Towards improved glycan and glycoprotein biomarkers in colorectal cancer by mass spectrometry

Jenny Hiu Lam Chik

Bachelor of Science (Hons) (Macquarie University)

A thesis submitted for the degree of

Doctor of Philosophy

Faculty of Science

Department of Chemistry and Biomolecular Sciences

Macquarie University, Sydney, Australia

October, 2014

Declaration

I hereby certified that the work presented in this thesis entitled, "Towards improved glycan and glycoprotein biomarkers in colorectal cancer using mass spectrometry" is the result of my own work except where acknowledged and is not being submitted for higher degree to any other university or institution. I consent to a copy of this thesis being available in the Macquarie University library for consultation, loan and photocopying henceforth.

Jenny Hiu Lam Chik
MQ ID: 40746852

October 2014

Abstract

Colorectal cancer (CRC) is a common malignancy with an estimated incidence of over one million new cases worldwide per annum. The classical diagnostic biomarker carcinoembryonic antigen (CEA) has limited sensitivity, so is not well suited for population-based screening. Therefore, there remains a need for novel biomarkers for the early detection of CRC. As protein glycosylation and glycans are underexplored as sources of potential biomarkers, this thesis uses mass spectrometry to investigate these as possible novel sources of CRC biomarkers.

In the first phase of the work, *N*- and *O*-glycans from five CRC cell lines were compared with those recovered from colonic tumours (Chapter 2). To provide a *bona fide* comparison, the epithelial cells from the tumours were enriched using an immunoaffinity method based on EpCAM expression. Glycosylation differences were found in both *N*- and *O*-glycans. However, the differences were especially striking in the *O*-glycans. Amongst these differences, the well-known glycan cancer marker, sialyl-Tn was expressed in high abundance in the tumours, but it was not detected in the cell lines, except in the mucin-producing LS174T cells. Additional work using qRT-PCR implicated the regulation of glycosyltransferase and mucin expression in the formation of sialyl-Tn. Overall, the glycosylation differences found between cell line model systems and tumours, indicates that glycan markers may be differentially expressed in these two systems. This finding cautions against the use of cell lines only for biomarker discovery as glycan biomarkers may be missed.

Expanding on this work, colon epithelial cells enriched from pair-matched normal and tumour tissues were investigated (Chapter 3 and 4). In Chapter 3, the *N*- and *O*-glycan profiles of epithelial cells from non-neoplastic and cancer tissues were compared. As expected, sialyl-Tn was increased in the tumours. Additionally, 13 other glycans (12 *N*-glycans, 1 *O*-glycan) were found differentially expressed in the tumours; 12 of these have not been reported previously. The changes in glycosylation also implicate changes in the corresponding biosynthetic pathways and are potential new markers for CRC.

Given that glycosylation changes were associated with membrane proteins, label-free membrane proteomics was also carried out on the epithelial-enriched normal and tumour colonic tissues (Chapter 4). Over 1600 proteins were identified with 52 membrane proteins differentially expressed, which included 9 glycoproteins. Proteins previously linked with CRC were identified as well as proteins not formerly associated with CRC and these altered proteins are implicated in nucleotide binding, protein transport, immunity and cell proliferation. The *N*-glycosylation sites of membrane glycoproteins were also mapped, which included the confirmation of 60 potential sites and identification of 16 unannotated sites. Intriguingly, quantification of formerly glycosylated and non-

glycosylated peptides from CEA suggests *N*-glycosylation site differences which may be used to differentiate normal and tumour tissues.

In summary, this thesis examines the alterations in glycan and glycoproteins of epithelial cells in colorectal cancer using mass spectrometry. The study has developed a comprehensive workflow that has enabled new insights to be revealed about glycan, protein and glycosylation site changes in tumours. These molecular changes may provide novel leads in the ongoing search for new biomarkers of improved sensitivity and specificity in colorectal cancer.

Acknowledgements

I am indebted to the following people for their guidance and support during the course of experimental work and thesis writing:

First of all, I would like to thank my supervisors, Professor Nicki Packer and Professor Mark Molloy, for their scientific guidance and mentorship, in particular, Nicki for her expertise in glycans and Mark for proteomics. Their vast knowledge, experience and constructive criticisms frequently propel me to improve all aspects of being a scientist such as vigorous data interpretation, questioning of underlying assumptions and succinct scientific writing. It would not have been possible to produce this work without their patience and persistent support throughout many years, especially during times when I felt discouraged. Clearly, they have played a major role in my development as the scientist I am today.

The lab members of Cancer Proteomics group (Molloy lab) and Glyco@MQ group (Packer lab) have also been instrumental in the production of this thesis. I thank them for the numerous scientific as well as everyday mundane conversations. The feedback they provide during lab meetings also challenge me to improve my work and my scientific knowledge. There are too many people to mention by name but I am especially indebted to the following people: Miyako Nakano for teaching me foundational scientific techniques for MS analysis; Robyn Peterson for her proof-reading of parts of this thesis; Matthew McKay, Morten Andersen, Chi-Hung Lin, Matthew Campbell, Albert Lee for their scientific suggestions and general mentorship; Dana Pascovici for her guidance on statistics; Edward Moh for his experimental work in LS174T cell lines; Jerry Zhou for his single-celled colon tissue preparations; Arun Dass, Maja Christiansen, Katherine Wongtrakul-kish, Merrina Anugraham, Jodie Abrahams, Ling Lee, Christoph Krisp, Sarah Randall, Crystal Semaan, Juan Martinez-Aguilar, for witnessing my journey over many years and helping me along the way.

I would also like to thank Professor Ian Paulsen and members of the Paulsen group, especially the following: Daniel Farrugia, Liping Li, Hasinka Ariyaratne, Ani Penesyan and Karl Hassan for allowing me to share their lab space and teaching me the technique of qRT-PCR.

Lastly, I would like to thank other friends, colleagues and family who have provided encouragement and support. I especially thank my parents, Anita and Peter, for their patience and all-round care during the whole endeavour; my sister, Sally, for providing perspective and fun times. I would also like to thank Angela Sun and Matt Cabanag for being my friends.

The time the above-mentioned people have spent with me is invaluable and priceless on many levels. As such, I sincerely thank them for the many good memories regardless of where their scientific and personal journey takes them next.

Publications

Part of this thesis that have been published or submitted for review:

1.   **Chik, J. H. L**., Zhou, J., Moh, E. S. X., Christopherson, R., Clarke, S. J., Molloy, M. P., & Packer, N. H. (2014). Comprehensive glycomics comparison between colon cancer cell cultures and tumours: Implications for biomarker studies. Journal of Proteomics, 108, 146-162. doi: http://dx.doi.org/10.1016/j.jprot.2014.05.002

I, Jenny Hiu Lam Chik, was the lead author of the paper 1 (presented in Chapter 2). All glycoprofiling and qRT-PCR work was performed by myself except for glycan work in the LS174T cell line which was conducted by Edward Moh (glycoprofile data analysed by myself). My supervisors Prof. Packer and A/Prof Mark Molloy supervised, advised and edited the manuscripts.

2.   Christiansen, M. N., **Chik, J.***, Lee, L., Anugraham, M., Abrahams, J. L., & Packer, N. H. (2014). Cell surface protein glycosylation in cancer. Proteomics, 14(4-5), 525-546. doi: 10.1002/pmic.201300387. *Equal first author

For paper 2 (presented in Chapter 1), all authors including myself contributed equally to the manuscript preparation and Prof. Packer edited and reviewed the manuscript.

Other publications (accepted or submitted for review) during the period of doctoral candidature and not included in this thesis:

1.   Martinez-Aguilar, J., **Chik, J**., Nicholson, J., Semaan, C., McKay, M. J., & Molloy, M. P. (2013). Quantitative mass spectrometry for colorectal cancer proteomics. Proteomics Clin Appl, 7(1-2), 42-54. doi: 10.1002/prca.201200080

I contributed towards writing the section on labelling strategies used in colorectal cancer proteomics. As labelling strategies were not utilised for the proteomics analysis (Chapter 4), this has not been included in the thesis.

2.   Lin, C. H., **Chik, J. H. L**., Packer, N. H., Molloy, M. P., Multidimensional fractionation is a requirement for quantitation of Golgi-resident glycosylation enzymes from cultured human cells. J Proteome Res 2015, 14, 747-755.

I provided data from my qRT-PCR experiments (presented in Chapter 2).

Presentations

Parts of this thesis have also been presented at national and international conferences:

1. Sydney Protein Group Thompson Prize evening 2013, oral presentation (finalist), "Glycosylation differences between colorectal cancer cell lines and tumours: correlation with glycosyltransferases and mucins"

   Chik, J., Zhou, J., Moh, E., Christopherson, R., Packer, N. & Molloy, M.

2. 18th Lorne Proteomics Symposium 2013, oral presentation, "Sugars in the real world: comparing protein glycosylation of epithelial colon cancer cells from cultured cells and tumours".

   Chik, J., Zhou, J., Moh, E., Christopherson, R., Packer, N. & Molloy, M.

3. 1st Proteomics and Beyond Symposium Sydney 2012, lighting talk and poster presentation, "Sugars in the real world: is the protein glycosylation of colon cancer cultured cells similar to cell-type partitioned colon cancer tissue?"

   Chik, J., Zhou, J., Christopherson, R., Packer, N. & Molloy, M.

4. 17th Lorne Proteomics Symposium 2012 and 4th Charles Warren Workshop 2012 (Athens, Georgia, United States), poster presentation, "Comparative glycan profiling of colon cancer cell lines and cell-type partitioned colon tumours"

   Chik, J., Zhou, J., Christopherson, R., Packer, N. & Molloy, M.

5. 16th Lorne Proteomics Symposium 2011, poster presentation, "Glycosylation differences between *in vitro* and *ex vivo* systems of investigation in colorectal cancer"

   Chik, J., Nakano, M., Clarke, S., Baker, M. & Packer, N.

Awards and Prizes

1. Australasian Proteomics Society Student Oral Presentation Award, 18th Lorne Proteomics Symposium (2013)

2. Northern Translational Cancer Research Scholar Award, Cancer Institute of NSW (2012)

3. ATA Scientific Lorne Travel Scholarship, Sydney Protein Group (2012)

4. People's Choice Best Poster, 1st Proteomics & Beyond Symposium (2012)

5. Postgraduate Research Fund Award, Macquarie University (2012)

6. Australian Postgraduate Award (2010-2013)

Common Abbreviations

CRC	Colorectal cancer
EIC	Extracted Ion chromatogram
ESI	Electrospray
Fuc	Fucose
GalNAc	*N*-acetylgalactosamine
GlcNAc	*N*-acetylglucosamine
HexNAc	*N*-acetylhexosamine
LC	Liquid chromatography
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
Man	Mannose
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NeuAc	*N*-acetylneuraminic acid
NSAF	Normalised spectral abundance factor(s)
SRM	Selected reaction monitoring

Table of Contents

Chapter 1

Introduction

1   Introduction

Altered glycosylation is a common phenotype exhibited by various cancers including cancers with high incidence such as breast, prostate, lung and colon [1, 2]. Since glycosylation is changed during cancer, it can be utilised as a potential source of novel biomarkers needed for early cancer detection. In this thesis, the focus will be on glycan and glycan-related changes in colorectal cancer. In this introduction, the basics of glycosylation and colorectal cancer will first be reviewed. This will be followed by an overview of the approaches used to analyse glycans and then an in-depth analysis of reported glycan and glycoprotein changes in the context of colorectal cancer. Finally, the rationale and the research aims of this work will be described.

1.1   Glycosylation

Glycosylation is usually defined as the enzymatic process of covalent attachment of carbohydrate chains, also known as glycans, onto biomolecules such as proteins or lipids [3]. Glycosylation is a protein post-translational modification (PTM) of proteins following their synthesis. Other notable examples of PTMs include phosphorylation [4, 5], acetylation [6, 7] and methylation [8]. Glycosylation is estimated to be one of the most common protein PTMs with approximately 50% of proteins proposed to be glycosylated [9].

Each glycan is synthesised from a number of monosaccharides, which are the most basic units of a carbohydrate. For ease of representation, two major systems of nomenclature are commonly used to represent glycans [10]: Oxford [11], and Consortium of Functional Glycomics nomenclature (CFG, described at http://glycomics.scripps.edu/coreD/PGAnomenclature.pdf). A hybrid of these systems can also be used. Using CFG nomenclature, a table of 20 natural monosaccharides is shown below (Table 1). Commonly-used abbreviations for each monosaccharide are also included. A subset (shaded in orange) is known to be commonly expressed in eukaryotes, although the additional monosaccharides (not shaded) can also be expressed in plants and bacteria.

Table 1. Representation of common naturally-occurring monosaccharides using Consortium of Functional Glycomics (CFG) notation. A subset (shaded) is more frequently observed in eukaryotes.

