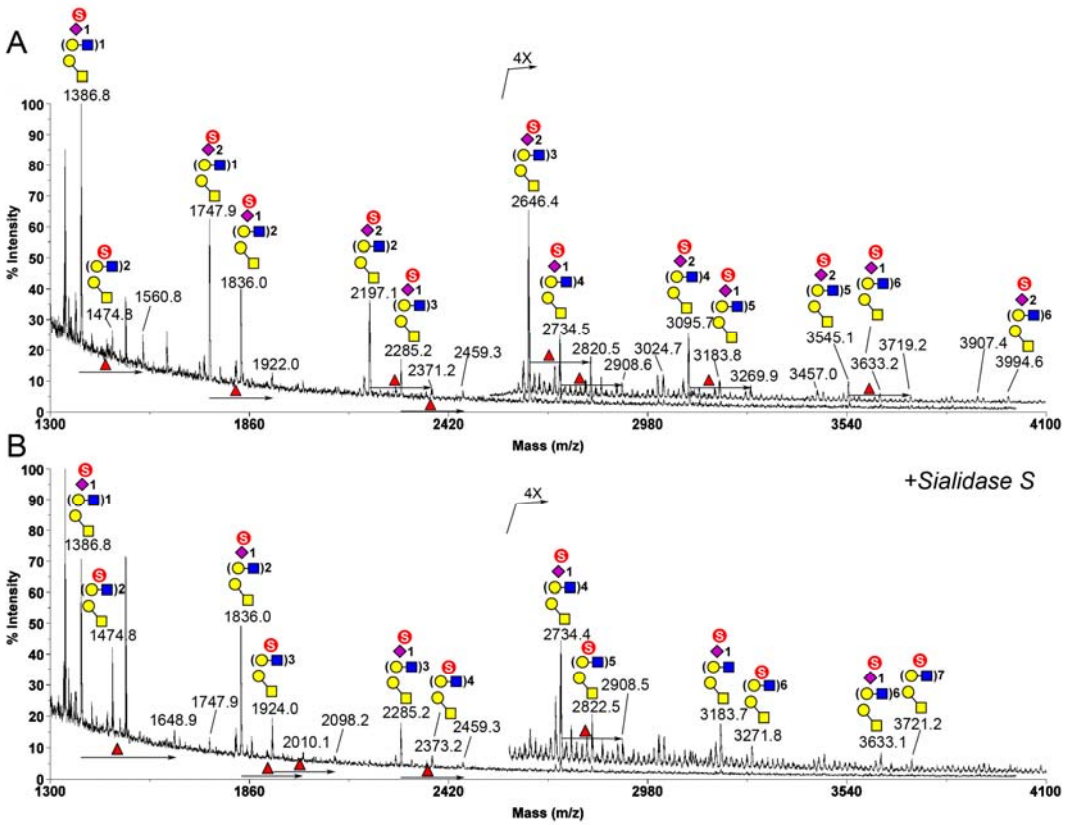



**Figure 3-25. Negative ion mode MALDI MS profiles of permethylated sulfated O-glycans from activated human B cells before (A) and after (B) sialidase S digestion.** The permethylated sample was processed similarly to N-glycans. All major peaks were tentatively assigned as annotated. Further confirmation by MS/MS is needed to validate the assignment.

**Table 3-1 Identification of candidate endothelial proteins carrying *N*-glycans extended with polylactosminoglycans, based on LC-MS/MS analysis of de-*N*-glycosylated peptides derived from tomato lectin (LEL)-enriched glycopeptides**

| Protein name | Description | *m/z* (observed) | *z* | Peptide score[1] | Matched peptide sequence[2] | Variable modifications[3] |
|---|---|---|---|---|---|---|
| EMP1_HUMAN | Epithelial membrane protein 1 | 1118.4718 | 2 | 111.22 | **N**CT**N**ISCSDSLSYASEDALK | 2 CAM[a)]; 1 (NQ)[b)] |
| | | 1118.9669 | 2 | 77.88 | **N**CT**N**ISCSDSLSYASEDALK | 2 CAM; 2 (NQ) |
| I13R2_HUMAN | Interleukin-13 receptor alpha-2 chain | 1072.5095 | 2 | 105.93 | EDDTTLVTATVE**N**ETYTLK | 1 (NQ) |
| PTPRJ_HUMAN | Receptor-type tyrosine-protein phosphatase eta | 748.3073 | 2 | 95.66 | VSD**N**ESSS**N**YTYK | 2 (NQ) |
| SCRB2_HUMAN | Lysosome membrane protein 2 | 919.4492 | 2 | 94.64 | ANIQFGD**N**GTTISAVSNK | 1 (NQ) |
| TM181_HUMAN | Transmembrane protein 181 | 798.8903 | 2 | 94.21 | VIQTSAA**N**FSL**N**NSK | 3 (NQ) |
| 4F2_HUMAN | 4F2 cell-surface antigen heavy chain | 719.3703 | 2 | 92.23 | SLVTQYL**N**ATGNR | 1 (NQ) |
| | | 805.8879 | 2 | 53.37 | DASSFLAEWQ**N**ITK | 1 (NQ) |
| S43A3_HUMAN | Solute carrier family 43 member 3 | 1009.4329 | 2 | 87.65 | DLCGPDAGPIG**N**ATGQADCK | 2 CAM; 1 (NQ) |
| ICAM2_HUMAN | Intercellular adhesion molecule 2 | 910.4265 | 2 | 84.51 | AAPAPQEATATF**N**STADR | 1 (NQ) |
| | | 698.8224 | 2 | 54.14 | G**N**ETLHYETFGK | 1 (NQ) |
| PSL2_HUMAN | Signal peptide peptidase-like 2A | 733.8373 | 2 | 84.38 | DM**N**QTLGD**N**ITVK | 2 (NQ); 1 Ox (M) |
| S12A2_HUMAN | Solute carrier family 12 member 2 | 1229.0406 | 2 | 80.93 | DATGNV**N**DTIVTELT**N**CTSAACK | 2 CAM; 2 (NQ) |
| S29A1_HUMAN | Equilibrative nucleoside transporter 1 | 868.4425 | 2 | 80.25 | LDMSQ**N**VSLVTAELSK | 1 (NQ) |
| LAMP1_HUMAN | Lysosome-associated membrane glycoprotein 1 | 905.9298 | 2 | 80.21 | **N**MTFDLPSDATVVL**N**R | 2 (NQ); 1 Ox (M) |
| | | 753.8756 | 2 | 64.03 | E**N**TSDPSLVIAFGR | 1 (NQ) |
| CTR1_HUMAN | High affinity cationic amino acid transporter 1 | 877.8841 | 2 | 79.01 | NWQLTEEDFG**N**TSGR | 1 (NQ) |
| | | 490.2235 | 2 | 48.02 | LCLN**N**DTK | 1 CAM; 2 (NQ) |
| NPC1_HUMAN | Niemann-Pick C1 protein | 1141.5058 | 2 | 75.59 | VD**N**ITDQFC**N**ASVVDPACVR | 2 CAM; 2 (NQ) |
| | | 957.9580 | 2 | 37.26 | LIAS**N**VTETMGI**N**GSAYR | 2 (NQ); 1 Ox (M) |
| CD59_HUMAN | CD59 glycoprotein | 901.9001 | 2 | 73.46 | TAV**N**CSSDFDACLITK | 2 CAM; 1 (NQ) |

| | | | | | | |
|---|---|---|---|---|---|---|
| TSN3_HUMAN | Tetraspanin-3 | 634.2811 | 2 | 73.43 | TY**N**GTNPDAASR | 1 (NQ) |
| MPZL2_HUMAN | Myelin protein zero-like protein 2 | 573.2935 | 2 | 64.75 | VLEAV**N**GTDAR | 1 (NQ) |
| HGNAT_HUMAN | Heparan-α-glucosaminide *N*-acetyltransferase | 758.3303 | 2 | 64.45 | YP**N**CTGGAAGYIDR | 1 CAM; 1 (NQ) |
| SYPL1_HUMAN | Synaptophysin-like protein 1 | 943.4515 | 2 | 63.58 | GQTEIQVNCPPAVTE**N**K | 1 CAM; 1 (NQ) |
| S36A4_HUMAN | Proton-coupled amino acid transporter 4 | 857.3727 | 2 | 61.6 | VFIS**N**ST**N**SSNPCER | 1 CAM; 2 (NQ) |
| PRIO_HUMAN | Major prion protein | 570.7562 | 2 | 61.57 | GE**N**FTETDVK | 1 (NQ) |
| SEPR_HUMAN | Seprase | 690.8228 | 2 | 59.64 | SV**N**ASNYGLSPDR | 1 (NQ) |
| NRAM2_HUMAN | Natural resistance-associated macrophage protein 2 | 620.3085 | 2 | 59.18 | D**N**STLAVDIYK | 1 (NQ) |
| LPP2_HUMAN | Lipid phosphate phosphohydrolase 2 | 670.3316 | 2 | 58.86 | V**N**CSVYVQLEK | 1 CAM; 1 (NQ) |
| BASI_HUMAN | Basigin | 590.6001 | 3 | 58.75 | ITDSEDKALM**N**GSESR | 1 (NQ); 1 Ox (M) |
| | | 491.2189 | 2 | 44.86 | ALM**N**GSESR | 1 (NQ); 1 Ox (M) |
| TM87A_HUMAN | Transmembrane protein 87A | 655.3169 | 2 | 58.02 | ENGT**N**LTFIGDK | 1 (NQ) |
| | | 643.8078 | 2 | 28.56 | LFQ**N**CSELFK | 1 CAM; 1 (NQ) |
| NPTN_HUMAN | Neuroplastin | 769.8266 | 2 | 57.78 | ENGMPMDIV**N**TSGR | 2 (NQ); 1 Ox (M) |
| SO3A1_HUMAN | Solute carrier organic anion transporter family member 3A1 | 1088.4488 | 2 | 57.77 | DVCAA**N**GSGGDEGPDPDLICR | 2 CAM; 1 (NQ) |
| SL9A7_HUMAN | Sodium/hydrogen exchanger 7 | 618.8454 | 2 | 56.78 | AFSTLLV**N**VSGK | 1 (NQ) |
| PCFT_HUMAN | Proton-coupled folate transporter | 601.2776 | 2 | 56.23 | FSADLGY**N**GTR | 1 (NQ) |
| NCKX6_HUMAN | Sodium/potassium/calcium exchanger 6 | 510.7423 | 2 | 54.67 | VCGL**N**VSDR | 1 CAM; 1 (NQ) |
| S15A4_HUMAN | Solute carrier family 15 member 4 | 714.3508 | 2 | 49.95 | LL**N**CTAPGPDAAAR | 1 CAM; 1 (NQ) |
| CD63_HUMAN | CD63 antigen | 571.7753 | 2 | 46.91 | N**N**HTASILDR | 2 (NQ) |
| ELTD1_HUMAN | EGF, latrophilin and seven transmembrane domain-containing protein 1 | 778.8700 | 2 | 45.13 | TTEFDT**N**STDIALK | 1 (NQ) |
| LAMP2_HUMAN | Lysosome-associated membrane glycoprotein 2 | 676.9969 | 3 | 44.36 | VASVININP**N**TTHSTGSCR | 1 CAM; 1 (NQ) |
| MRP4_HUMAN | Multidrug resistance-associated protein 4 | 883.9090 | 2 | 43.9 | QSML**N**VTVNGGG**N**VTEK | 3 (NQ); 1 Ox (M) |