◯ Galactose (Gal)	▲ Fucose (Fuc)	
◼ *N*-acetylgalactosamine (GalNAc)	◆ *N*-acetylneuraminic acid (NeuAc)	
◻ Galactosamine (GalN)	◇ *N*-glycolylneuraminic acid (NeuGc)	
● Glucose (Glc)	★ Xylose (Xyl)	
◼ *N*-acetylglucosamine (GlcNAc)	◆ 2-Keto-3-deoxynononic acid (Kdn)	
◣ Glucosamine (GlcN)	◈ Glucuronic acid (GlcA)	
● Mannose (Man)	◈ Iduronic acid (IdoA)	
◼ *N*-Acetylmannosmine (ManNAc)	◈ Galacturonic acid (GalA)	
◣ Mannosamine (ManN)	◈ Mannuronic acid (ManA)	

Unlike protein synthesis, glycan synthesis is non-template driven, producing complex structures. Structural variability arises from the glycosidic linkages which allows each monosaccharide to be covalently linked to every other via the series of hydroxyl groups present [12]. In order to specify the glycosidic linkage involved between monosaccharides, IUPAC nomenclature includes both anomerisation type (α or β) and linkage (illustrated in Fig 1) [13]. For example, α2-3 linked sialic acid (NeuAc) indicates that the glycosidic linkage is α-anomeric and the number of the anomeric carbon atom on sialic acid is 2 (Fig 1b). Moreover, the glycosidic bond occurs between the 2nd carbon (from sialic acid) to the 3rd carbon atom of the adjacent monosaccharide residue (galactose, Fig 1b).

For Oxford notation, alphanumeric characters are combined with angled linking between monosaccharides to specify glycosidic linkages (Fig 1c). CFG nomenclature is similar, although there is no angled linking between monosaccharides. Linkage information is specified by alphanumeric characters only (example in Fig 2). In the hybrid system, CFG notation is combined with angled linking (Fig 1d). In this thesis, CFG notation will be generally used unless emphasis on glycosidic linkages is required. When this occurs, hybrid notation will be used.

Due to differences in glycosidic linkages, glycan structures can have the same chemical composition and mass but different stereochemistry – these glycans are known as isomers [12]. Separation and characterisation of different glycan isomers is an important consideration in glycan structural and functional analysis [14, 15].
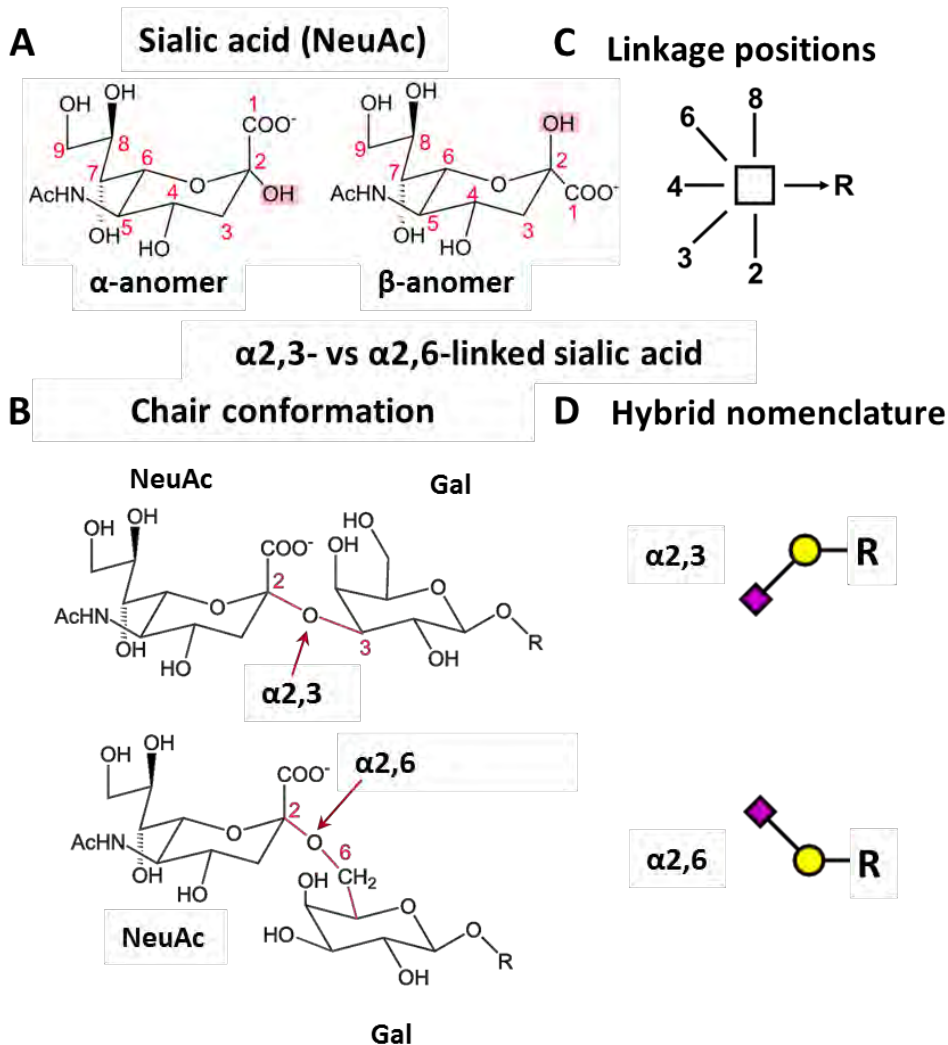
Figure 1. Stereochemistry of carbohydrates (illustrated using sialic acid/NeuAc) represented by chair conformation and hybrid nomenclature (CFG depiction with linkage positions). A) α and β anomers of sialic acid with carbons numbered. Note the differences in position of the hydroxyl (-OH) group at the anomeric carbon (carbon 2). Differences in α2,3- and α2,6-linked sialic acids represented by chain conformation **(B)** and hybrid nomenclature **(D)**. For example, α2-3 linked sialic acid indicates that the glycosidic linkage is α-anomeric and the number of the anomeric carbon atom on sialic acid is 2. Moreover, the glycosidic bond occurs between the 2nd carbon (from sialic acid) to the 3rd carbon atom of the adjacent monosaccharide residue (galactose). R represents *N*- or *O*-glycan or glycolipid. **C)** Linkage positions used in the hybrid/Oxford nomenclature. Adapted from [16, 17] and [18].

The most common attachment of glycans to proteins is through the amide group on asparagine amino acid (*N*-linked) or the hydroxyl group on serine (Ser) or threonine (Thr; *O*-linked). These are discussed below in detail.
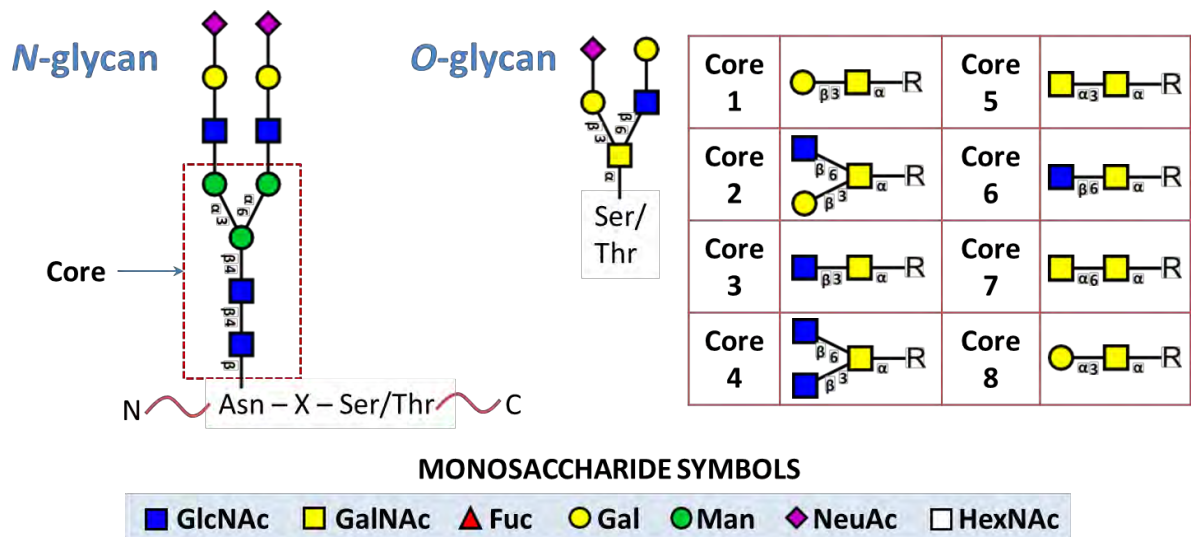
Figure 2. *N*- and *O*-glycans. The chitobiose core (or core) is shared amongst all known *N*-glycans. Cores 1-8 of *O*-glycans are also shown. X represents any amino acid except proline.

1.1.1 *N*-glycans

N-glycans are defined as carbohydrate chains covalently linked to asparagine (Asn) amino acids through the amide nitrogen (Fig 2). This bond is known as the *N*-glycosidic bond. Although five types of *N*-glycosidic bonds are known to exist in nature, the most common is the joining of an *N*-acetylglucosamine (GlcNAc) monosaccharide to the asparagine amide nitrogen. All *N*-linked glycans can be found attached to an asparagine which resides within a consensus sequence (sequon) consisting of asparagine-X-serine/threonine/cysteine (where X is any amino acid except proline, Fig 2)[19]. Because of this consensus sequence, the *N*-glycan sites on a polypeptide can be easily predicted, unlike the situation for *O*-glycans. *N*-glycans have a common structural motif known as the chitobiose core (boxed in Fig 2) which consists of two GlcNAc residues and three mannose arranged in a conserved branched configuration.

N-glycans can be further sub-divided into three categories: (i) high-mannose, (ii) complex, and (iii) hybrid (Fig 3). High-mannose glycans have their chitobiose cores extended with only mannose residues. In contrast, hybrid glycans have one branch of their chitobiose core extended with non-mannose residues such as GlcNAc, GalNAc, galactose and sialic acids (NeuAc or NeuGc); these residues comprise all branches of the complex glycans. Additionally, hybrid/complex glycans can be modified with the addition of fucose residues to the core (known as core-fucosylation) and the non-reducing terminus (outer-arm fucosylation). A notable subtype of complex glycans are bisecting *N*-glycans, which are characterised by a β1,4-linked GlcNAc to the central mannose of the *N*-glycan core (Fig 3).
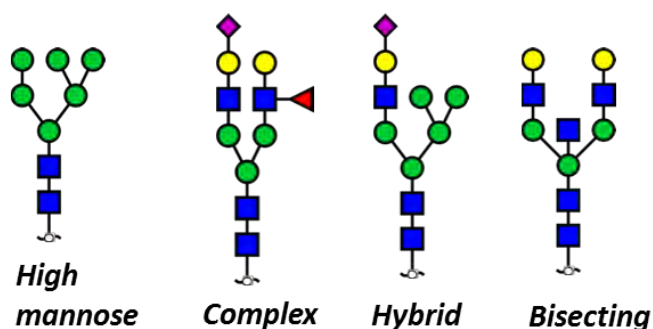
Figure 3. Representative types of *N*-glycans – high mannose, complex, hybrid and bisecting.

1.1.2   *O*-GalNAc glycans (*O*-glycans)

O-linked carbohydrates are known to be covalently linked to serine or threonine residues via hydroxyl groups. There are different types of *O*-linked glycans such as *O*-GalNAc, *O*-fucose and *O*-glucose glycans, which have been named depending on the first monosaccharide attached via the *O*-linked glycosidic bond [20]. In this thesis, the focus will be on *O*-GalNAc glycans (Fig 2) which are frequently found in mammals [21]. Henceforth, the term *O*-glycans will refer specifically to *O*-GalNAc glycans. *O*-glycans are more structurally variable relative to *N*-glycans as they are attached in one of eight structural "cores" (Fig 2). Four of these core structures (1, 2, 3, and 4) are particularly widespread in mammalian glycoproteins [21]. Unlike *N*-glycans, *O*-glycans are not known to be attached to proteins following rules of a strict structural motif or a consensus sequence [22] making the prediction of *O*-glycan sites on proteins much more difficult.