| | | | | | | |
|---|---|---|---|---|---|---|
| VPP2_HUMAN | V-type proton ATPase 116 kDa subunit a isoform 2 | 618.3079 | 2 | 41.57 | MVLW**N**DSVVR | 1 (NQ); 1 Ox (M) |
| MPZL1_HUMAN | Myelin protein zero-like protein 1 | 582.7982 | 2 | 40.81 | EIFVA**N**GTQGK | 1 (NQ) |
| LCAP_HUMAN | Leucyl-cystinyl aminopeptidase | 924.8939 | 2 | 39.99 | EETLLYDS**N**TSSMADR | 1 (NQ); 1 Ox (M) |
| CA085_HUMAN | Uncharacterized protein C1orf85 | 542.7747 | 2 | 37.54 | TFA**N**GSLAFR | 1 (NQ) |
| EMP3_HUMAN | Epithelial membrane protein 3 | 1030.4383 | 2 | 37.5 | ESLNLWYDCTWN**N**DTK | 1 CAM; 1 (NQ) |
| TM206_HUMAN | Transmembrane protein 206 | 764.8620 | 2 | 36.91 | I**N**YTDPFS**N**QTVK | 2 (NQ) |
| TNR5_HUMAN | Tumor necrosis factor receptor superfamily member 5 | 587.3093 | 2 | 33.7 | DLVVQQAGT**N**K | 1 (NQ) |
| CLDN1_HUMAN | Claudin domain-containing protein 1 | 659.8095 | 2 | 31.95 | SPVQE**N**SSDLNK | 1 (NQ) |
| AT1B3_HUMAN | Sodium/potassium-transporting ATPase subunit beta-3 | 796.8850 | 2 | 31.79 | **N**LTVCPDGALFEQK | 1 CAM; 1 (NQ) |
| CD47_HUMAN | Leukocyte surface antigen CD47 | 629.3036 | 2 | 31.25 | DIYTFDGAL**N**K | 1 (NQ) |

1. With a Mascot P<0.05 significant threshold value (peptide ion score>27, no false positive hit from reverse database), all redundantly identified de-*N*-glycosylated peptides have been manually removed. Only the highest peptide scores corresponding to each of the unique sequences are tabulated here.

2. The Asn (N) residues presented in bold print correspond to the Asn in the consensus *N*-glycosylation site sequon, Asn-Xxx-Ser/Thr. For matched peptide sequences carrying more than one N or Q, no attempt was made to verify the deamidated residue. In most cases, the MS/MS data is insufficient to discriminate among the probable sites.

3. CAM denotes carbamidomethylation at Cys, (NQ) denotes deamidation at either Asn or Gln, Ox (M) denotes methionine oxidation. These variable modifications were identified to be present based on the MS and MS/MS data by Mascot and directly reported here without further validation.

# Chapter IV Discussion

Advanced mass spectrometry is the single most important enabling technique in both systematic and targeted glycomic approaches to define the structural basis of a diverse range of fundamental issues in glycobiology. Despite many recent work demonstrating its feasibility, practical utilities and potentials, current limitations remain in identifying less abundant but critically important epitopes/glyco-structures from a rapid first-screen glycomic analysis and to subsequently locate the glycotopes to particular glycan and protein carriers. This thesis work set out to address one of these analytical challenges, namely the polyLacNAcs. A full range of MS-based methodologies were developed and optimized using the endothelial cells EA.hy926 and HUVEC. In parallel, the occurrence of polyLacNAcs and other glycotopes on mouse and human B cells were investigated.

## 4.1 Technical Accomplishments

In summary, it was demonstrated that direct glycomic MS and MS/MS analysis of permethylated N-glycans by MALDI-TOF/TOF can be complemented by nano-ESI-$MS^n$ analysis on LTQ-Orbitrap. The former provides a direct picture detailing the glycomic complexity whereas the latter affords diagnostic neutral losses that can be utilized subsequently in total ion mapping. In the context of polyLacNAcs, implementation of TIM was shown to be highly efficient in filtering out noises for a clearer mapping of the carrier glycans. However, neither methods of direct MS analysis afforded additional information with respect to the branching pattern and attachment sites of the polyLacNAcs. Both endo-β-galactosidase digestion and Smith degradation were demonstrated here to be useful in allowing selective cleavages of the polyLacNAc-carrying N-glycans into fragments that can be individually characterized further in more details to obtain the missing information.

In particular, endo-β-galactosidase digestion allows identification of branched polyLacNAcs, which was shown to be present only in EA.hy926 but not in HUVEC or BCL1. It also provided the fragments carrying the non-reducing terminal glycotopes on the polyLacNAc chains. In the case of BCL1, these corresponded to terminal LacNAc-Gal with and without α-Gal capping, and monosialylated and disialylated LacNAc-Gal with and without additional sulfate on the GlcNAc, which could not have been easily recognized in the direct first screen MS profiling of the N-glycans. The remaining N-glycan core structures recovered indicated that polyLacNAcs extension was not restricted to C6 of Man on the 6-arm in all cases examined. Through Smith degradation, the β6-arm polyLacNAc chains could be specifically cleaved off, identified and distinguished from that located at other positions by MS due to its unique mass. Extending to glycoproteomics, the Lycopersicon esculentum lectin based enrichment strategy was optimized at glycoprotein, glycopeptide and glycan

levels, leading to identification in EA.hy926 of several protein carriers including endothelial membrane protein 1, LAMP1, LAMP2, 4F2, and I-CAM2.

### 4.2 Outstanding Technical Limitations and Prospects

#### 4.2.1 Application of endo-β-galactosidase to quantify the length of polyLacNAcs

Branched polyLacNAcs were detected in EA.hy926 but not in HUVEC and BCL1 although EA.hy926 was derived from fusion of HUVEC with A549 and their biological behaviors are almost the same. The internal sulfation at C6 of Gal, fucosylation at C3 of GlcNAc, or other modifications might have interfered with the digestion of this enzyme but further work is needed to distinguish this possibility from a simple lack of branching. In this context, a non-enzymatic method such as Smith degradation will be ideal to release intact the entire polyLacNAc chain without the concern of incomplete enzyme digestion due to extra modifications. In principle, this approach will also allow us to determine the length of the polyLacNAc chain, especially the one attached to C6 of the mannose on the 6-arm, which is accorded with many biological functions. However, the results of Smith degradation are still too complex to allow a full assignment of the many reaction products. The conditions of reaction, theoretical and side products may be needed further verification using purified glycans as standard, and it would be a good method to help identify polyLacNAc although it will be hard to get the pure polyLacNAc standards.

The length of polyLacNAc and the percentage of each terminal epitope are very important issues to help us figuring out the potential biological functions of these epitopes. To derive this structural information, we need to quantify the individual glycan fragments released by endo-β-galactosidase. Permethylation of glycans can stabilize the sialic acids residues in acidic glycans and enable all glycans to become more uniformly ionized by reducing the variation of the chemical properties among the glycans such as the hydroxyl or amino groups. It can facilitate the relative quantitation of the individual glycans in a mixture. In addition, methylated glycans ionize more efficiently than native glycans, and due to their hydrophobic nature, are easily separated from salts and other contaminations that may affect the MS analysis (Dell et al., 1994; Kang et al., 2005). However, there are still some differences in ionization efficiencies among the various components present in the sample even after permethylation (Chen and He, 2005), as well as the possible presence of different amounts of interfering contaminants among the samples. Although there are some reports on the use of MALDI-TOF MS for quantitative studies of oligosaccharides and protein glycosylation (Pitt and Gorman, 1997), there are still substantial limitations on the use of MS for quantitative purposes.

As an alternative to MS, various chromophore or fluorophore have been used to label glycans via reductive amination of. An advantage of this labeling approach is the

stoichiometric attachment of one label per glycan, allowing a direct quantitation based on fluorescence or UV-absorbance intensity. The most widely applied labels are 2-AB, 2-AA, and PA. Among them, pyridylamination, as a fluorescent-tagging method for glycans, is widely used for structural analysis and measurement of glycans. The databases for structural assignment based on standardized elution positions have been developed as well. In this thesis work, endo-β-galactosidase was mostly applied at the glycan level for the glycans released from endothelia cells or B cells. However, glycan compositions after enzyme digestion are too complicated to allow a direct and efficient fluorometric HPLC mapping by tagging the whole mixtures indiscriminately. Although crude fractionation of the smaller fragments from the remaining N-glycan core structures by P2 gel filtration is possible, it is also rather inefficient, involving multiple steps of sample collections and screening.

To overcome such uncertainties, endo-β-galactosidase digestion may be applied at the cell level or tryptic peptide level in the future. In theory, we can easily fractionate the released terminal fragments from the proteins or glycopeptides after enzyme digestion. This has indeed been accomplished here by passing through a Sep-pak C18 cartridge (Section 3.4.3) although further fluorescent-tagging was not undertaken. In addition, the remaining glycopeptides can also be further characterized and identified. There are still some difficulties needed to overcome such as maximizing the efficiency of enzyme on extracted membrane proteins which may have precipitated and hard to re-dissolve, and calibrating the elution time of each specific fragment on chosen HPLC system, etc. Take the branched fragments released from EA.hy926 or the disialyl-LacNAc-Gal with or without sulfation from BCL1 as examples, there is still no commercially available standard and it is hard or impossible to synthesize all of them. Thus, we may need to purify each peak separately and verify by MALDI MS. However, once fully set-up, the intensity of each terminal fragment may tell us directly the relative amounts of every epitope, whereas the amounts of internal GlcNAc-Gal relative to total amount of terminal epitopes may allow the average length of polyLacNAcs in cells to be estimated for comparative analysis at the highest sensitivity possible.

### 4.2.2 The glycoproteomics of polyLacNAcs

The methodologies for detection of cancer and cancer risk based on analysis of human tumor and control tissues employing lectin separation, followed by identification of bound proteins using shotgun proteomics has been developed for identification of potential cancer biomarkers (Abbott et al., 2008a; Abbott et al., 2008b). The commonly used lectin affinity chromatography protocols involves immobilization of lectins onto various forms of solid supports such as agarose and silica in a number of chromatographic formats, including tubes, columns and microfluidic channels (Mao et al., 2004; Xiong et al., 2003 ). Although membrane

fractions are already enriched in glycoproteins, they normally still contain high numbers of non-glycosylated proteins such as the highly abundant structural components like actin. A further purification of plasma membrane glycopeptides is necessary to determine N-glycosylation sites in larger scale by avoiding suppression effects during mass spectrometric analysis. Comparing the results between one- and two-step enrichment, there are not only more candidates but, importantly, more non-specific bindings resulting from one-step enrichment. Based on analysis of the released N-glycans, the glycan profile of bound fraction from one-step enrichment was found to be almost the same as that with two-step enrichment except that there were some high mannose and smaller complexed type N-glycans such as bi-antennary in bound fractions.

Unambiguous identification of glycopeptides based on direct LC-MS/MS analysis was not possible because, in general, peptide sequence-informative b and y ions were hardly detectable, and delineation of the MS/MS spectra to identify the m/z of the peptide core was problematic even when the quality of the spectra was good, especially in the case of the polyLacNAcs because of the glycopeptide may be too large to be detected in the first place. At the glycan level, N-glycans having polyLacNAcs may simply be isolated by virtue of their larger size but this is not applicable to common glycoproteomic workflow which may generate very large non-glycosylated peptides. In this context, the application of endo-β-galactosidase to glycopeptides may help us to minimize the size of glycopeptides by trimming away the attached polyLacNAcs from the core. This would facilitate identification of the glycosylation site for glycans carrying polyLacNAc, provided the samples do not contain significant amount of glycoforms terminating with agalactosylated GlcNAc which cannot be distinguished from termini created by endo-β-galactosidase digestion.

While an efficient, automated direct glycopeptide sequencing and identification at the glycoproteomic scale has yet emerged, current approach including this thesis work has relied mostly on analysis of de-N-glycosylated peptides for glycosylation site and protein ID. Often, the proteomic identification results include several glycoproteins that are commonly identified by a range of very different approaches. For example, 4F2 identified in this work from both one- and two-step LEL enrichment was also identified through many other enrichments targeting different glycan epitope and/or different kinds of cells. This may result from some glycoproteins being most abundantly expressed, most readily extracted, or the particular deglycosylated peptides most readily detected by LC-MS/MS, coupled with insufficient lectin capture specificities, or simply contributed by non-specific bindings. In addition, there are only 12 proteins identified by both one- and two-step enrichment from a pool of 91 proteins identified in one-step enrichment and 52 proteins identified in two-step enrichment. This reflects the somewhat randomness of the process and considerable

experimental variations, which in turn requires repetitive LC-MS/MS runs to lend credibility and statistical significance. In conclusion, the ID results only give us some clues about the "candidate" and not the "verified" carrier of polyLacNAc for us to confidently develop a functional model.

### 4.3 The biological implications of polyLacNAc

In this thesis work, polyLacNAcs were identified to be present in endothelial cell and B cells. More specifically, direct structural evidences were obtained for the N-glycans of EA.hy926, HUVEC, murine B lymphoma (BCL1), N- and O-glycans of human activated B cells. The occurrence of polyLacNAc branching was found to be different between HUVEC and EA.hy926 although the immortalized EA.hy926 cell line was derived from HUVEC and that the biological behaviors between these two cell lines are similar.

The human blood group i and I antigens are determined by linear and branched polyLacNAc structures, respectively. The acquisition of I-branches is important, since the additional N-acetyllactosamine side chains can have extra functional terminal structures. It has been demonstrated that multivalent sialyl Lewis X polyLacNAcs inhibit L-selectin-mediated binding and the rejection of organ transplants with much better efficacy than monovalent sialyl Lewis X polyLacNAcs (Toppila et al., 1997; Turunen et al., 1995). In early mouse embryonic development, embryos express I antigen, which is gradually replaced with i antigen during development. Whether the presence of branched polyLacNAc will influence the ability of tube formation in endothelial cells or not may be an interesting issue, and whether the inflammation signal such as TNF-α or LPS can influence the incidence of branching in HUVEC still needs to be verified. The relationships and biological behaviors of HUVEC and EA.hy926 may be good models to study the activity and specificity of glycosyltransferases responsible for the biosynthesis of branched polyLacNAcs besides the commonly used human PA-1 embryonic carcinoma cells. In addition, comparing the relationship of glycan structures and ability of angiogenesis between isolated viable tumor endothelial cells and normal endothelial cells of uninvolved tissue may also provide an *in vitro* model system to elucidate epigenetic effects on angiogenesis in cancer and to optimize antiangiogenic therapy.

Apart from the spatial and valency or density effects in presenting the terminal epitopes on extended and/or branched polyLacNAc, it is unclear if nature evolves specific lectins that will specifically recognize the polyLacNAc or LacNAc elements *per se* in any biological process. The most discussed and studied in this context are the galectins, which are known to have a basic affinity for LacNAc. Galectins are animal lectins that bind β-galactosides through their conserved carbohydrate recognition domains (CRDs) (Hirabayashi et al., 2002; Yang et al., 2008). To date, 15 mammalian

galactins (Galectin-1 to -15) have been identified formally in a wide variety of tissues from several species and more are likely to be discovered. Galectins have been implicated in numerous biologic processes, including cell adhesion, migration, proliferation, differentiation, transformation, apoptosis, angiogenesis, and immune responses . They can be found in the cytoplasm, nucleus, and extracellular environment (Liu et al., 2002).

In the immune system, galectin-1, -3, -4, -8, -9, -10 may affect the function of T cells. Galectin-1, -3, -8, and -9 could also modulate the differentiation or function of B cells (Liu and Rabinovich, 2010). In addition to B and T cells, they can also influence the function of neutrophils, monocytes, macrophages, dentritic cells, mast cells, and eosinophils. It seems that almost all the immune cells can interact with galectins in the immune system which may result from the widely expressed β-galactosides or LacNAc which are the presumed ligands of galectins in many kinds of cells. However, there seldom is any direct evident for the presence of polyLacNAc in immune cells especially primary cells from human or mice. Importantly, the functional interacting motifs between glycans on immune cells with galectins are not only limited to or dependent on polyLacNAc but also the number of LacNAc units within a polyLacNAc chain, and the presence and linkage of sialic acids or location of sulfates on the glycans.

In addition, the binding affinity between an individual galectins and its minimal glycan ligand is typically low, so the propensity of oligomerization of galectins can enhance avidity and lead to the formation of stabilized lattice. Most galectins exhibit diminished binding to β-galactosides capped with α2,6-sialic acid. For example, galectin-1 could bind unsialylated and α2,3-sialylated polyLacNAcs with approximately equal affinity, whereas binding was completely inhibited by α2,6-sialylation (Leppanen et al., 2005). It was also reported that α2,6-sialylation blocked the interaction of Gal-1, -2, and -3 with LacNAc in glycan microarrays (Stowell et al., 2008). To date, there is still lack of information about the affinity between branched and non-branched polyLacNAcs interacting with galectins. According to the theory of lattice formation, the branched one should provide more ligands to be bound and recognized by galectins, but this should be further verified.

For B cells, it has been reported that galectin-1 plays an important role during B cell development, differentiation, and survival (Espeli et al., 2009; Gauthier et al., 2002; Rossi et al., 2006; Tabrizi et al., 2009; Tsai et al., 2008; Zuniga et al., 2001). Galectin-3 contributes to IL-4-induced downregulation of Blimp-1, which is essential for the development of memory B cells (Acosta-Rodriguez et al., 2004). Besides, recombinant galectin-9 also induces apoptosis in a human B cell line (Okudaira et al., 2007). However, just as in any other physiological settings, many different lectins are expected to interact in concert with a plethora of glycan ligands on B cells. After the

initial one to one study in detail for the induced immune responses such as the antibody secretion, cell differentiation, and possibly subsequent signaling, a full delineation of the interacting partners including the glyco-epitope carrier will be essential to further our understanding beyond a reductionist's approach.

**4.4 Disialylation and sulfation on B lymphocytes**

In addition to the presence of linear polyLacNAcs on B cells, this work has made an important contribution to the understanding of B cell glycome by identifying an unusual terminal disiayl motif capping on both the polyLacNAc extended and non-extended N-glycans and the simple core 1 O-glycans of BCL1. BCL1 is the first reported spontaneous B cell tumor line that arose in a 2-year-old BALB/c mouse (Knapp et al., 1979a; Knapp et al., 1979b; Slavin and Strober, 1978; Strober et al., 1979; Vitetta et al., 1979). It expresses CD5, surface IgM, Mac-1, CD43 and low level of B220, and is likely to have B-1a cell origin. Whether the primary B-1a cells express disialyl motif is still unknown and need to be further investigated. On the other hand, using the DMB-sialic acid fluorometric assay, this work has further shown that the expressed level of disialyl motif on naïve primary mouse B cells increases when stimulated to differentiate by LPS-induced T-independent (TI) or IL4/IL5/CD40L-induced T-dependent (TD) pathway. Coinciding with this increase in the disialyl content, the sialyltransferase thought to be responsible for making the disialyl motif, ST8Sia VI, was also shown to be the most significantly up-regulated among all other ST8s probed in this induced plasma cell differentiation experimental model. A further knockdown of ST8Sia VI in BCL1 indeed led to a reduced expression of disialylation, on both N- and O-glycans.

For the important linkage of sialic acids, results from MS/MS analysis on the endo-β-galactosidase released fragment and susceptibility to sialidase S digestion both suggested that the first sialic acid of the disialyl acid is α2,3-linked to the Gal. However, both 3- and 6-linked Gal were present in the total N-glycans in BCL1 based on GC-MS linkage analysis. It suggested that the terminal sialylation of polyLacNAc occurred preferentially via α2,3-linkage, whereas the non-extended LacNAc units can also be α2-6 sialylated. This situation has previously been reported, namely the termini of long polyLacNAc was α2,3-sialylated, whereas the short side chain was mainly α2,6-sialylated (Fukuda et al., 1984a). In addition, the *in vitro* reactivity of expressed mouse ST8Sia VI was 4.6 fold higher with 3’-sialyl-N-acetyllactosamine than 6’-sialyl-N-acetyllactosamine (Takashima et al., 2002). The results reported here are thus consistent with previous data although the linkage preference of ST8Sia VI in vivo and its physiological relevance still need to be investigated.

Our DMB-sialic acid analysis suggests that the expression of disialic acid increases during mouse primary B cell differentiation. The congregation of sialic acids

and fluorescent detection of DMB labeled sialic acids is more sensitive than MALDI MS detection of dispersed permethylated glycan due to the limitation of sample amount. On one hand, efforts are ongoing to try to obtain MS data on the glycan structures so as to better define the actual glycan carriers by further improvement in analytical methodology. On the other hand, it may be more relevant to directly go for human system. We have therefore also started to look at human B cells. In the first screen of N- and O-glycans from activated human B cells by MALDI MS, there was indeed evidence for polyLacNAc in both N- and O-glycans, as well as obvious signals for core 1 O-glycans conjugated with three Neu5Ac. The latter is indicative of the presence of disialyl motif in O-glycans of human activated B cells although the corresponding information on the probable presence of disialic acid on the N-glycans is still lacking. The use of the more sensitive DMB-sialic acid analysis on human B cells may help us to identify the expression of disialyl epitopes in the future and extend the biological relevance to human system.

The sialic acid-binding immunoglobulin-like lectins (Siglecs) are a recently designated superfamily of cell surface molecules. It is clear that Siglecs in the immune system have the potential to mediate cell-cell interactions and signaling functions. However, defining their precise functions and determining which ligands are biologically relevant post an important challenge. It is known that human Siglec-7 and Siglec-11 prefer sialosides with the Neu5Acα2-8Neu5Ac structure (Yamaji et al., 2002). Similar to Siglec-7, recombinant murine Siglec-E (mSiglec-E) preferred α2,8-linked disialic acid over α2,3- and α2,6-linked sialic acids from the glycan binding assay (Zhang et al., 2004). Depending on the expression sites of mSiglec-E, it is likely to modulate the functions of several types of effector cells such as peritoneal cavity macrophages, subsets of mature NK cells, splenic dendritic cells, peritoneal cavity B-1a-like cells, transitional T2, and marginal zone B cells. Any interaction and regulation between mSiglec-E and disialyl epitope *in vivo* in immune cells is anticipated to be a complicated and important story, especially in innate immunity.