1.2   Glycan biosynthesis

Glycosidases and glycosyltranferases are the two main classes of glycan-processing enzymes involved in glycan synthesis. Glycosidases function to break down glycans into individual residues while glycosyltransferases have the opposite role of elongating (or building) glycans [23]. Interestingly, it is believed that humans have a repertoire of approximately 250 glycosyltransferases and glycosidases and as many as 30 are involved in the co-ordinated synthesis of a single complex oligosaccharide [24].
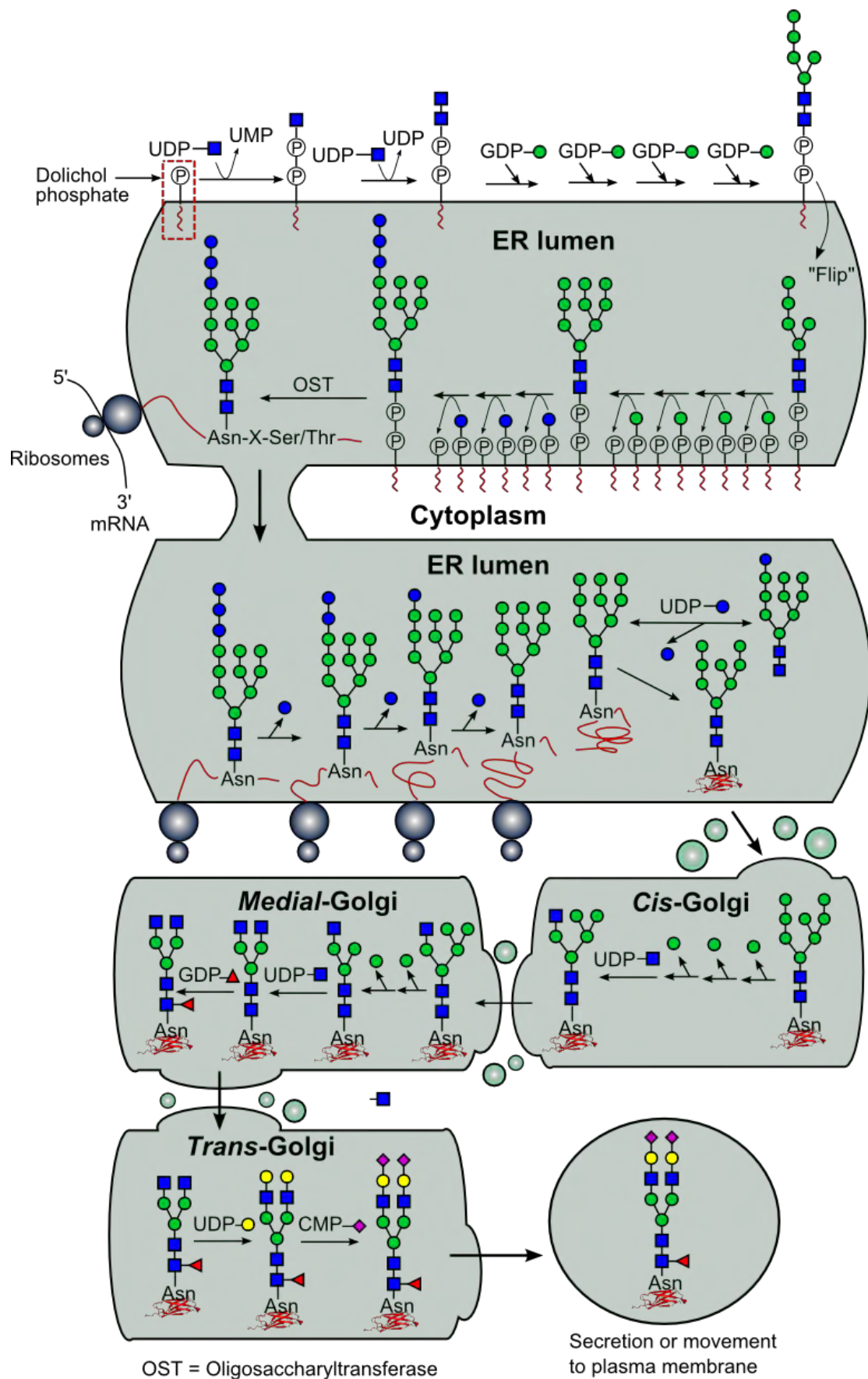
Figure 4. An overview of the _N_-glycosylation pathway adapted from Essentials in Glycobiology [25].

This schematic shows the nascent _N_-glycan in the ER through to its maturation in the Golgi network.

Glycosidases are located both in the endoplasmic reticulum (ER) and Golgi apparatus. The majority of glycosyltransferases are restricted to the Golgi apparatus [24]. However, the transmembrane domain of glycosyltransferases may be proteolytically cleaved, leading to the secretion of these enzymes in soluble form in body fluids [26, 27]. The biosynthetic pathways and enzymes involved in *N*-glycosylation, as well as those involved in the Core 1-4 based *O*-glycosylation have been identified, although limited information is known about the biosynthesis of Core 5-8 *O*-glycans. The overall activity of glycosyltransferases is likely affected by a complex interplay between factors such as substrate availability (nucleotide sugar donors)[28], sugar transporters[28], mRNA stability[29] and protein turnover [30].

1.2.1   *N*-glycan biosynthesis

The eukaryotic *N*-glycosylation pathway is generally well-understood. A schematic representing an overview of this pathway is shown in Fig 4. Briefly, *N*-glycosylation begins with the transfer of GlcNAc onto a lipid-phosphate precursor (dolichol phosphate, boxed in Fig 4) located on the cytoplasmic face of the ER. This is gradually extended to a 7-sugar compound (Dol-P-P-GlcNAc$_2$Man$_5$) by a series of GlcNAc transferases and mannosyltransferases, before the sugar is "flipped" over to the luminal side [31]. The glycan continues to be extended to a 14-sugar compound by the addition of four mannose and three glucose residues [25]. This compound is transferred 'en bloc' onto an asparagine residue by oligosaccharyltransferase (OST) and is truncated by glucosidases and mannosidases as it is processed through the *cis-* and *medial-*Golgi. It can then be extended by various glycosyltransferases to become a mature hybrid/complex *N*-glycan in the *medial-* and *trans-*Golgi network [25]. Following the maturation of the protein and glycan, the glycoprotein is sorted into specialised vesicles for secretion (by exocytosis) or transportation to plasma membrane (by endocytosis) [32].

1.2.2   *O*-GalNAc glycan (*O*-glycan) biosynthesis

A summary of the *O*-GalNAc glycosylation pathway for core 1-4 structures with the associated glycosyltransferases is shown in Fig 5. *O*-glycan synthesis differs from *N*-glycan synthesis as each residue is added sequentially entirely within the Golgi apparatus. The transfer of GalNAc onto a serine/threonine residue by the polypeptide GalNAc transferase (ppGalNAcT) is the initial step in *O*-glycan synthesis [22]. This forms the Tn antigen (*O*-GalNAc) which is the basis for all core structures (shaded, centre, Fig 5). Tn antigen can be sialylated to form sialyl-Tn, galactosylated to form Core 1 or *N*-acetylglucosaminylated to form core 3. Core 1 and core 3 can be further extended [22]. For example, core 1 can also be extended to form Core 2 while Core 3 can be elongated to form core 4. Various glycosyltransferases, same as those that extend complex *N*-glycans [33], can elongate cores 1 to 4 structures to more mature *O*-glycans (not shown).
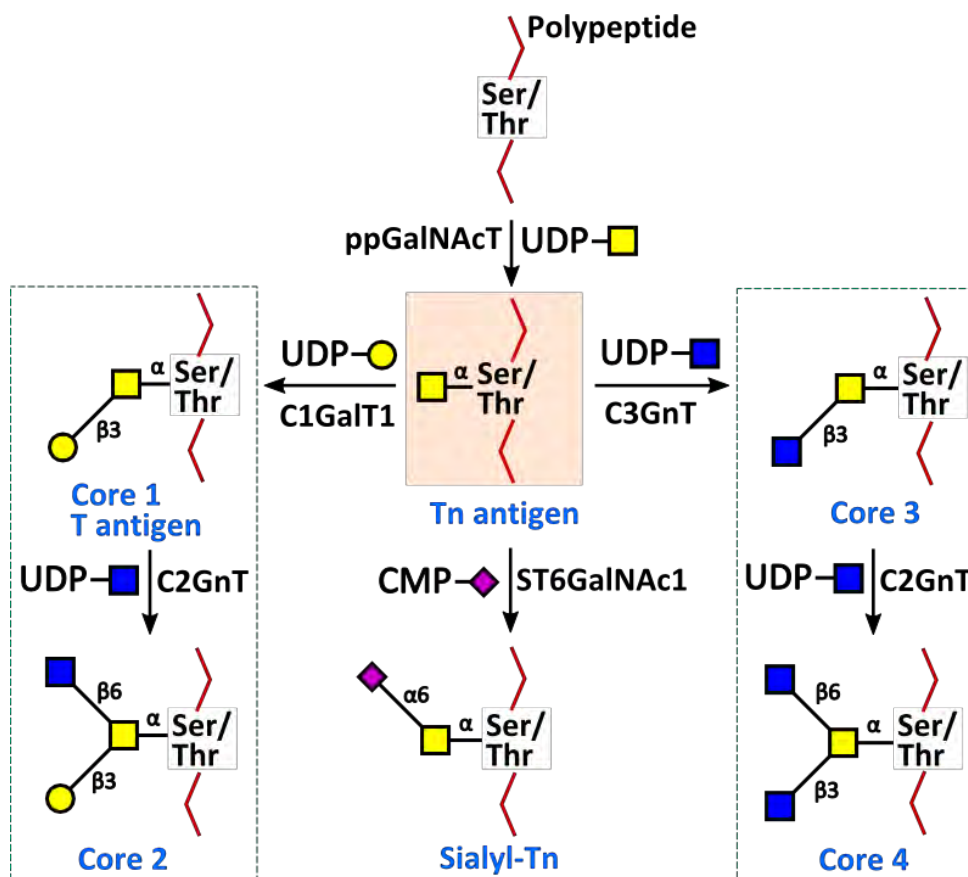
Figure 5. Major pathways for synthesis of *O*-GalNAc glycan (*O*-glycan) cores. More complex *O*-glycans are extended from these core structures. The biosynthesis of sialyl-Tn is also shown. Adapted from [22].

1.2.3 Glycosidase and glycosyltransferase specificity

Generally, glycosidases and glycosyltransferases are known to be highly specific with only one enzyme responsible for the formation or hydrolysis of a specific glycosidic linkage [23]. In this thesis, the gene of a glycosyltransferase will be referred to using all capital letters while the enzyme product (protein) will be abbreviated using small letters. For example, fucosyltransferase 8 (the enzyme) is the protein product of FUT8 gene and can be abbreviated as Fut8. Like other glycosyltransferases generally, Fut8 (Fig 6) catalyses formation on only one linkage, which is the addition of α1,6-linked fucose to the *N*-glycan core. Another example is GlcNAcT-III (the enzyme product of MGAT3 gene) (Fig 6), which covalently attaches β1,4-linked GlcNAc to the *N*-glycan core to form bisecting *N*-glycans.

However, multiple genes can encode for glycosyltransferases synthesising the same glycosidic linkage, with members of the same family having different substrate specificities. Notable examples include ST3GAL1-6 genes, which encode for α2,3-sialyltransferases (Fig 6). Although the gene products are known to catalyse addition of α2,3-linked sialic acids, the ST3Gal1 and ST3Gal2 enzymes