Moreover, this is to be layered on top the well known regulatory roles of CD22, another Siglec that recognizes α2-6 sialyl LacNAc. CD22 is an important inhibitory co-receptor of B cell receptor-mediated signaling. Current literature and understanding imply that the preferred ligand for CD22, α2-6 sialyl LacNAc, is mostly carried on itself, forming *cis*-interactions. More recently, it was shown that 6-sulfated α2-6 sialyl LacNAc is, in fact a better ligand in human (Kimura et al., 2007). Our MALDI MS detection of sulfo sialylated glycans from activated human B cells after digestion with sialidase S indicated that there are indeed sulfated α2-6 sialyl LacNAc on both N- and O-glycans (Fig. 3-24, 25), possibly also on the termini of polyLacNAcs. This is the first structural evidence for the presence of this preferred CD22 ligand on human B cells although detailed MS/MS analysis has not been performed due to the limitation of sample amount.

A series of sialoglycoprotein ligands on activated T and B cells are recognized by recombinant CD22, including CD45, a major leukocytes tyrosine phosphatase (Baum et al., 1996; Kelm et al., 1994; Sgroi et al., 1993; Stamenkovic et al., 1991). Although CD45 is a putative ligand of CD22, it is neither required as a *cis* ligand for masking the binding of multivalent sialoside probes, nor as a *trans* ligand for recruitment of CD22 to sites of cell contact in mouse (Collins et al., 2004). Yet, it has been widely implicated in cell proliferation, signaling and differentiation and is associated with the B cell receptor during signaling. B cells predominantly express the higher molecular weight (200-220kDa) form, which contains exon 4, 5, and 6 (RA, RB, RC, respectively). Recently, it was reported that regulated O-linked glycosylation of CD45RB especially sialylation can be used to follow B cell differentiation and that this regulation may be involved in fine-tuning antigen signaling in human (Koethe et al., 2011). In this work, DMB-sialic acid analysis of the immunoprecipitated CD45 from BCL1 suggested that it carried the disialyl motif on its O-glycans. It remains to be demonstrated if similar disialyl motif is also found on the CD45 from primary mouse or human B cells, or any other glycoprotein carriers. In addition, whether the elevated disialyl content in the process of differentiated human B cells or whether the elevated expression of disialic acids on CD45 would influence the interaction with CD22 in mature B cells to modulate the immune response of T or B cells, await further investigation.

During B cell differentiation, it has been reported that extracellular galectins-1 and -8 bind more significantly to mature B cells than to plasma B cells (Tsai et al., 2008). In addition to the traditional ligands such as Galβ1-4GlcNAc epitope, the sulfated LacNAc can also be the good ligands for galectin-1 including 6-sulfated LacNAc and 3'-sulfated LacNAc (Tsai et al., 2011). Both the sulfate and sialic acids can provide a negative charge on the termini of glycans although the steric effect between sialic acid and sulfate is different. The interaction between galectins and terminal negative charge, with or without conformational specificity, still needs to be studied. In addition, we have shown that there are both α2,3- and α2,6-linked sialic acids in human and mouse B cells but we still have no structural evidence about the expression or length of polyLacNAc in primary B cells during B cell differentiation in mouse due to the limited sample amount. The elevated α2-6 sialic acid modification during B cell differentiation would interfere with the interaction of B cells and galectins. However, the α2,8-disialic acids increase along with the elevated expression of α2,6-sialic acids. Whether the decreased interaction with galectins correlates with the increased interaction with Siglecs such as CD22 and Siglec-7 (mSiglec-E in mouse) remain unknown. Galectins may play an important role in early part of B cell differentiation and Siglecs may participate the later part of B cell differentiation. In addition, the expression of 6-sulfated α2,6-sialyl LacNAc may further complicate the picture by serving as competitive and better ligands for galectins and Siglecs in modulating immune response. In conclusion, the structural information on the glycans as supported by advanced mass spectrometry and other

chemical analysis developed here will provide a good basis for future glycobiology studies.

## Chapter V References


Abbott, K.L., Aoki, K., Lim, J.M., Porterfield, M., Johnson, R., O'Regan, R.M., Wells, L., Tiemeyer, M., and Pierce, M. (2008a). Targeted glycoproteomic identification of biomarkers for human breast carcinoma. J Proteome Res *7*, 1470-1480.

Abbott, K.L., Nairn, A.V., Hall, E.M., Horton, M.B., McDonald, J.F., Moremen, K.W., Dinulescu, D.M., and Pierce, M. (2008b). Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. Proteomics *8*, 3210-3220.

Acosta-Rodriguez, E.V., Montes, C.L., Motran, C.C., Zuniga, E.I., Liu, F.T., Rabinovich, G.A., and Gruppi, A. (2004). Galectin-3 mediates IL-4-induced survival and differentiation of B cells: functional cross-talk and implications during Trypanosoma cruzi infection. J Immunol *172*, 493-502.

Agrawal, B.B., and Goldstein, I.J. (1965). Specific binding of concanavalin A to cross-linked dextran gels. Biochem J *96*, 23contd-25c.

Amano, M., Galvan, M., He, J., and Baum, L.G. (2003). The ST6Gal I sialyltransferase selectively modifies N-glycans on CD45 to negatively regulate galectin-1-induced CD45 clustering, phosphatase modulation, and T cell death. J Biol Chem *278*, 7469-7475.

Angata, K., and Fukuda, M. (2003). Polysialyltransferases: major players in polysialic acid synthesis on the neural cell adhesion molecule. Biochimie *85*, 195-206.

Angata, K., Suzuki, M., and Fukuda, M. (1998). Differential and cooperative polysialylation of the neural cell adhesion molecule by two polysialyltransferases, PST and STX. J Biol Chem *273*, 28524-28532.

Angata, T., and Varki, A. (2002). Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chem Rev *102*, 439-469.

Aoki, K., Perlman, M., Lim, J.M., Cantu, R., Wells, L., and Tiemeyer, M. (2007). Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. J Biol Chem *282*, 9127-9142.

Atrih, A., Richardson, J.M., Prescott, A.R., and Ferguson, M.A. (2005). Trypanosoma brucei glycoproteins contain novel giant poly-N-acetyllactosamine carbohydrate chains. J Biol Chem *280*, 865-871.

Avril, T., North, S.J., Haslam, S.M., Willison, H.J., and Crocker, P.R. (2006). Probing the cis interactions of the inhibitory receptor Siglec-7 with alpha2,8-disialylated ligands on natural killer cells and other leukocytes using glycan-specific antibodies and by analysis of alpha2,8-sialyltransferase gene expression. J Leukoc Biol *80*, 787-796.

Babu, P., North, S.J., Jang-Lee, J., Chalabi, S., Mackerness, K., Stowell, S.R., Cummings, R.D., Rankin, S., Dell, A., and Haslam, S.M. (2009). Structural characterisation of neutrophil glycans by ultra sensitive mass spectrometric glycomics methodology. Glycoconj J *26*, 975-986.

Bateman, A.C., Karamanska, R., Busch, M.G., Dell, A., Olsen, C.W., and Haslam, S.M. (2010). Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAc{alpha}2-6 glycans. J Biol Chem *285*, 34016-34026.

Baum, L.G., Derbin, K., Perillo, N.L., Wu, T., Pang, M., and Uittenbogaart, C. (1996). Characterization of terminal sialic acid linkages on human thymocytes.

Correlation between lectin-binding phenotype and sialyltransferase expression. J Biol Chem *271*, 10793-10799.
Bierhuizen, M.F., Mattei, M.G., and Fukuda, M. (1993). Expression of the developmental I antigen by a cloned human cDNA encoding a member of a beta-1,6-N-acetylglucosaminyltransferase gene family. Genes Dev *7*, 468-478.
Brewer, C.F., Miceli, M.C., and Baum, L.G. (2002). Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. Curr Opin Struct Biol *12*, 616-623.
Callaghan, J.M., Toh, B.H., Pettitt, J.M., Humphris, D.C., and Gleeson, P.A. (1990). Poly-N-acetyllactosamine-specific tomato lectin interacts with gastric parietal cells. Identification of a tomato-lectin binding 60-90 X 10(3) Mr membrane glycoprotein of tubulovesicles. J Cell Sci *95 ( Pt 4)*, 563-576.
Callaghan, J.M., Toh, B.H., Simpson, R.J., Baldwin, G.S., and Gleeson, P.A. (1992). Rapid purification of the gastric H+/K(+)-ATPase complex by tomato-lectin affinity chromatography. Biochem J *283 ( Pt 1)*, 63-68.
Canis, K., McKinnon, T.A., Nowak, A., Panico, M., Morris, H.R., Laffan, M., and Dell, A. (2010). The plasma von Willebrand factor O-glycome comprises a surprising variety of structures including ABH antigens and disialosyl motifs. J Thromb Haemost *8*, 137-145.
Carlsson, S.R., and Fukuda, M. (1990). The polylactosaminoglycans of human lysosomal membrane glycoproteins lamp-1 and lamp-2. Localization on the peptide backbones. J Biol Chem *265*, 20488-20495.
Chakraborty, A.K., and Pawelek, J.M. (2003). GnT-V, macrophage and cancer metastasis: a common link. Clin Exp Metastasis *20*, 365-373.
Chen, H., and He, M. (2005). Quantitation of synthetic polymers using an internal standard by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Am Soc Mass Spectrom *16*, 100-106.
Chen, S.C., Huang, B., Liu, Y.C., Shyu, K.G., Lin, P.Y., and Wang, D.L. (2008). Acute hypoxia enhances proteins' S-nitrosylation in endothelial cells. Biochem Biophys Res Commun *377*, 1274-1278.
Cheresh, D.A., Pierschbacher, M.D., Herzig, M.A., and Mujoo, K. (1986). Disialogangliosides GD2 and GD3 are involved in the attachment of human melanoma and neuroblastoma cells to extracellular matrix proteins. J Cell Biol *102*, 688-696.
Chokhawala, H.A., Yu, H., and Chen, X. (2007). High-throughput substrate specificity studies of sialidases by using chemoenzymatically synthesized sialoside libraries. Chembiochem *8*, 194-201.
Ciucanu, I., and Kerek, F. (1984). A simple and rapid method for the permethylation of carbohydrates. Carbohydrate Research *131*, 209-217.
Collins, B.E., Blixt, O., DeSieno, A.R., Bovin, N., Marth, J.D., and Paulson, J.C. (2004). Masking of CD22 by cis ligands does not prevent redistribution of CD22 to sites of cell contact. Proc Natl Acad Sci U S A *101*, 6104-6109.
Costello, B.D.a.C.E. (1988). A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates Glycoconjugate Journal *5*, 397-409.
Cummings, R.D., and Kornfeld, S. (1982). Characterization of the structural determinants required for the high affinity interaction of asparagine-linked