preferentially transfer sialic acids to the terminal Galβ1-3GalNAc oligosaccharide [34] while ST3Gal3 enzyme preferentially acts on type one chain (Galβ1-3GlcNAc) instead [34]. Another example is ST6GALNAC1 and ST6GALNAC2, both of which catalyse covalent addition of α2,6-linked sialic acid to GalNAc (Fig 6). ST6GALNAC1 acts preferentially towards Tn antigen substrate, while ST6GALNAC2 demonstrates preferences towards core 1 (*O*-GalNAc-Gal, T antigen)[35].
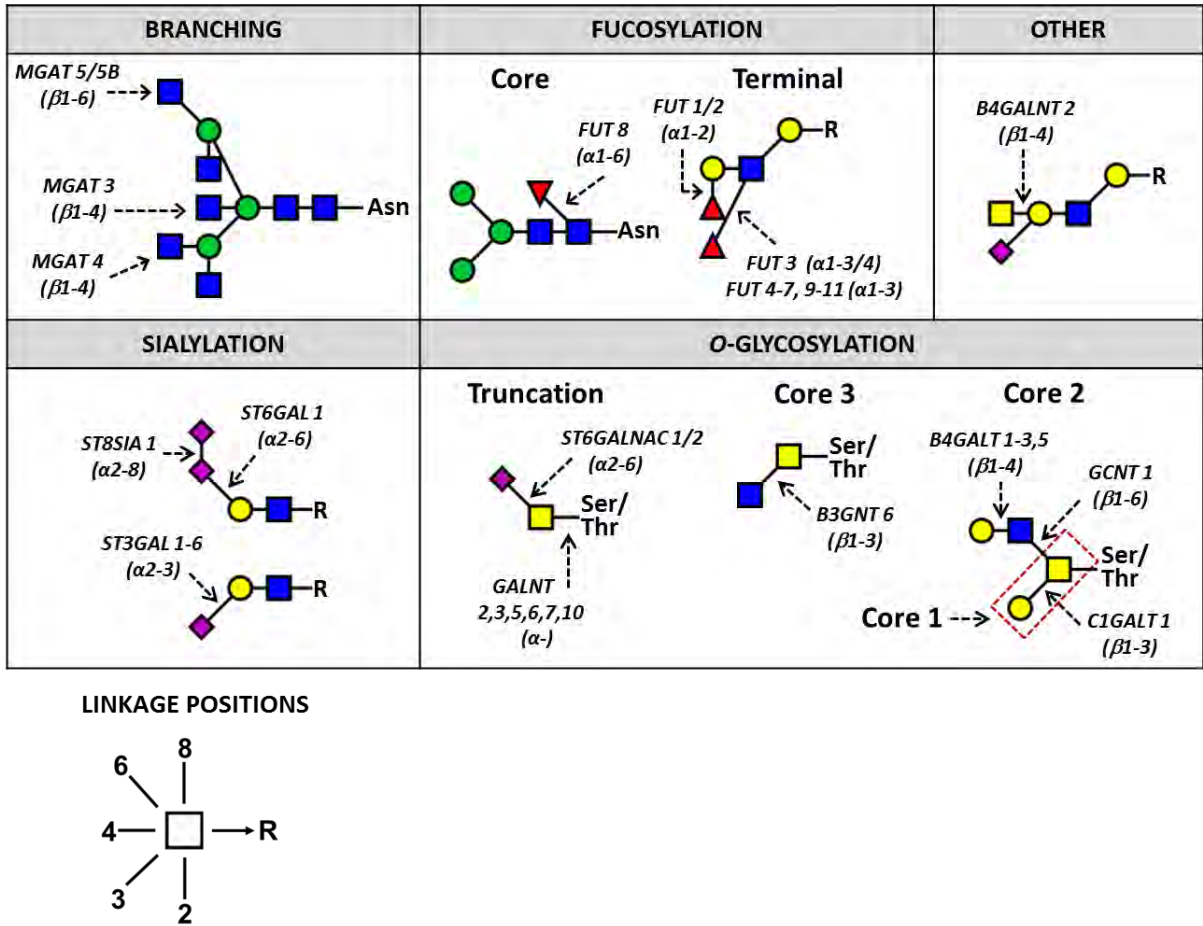
Figure 6. Representative diagram illustrating the glycosyltransferase genes involved in synthesis of various *N*- and *O*-glycans altered during different cancer types including colorectal cancer. R indicates *N*- and *O*-glycans. Taken from co-authored publication [36]**.**

1.3 Colorectal cancer (CRC)

There is a vast body of literature since the late 1980s [37] which strongly suggests that glycosylation and glycosylation-related changes occur during colorectal cancer (CRC) [38, 39] (see section 1.5). Hence, an overview on CRC will now be provided prior to further discussion of CRC-associated glycan changes. CRC is defined as cancer of the large bowel and/or rectum [40]. It is a highly prevalent type of malignancy worldwide, being the third and second most common cancer in males and females respectively [41]. Globally, over one million annual new cases of CRC are reported with half a million deaths each year [41]. In Australia the incidence reflects the worldwide trend with CRC being second

most commonly diagnosed cancer (15,840 cases/year) or cause of death from cancer (approximately 4000/year), according to the most recent estimates [42]. In the following sections, the anatomy and histology of the colon, the Australian clinicopathological staging system (ACPS) and the classical genetic model of CRC carcinogenesis is briefly reviewed.

1.3.1   Normal colon anatomy and histology

The colon is located in the terminal 1-1.5 m of the gastrointestinal tract between the ileum of the small intestine and the anus [43]. A schematic shown in Fig 7 illustrates the major regions of the colon: cecum; appendix; ascending, transverse, descending and sigmoid colon; rectum and anus. The colon functions to transport ions to maintain whole-body electrolyte homeostasis [44] and to desiccate partially digested semifluid food (chyme) to faeces [45].
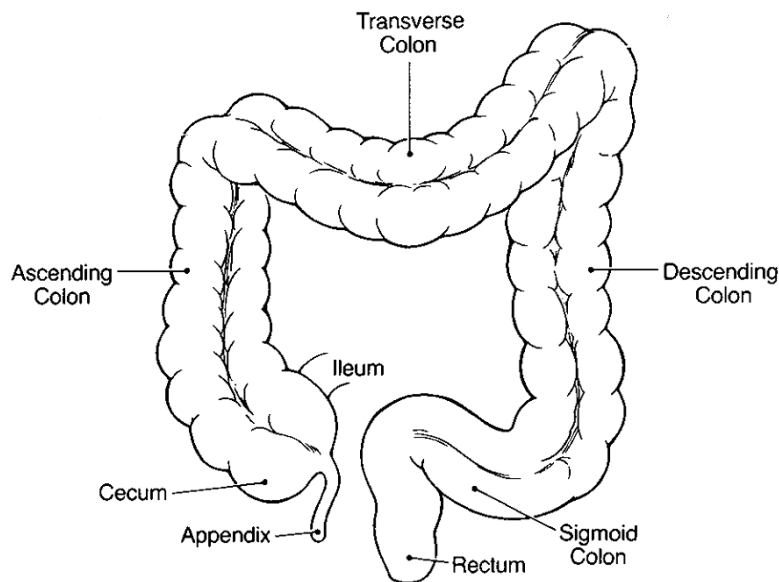
Figure 7. Schematic of the major regions of the colon. Adapted from [43].

Histologically, the colon contains layers similar to that of the small intestine but does not have villi [46]. In the normal colon, the major layers consist of the mucosa, submucosa, muscular propria and the serosa. Fig 8 shows a cross-section of a normal colonic mucosa. Numerous tubular glands, also known as crypts, can be observed. The mucosa is formed by the epithelium, lamina propria and muscularis mucosae [46]. The epithelium is the inner lining of the colon and contains epithelial cells of the simple columnar or cuboidal type [43]. Mucus-filled goblet cells also line the epithelium [43]. A major role of these cells is to synthesise, store and secrete mucus, which lubricates the colon to facilitate passage of faeces [46, 47].

Figure 8. Histology of normal colonic mucosae (transverse section, 30x magnification). Simple (absorptive) columnar cells and goblet cells form the colonic epithelium and lamina propria (the colonic stroma). The muscularis mucosae, submucosa and a lymphatic nodule are also shown. Image taken from [46].

The lamina propria extends from the basement membrane of the epithelial cells to the muscularis mucosae and forms part of the mucosal stroma (Fig 8). It contains a variety of cells such as fibroblasts and lymphatic cells (B cells, T cells and macrophages) [43]. The latter is involved in immunological defence and surveillance against noxious agents present in chyme [48]. Underlying the lamina propria is the muscularis mucosae which is a thin smooth muscle layer which separates the mucosa from the submucosa [43].

The submucosa is comprised of essentially similar constituents to the lamina propria, for example, lymphocytes, B cells, fibroblasts and macrophages [43]. Lymphatic cells may become aggregated to form lymphatic nodules [46]. It also contains connective tissues, blood vessels and nerves [46]. The muscularis propria, also known as muscularis externa, is a thick smooth muscle layer with a circular inner layer and a longitudinal outer layer. It functions to propel chyme through colon by neuronally and hormonally controlled mass movements called peristalsis [43].

Altered morphology and histology of the colon are observed in formation of benign and malignant tumours [49]. For adenomas, morphological changes include mild dysplasia, where there is a disordered cellular organisation and abnormal growth, increased apoptosis and mucin depletion. The appearance of the cell nuclei is also altered to become more enlarged and elongated. Depending on their altered morphology, adenomas are classified as tubular, villous or tubulovillous [50]. For cancers, the main histopathological criterion is invasion of neoplastic cells through muscularis

mucosae into submucosa [49, 50]. Despite this pathology, immune and stromal cells such as myofibroblasts may be present within the tumour [51]. A tumour can be classified or staged depending on the extent of invasion.

1.3.2 Australian clinicopathological staging system (ACPS)

The Australian clinicopathological staging system (ACPS) has been proposed for reporting colorectal adenocarcinoma and is comparable to other commonly used staging systems such as TNM (Tumour/Node/Metastasis) staging [52]. It utilises a combination of clinical, operative and pathological information when staging colonic tumours [53]. A variant of ACPS, from Concord hospital (Sydney, Australia), has also been used and all samples used in this investigation have been staged according to the Concord variant of ACPS. Table 2 compares sub-stages of ACPS variants with Dukes' and the TNM staging systems [54].

Table 2. Comparison of TNM, Dukes and ACPS variants of colorectal cancer staging systems [52-54].

Maximum spread	Stage group	TNM stage	Dukes	ACPS	Concord variant
Mucosa	0	Tis, N0, M0		A0	A1
Submucosa	I	T1, N0, M0	A	A	A2
Muscularis Propria		T2, N0, M0	B		A3
Beyond muscularis propria	IIA	T3, N0, M0	B	B	B1
Serosa	IIB	T4, N0, M0	B		B2
Regional lymph nodes	IIIA	T1-T2, N1, M0	C	C	C1
	IIIB	T3-T4, N1, M0	C		C2
	IIIC	T4, N2, M0	C	D	D1
Metastasis	IV	Any T, Any N, M1	D		D2

1.3.3 The genetic model of colon cancer

Many molecular aspects of CRC have been found to be dysregulated in this complex disease, including DNA [55, 56], protein [57, 58] and cell signalling pathways [59-62]. In this section, a general overview on the well-known genetic model of colon cancer carcinogenesis will be presented, prior to a more in-depth review of the known altered glycosylation in CRC.

The classical evolution of colon cancer is well understood because most colonic tumours progress through a series of distinct morphological stages. Many arise from adenomas [49], which are benign tumours of colonic epithelium. These adenomas may progress over decades to become fully malignant tumours. From the genetic viewpoint, the adenomas have gradually accumulated a series

of genetic mutations that confer selective growth advantage (Fig 9) [63]. Through clonal expansion, this allows the cells to outgrow sister cells and eventually become malignant tumours [64, 65].
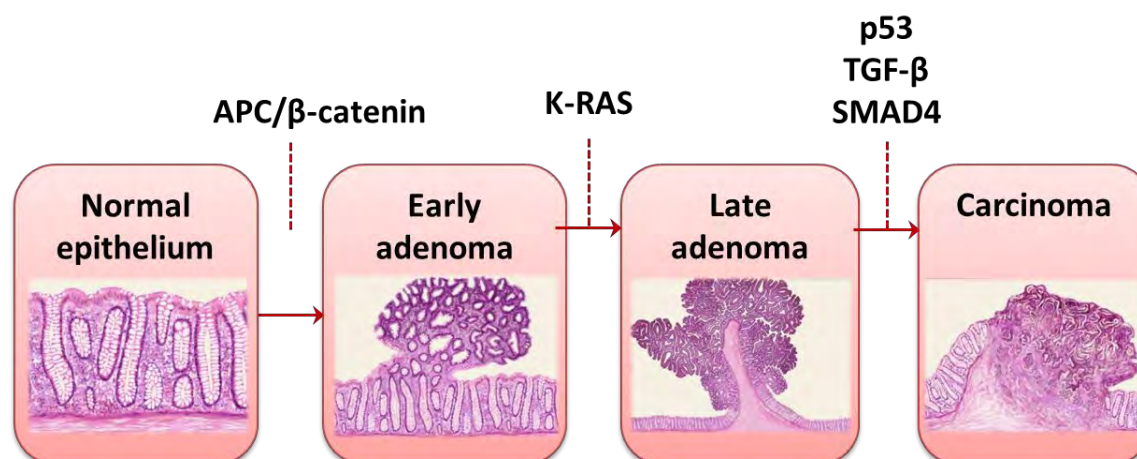
Figure 9. The classical genetic model of colorectal cancer. In this model, a series of accumulated genetic mutations eventually lead to the formation of malignant carcinomas. It is estimated that 50-85% of colorectal cancers initiate by this pathway. Compiled from [64-66].

The accepted classical genetic model suggests that a preferred order of genetic mutations exists, although it has also been acknowledged that it is progressive accumulation, rather than the sequence, which is the most crucial for progression to cancer [64]. Nevertheless, in the classical model, the first mutation is likely to result in only limited clonal expansion, such as a small adenoma (Fig 9). The first mutation often occurs in the APC gene [65] (Fig 9). APC mutation can occur somatically [67] or in the germ line [68], the latter of which is called Familial Adenomatous Polyposis [68]. APC mutations usually results in protein truncation [67] and loss of tumour suppressor function, thereby initiating the process of tumour development [63, 65]. A second mutation, in KRAS, for example, can allow the adenomatous cells to accelerate their growth to form a larger adenoma [65] (Fig 9). Additional mutations in genes such as p53, TGF-β and SMAD4 provide further selective growth advantage, until malignant cells develop which can invade and metastasise into other tissues [64, 65]. These mutations, which confer a selective growth advantage, have been reported as 'driver' mutations [65]. Although the estimated selective growth advantage of 'driver' mutations is only 0.4% between cell birth and cell death [69], the growth advantage may be compounded over many years, resulting in an abnormal and sizable cell mass.

1.3.4   Microsatellite instability (MSI) and DNA hypermethylation

In addition to the reported classical pathway for CRC carcinogenesis, the two major alternate pathways for tumorigenesis are also known and will be briefly mentioned here. These are microsatellite instability and hypermethylation of DNA.

It has been estimated that 15% of colorectal cancers have microsatellite instability (MSI) [70]. These tumours have mutations in their mismatch repair (MMR) genes. MMR genes form a highly conserved system which repairs erroneous insertion, deletion and mis-incorporation of bases that may arise during DNA replication. Dysregulation of MMR system is known to increase mutations occurring in short tandem repeats of DNA, called microsatellite regions [66, 71]. Four MMR genes have been identified which are commonly mutated in colorectal cancer: MLH1, MSH2, MSH6 or PMS2 [70]. When compared to microsatellite stable (MSS) colorectal cancers, differences in morphology include higher mucin production and poorer differentiation [72]. In addition, MSI cancers are more likely to arise from the proximal colon (that is, ascending and transverse colon), be more common in females and have a better prognosis [72].

On the other hand, DNA hypermethylation is also estimated to be associated with 40% of all colorectal cancers [66]. In this alternative pathway, tumorigenesis commences through accumulation of gene changes from hypermethylation of promoter regions, which are normally unmethylated in healthy cells. Hypermethylation generally leads to transcriptional silencing. In colorectal cancer, tumour suppressor genes such as p14, p16 and HLTF are frequently methylated and silenced [73]. Moreover, hypermethylation of DNA mismatch repair genes such as hMLH1 and MGMT can occur [72]. Therefore, there is an association between MSI and DNA methylation although the extent of overlap is not precise. For example, some cancers may exhibit extensive DNA methylation but not MMR mutation [74]. Despite this, DNA hypermethylation is an important tumorigenic factor in a minor proportion of colorectal cancers. The methylation-induced silencing of tumour suppressor genes and/or DNA repair genes may reduce apoptosis and elevate somatic mutation rate, eventually leading to tumour development [66].

1.4   Glycosylation and glycosylation-related changes in colorectal cancer

An existing and continuously growing body of literature strongly indicates that glycosylation is often altered during colorectal cancer (CRC) progression and this connection between glycosylation and CRC is the subject of investigation in this thesis. In the following sections, the various scientific approaches used for glycosylation studies will be briefly discussed, namely *in vitro, ex vivo* or *in vivo* systems of investigation (section 1.4.1) and the suite of currently available glycan analysis techniques (section 1.4.2) will be reviewed, before launching into a more comprehensive discussion of changed protein glycosylation in CRC in terms of glycan structural (section 1.5) and glycoprotein (section 1.6) alterations.

1.4.1 Systems of scientific investigation

There are three broad systems of scientific investigation which have been applied to the study of glycosylation and glycosylation-related changes. These are *in vitro* (Latin: in glass), *ex vivo* (Latin: out of the living), or *in vivo* (Latin: within the living) systems of investigation. *In vitro* systems are usually represented by cultured cells, *ex vivo* systems by human tissue samples and *in vivo* systems by animal models. Each of these investigative systems has their own advantages and limitations. Thus, optimal use is dependent on the particular scientific question of interest, which is elaborated further below.

Human colon tissues, along with their related biofluids such as serum, are undisputed as ideal samples in terms of clinical relevance [75, 76]. Since samples are of human origin, there are no species-specific differences, although there may be genetic variability amongst different subpopulations [77]. In addition, they are not usually propagated in culture, removing the potential for changes arising from culture conditions. Given efficient sampling, collection and storage procedures, these samples are likely to be highly reflective of the true human *in vivo* status. This valuable data can be further compared with clinicopathological parameters, to make statistically relevant correlations and provide diagnostic, prognostic or predictive information. However, sample enrichment or fractionation may be necessary to reduce their enormous biological complexity prior to analysis [76]. Nonetheless, analysis of human tissues is undoubtedly an extremely valuable and clinically relevant resource, despite their limited quantity and inherent complexity. Multiple studies involving human tissues have already identified glycan and glycoprotein changes associated with cancer which have been correlated with patient outcome [78-81].

Although tissues samples are highly clinically relevant, and correlations towards patient outcome can be made with matching clinicopathological data, it is frequently difficult to determine the underlying molecular regulation using this approach, since dose-response experiments cannot be easily performed due to legitimate ethical concerns. An additional consideration is the necessity to perform experiments where tumours are located within a cellular milieu, to reflect the full complexity of agent-host interactions. To this end, *in vivo* animal models are frequently employed to overcome limitations of an *ex vivo* approach, despite the potential for species-specific differences [82-84]. Nevertheless, there are several mouse and rat models developed to investigate pre-malignant colon conditions, such as adenomas, as well as early stage carcinomas [85]. Investigations into tumour properties, such as tumour growth [86-88] and metastatic potential [89, 90] are also relatively common in glycan studies, which has revealed some insights about glycosylation-related changes affecting tumour behaviour within a physiological system.

On the other hand, *in vitro* models are cancer cells which are routinely grown in the laboratory, having originated from tumours. Hundreds of colon and rectal cancer cell lines have been reported [86]. Some are more studied than others, for example, SW480 and SW620, since they are obtained from the same patient and therefore, considered an isogenic model of CRC metastasis [58, 91]. An obvious advantage of utilising cell lines is the ability to generate abundant sample material, which facilitates glycosylation studies. Also, investigations into genetic mutations, drug efficacy and cell function assays [87, 92] can be performed with minimal ethical issues. However, notable disadvantages include the lack of easily available non-cancer cell lines representing the normal healthy colon and the potential of well-established cell lines to be different from their parent tumours due to repeated subculturing [93]. Nonetheless, the many advantages of cell lines have made them a commonly used approach, particularly in the study of the regulation of glycan expression [94-98].

1.4.2   Analytical Techniques

Both targeted and non-targeted approaches have been utilised to study protein glycosylation and glycoprotein expression in CRC. Targeted approaches have consisted mainly of immunohistochemistry with antibodies (glycan- or protein-specific) or lectins, while non-targeted approaches primarily use mass-spectrometry. Both of these approaches have various advantages and disadvantages in terms of their throughput, glycan structural determination and degree of quantitation (summarised in Table 3).

Table 3. Major advantages and disadvantages of targeted and non-targeted glycoanalytical techniques

Approach	Typical techniques	Pros	Cons
Targeted	Immunohistochemistry Lectin/Western Blotting/ Flow cytometry	• High throughput • Amenable to large patient cohorts (100+)	• Semi-quantitative. • Dependent on availability and specificity of antibody/lectin
Non-targeted	Discovery-based mass spectrometry usually interfaced with orthogonal separation technique	• Ideal for novel biomarker discovery • Sensitive • Suited to detailed structural characterisation	• Low throughput. Small patient cohorts. • Sample preparation (enrichment) usually required.

For targeted approaches, the suite of antibody-based techniques is large and includes, but is not limited to, Western blotting [89], flow cytometry [98, 99] and immunohistochemistry [37, 100]. One clear advantage with antibody-based targeted approaches is their throughput, which is typically higher than their non-targeted counterparts. Immunohistochemistry is one technique frequently employed in large clinical cohorts, some exceeding 100 patients [100, 101] for investigation of the cellular expression of cancer-associated glycan epitopes, such as sialyl-Tn [78], Lewis [100] and sialyl Lewis antigens [100]. In contrast, many MS-based discovery glycan investigations use much smaller cohorts, for example, a recent study on *N*-glycosylation of proteins in CRC used only 13 patients [38] while an earlier study of mucin *O*-glycosylation analysed samples from only 3 patients[102]. Having mentioned this, non-targeted MS approaches can be made targeted by use of selected reaction monitoring (SRM) or related methods. This has been demonstrated on bovine milk oligosaccharides [103], *O*-glycans from salivary mucins [104] and *N*-glycans from a pancreatic cell line [105]. Though targeted SRM glycan analysis has not yet been utilised in CRC, the application of SRM in various glycan studies analyses shows the potential application of this quantitative technique for routine glycan analysis, which may also increase throughput. Taken together, the use of targeted approaches thus facilitates timely processing of statistically powerful sample populations, thereby increasing potential clinical relevance and correlation to patient outcome.

Additionally, the relative specificity of targeted approaches for particular glycan structures/motifs has made them a commonly exploited tool to detect specific glycan changes. In particular, lectins (glycan-targeting proteins) have been used in a variety of formats, such as lectin histochemical staining [106-108], lectin affinity chromatography [109, 110], lectin blotting [89] and lectin array [111]. Lectins can also be used for sample enrichment prior to subsequent analyses, including liquid chromatography tandem mass spectrometry (LC-MS/MS) [112]. The use of both glycan-targeted antibodies and lectins has yielded useful information regarding glycan changes during CRC,

particularly alterations in structural motifs such as sialyl Lewis X [98]. However, a limitation of these approaches is that the structure of the entire glycan moiety is difficult to determine, with the exception of truncated *O*-glycans such as Tn and sialyl-Tn, where the structure is known. Two approaches which can overcome this limitation towards more comprehensive structural characterisation are nuclear magnetic resonance (NMR) and mass spectrometry (MS). NMR relies on the absorption of electromagnetic energy by atomic nuclei when placed in a strong magnetic field while MS separates ions using electromagnetic fields, based on their mass-to-charge ratio (*m/z*). Of these, MS is relatively more common when studying human glycosylation, due to its higher sensitivity (especially for biological samples where sample quantity is limiting), although NMR has also been used [102, 113, 114]. MS is also frequently coupled with a separate pre-fractionation technique such as capillary electrophoresis, gas or liquid chromatography to further reduce sample complexity prior to MS fragmentation [115]. As liquid chromatography tandem mass spectrometry will be used as a mainstay technique in this thesis, liquid chromatography techniques will be further reviewed below. This will be followed by general overview of the various mass spectrometry architecture that can be employed to analyse glycans.

1.4.2.1 Liquid chromatography separation techniques for glycans

The main principle behind separation of glycans using liquid chromatography techniques is their retention (or adsorption) to a stationary phase, which competes with their solubility in the mobile phase. Thus, the material comprising the stationary phase is a critical consideration in LC separations, given that they display different characteristics and pH stabilities (Table 5). For glycan analysis in complex biological mixtures, three of the major separation techniques include (but are not limited to) porous graphitised carbon (PGC), hydrophilic interaction liquid chromatography (HILIC), and reverse phase (such as C18). A brief description of each of these separation techniques is provided in Table 4. Following this, the various advantages and disadvantages of these techniques for glycan separation are shown in Table 5.

Table 4. Major types of liquid chromatography separation techniques and their application to glycan analysis.

Technique	Description
Porous Graphitised Carbon	Glycans (labelled or unlabelled) are bound to a column with highly structured graphite surface and eluted with a mainly organic mobile phase [116]. Glycans retained based on planarity, size and charge [15].
Hydrophilic Interaction Liquid Chromatography (HILIC)	A variant of normal phase chromatography where glycans (usually labelled) are bound to a hydrophilic column and eluted with a hydrophobic (mainly organic) mobile phase [117]. Glycans are retained based on the number and size of polar groups which is usually correlated with glycan size [118].
Reverse phase	Glycans (labelled or derivatised for retention) are bound on a hydrophobic column (such as C18) and eluted with polar (aqueous) mobile phase [119].

	Charged glycans (such as those which are sialylated) are eluted earlier than neutral glycans [120].

Table 5. Comparison of the major types of stationary phases used in glycomics investigations.