oligosaccharides with immobilized Phaseolus vulgaris leukoagglutinating and erythroagglutinating lectins. J Biol Chem *257*, 11230-11234.
Cummings, R.D., and Kornfeld, S. (1984). The distribution of repeating [Gal beta 1,4GlcNAc beta 1,3] sequences in asparagine-linked oligosaccharides of the mouse lymphoma cell lines BW5147 and PHAR 2.1. J Biol Chem *259*, 6253-6260.
Cummings, R.D., Trowbridge, I.S., and Kornfeld, S. (1982). A mouse lymphoma cell line resistant to the leukoagglutinating lectin from Phaseolus vulgaris is deficient in UDP-GlcNAc: alpha-D-mannoside beta 1,6 N-acetylglucosaminyltransferase. J Biol Chem *257*, 13421-13427.
Dang, A.M., Phillips, J.A., Lin, T., and Raveche, E.S. (1996). Altered CD45 expression in malignant B-1 cells. Cell Immunol *169*, 196-207.
Dell, A. (1987). F.A.B.-mass spectrometry of carbohydrates. Adv Carbohydr Chem Biochem *45*, 19-72.
Dell, A., Reason, A.J., Khoo, K.H., Panico, M., McDowell, R.A., and Morris, H.R. (1994). Mass spectrometry of carbohydrate-containing biopolymers. Methods Enzymol *230*, 108-132.
Demetriou, M., Granovsky, M., Quaggin, S., and Dennis, J.W. (2001). Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. Nature *409*, 733-739.
Denecke, J., Kranz, C., Nimtz, M., Conradt, H.S., Brune, T., Heimpel, H., and Marquardt, T. (2008). Characterization of the N-glycosylation phenotype of erythrocyte membrane proteins in congenital dyserythropoietic anemia type II (CDA II/HEMPAS). Glycoconj J *25*, 375-382.
Dennis, J.W., Granovsky, M., and Warren, C.E. (1999). Glycoprotein glycosylation and cancer progression. Biochim Biophys Acta *1473*, 21-34.
Dennis, J.W., Laferte, S., Waghorne, C., Breitman, M.L., and Kerbel, R.S. (1987). Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. Science *236*, 582-585.
Dennis, J.W., Nabi, I.R., and Demetriou, M. (2009). Metabolism, cell surface organization, and disease. Cell *139*, 1229-1241.
Diamandis, E.P. (2004). Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol Cell Proteomics *3*, 367-378.
Donnelly, E.H., and Goldstein, I.J. (1970). Glutaraldehyde-insolubilized concanavalin A: an adsorbent for the specific isolation of polysaccharides and glycoproteins. Biochem J *118*, 679-680.
Earl, L.A., Bi, S., and Baum, L.G. (2010). N- and O-glycans modulate galectin-1 binding, CD45 signaling, and T cell death. J Biol Chem *285*, 2232-2244.
Edgell, C.J., McDonald, C.C., and Graham, J.B. (1983). Permanent cell line expressing human factor VIII-related antigen established by hybridization. Proc Natl Acad Sci U S A *80*, 3734-3737.
Emeis, J.J., and Edgell, C.J. (1988). Fibrinolytic properties of a human endothelial hybrid cell line (Ea.hy 926). Blood *71*, 1669-1675.
Espeli, M., Mancini, S.J., Breton, C., Poirier, F., and Schiff, C. (2009). Impaired B-cell development at the pre-BII-cell stage in galectin-1-deficient mice due to inefficient pre-BII/stromal cell interactions. Blood *113*, 5878-5886.
Fan, Y.Y., Yu, S.Y., Ito, H., Kameyama, A., Sato, T., Lin, C.H., Yu, L.C., Narimatsu, H., and

Khoo, K.H. (2008). Identification of further elongation and branching of dimeric type 1 chain on lactosylceramides from colonic adenocarcinoma by tandem mass spectrometry sequencing analyses. J Biol Chem *283*, 16455-16468.
Finne, J., Krusius, T., and Rauvala, H. (1977a). Occurrence of disialosyl groups in glycoproteins. Biochem Biophys Res Commun *74*, 405-410.
Finne, J., Krusius, T., Rauvala, H., and Hemminki, K. (1977b). The disialosyl group of glycoproteins. Occurrence in different tissues and cellular membranes. Eur J Biochem *77*, 319-323.
Freeze, H.H., and Aebi, M. (2005). Altered glycan structures: the molecular basis of congenital disorders of glycosylation. Curr Opin Struct Biol *15*, 490-498.
Fukuda, M., Bothner, B., Ramsamooj, P., Dell, A., Tiller, P.R., Varki, A., and Klock, J.C. (1985). Structures of sialylated fucosyl polylactosaminoglycans isolated from chronic myelogenous leukemia cells. J Biol Chem *260*, 12957-12967.
Fukuda, M., Dell, A., and Fukuda, M.N. (1984a). Structure of fetal lactosaminoglycan. The carbohydrate moiety of Band 3 isolated from human umbilical cord erythrocytes. J Biol Chem *259*, 4782-4791.
Fukuda, M., Dell, A., Oates, J.E., and Fukuda, M.N. (1984b). Structure of branched lactosaminoglycan, the carbohydrate moiety of band 3 isolated from adult human erythrocytes. J Biol Chem *259*, 8260-8273.
Fukuda, M., Fukuda, M.N., and Hakomori, S. (1979). Developmental change and genetic defect in the carbohydrate structure of band 3 glycoprotein of human erythrocyte membrane. J Biol Chem *254*, 3700-3703.
Fukuda, M., Guan, J.L., and Rose, J.K. (1988). A membrane-anchored form but not the secretory form of human chorionic gonadotropin-alpha chain acquires polylactosaminoglycan. J Biol Chem *263*, 5314-5318.
Fukuda, M., Lauffenburger, M., Sasaki, H., Rogers, M.E., and Dell, A. (1987). Structures of novel sialylated O-linked oligosaccharides isolated from human erythrocyte glycophorins. J Biol Chem *262*, 11952-11957.
Fukuda, M., Spooncer, E., Oates, J.E., Dell, A., and Klock, J.C. (1984c). Structure of sialylated fucosyl lactosaminoglycan isolated from human granulocytes. J Biol Chem *259*, 10925-10935.
Fukuda, M.N., and Matsumura, G. (1976). Endo-beta-galactosidase of Escherichia freundii. Purification and endoglycosidic action on keratan sulfates, oligosaccharides, and blood group active glycoprotein. J Biol Chem *251*, 6218-6225.
Fukuda, M.N., Sasaki, H., Lopez, L., and Fukuda, M. (1989). Survival of recombinant erythropoietin in the circulation: the role of carbohydrates. Blood *73*, 84-89.
Garcia-Vallejo, J.J., Van Dijk, W., Van Het Hof, B., Van Die, I., Engelse, M.A., Van Hinsbergh, V.W., and Gringhuis, S.I. (2006). Activation of human endothelial cells by tumor necrosis factor-alpha results in profound changes in the expression of glycosylation-related genes. J Cell Physiol *206*, 203-210.
Garcia, G.G., Berger, S.B., Sadighi Akha, A.A., and Miller, R.A. (2005). Age-associated changes in glycosylation of CD43 and CD45 on mouse CD4 T cells. Eur J Immunol *35*, 622-631.
Garner, O.B., and Baum, L.G. (2008). Galectin-glycan lattices regulate cell-surface glycoprotein organization and signalling. Biochem Soc Trans *36*, 1472-1477.
Gauthier, L., Rossi, B., Roux, F., Termine, E., and Schiff, C. (2002). Galectin-1 is a

stromal cell ligand of the pre-B cell receptor (BCR) implicated in synapse formation between pre-B and stromal cells and in pre-BCR triggering. Proc Natl Acad Sci U S A *99*, 13014-13019.

Gillespie, W., Kelm, S., and Paulson, J.C. (1992). Cloning and expression of the Gal beta 1, 3GalNAc alpha 2,3-sialyltransferase. J Biol Chem *267*, 21004-21010.

Gu, J., Nishikawa, A., Fujii, S., Gasa, S., and Taniguchi, N. (1992). Biosynthesis of blood group I and i antigens in rat tissues. Identification of a novel beta 1-6-N-acetylglucosaminyltransferase. J Biol Chem *267*, 2994-2999.

Guo, H.B., Randolph, M., and Pierce, M. (2007). Inhibition of a specific N-glycosylation activity results in attenuation of breast carcinoma cell invasiveness-related phenotypes: inhibition of epidermal growth factor-induced dephosphorylation of focal adhesion kinase. J Biol Chem *282*, 22150-22162.

Hakomori, S. (1964). A Rapid Permethylation of Glycolipid, and Polysaccharide Catalyzed by Methylsulfinyl Carbanion in Dimethyl Sulfoxide. J Biochem *55*, 205-208.

Hara, S., Takemori, Y., Yamaguchi, M., Nakamura, M., and Ohkura, Y. (1987). Fluorometric high-performance liquid chromatography of N-acetyl- and N-glycolylneuraminic acids and its application to their microdetermination in human and animal sera, glycoproteins, and glycolipids. Anal Biochem *164*, 138-145.

Harvey, D.J. (2000). Postsource decay fragmentation of N-linked carbohydrates from ovalbumin and related glycoproteins. J Am Soc Mass Spectrom *11*, 572-577.

Harvey, D.J., Bateman, R.H., Bordoli, R.S., and Tyldesley, R. (2000). Ionisation and fragmentation of complex glycans with a quadrupole time-of-flight mass spectrometer fitted with a matrix-assisted laser desorption/ionisation ion source. Rapid Commun Mass Spectrom *14*, 2135-2142.

Heinz Egge, J.P.-K. (1987). Fast atom bombardment mass spectrometry for structural elucidation of glycoconjugates. *6*, 331-393.

Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Muller, W.E.*, et al.* (2002). Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. Biochim Biophys Acta *1572*, 232-254.

Hirabayashi, J., Satoh, M., and Kasai, K. (1992). Evidence that Caenorhabditis elegans 32-kDa beta-galactoside-binding protein is homologous to vertebrate beta-galactoside-binding lectins. cDNA cloning and deduced amino acid sequence. J Biol Chem *267*, 15485-15490.

Hirabayashi, J., Ubukata, T., and Kasai, K. (1996). Purification and molecular characterization of a novel 16-kDa galectin from the nematode Caenorhabditis elegans. J Biol Chem *271*, 2497-2505.

Huang, M.T., Mason, J.C., Birdsey, G.M., Amsellem, V., Gerwin, N., Haskard, D.O., Ridley, A.J., and Randi, A.M. (2005). Endothelial intercellular adhesion molecule (ICAM)-2 regulates angiogenesis. Blood *106*, 1636-1643.

Ide, Y., Miyoshi, E., Nakagawa, T., Gu, J., Tanemura, M., Nishida, T., Ito, T., Yamamoto, H., Kozutsumi, Y., and Taniguchi, N. (2006). Aberrant expression of N-acetylglucosaminyltransferase-IVa and IVb (GnT-IVa and b) in pancreatic cancer. Biochem Biophys Res Commun *341*, 478-482.

Inaba, N., Hiruma, T., Togayachi, A., Iwasaki, H., Wang, X.H., Furukawa, Y., Sumi, R.,

Kudo, T., Fujimura, K., Iwai, T.*, et al.* (2003). A novel I-branching beta-1,6-N-acetylglucosaminyltransferase involved in human blood group I antigen expression. Blood *101*, 2870-2876.

Inoue, S., Lin, S.L., and Inoue, Y. (2000). Chemical analysis of the developmental pattern of polysialylation in chicken brain. Expression of only an extended form of polysialyl chains during embryogenesis and the presence of disialyl residues in both embryonic and adult chicken brains. J Biol Chem *275*, 29968-29979.

Inoue, S., Lin, S.L., Lee, Y.C., and Inoue, Y. (2001). An ultrasensitive chemical method for polysialic acid analysis. Glycobiology *11*, 759-767.

Ishida, H., Togayachi, A., Sakai, T., Iwai, T., Hiruma, T., Sato, T., Okubo, R., Inaba, N., Kudo, T., Gotoh, M.*, et al.* (2005). A novel beta1,3-N-acetylglucosaminyltransferase (beta3Gn-T8), which synthesizes poly-N-acetyllactosamine, is dramatically upregulated in colon cancer. FEBS Lett *579*, 71-78.