Stationary phase for glycan separation	Advantages	Disadvantages	Ref(s)
Porous graphitised carbon (PGC)	- Separation of isomeric glycans. - Wide pH compatibility (0-14) and chemical stability. - Glycans may be labelled or unlabelled with a reducing end tag.	- Reduction of glycans required to prevent peak splitting due to reducing end anomers. - Increased retention for glycans of high molecular weight/charge.	[15, 116, 118, 121-123]
Hydrophilic Interaction Liquid Chromatography (HILIC)	- Separation of isomeric glycans. - Glycans may be labelled or unlabelled with a reducing end tag.	Lower pH compatibility compared to PGC (pH stability of 2 – 7.5) if silica-bonded phase used.	[118, 124-126]
Reverse phase	- Compatible with many standard reverse phase nano LC-MS/MS systems already used in proteomics laboratories. - Can be added as a second dimension following normal phase HPLC for higher chromatographic resolution.	- Isomers not as well resolved compared to HILIC and PGC. - Lower pH compatibility compared to PGC (pH stability of 2 – 7.5) due to silica-based particles.	[118, 119, 123]

As the biological function of glycans may be affected by their structure, the separation of isomers is an important feature required for LC technique to be utilised in this study. In this regard, both PGC and HILIC display superior capability to separate isomeric glycans. However, in our laboratory, PGC was shown to have a longer column life compared to HILIC [121], which may be attributed to its wider pH stability and therefore, PGC separations have been performed in this study. Nevertheless, it is acknowledged that HILIC may be employed to characterise glycans related to colorectal cancer and indeed, a glycomics study was performed by Balog et. al.[38], thus illustrating that multiple LC separation techniques may be applied and are complementary in related glycomics investigations. In addition, when PGC is interfaced with mass spectrometry, both glycan mass and the various retention times of separated isomers can be used to create a structure-retention time library of

known glycans [127]. However, the use of internal standards have been recommended in LC systems using capillary (or micro) flow to correct for any retention time shifts, as the reproducibility of capillary LC runs is not adequately robust [127].

Having mentioned the above, there remain some technical limitations to LC-MS/MS systems using PGC (and HILIC). Although different glycan isomers can be separated, they may not completely and unambiguously identified. In order to clearly identify these structures, additional information from exoglycosidase digestions and/or MSn fragmentation is extremely useful [128, 129]. Exoglycosidases (used in combination or singly) can be used to cleave specific glycosidic bonds, resulting in predictable and reproducible shifts in retention times, thus enabling the clear assignment of glycosidic linkages on glycans [129]. While this retention information can ultimately be utilised in the structure-retention time library (as above), it is noted that the initial building of the library may still require the use of exoglycosidases, especially when glycan quantity is limiting (and therefore, alternative technique such as NMR cannot be used). On the other hand, additional fragmentation stages, such as MS^3, may also provide valuable structural information, especially when specific exoglycosidases are unavailable. Using a spectral matching approach, for example, an oligosaccharide with 1-4 linked GlcNAc may be distinguished from another with 1-4 linked GalNAc by the correlation and intensity of MS3 fragments of a cross-ring cleavage ion [128]. Taken together, the use of exoglycosidases and/or MS3 may add a higher level of detail to glycan characterisation and overcome some of the technical limitations of LC MS/MS with PGC. Nonetheless, this degree of detail may not be required for all glycomics studies or even all glycans within the same investigation, allowing for uncertainty ('fuzziness') in structural assignments. In this thesis, glycan fuzziness is denoted by the use of brackets in the context of CFG nomenclature.

1.4.2.2 Mass spectrometric analysis of glycans

There are three main aspects of mass spectrometry architecture that can be considered when designing experiments for optimal glycan analysis: 1) Ionisation method, 2) Type of mass analyser and 3) Fragmentation for glycan sequencing. These will be discussed further below. As the low throughput-nature of data analysis is a notable bottleneck in glycomics, this issue will also be briefly discussed.

Ionisation methods

Recently, the power of MS has been utilised more than ever previously with the advent of 'soft' electrospray ionisation (ESI) [130] which allows more optimal information-rich fragmentation without excessive' destruction. The other predominant ionisation method in glycomics is MALDI (Matrix assisted laser desorption/ionisation and ESI (electrospray ionisation), although fast atom bombardment has also been historically used for glycan analysis [21, 131]. For

MALDI, the analytes (in this case glycans) are co-crystallised in a matrix on a target plate and analytes are ionised by laser induced desorption [132]. The resulting ions are accelerated through the mass spectrometer for analysis. In contrast, analytes are dissolved in a suitable solvent rather than crystallised in a matrix in ESI [133]. The solution containing the analyte is sprayed through a needle applying a high voltage which results in the generation of fine droplets and solvent evaporation. This leads to the formation of multiply-charged gas phase ions containing the analyte that can be further analysed in the mass spectrometer. Not surprisingly, these two different ionisation mechanisms have various advantages and disadvantages which are tabulated in Table 6 below.

Table 6. Ionisation methods commonly used for glycomics analysis. Both of these approaches detailed below as considered 'soft ionisation' techniques which allow for glycans to be ionised without their excessive destruction.

Ionisation method	Advantages	Disadvantages	Ref(s)
MALDI (Matrix assisted laser desorption/ionisation)	• Simple and less labour intensive sample preparation compared to ESI. • Analysis rapid when performed without prior fractionation. • Provides quick snapshot for monitoring exoglycosidase digestions or estimating glycan variety in biological mixture.	• Fractionation (if desired) needs to be performed offline. • No separation of isomeric glycans if no prefractionation performed. • Dissociation of labile glycosidic bonds for native/reduced glycans. These are bonds involving acidic monosaccharides and their substituents (sulfate/phosphate).	[132-134]
ESI (Electrospray ionisation)	• Commonly interfaced with online liquid chromatography (LC) to obtain a highly detailed glycan profile. • Can be used to profile native or reductively aminated glycans without dissociation of labile glycosidic bonds.	• Samples need to be free of salts to prevent formation of adduct peaks. • More analysis/instrument time required when coupled with online LC.	[132-134]

Despite the noted disadvantages of ESI, such as more labour-intensive sample preparation, this ionisation method has been utilised for glycan characterisation in this thesis as it captures a more highly detailed glycan profile for the analysed samples. In addition, ESI is frequently coupled with liquid chromatography (LC) and this is an approach which will be used in this study. The coupling of

ESI with LC allows for separation of glycan isomers and improved analysis of low abundance glycans relative to unfractionated mixtures, which is highly suitable for data-dependent discovery of novel glycan biomarkers that is an aim of this investigation.

Mass analysers

There are a variety of mass analysers available (for detail review please see Yates) although several mass analysers are preferred for glycomics analysis. These include quadrupole ion traps (IT) and TOFs (time-of-flight) analysers although Orbitraps and FTICR (Fourier Transform Ion Cyclotron Resonance) can also be used. A brief overview of the working principle behind these mass analysers is provided below. Each type of mass analyser has unique properties such as differences in resolution, mass accuracy, sensitivity and scan rate. Multiple mass analysers can also be combined to create hybrid mass analysers for specific analytical requirements. The basic performance parameters of frequently used mass analyzers (singly or in hybrid) are shown in Table 7 for further comparison.

Quadrupole Ion traps

Quadrupole Ion traps are known to exist in two types – the three dimensional quadrupole ion trap (also known as Paul trap) and the two-dimensional linear ion trap (LIT). Both operate on the principle that charged ions are trapped by a combination electric and magnetic fields produced by four electrodes [135]. Despite their limited mass accuracy and resolution (Table 7), these instruments are still used standalone because of their high sensitivity, especially compared to beam type instruments [136]. Linear Ion Traps can also be used at the front end of hybrid instruments and combined with other mass analysers [135].

Time-of-flight (TOF)

Time-of-flight (TOF) analyzers produce mass spectrum by measuring the flight time of charged ion down a vacuum and field-free flight tube. A limitation of initial TOF instruments is inability to perform high quality MS/MS as the ability to select ions for fragmentation is limited [136]. To overcome this limitation, a tandem TOF instrument (TOF/TOF) may be used. This instrument uses the first TOF analyser as an ion selector and after ion pass through a collision cell, the second TOF produces MS/MS.

Orbitraps

Orbitraps captures ions by influencing their motion in a static electromagnetic field, where they move in a circular (or orbital) motion around a central electrode and oscillate in the axial direction. The *m/z* value is then determined by Fourier Transform algorithm to convert the ion's frequency of motion into mass spectrum [137]. These are frequently used as back end of hybrid instruments

(where the front end is composed of mass analysers for ion selection and fragmentation such as a linear ion trap (LTQ)).

*Fourier transform ion cyclotron resonance (FTICR)*

These mass analysers are known to have the greatest capacity for mass resolution (up to 100,000) and mass accuracy (1-2 ppm) [136]. They operate by trapping ions in a static electromagnetic field where they move in circular motion and the *m/z* value of an ion is derived by the ion's frequency of motion. Similar to Orbitraps, FTICRs can also be utilised as back end of hybrid instruments (where the front end is composed may be a LTQ).

Table 7. Performance comparisons of mass spectrometers frequently used for glycan analysis. Adapted from [136] and [138].

Instrument	Resolution	Mass accuracy	Scan Rate
Linear Ion Trap (LTQ)	2000	100-300 ppm	Fast
TOF-TOF	20-25,000	10-20 ppm	Fast
Quadrupole Time-of-flight (QTOF)	10,000	2-5 ppm	Moderate
LTQ-Orbitrap	100,000	2 ppm	Moderate
Fourier Transform Ion Cyclotron Resonance (FTICR)	50-100,000	1-2 ppm	Moderate
LTQ-FTICR	500,000	< 2ppm	Slow

Tandem mass spectrometry of glycans (MS^n)

Tandem mass spectrometry involves the selection and fragmentation of precursor (or parent) ions to generate product ions which are further analysed in the mass spectrometer. It can be performed in multiple stages, with fragmentation occurring in between each stage. Methods used to induce fragmentation of biological compounds (such as glycans) include collision induced dissociation (CID), electron-based dissociation and infrared multiphoton dissociation (IRMPD). Regardless of dissociation method employed, structural glycan information can be obtained depending on the fragmentation pattern since glycans fragment reproducibly along specific pathways to produce "diagnostic ions". In particular, these ions may be used to distinguish glycan structural isomers from one another [139]. A widely adopted method to specify glycan ions is the nomenclature originally proposed by Domon and Costello (shown in Fig 10).
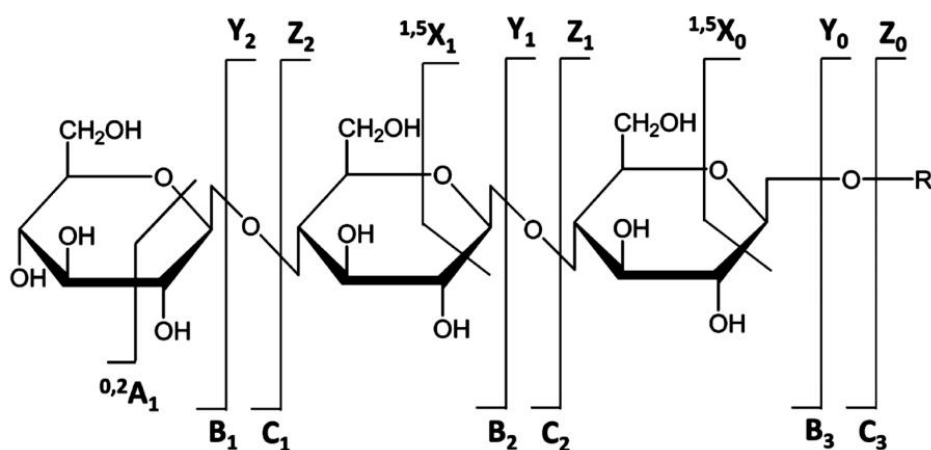
Figure 10. Nomenclature for glycan fragments obtained from tandem mass spectra. Originally published in [140].

In terms of fragmentation for CID, there are two major types of cleavages that can occur in glycans: glycosidic cleavages where the linkages between two adjacent monosaccharides are broken and cross-ring cleavages, there is rupture within a monosaccharide residue's sugar ring. Glycosidic cleaves provide structural details regarding a glycan's composition and sequence whereas cross-ring cleavages reveal information about the glycan's various linkage types [140]. In addition to these, a third type of cleavage, known as internal cleavage may also occur. These also potentially useful fragments are generated when two (or more) cleavages (either glycosidic or cross-ring) occur on different regions on the glycan molecule [140, 141].

Glycans may be analysed in either positive or negative ion mode. In positive ion mode, glycan may be either protonated or observed as a metal (e.g. sodium) adduct [140]. Various studies have shown that protonated glycans produce glycosidic cleavages, which provides sequence but not linkage information [141]. Due to limited ionisation efficiency and lability of sialic acid groups in positive ion mode, glycans may be permethylated, to increase their chemical stability. This involves the conversion of every hydroxyl group (-OH) to an O-Methyl group [132] and allows for sialylated glycans to be analysed alongside non-sialylated glycans in positive mode.

It has been argued that analysis of permethylated glycans in positive mode is preferable to that of native glycans, since they ionise better and give more definitive sequencing, based on the predictable and well-characterised series of fragment ions [132]. This includes the ability to generate fragment ions that provide detail information about glycan epitopes at the non-reducing end terminus, such as sialyl and non-sialyl Lewis A/X epitopes (for more information on Lewis epitopes, please see section 1.5.1.2). However, this level of information may also be obtained in the negative ion mode, which also has a substantial number of conserved fragment ions [139, 142]. Moreover, cross-ring fragments

appear to be more abundant in negative ion mode - these may provide additional information on structural features such as branching on *N*-glycan antennae.

Nevertheless, a degree of ambiguity can arise for both negative and positive ion mode glycan analysis, either due to limited availability of diagnostic fragment ions or the complex nature of the glycan (for instance, if different terminal epitopes exist on different antennae). To overcome this limitation, pertinent precursor ions from MS2 may be selected for further fragmentation (MS3) in order to specifically generate diagnostic fragments which can distinguish different epitopes (such between sialyl Lewis A/X and blood group H). Although this approach of multi-stage fragmentation is likely to provide more comprehensive structural information, it is also acknowledged that the use of exoglycosidase digestions can also give further structural details, in particular, information regarding specific glycan linkages. The incubation of glycans with exoglycosidases of known specificity that cleave monosaccharide residues from the nonreducing terminus results in reproducible changes in retention time/mass and this classical approach may also be employed to provide complementary information to mass spectral analysis (as previously mentioned in section 1.4.2.1).

1.4.2.3 Perspectives and limitations of mass spectrometric glycan analysis

Together with knowledge of known glycan biosynthetic pathways, glycan structures can be highly characterised using MS and MS/MS data, including low-abundance structures at 100 fmol/µL [143]. This sensitivity also makes it a powerful analytical technique for glycan analysis of biopsy-sized tissue samples (approximately 5 mg of material wet weight [102, 122]) which are considered clinically relevant but consisting of only small glycan quantities. Due to its non-targeted nature, it also has higher capacity in the determination of exact glycan structure such that structures not previously predicted to be associated with prognosis may be uncovered [102]. Therefore, its sensitivity and higher coverage enhances the potential to identify novel glycan biomarker changes in low-abundance glycans.

Despite its sensitivity and structural characterisation capacity, a significant drawback to current MS techniques is the current low throughput nature of glycan data analysis. At present, data analysis of the mass spectra of glycan structures remains highly manual and time-consuming. Several bioinformatics tools are in development in order to overcome this substantial bottleneck in glycan profiling [144, 145]. Furthermore, data deposition into curated databases such as UniCarb-DB [146, 147] (http://unicarb-db.biomedicine.gu.se/) is becoming more common and indeed, increasingly encouraged by editors of reputable proteomics journals [148]. The two-pronged approach of developing algorithms for automated spectral matching [147] (similar to those routinely used in proteomics), as well as increasing high-quality glycan spectra available for matching, is envisioned to increase throughput and automation of glycan analysis in the near future.

40

As we continue improving our current knowledge of glycan expression, there are also concurrent investigations into the underlying regulation of glycan biosynthesis. The majority have mainly focused on the regulation of glycosyltransferases, although there are a few focused on glycosidases such as sialidases [149, 150]. In terms of analytical techniques, mRNA assays, such as qRT-PCR (quantitative reverse transcriptase polymerase chain reaction), at present remain a mainstay in both tumour tissues and cell lines to analyse the expression of glycan processing enzymes [98, 151-154]. The sensitive, specific and quantitative nature of qRT-PCR may have contributed to the popularity this technique. The sensitivity is especially critical since many glycosyltransferases are extremely low in abundance and specialised enrichment strategies are frequently necessary for their protein detection. Thus, there are few studies measuring their expression at the protein level (examples include [155] and [156]). Finally, a small number of studies have used enzyme activity assays to investigate glycosyltransferase regulation [157, 158]. Previously, the application of this technique was relatively limited as in-house chemical synthesis of assay reagents, such as donor and acceptor substrates, was required due to lack of commercial availability. However, this may alter in the near future with increasing availability of commercial kits for glycosyltransferase activity.

1.5 Altered glycosylation and glycan-processing enzymes in colorectal cancer

Generally, when investigating glycosylation-related changes in CRC, there have been two main *ex vivo* sample types used – colon tissue and blood serum/plasma [76, 159], although a number of studies have used faecal [160, 161] and urine samples [162]. Tissues have been obtained from primary colon tumours or from their metastases. Colon primary cancer tissues are usually procured from surgical resections, sigmoidoscopies, or colonoscopies [163], while serum is frequently obtained by minimally invasive venipuncture. Glycosylation changes occurring in serum have been reviewed elsewhere [1] and are outside the scope of this study. Here, the changes in glycosylation that have been reported to occur on the cell surface of *in vitro* cell lines, *ex vivo* tissues or animal models of CRC are the major focus. A summary of findings on the glycosylation changes in CRC carcinogenesis is collated in Table 8 (below). Moreover, alterations accompanied by changes in the glycan biosynthetic machinery such as the expression of corresponding glycosyltransferases will also be discussed. These have been summarised in Table 9. A schematic (Fig 6) summarising the major glycan changes associated with CRC and their corresponding glycosyltransferase genes is also shown. The major types of reported glycosylation alterations are sialylation, fucosylation, core 3 *O*-glycans and bisecting *N*-glycans and these are discussed in detail below. The information presented in this section has been incorporated into a peer-reviewed publication (located at the end of this chapter) [36].

Table 8. Cell surface glycan changes in colorectal cancer

Glycan change	Regulation	Sample Type (H/M)		Disease stage		Analytical method			Ref(s)
		Tissue	Cell line	Cancer vs non cancer	Early/late/ metastatic	MAb	Lectin	HPLC/MS	
Sialylation									
Sialyl Lewis A	d	H			x	x			[100, 164, 165]
Sialyl Lewis X	↑, d	H		x	x	x	x	x	[100, 102, 157, 164, 166, 167]
Sialyl Lewis-type	↑	H		x				x	[38]
α2-3	↑	H		x			x		[106]
Fucosylation									
Lewis A	d	H		x		x			[100]
Lewis X	↓, d	H		x	x	x		x	[100, 102, 166]
Branching									
Bisecting GlcNAc	↓	H		x				x	[38]
Truncation									
T	↑, ↓	H		x	x	x	x	x	[37, 166, 168]
Sialyl Tn	↑, d	H		x	x	x			[37, 78, 80, 164, 166, 169, 170]
Paucimannose-type	↑	H		x				x	[38]

Legend Key and abbreviations: ↑, Increase; ↓, decrease; d, detected; nd, not detected; H, human; MAb, monoclonal antibody; HPLC/MS, high performance liquid chromatography mass spectrometry

Table 9. Glycosyltransferase regulation of cell surface glycans in colorectal cancer

Glycosyltransferase change	Regulation	Sample Type (H/M)		Disease stage		Analytical method		Ref (s)
		Tissue	Cell line	Cancer vs non cancer	Early/late/ metastatic	RT-PCR	Enzyme Assay[a]	
Sialylation								
ST3GAL 1	↑	H		x		x		[151]
ST3GAL 3	↑	H			x	x		[81]
ST3GAL 4	↑, ↓	H		x	x	x, NB		[151, 153]
ST6GAL 1	↑	H		x		x		[81]
ST3GAL 2	↑	H		x		x		[153]
ST3GAL Transferase	↑	H		x			x	[158]
ST6GALNAC Transferase	↓	H	H	x			x	[96, 158]
Fucosylation								
FUT 8	↑	H		x			WB	[155]
FUT 3	↑, ↓, d	H	H	x	x	x		[81, 98, 153]
FUT 1, 4, 6	↑	H		x		x	x	[81, 151, 153, 157]
Bisecting								
MGAT 3	↑	H		x	x	x		[81]
Truncation								
PPGALNAC Transferase	↓	H		x			x	[158]
Other								
B4GALNACT 2	↓	H		x			x	[171]
B3GNT6, B6GN transferase	↓		H	x			x	[96]
C1GALT 1	↑	H		x			x	[158]

Legend Key and abbreviations: ↑, Increase; ↓, decrease; d, detected; nd, not detected; H, human; M, mouse; WB, Western Blot; NB, Northern Blot

1.5.1 Sialylation

Sialylation on the cell surface in general has been associated with poor prognosis in CRC. An early study reported significantly increased sialylation in metastases compared to corresponding primary colorectal tumours (n=23) by using lectins targeted towards α2,3-sialylation (from *Maackia amurensis,* MAA) and α2,6-sialylation (from *Sambucus nigra,* SNA) [166]. A separate study also indicated that the type of sialylation (α2,3- vs. α2,6 sialylation) was also important for metastasis in CRC. In this study, α2,3-sialylation was suggested by reactivity towards MAA lectin; while α2,6-sialylation was indicated by SSA (*Sambucus sieboldiana*) lectin which has also shown preference for α2,6-linked sialic acids like the SNA lectin [172]. For MAA lectins, 45% (29/65) of cancer colonic tissues were stained positive whereas there was only weak staining in the corresponding normal mucosa [106] which indicated an increase of α2,3-sialylation in the tumours. Moreover, α2,3-sialylation was also significantly associated with lymph node metastases and lymphatic invasion (p < 0.01). In contrast, for SSA lectins, only 23.1% (15/65) of tumours showed positive staining and SSA reactivity was not associated with lymph node metastases or lymphatic invasion, suggesting that α2,6-sialylation was not markedly alter during CRC progression [106].

Interestingly, this contradicts the findings of other studies where elevated α2,6-sialylation expression is implicated in human colonic tumours [81, 153]. For example, there is some evidence to suggest that α2,6-sialylation may enhance tumour malignancy by mediating metastasis-related cell properties. A salient example is cell adhesion where sialylation may play a functional role. Using LS174T colon cancer cells, cell adhesion was found to be inhibited following pre-treatment with sialylation neutralising lectins (SNA) or desialylation with neuraminidase [166]. This suggests that sialylation, particularly α2,6-sialylation, is involved in cell attachment, which may have implications for metastasis [173]. Thus, there are some minor discrepancies in the literature as to whether α2,6-sialylation is elevated or remains unchanged in CRC and the association of CRC with specific sialic acid linkages therefore requires further investigation. Nevertheless, the various lines of evidence generally indicate that increased overall sialylation is associated with CRC progression. Moreover, additional investigations are needed to verify whether reduced cell adhesion inhibits metastasis in animal models and humans.

1.5.1.1 Sialyl-Tn

In addition to alterations in sialylation in general, it was also observed that specific sialylated antigens (such sialyl-Tn, Fig 5; and sialyl-Lewis antigens, Fig 11) can be increased in CRC tumours [166]. A well-known and prominent example is sialyl-Tn (Fig 5). Sialyl-Tn (NeuAcα2-6GalNAc) is a disaccharide glycan antigen that has been known as a pan-carcinoma cancer marker since the 1980s [37]. Sialyl-Tn expression in CRC tissues has been relatively well-studied using monoclonal antibodies, and more recently using mass spectrometry [38, 102]. Specifically, there have been a handful of studies on sialyl-Tn (STn) in CRC tissues using four well-known anti-STn antibodies (MLS102, B72.3, TKH2, HB-STn1) with patient cohorts ranging from 29 to 137

patients [37, 78, 80, 164, 169, 170, 174]. These studies demonstrate that this antigen is overexpressed in colon tumours, with a majority of colorectal cancer patients (83.4%, 407 patients across 6 studies) reported as STn-positive whereas this antigen is not generally observed in non-neoplastic tissues [94, 169]. In addition, the increased expression of this marker has also been significantly associated with shorter disease-free survival and overall 5 year survival in CRC ($p < 0.001$)[78].

Although a majority of sialyl-Tn studies have been undertaken in tissues, a small number have been conducted using the LS174T cell line. LS174T is a human mucinous and epithelial colonic cell line reported to produce sialyl-Tn [94] and therefore previously used as an *in vitro* model to understand the enzymatic basis of sialyl-Tn expression [95]. Among the most critical glycosyltransferases for sialyl-Tn expression are core 1 synthase (C1GalT1), α2,6-sialyltransferase 1 (St6GalNAc1), and core 3 synthase (C3GnT) as shown in Fig 5. Interestingly, complexities have emerged when comparing sialyl-Tn regulation in LS174T cells and tumours. When clones of LS174T with differential sialyl-Tn expression were compared, the activity of core 1 synthase (C1GalT, Fig 5) was found to be extremely limited in low sialyl-Tn expressing LS174T clone [95]. The activity of other relevant enzymes such as α2,6-sialyltransferases (for example, ST6GalNac1, Fig 5) and core 3 synthase (C3GnT, Fig 5) were also assayed but activities were comparable between the high and low-sTn expressing clones. This suggests that reduced core 1 synthase expression is an important factor in increased sialyl-Tn. More importantly, the activity of core 1 synthase was associated with sialyl-Tn expression in these cell lines. This stands in stark contrast to human tissue where a similar study indicates that activities of relevant glycosyltransferases are not always associated with the expression of sialyl-Tn in human tumours [158]. This included the activities of core 1 synthase, α2,6-sialyltransferase and core 3 synthase which were altered, although not always correlated with sialyl-Tn [158]. In particular, α2,6-sialyltransferase activity was found to be generally decreased in tumours compared to their pair-matched normal mucosae, which was unexpected. Overall, these discrepancies highlight the additional complexity in tissues compared to cell lines and suggest differences in sialyl-Tn regulation in human tissues which warrants further investigation.