Jain, A., Tindell, C.A., Laux, I., Hunter, J.B., Curran, J., Galkin, A., Afar, D.E., Aronson, N., Shak, S., Natale, R.B.*, et al.* (2005). Epithelial membrane protein-1 is a biomarker of gefitinib resistance. Proc Natl Acad Sci U S A *102*, 11858-11863.

Jung, K., Cho, W., and Regnier, F.E. (2009). Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography. J Proteome Res *8*, 643-650.

Kang, P., Mechref, Y., Klouckova, I., and Novotny, M.V. (2005). Solid-phase permethylation of glycans for mass spectrometric analysis. Rapid Commun Mass Spectrom *19*, 3421-3428.

Kanter, J.L., Narayana, S., Ho, P.P., Catz, I., Warren, K.G., Sobel, R.A., Steinman, L., and Robinson, W.H. (2006). Lipid microarrays identify key mediators of autoimmune brain inflammation. Nat Med *12*, 138-143.

Kasahara, K., Watanabe, Y., Yamamoto, T., and Sanai, Y. (1997). Association of Src family tyrosine kinase Lyn with ganglioside GD3 in rat brain. Possible regulation of Lyn by glycosphingolipid in caveolae-like domains. J Biol Chem *272*, 29947-29953.

Kelm, S., Schauer, R., Manuguerra, J.C., Gross, H.J., and Crocker, P.R. (1994). Modifications of cell surface sialic acids modulate cell adhesion mediated by sialoadhesin and CD22. Glycoconj J *11*, 576-585.

Kiang, W.L., Krusius, T., Finne, J., Margolis, R.U., and Margolis, R.K. (1982). Glycoproteins and proteoglycans of the chromaffin granule matrix. J Biol Chem *257*, 1651-1659.

Kimura, N., Ohmori, K., Miyazaki, K., Izawa, M., Matsuzaki, Y., Yasuda, Y., Takematsu, H., Kozutsumi, Y., Moriyama, A., and Kannagi, R. (2007). Human B-lymphocytes express alpha2-6-sialylated 6-sulfo-N-acetyllactosamine serving as a preferred ligand for CD22/Siglec-2. J Biol Chem *282*, 32200-32207.

Kitajima, K., Kuroyanagi, H., Inoue, S., Ye, J., Troy, F.A., 2nd, and Inoue, Y. (1994). Discovery of a new type of sialidase, "KDNase," which specifically hydrolyzes deaminoneuraminyl (3-deoxy-D-glycero-D-galacto-2-nonulosonic acid) but not N-acylneuraminyl linkages. J Biol Chem *269*, 21415-21419.

Knapp, M.R., Jones, P.P., Black, S.J., Vitetta, E.S., Slavin, S., and Strober, S. (1979a). Characterization of a spontaneous murine B cell leukemia (BCL1). I. Cell surface expression of IgM, IgD, Ia, and FcR. J Immunol *123*, 992-999.

Knapp, M.R., Severinson-Gronowicz, E., Schroder, J., and Strober, S. (1979b).

Characterization of a spontaneous murine B cell leukemia (BCL1). II. Tumor cell proliferation and IgM secretion after stimulation by LPS. J Immunol *123*, 1000-1006.

Kobata, A. (1988). Structures, function, and transformational changes of the sugar chains of glycohormones. J Cell Biochem *37*, 79-90.

Kochibe, N., and Furukawa, K. (1980). Purification and properties of a novel fucose-specific hemagglutinin of Aleuria aurantia. Biochemistry *19*, 2841-2846.

Koenderman, A.H., Koppen, P.L., and Van den Eijnden, D.H. (1987). Biosynthesis of polylactosaminoglycans. Novikoff ascites tumor cells contain two UDP-GlcNAc:beta-galactoside beta 1----6-N-acetylglucosaminyltransferase activities. Eur J Biochem *166*, 199-208.

Koethe, S., Zander, L., Koster, S., Annan, A., Ebenfelt, A., Spencer, J., and Bemark, M. (2011). Pivotal Advance: CD45RB glycosylation is specifically regulated during human peripheral B cell differentiation. J Leukoc Biol.

Kojima, N., Yoshida, Y., Kurosawa, N., Lee, Y.C., and Tsuji, S. (1995). Enzymatic activity of a developmentally regulated member of the sialyltransferase family (STX): evidence for alpha 2,8-sialyltransferase activity toward N-linked oligosaccharides. FEBS Lett *360*, 1-4.

Kono, M., Yoshida, Y., Kojima, N., and Tsuji, S. (1996). Molecular cloning and expression of a fifth type of alpha2,8-sialyltransferase (ST8Sia V). Its substrate specificity is similar to that of SAT-V/III, which synthesize GD1c, GT1a, GQ1b and GT3. J Biol Chem *271*, 29366-29371.

Kornfeld, K., Reitman, M.L., and Kornfeld, R. (1981). The carbohydrate-binding specificity of pea and lentil lectins. Fucose is an important determinant. J Biol Chem *256*, 6633-6640.

Kremser, M.E., Przybylo, M., Hoja-Lukowicz, D., Pochec, E., Amoresano, A., Carpentieri, A., Bubka, M., and Litynska, A. (2008). Characterisation of alpha3beta1 and alpha(v)beta3 integrin N-oligosaccharides in metastatic melanoma WM9 and WM239 cell lines. Biochim Biophys Acta *1780*, 1421-1431.

Lagana, A., Goetz, J.G., Cheung, P., Raz, A., Dennis, J.W., and Nabi, I.R. (2006). Galectin binding to Mgat5-modified N-glycans regulates fibronectin matrix remodeling in tumor cells. Mol Cell Biol *26*, 3181-3193.

Larsen, M.R., Jensen, S.S., Jakobsen, L.A., and Heegaard, N.H. (2007). Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. Mol Cell Proteomics *6*, 1778-1787.

Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P., and Jorgensen, T.J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics *4*, 873-886.

Lau, K.S., Partridge, E.A., Grigorian, A., Silvescu, C.I., Reinhold, V.N., Demetriou, M., and Dennis, J.W. (2007). Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. Cell *129*, 123-134.

Lee, N., Wang, W.C., and Fukuda, M. (1990). Granulocytic differentiation of HL-60 cells is associated with increase of poly-N-acetyllactosamine in Asn-linked oligosaccharides attached to human lysosomal membrane glycoproteins. J Biol Chem *265*, 20476-20487.

Lee, Y.C., Kim, Y.J., Lee, K.Y., Kim, K.S., Kim, B.U., Kim, H.N., Kim, C.H., and Do, S.I. (1998). Cloning and expression of cDNA for a human Sia alpha 2,3Gal beta 1,

4GlcNA:alpha 2,8-sialyltransferase (hST8Sia III). Arch Biochem Biophys *360*, 41-46.

Leppanen, A., Stowell, S., Blixt, O., and Cummings, R.D. (2005). Dimeric galectin-1 binds with high affinity to alpha2,3-sialylated and non-sialylated terminal N-acetyllactosamine units on surface-bound extended glycans. J Biol Chem *280*, 5549-5562.

Leppanen, A., Zhu, Y., Maaheimo, H., Helin, J., Lehtonen, E., and Renkonen, O. (1998). Biosynthesis of branched polylactosaminoglycans. Embryonal carcinoma cells express midchain beta1,6-N-acetylglucosaminyltransferase activity that generates branches to preformed linear backbones. J Biol Chem *273*, 17399-17405.

Lin, S.L., Inoue, Y., and Inoue, S. (1999). Evaluation of high-performance anion-exchange chromatography with pulsed electrochemical and fluorometric detection for extensive application to the analysisof homologous series of oligo- and polysialic acids in bioactive molecules. Glycobiology *9*, 807-814.

Lindberg, B., and Lonngren, J. (1978). Methylation analysis of complex carbohydrates: general procedure and application for sequence analysis. Methods Enzymol *50*, 3-33.

Liu, F.T., Patterson, R.J., and Wang, J.L. (2002). Intracellular functions of galectins. Biochim Biophys Acta *1572*, 263-273.

Liu, F.T., and Rabinovich, G.A. (2010). Galectins: regulators of acute and chronic inflammation. Ann N Y Acad Sci *1183*, 158-182.

Liu, H., Kojima, N., Kurosawa, N., and Tsuji, S. (1997). Regulated expression system for GD3 synthase cDNA and induction of differentiation in Neuro2a cells. Glycobiology *7*, 1067-1076.

Loboda, A.V., Krutchinsky, A.N., Bromirski, M., Ens, W., and Standing, K.G. (2000). A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. Rapid Commun Mass Spectrom *14*, 1047-1057.

Ma, B.Y., Yoshida, K., Baba, M., Nonaka, M., Matsumoto, S., Kawasaki, N., Asano, S., and Kawasaki, T. (2009). The lectin Jacalin induces human B-lymphocyte apoptosis through glycosylation-dependent interaction with CD45. Immunology *127*, 477-488.

Mao, X., Luo, Y., Dai, Z., Wang, K., Du, Y., and Lin, B. (2004). Integrated lectin affinity microfluidic chip for glycoform separation. Anal Chem *76*, 6941-6947.

Merkle, R.K., and Cummings, R.D. (1987). Relationship of the terminal sequences to the length of poly-N-acetyllactosamine chains in asparagine-linked oligosaccharides from the mouse lymphoma cell line BW5147. Immobilized tomato lectin interacts with high affinity with glycopeptides containing long poly-N-acetyllactosamine chains. J Biol Chem *262*, 8179-8189.

Muramatsu, H., Kusano, T., Sato, M., Oda, Y., Kobori, K., and Muramatsu, T. (2008). Embryonic stem cells deficient in I beta1,6-N-acetylglucosaminyltransferase exhibit reduced expression of embryoglycan and the loss of a Lewis X antigen, 4C9. Glycobiology *18*, 242-249.

Muramatsu, T. (1988). Developmentally regulated expression of cell surface carbohydrates during mouse embryogenesis. J Cell Biochem *36*, 1-14.

Nan, B.C., Shao, D.M., Chen, H.L., Huang, Y., Gu, J.X., Zhang, Y.B., and Wu, Z.G. (1998).

Alteration of N-acetylglucosaminyltransferases in pancreatic carcinoma. Glycoconj J *15*, 1033-1037.

Nangia-Makker, P., Honjo, Y., Sarvis, R., Akahani, S., Hogan, V., Pienta, K.J., and Raz, A. (2000). Galectin-3 induces endothelial cell morphogenesis and angiogenesis. Am J Pathol *156*, 899-909.

Naven, T.J., Harvey, D.J., Brown, J., and Critchley, G. (1997). Fragmentation of complex carbohydrates following ionization by matrix-assisted laser desorption with an instrument fitted with time-lag focusing. Rapid Commun Mass Spectrom *11*, 1681-1686.

Nemansky, M., Schiphorst, W.E., and Van den Eijnden, D.H. (1995). Branching and elongation with lactosaminoglycan chains of N-linked oligosaccharides result in a shift toward termination with alpha 2-->3-linked rather than with alpha 2-->6-linked sialic acid residues. FEBS Lett *363*, 280-284.

Neurohr, K.J., Young, N.M., and Mantsch, H.H. (1980). Determination of the carbohydrate-binding properties of peanut agglutinin by ultraviolet difference spectroscopy. J Biol Chem *255*, 9205-9209.

North, S.J., Hitchen, P.G., Haslam, S.M., and Dell, A. (2009). Mass spectrometry in the analysis of N-linked and O-linked glycans. Curr Opin Struct Biol *19*, 498-506.

North, S.J., Huang, H.H., Sundaram, S., Jang-Lee, J., Etienne, A.T., Trollope, A., Chalabi, S., Dell, A., Stanley, P., and Haslam, S.M. (2010). Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. J Biol Chem *285*, 5759-5775.

Ohyama, Y., Kasai, K., Nomoto, H., and Inoue, Y. (1985). Frontal affinity chromatography of ovalbumin glycoasparagines on a concanavalin A-sepharose column. A quantitative study of the binding specificity of the lectin. J Biol Chem *260*, 6882-6887.

Okudaira, T., Hirashima, M., Ishikawa, C., Makishi, S., Tomita, M., Matsuda, T., Kawakami, H., Taira, N., Ohshiro, K., Masuda, M.*, et al.* (2007). A modified version of galectin-9 suppresses cell growth and induces apoptosis of human T-cell leukemia virus type I-infected T-cell lines. Int J Cancer *120*, 2251-2261.

Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics *4*, 2010-2021.

Pan, S., Wang, Y., Quinn, J.F., Peskind, E.R., Waichunas, D., Wimberger, J.T., Jin, J., Li, J.G., Zhu, D., Pan, C.*, et al.* (2006). Identification of glycoproteins in human cerebrospinal fluid with a complementary proteomic approach. J Proteome Res *5*, 2769-2779.

Pang, P.C., Tissot, B., Drobnis, E.Z., Morris, H.R., Dell, A., and Clark, G.F. (2009). Analysis of the human seminal plasma glycome reveals the presence of immunomodulatory carbohydrate functional groups. J Proteome Res *8*, 4906-4915.

Paulson, J.C., and Rademacher, C. (2009). Glycan terminator. Nat Struct Mol Biol *16*, 1121-1122.

Peracaula, R., Barrabes, S., Sarrats, A., Rudd, P.M., and de Llorens, R. (2008). Altered glycosylation in tumours focused to cancer diagnosis. Dis Markers *25*, 207-218.

Peterman, S.M., and Mulholland, J.J. (2006). A novel approach for identification and

characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. J Am Soc Mass Spectrom *17*, 168-179.
Pierce, M., and Arango, J. (1986). Rous sarcoma virus-transformed baby hamster kidney cells express higher levels of asparagine-linked tri- and tetraantennary glycopeptides containing [GlcNAc-beta (1,6)Man-alpha (1,6)Man] and poly-N-acetyllactosamine sequences than baby hamster kidney cells. J Biol Chem *261*, 10772-10777.
Piller, F., and Cartron, J.P. (1983). UDP-GlcNAc:Gal beta 1-4Glc(NAc) beta 1-3N-acetylglucosaminyltransferase. Identification and characterization in human serum. J Biol Chem *258*, 12293-12299.
Piller, F., Cartron, J.P., Maranduba, A., Veyrieres, A., Leroy, Y., and Fournet, B. (1984). Biosynthesis of blood group I antigens. Identification of a UDP-GlcNAc:GlcNAc beta 1-3Gal(-R) beta 1-6(GlcNAc to Gal) N-acetylglucosaminyltransferase in hog gastric mucosa. J Biol Chem *259*, 13385-13390.
Pitt, J.J., and Gorman, J.J. (1997). Oligosaccharide characterization and quantitation using 1-phenyl-3-methyl-5-pyrazolone derivatization and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Biochem *248*, 63-75.
Renkonen, O. (2000). Enzymatic in vitro synthesis of I-branches of mammalian polylactosamines: generation of scaffolds for multiple selectin-binding saccharide determinants. Cell Mol Life Sci *57*, 1423-1439.
Rosenberg, A., ed. (1995). Biology of the sialic acids (New York, Plenum Press, New York).
Rossi, B., Espeli, M., Schiff, C., and Gauthier, L. (2006). Clustering of pre-B cell integrins induces galectin-1-dependent pre-B cell receptor relocalization and activation. J Immunol *177*, 796-803.
Saijonmaa, O., Nyman, T., Hohenthal, U., and Fyhrquist, F. (1991). Endothelin-1 is expressed and released by a human endothelial hybrid cell line (EA.hy 926). Biochem Biophys Res Commun *181*, 529-536.
Saito, H., Nishikawa, A., Gu, J., Ihara, Y., Soejima, H., Wada, Y., Sekiya, C., Niikawa, N., and Taniguchi, N. (1994). cDNA cloning and chromosomal mapping of human N-acetylglucosaminyltransferase V+. Biochem Biophys Res Commun *198*, 318-327.
Saitoh, O., Wang, W.C., Lotan, R., and Fukuda, M. (1992). Differential glycosylation and cell surface expression of lysosomal membrane glycoproteins in sublines of a human colon cancer exhibiting distinct metastatic potentials. J Biol Chem *267*, 5700-5711.
Sasaki, K., Kurata-Miura, K., Ujita, M., Angata, K., Nakagawa, S., Sekine, S., Nishi, T., and Fukuda, M. (1997). Expression cloning of cDNA encoding a human beta-1,3-N-acetylglucosaminyltransferase that is essential for poly-N-acetyllactosamine synthesis. Proc Natl Acad Sci U S A *94*, 14294-14299.
Sasaki, K., Kurata, K., Kojima, N., Kurosawa, N., Ohta, S., Hanai, N., Tsuji, S., and Nishi, T. (1994). Expression cloning of a GM3-specific alpha-2,8-sialyltransferase (GD3 synthase). J Biol Chem *269*, 15950-15956.
Sato, C. (2004). Chain length diversity of sialic acids and its biological significance. Trends Glycosci Glycotechnol *16*, 331-344.
Sato, C., Fukuoka, H., Ohta, K., Matsuda, T., Koshino, R., Kobayashi, K., Troy, F.A., 2nd,

and Kitajima, K. (2000). Frequent occurrence of pre-existing alpha 2-->8-linked disialic and oligosialic acids with chain lengths up to 7 Sia residues in mammalian brain glycoproteins. Prevalence revealed by highly sensitive chemical methods and anti-di-, oligo-, and poly-Sia antibodies specific for defined chain lengths. J Biol Chem *275*, 15422-15431.

Sato, C., Inoue, S., Matsuda, T., and Kitajima, K. (1998a). Development of a highly sensitive chemical method for detecting alpha2-->8-linked oligo/polysialic acid residues in glycoproteins blotted on the membrane. Anal Biochem *261*, 191-197.

Sato, C., Inoue, S., Matsuda, T., and Kitajima, K. (1999). Fluorescent-assisted detection of oligosialyl units in glycoconjugates. Anal Biochem *266*, 102-109.

Sato, C., Kitajima, K., Inoue, S., and Inoue, Y. (1998b). Identification of oligo-N-glycolylneuraminic acid residues in mammal-derived glycoproteins by a newly developed immunochemical reagent and biochemical methods. J Biol Chem *273*, 2575-2582.

Sato, C., Matsuda, T., and Kitajima, K. (2002). Neuronal differentiation-dependent expression of the disialic acid epitope on CD166 and its involvement in neurite formation in Neuro2A cells. J Biol Chem *277*, 45299-45305.

Sato, C., Yasukawa, Z., Honda, N., Matsuda, T., and Kitajima, K. (2001). Identification and adipocyte differentiation-dependent expression of the unique disialic acid residue in an adipose tissue-specific glycoprotein, adipo Q. J Biol Chem *276*, 28849-28856.

Schwientek, T., Nomoto, M., Levery, S.B., Merkx, G., van Kessel, A.G., Bennett, E.P., Hollingsworth, M.A., and Clausen, H. (1999). Control of O-glycan branch formation. Molecular cloning of human cDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core 2 and core 4. J Biol Chem *274*, 4504-4512.

Seko, A., and Yamashita, K. (2005). Characterization of a novel galactose beta1,3-N-acetylglucosaminyltransferase (beta3Gn-T8): the complex formation of beta3Gn-T2 and beta3Gn-T8 enhances enzymatic activity. Glycobiology *15*, 943-951.

Sgroi, D., Varki, A., Braesch-Andersen, S., and Stamenkovic, I. (1993). CD22, a B cell-specific immunoglobulin superfamily member, is a sialic acid-binding lectin. J Biol Chem *268*, 7011-7018.

Shiraishi, N., Natsume, A., Togayachi, A., Endo, T., Akashima, T., Yamada, Y., Imai, N., Nakagawa, S., Koizumi, S., Sekine, S.*, et al.* (2001). Identification and characterization of three novel beta 1,3-N-acetylglucosaminyltransferases structurally related to the beta 1,3-galactosyltransferase family. J Biol Chem *276*, 3498-3507.

Slavin, S., and Strober, S. (1978). Spontaneous murine B-cell leukaemia. Nature *272*, 624-626.

Spillmann, D., and Finne, J. (1994). Identification of a major poly-N-acetyllactosamine-containing cell-surface glycoprotein of mouse teratocarcinoma cells. Appearance on cells induced to primitive endoderm but not parietal endoderm differentiation. Eur J Biochem *220*, 385-394.

Spooncer, E., Fukuda, M., Klock, J.C., Oates, J.E., and Dell, A. (1984). Isolation and characterization of polyfucosylated lactosaminoglycan from human granulocytes. J Biol Chem *259*, 4792-4801.

Srinivasan, N., Bane, S.M., Ahire, S.D., Ingle, A.D., and Kalraiya, R.D. (2009). Poly N-acetyllactosamine substitutions on N- and not O-oligosaccharides or Thomsen-Friedenreich antigen facilitate lung specific metastasis of melanoma cells via galectin-3. Glycoconj J *26*, 445-456.

Stamenkovic, I., Sgroi, D., Aruffo, A., Sy, M.S., and Anderson, T. (1991). The B lymphocyte adhesion molecule CD22 interacts with leukocyte common antigen CD45RO on T cells and alpha 2-6 sialyltransferase, CD75, on B cells. Cell *66*, 1133-1144.

Stowell, S.R., Arthur, C.M., Mehta, P., Slanina, K.A., Blixt, O., Leffler, H., Smith, D.F., and Cummings, R.D. (2008). Galectin-1, -2, and -3 exhibit differential recognition of sialylated glycans and blood group antigens. J Biol Chem *283*, 10109-10123.

Strober, S., Gronowicz, E.S., Knapp, M.R., Slavin, S., Vitetta, E.S., Warnke, R.A., Kotzin, B., and Schroder, J. (1979). Immunobiology of a spontaneous murine B cell leukemia (BCL1). Immunol Rev *48*, 169-195.

Suggs, J.E., Madden, M.C., Friedman, M., and Edgell, C.J. (1986). Prostacyclin expression by a continuous human cell line derived from vascular endothelium. Blood *68*, 825-829.

Sutton-Smith, M., Wong, N.K., Khoo, K.H., Wu, S.W., Yu, S.Y., Patankar, M.S., Easton, R., Lattanzio, F.A., Morris, H.R., Dell, A.*, et al.* (2007). Analysis of protein-linked glycosylation in a sperm-somatic cell adhesion system. Glycobiology *17*, 553-567.

Tabrizi, S.J., Niiro, H., Masui, M., Yoshimoto, G., Iino, T., Kikushige, Y., Wakasaki, T., Baba, E., Shimoda, S., Miyamoto, T.*, et al.* (2009). T cell leukemia/lymphoma 1 and galectin-1 regulate survival/cell death pathways in human naive and IgM+ memory B cells through altering balances in Bcl-2 family proteins. J Immunol *182*, 1490-1499.

Takamatsu, S., Antonopoulos, A., Ohtsubo, K., Ditto, D., Chiba, Y., Le, D.T., Morris, H.R., Haslam, S.M., Dell, A., Marth, J.D.*, et al.* (2010). Physiological and glycomic characterization of N-acetylglucosaminyltransferase-IVa and -IVb double deficient mice. Glycobiology *20*, 485-497.

Takamatsu, S., Oguri, S., Minowa, M.T., Yoshida, A., Nakamura, K., Takeuchi, M., and Kobata, A. (1999). Unusually high expression of N-acetylglucosaminyltransferase-IVa in human choriocarcinoma cell lines: a possible enzymatic basis of the formation of abnormal biantennary sugar chain. Cancer Res *59*, 3949-3953.

Takashima, S., Ishida, H.K., Inazu, T., Ando, T., Ishida, H., Kiso, M., Tsuji, S., and Tsujimoto, M. (2002). Molecular cloning and expression of a sixth type of alpha 2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans. J Biol Chem *277*, 24030-24038.

Tang, W., Chang, S.B., and Hemler, M.E. (2004). Links between CD147 function, glycosylation, and caveolin-1. Mol Biol Cell *15*, 4043-4050.

Tao, S.C., Chen, C.S., and Zhu, H. (2007). Applications of protein microarray technology. Comb Chem High Throughput Screen *10*, 706-718.

Teintenier-Lelievre, M., Julien, S., Juliant, S., Guerardel, Y., Duonor-Cerutti, M., Delannoy, P., and Harduin-Lepers, A. (2005). Molecular cloning and expression of a human hST8Sia VI (alpha2,8-sialyltransferase) responsible for the synthesis of the diSia motif on O-glycosylproteins. Biochem J *392*, 665-674.

Thomas, S., Robinson, C.J., and Gray, E. (1997). Responses of HUVEC and EAhy926 to heparin and fibroblast growth factors. In Vitro Cell Dev Biol Anim *33*, 492-494.

Togayachi, A., Akashima, T., Ookubo, R., Kudo, T., Nishihara, S., Iwasaki, H., Natsume, A., Mio, H., Inokuchi, J., Irimura, T.*, et al.* (2001). Molecular cloning and characterization of UDP-GlcNAc:lactosylceramide beta 1,3-N-acetylglucosaminyltransferase (beta 3Gn-T5), an essential enzyme for the expression of HNK-1 and Lewis X epitopes on glycolipids. J Biol Chem *276*, 22032-22040.

Togayachi, A., Kozono, Y., Ishida, H., Abe, S., Suzuki, N., Tsunoda, Y., Hagiwara, K., Kuno, A., Ohkura, T., Sato, N.*, et al.* (2007). Polylactosamine on glycoproteins influences basal levels of lymphocyte and macrophage activation. Proc Natl Acad Sci U S A *104*, 15829-15834.

Toppila, S., Lauronen, J., Mattila, P., Turunen, J.P., Penttila, L., Paavonen, T., Renkonen, O., and Renkonen, R. (1997). L-selectin ligands in rat high endothelium: multivalent sialyl Lewis x glycans are high-affinity inhibitors of lymphocyte adhesion. Eur J Immunol *27*, 1360-1365.

Torres, C.R., and Hart, G.W. (1984). Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. J Biol Chem *259*, 3308-3317.

Townsend, R.R., and Hardy, M.R. (1991). Analysis of glycoprotein oligosaccharides using high-pH anion exchange chromatography. Glycobiology *1*, 139-147.

Troy, F.A.I. (1996). Sialobiology and the polysialic acid glycotope. In Biology of the Sialic Acids. (Rosenberg, A., ed. ) (New York and London, Plenem Press), pp. 95-144.

Tsai, C.-M., Guan, C.-H., Hsieh, H.-W., Hsu, T.-L., Tu, Z., Wu, K.-J., Lin, C.-H., and Lin, K.-I. (2011). Galectin-1 and Galectin-8 Have Redundant Roles in Promoting Plasma Cell Formation. The Journal of Immunology.

Tsai, C.M., Chiu, Y.K., Hsu, T.L., Lin, I.Y., Hsieh, S.L., and Lin, K.I. (2008). Galectin-1 promotes immunoglobulin production during plasma cell differentiation. J Immunol *181*, 4570-4579.

Turunen, J.P., Majuri, M.L., Seppo, A., Tiisala, S., Paavonen, T., Miyasaka, M., Lemstrom, K., Penttila, L., Renkonen, O., and Renkonen, R. (1995). De novo expression of endothelial sialyl Lewis(a) and sialyl Lewis(x) during cardiac transplant rejection: superior capacity of a tetravalent sialyl Lewis(x) oligosaccharide in inhibiting L-selectin-dependent lymphocyte adhesion. J Exp Med *182*, 1133-1141.

Ujita, M., McAuliffe, J., Hindsgaul, O., Sasaki, K., Fukuda, M.N., and Fukuda, M. (1999a). Poly-N-acetyllactosamine synthesis in branched N-glycans is controlled by complemental branch specificity of I-extension enzyme and beta1,4-galactosyltransferase I. J Biol Chem *274*, 16717-16726.

Ujita, M., McAuliffe, J., Suzuki, M., Hindsgaul, O., Clausen, H., Fukuda, M.N., and Fukuda, M. (1999b). Regulation of I-branched poly-N-acetyllactosamine synthesis. Concerted actions by I-extension enzyme, I-branching enzyme, and beta1,4-galactosyltransferase I. J Biol Chem *274*, 9296-9304.

Unger, R.E., Krump-Konvalinkova, V., Peters, K., and Kirkpatrick, C.J. (2002). In vitro expression of the endothelial phenotype: comparative study of primary isolated cells and cell lines, including the novel cell line HPMEC-ST1.6R. Microvasc Res *64*,

384-397.
Uttamchandani, M., and Yao, S.Q. (2008). Peptide microarrays: next generation biochips for detection, diagnostics and high-throughput screening. Curr Pharm Des *14*, 2428-2438.
Varki, A. (2008). Sialic acids in human health and disease. Trends Mol Med *14*, 351-360.
Vitetta, E.S., Yuan, D., Krolick, K., Isakson, P., Knapp, M., Slavin, S., and Strober, S. (1979). Characterization of the spontaneous murine B cell leukemia (BCL1). III. Evidence for monoclonality by using an anti-idiotype antibody. J Immunol *122*, 1649-1654.
Voduc, D., Kenney, C., and Nielsen, T.O. (2008). Tissue microarrays in clinical oncology. Semin Radiat Oncol *18*, 89-97.
Wada, Y., Azadi, P., Costello, C.E., Dell, A., Dwek, R.A., Geyer, H., Geyer, R., Kakehi, K., Karlsson, N.G., Kato, K.*, et al.* (2007). Comparison of the methods for profiling glycoprotein glycans--HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. Glycobiology *17*, 411-422.
Xiong, L., Andrews, D., and Regnier, F. (2003). Comparative proteomics of glycoproteins based on lectin selection and isotope coding. J Proteome Res *2*, 618-625.
Xu, D., and Hemler, M.E. (2005). Metabolic activation-related CD147-CD98 complex. Mol Cell Proteomics *4*, 1061-1071.
Yamaji, T., Teranishi, T., Alphey, M.S., Crocker, P.R., and Hashimoto, Y. (2002). A small region of the natural killer cell receptor, Siglec-7, is responsible for its preferred binding to alpha 2,8-disialyl and branched alpha 2,6-sialyl residues. A comparison with Siglec-9. J Biol Chem *277*, 6324-6332.
Yamashita, K., Ohkura, T., Tachibana, Y., Takasaki, S., and Kobata, A. (1984). Comparative study of the oligosaccharides released from baby hamster kidney cells and their polyoma transformant by hydrazinolysis. J Biol Chem *259*, 10834-10840.
Yamashita, K., Totani, K., Ohkura, T., Takasaki, S., Goldstein, I.J., and Kobata, A. (1987). Carbohydrate binding properties of complex-type oligosaccharides on immobilized Datura stramonium lectin. J Biol Chem *262*, 1602-1607.
Yang, E., Shim, J.S., Woo, H.J., Kim, K.W., and Kwon, H.J. (2007). Aminopeptidase N/CD13 induces angiogenesis through interaction with a pro-angiogenic protein, galectin-3. Biochem Biophys Res Commun *363*, 336-341.
Yang, R.Y., Rabinovich, G.A., and Liu, F.T. (2008). Galectins: structure, function and therapeutic potential. Expert Rev Mol Med *10*, e17.
Yasukawa, Z., Sato, C., and Kitajima, K. (2005a). Inflammation-dependent changes in alpha2,3-, alpha2,6-, and alpha2,8-sialic acid glycotopes on serum glycoproteins in mice. Glycobiology *15*, 827-837.
Yasukawa, Z., Sato, C., and Kitajima, K. (2005b). Inflammation-dependent changes in alpha2,3-, alpha2,6-, and alpha2,8-sialic acid glycotopes on serum glycoproteins in mice. Glycobiology *15*, 827-837.
Yasukawa, Z., Sato, C., and Kitajima, K. (2007). Identification of an inflammation-inducible serum protein recognized by anti-disialic acid antibodies as carbonic anhydrase II. J Biochem *141*, 429-441.
Yasukawa, Z., Sato, C., Sano, K., Ogawa, H., and Kitajima, K. (2006). Identification of

disialic acid-containing glycoproteins in mouse serum: a novel modification of immunoglobulin light chains, vitronectin, and plasminogen. Glycobiology *16*, 651-665.

Yeh, J.C., Ong, E., and Fukuda, M. (1999). Molecular cloning and expression of a novel beta-1, 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I branches. J Biol Chem *274*, 3215-3221.

Yoshida, Y., Kojima, N., and Tsuji, S. (1995). Molecular cloning and characterization of a third type of N-glycan alpha 2,8-sialyltransferase from mouse lung. J Biochem *118*, 658-664.

Yu, S.Y., Wu, S.W., Hsiao, H.H., and Khoo, K.H. (2009). Enabling techniques and strategic workflow for sulfoglycomics based on mass spectrometry mapping and sequencing of permethylated sulfated glycans. Glycobiology *19*, 1136-1149.

Yu, S.Y., Wu, S.W., and Khoo, K.H. (2006). Distinctive characteristics of MALDI-Q/TOF and TOF/TOF tandem mass spectrometry for sequencing of permethylated complex type N-glycans. Glycoconj J *23*, 355-369.

Zaia, J. (2008). Mass spectrometry and the emerging field of glycomics. Chem Biol *15*, 881-892.

Zhang, H., Li, X.J., Martin, D.B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol *21*, 660-666.

Zhang, J.Q., Biedermann, B., Nitschke, L., and Crocker, P.R. (2004). The murine inhibitory receptor mSiglec-E is expressed broadly on cells of the innate immune system whereas mSiglec-F is restricted to eosinophils. Eur J Immunol *34*, 1175-1184.

Zhang, L. (2010). Glycosaminoglycan (GAG) biosynthesis and GAG-binding proteins. Prog Mol Biol Transl Sci *93*, 1-17.

Zuniga, E., Rabinovich, G.A., Iglesias, M.M., and Gruppi, A. (2001). Regulated expression of galectin-1 during B-cell activation and implications for T-cell apoptosis. J Leukoc Biol *70*, 73-79.