1.5.1.2 Sialyl-Lewis and Lewis antigens
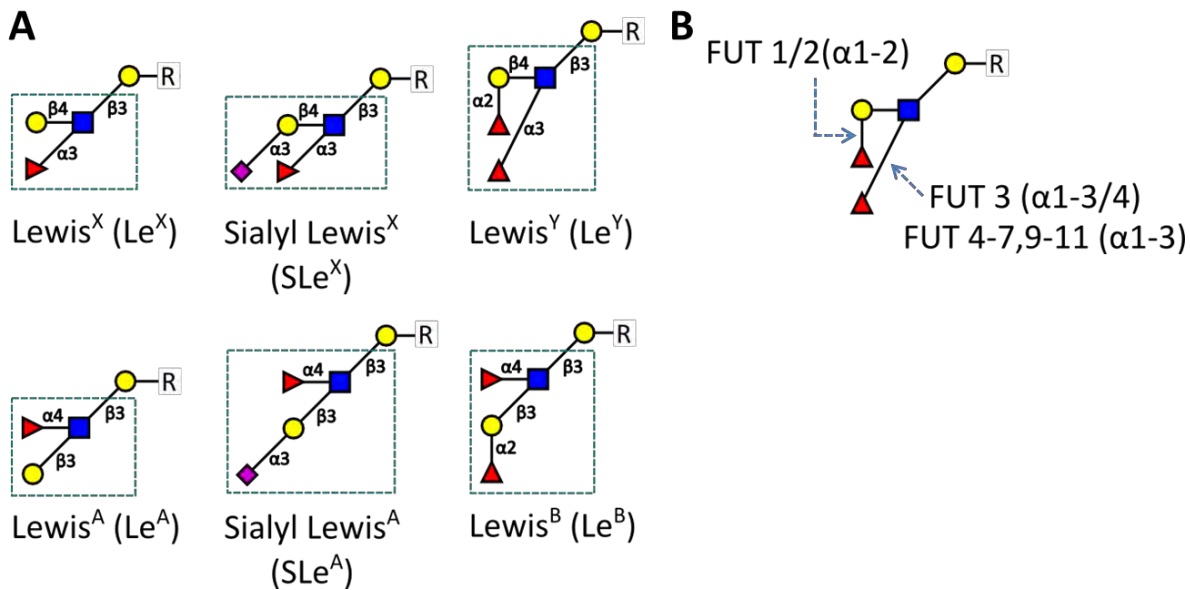
Figure 11. Lewis and sialyl-Lewis antigens (A) and the fucosyltransferase enzymes involved in their synthesis (B). R represents *N*- or *O*-glycan or glycolipid. Schematic adapted from [33] and [36].

The Lewis antigens are a series of related glycans containing α1-3/α1-4 fucose residues and well-known members of this group include Lewis^A, Lewis^B, Lewis^X and Lewis^Y (Fig **Error! Reference source not found.**). Two Lewis antigens can be sialylated – Lewis^A and Lewis^X, as shown Fig **Error! Reference source not found.** [33]. A number of studies have investigated the association between sialyl-Lewis antigens and CRC [100, 101]. In one study, immunohistochemistry was performed on 132 human colorectal adenocarcinomas, using anti-sialyl Lewis X (sLe^X) antibody, F6H [101]. SLe^X-positivity in tumours was significantly associated with depth of tumour invasion (p=0.0026), presence of lymph node metastasis (p=0.0002), lymphatic invasion (p=0.003) and disease stage (p=0.0013). There was also a highly significant difference in 5-year disease free survival rates between sLe^X-positive and sLe^X-negative patients (57.7 and 89.1% respectively, p=0.0002) as well as 5-year overall survival rates (sLe^X-positive, 58.3%; sLe^X-negative, 93%, p < 0.0001)[167]. A similar observation was found in another study where 81 normal colonic mucosa and 101 pair-matched colorectal carcinoma tissues were examined and sLe^X-positivity was significantly associated with more advanced cancer (p=0.0029) and shortened survival time post-surgery (p=0.001) [100]. Furthermore, these results were further supported by a recent study using LC-MS/MS which also found increased abundance of sialyl-Lewis type structures [38]. Together, these investigations indicate that increased expression of sLe^X is correlated with further cancer progression and overall poorer prognosis in CRC.

The regulation of sialyl-Lewis antigens is likely to be complicated given the array of different glycosidic linkages and sugar residues required for their synthesis. One of the many factors known to regulate sLe^X

expression is fucosyltransferase expression (Fig 10b). In one study using HT29 colon cancer cells, FUT3 gene expression, which corresponds to α1,3/4-fucosyltransferase (Fig 6), was found to be implicated [98]. When surface sLe^X expression was increased by sodium butyrate treatment, mRNA of FUT3 was correspondingly increased [98]. The converse was observed when sLe^X expression was reduced. These changes were not associated with other fucosyltransferase genes such as FUT4 and FUT6. Overall, these observations strongly suggest that FUT3 gene regulation is correlated with sLe^X expression in HT29 cells. However, some contradictions were found when FUT3 mRNA and sLe^X expression were determined in human colonic tissues in the same study [98]. Remarkably, the majority of tumours (6 of 7) exhibited a high level expression of sLe^X compared to their matched normal mucosa. Nonetheless, this contrasts with their FUT3 mRNA expression, which was mainly unchanged or decreased, rather than increased as expected. Taken together, this demonstrates that FUT3 mRNA expression can be correlated in culture colon carcinoma cells (HT29) but not necessarily in colon carcinoma tissue. This is redolent of the pattern previously mentioned for sialyl-Tn, where glycosyltransferase expression was more readily associated with the expressed glycan structure in cell lines, compared to human tumours. It also further underscores the differences in complexity between cell lines and tumours.

As alterations in fucosyltransferase expression can also be characteristic of Lewis antigens, it is not surprising that Lewis-type antigens such as Lewis X (Le^X) have been linked with CRC. When Lewis antigen expression was investigated using tissue immunohistochemistry (n=101)[100], Le^X-positivity was common amongst many patients (66/101, 65.4%) and survival time post-surgery was significantly shorter in Le^X-expressing patients compared to Le^X-negative patients (p=0.023). Using Cox's regression analysis, Le^X expression in tumours was also prognostic for decreased patient survival, independent of stage and histological type. All in all, this indicates Le^X-positive tumours are associated with poorer prognosis compared to Le^X-negative tumours. Interestingly, there is some evidence to suggest a different trend is observed when Le^X epitope is sulfated, as reported by a later study employing LC-MS/MS [102]. Here, the sulfo-Le^X antigen was carried as part of a pentasaccharide [SO3-3Galβ1-4(Fucα1-3)GlcNAcβ1-3(NeuAcα2-6)GalNAc], whose expression was decreased, rather than increased, in the tumours [102]. However, this was only observed in a preliminary study with three patients. Additional verification in a larger cohort is necessary to determine the statistical significance of reduced sulfo-Le^X expression and its biological role in tumours.

To conclude, there is a body of literature which strongly indicates that increased sialylation and sialylation-related antigens are associated with malignant CRC and poorer patient survival. There have been notable efforts to investigate the underlying regulatory mechanisms of sialylation, such as the association of FUT3 with sialyl-Lewis^X expression. However, it was also reported that association between glycosyltransferase gene expression and resultant glycan structures in human tissues is less clear than for cell lines. In addition,

increased Le[X] expression has also been correlated with CRC and Le[X] expression was found to be prognostic for decreased patient survival. Though the expression of enzymes related to Le[X] expression has not been investigated, it is speculated that a variety of fucosyltransferases (Fig 6) are involved as they are an essential feature of Lewis antigens. Overall, the regulation of sialylation and (sialyl) Lewis[X] antigen expression is important given that their increased expression has been correlated with indicators such as shortened survival, invasion and metastasis which may contribute to an adverse patient outcome.

1.5.2   Core-Fucosylation

There has been some investigation into the role of core-fucosylation in CRC progression. For example, its expression has been correlated with cell adhesion [155]. Core-fucosylation is a type of fucosylation which occurs only on *N*-glycans and consists of the α1,6-fucosylation of the chitobiose core by fucosyltransferase 8 (encoded by FUT8 gene, Fig 6). In one study, fucosyltransferase 8 (Fut8) and E-cadherin were reported to be elevated in five of six tumours from pair-matched normal and tumour colon tissues [155]. Since E-cadherin is a protein well-known to mediate cell adhesion [175, 176], it was hypothesised that abnormal cell-adhesion could result from increased core-fucosylation and E-cadherin expression. The impact of increased core-fucosylation on E-cadherin-dependent cell adhesion was confirmed by FUT8-transfection into WiDr colon cancer cells, which had low endogenous FUT8 expression [155]. Increased cell adhesion was observed following FUT8 transfection. In addition, the reverse occurred when FUT8 was knocked down in TGP49 cells [155]. Cells showed more loose cell-cell contacts and significant decrease in E-cadherin dependent cell aggregation. Taken together, these findings suggest that core-fucosylation has a role in E-cadherin-dependent cell adhesion in cancer and FUT8 expression is closely involved in this process. Thus, there is evidence to suggest that core-fucosylation and its regulation may have functionally important roles in CRC.

1.5.3   Core 3 *O*-glycans

Expression of core 3 *O*-glycans (Fig 5) are known to be relatively common in the healthy human colon with a majority of glycan structures based on the core 3 structure, as determined by LC-MS/MS [122, 177]. The majority of structures in CRC tumours are also reported to be based on the core 3 structure [102]. Interestingly, the activity of the glycosyltransferase integral to core 3 expression (core 3 synthase, also known as β3GnT6,Fig 6) is frequently reported as decreased in colon tumour tissues [152, 178] and below detection in cell lines [179] which implies a corresponding decrease in core 3 *O*-glycans in colon carcinoma. However, this contradicts with the above-mentioned mass spectrometry-based findings where core 3-based *O*-glycans are observed in abundance both in tumours and normal colonic tissues [102, 122, 177]. Thus, the association between core 3 structures and core 3 synthase has proven to be elusive in colonic tissues. One potential explanation for this discrepancy is the reported instability and insolubility of this enzyme in routinely used detergents, which may have reduced its activity during the assay [179]. Therefore,

future application of alternative techniques for measuring core 3 synthase expression such as qRT-PCR or selected reaction monitoring may resolve the apparent contradiction between the abundance of core 3 *O*-glycan structures and reduced core 3 synthase activity in CRC tissues.

1.5.4   Bisecting *N*-glycans

Bisecting *N*-glycans are a specific class of *N*-glycans characterised by the addition of an extra GlcNAc to the central mannose residue of the chitobiose core in a β1,4-linkage (Fig 3). The glycosyltransferase known for this covalent addition is GlcNAcT-III which is encoded by the gene MGAT3 (Fig 6). There is some evidence to indicate that bisecting *N*-glycans are reduced in colon tumours, as determined by LC-MS/MS in a recent study using pair-matched colon tissues from 13 patients [38] although further investigation is necessary to elucidate the biological role of bisecting *N*-glycans in CRC. Moreover, the expression of GlcNAcT-III was implied to be regulated by α-catenin expression and actin formation in the DLD-1 colon carcinoma cell line [180]. In this study, GlcNAcT-III activity was shown to be higher in α-catenin rescued DLD-1 cells compared to their controls. Furthermore, GlcNAcT-III activity was reduced by an inhibitor of actin polymerisation. These results suggest that actin cytoskeletal formation plays an important role in GlcNAcT-III induction in these cells. It also demonstrates that complex cross-talk may occur between bisecting *N*-glycosylation and other aspects of the cellular system.

1.5.5   Summary

Altered glycosylation patterns are frequently correlated with adverse patient outcomes such as tumour metastasis and shortened survival. The most outstanding examples of glycosylation changes that have been reported in CRC include changes in sialylation, fucosylation and their associated antigens such as sialyl-Tn and Lewis X. In addition, there is evidence to suggest bisecting *N*-glycans and core 3 *O*-glycans may also be changed in CRC. Some insights have been obtained regarding the regulation in membrane glycan expression by glycosyltransferases, although there were notable inconsistencies between cell line and tissues. Nevertheless, it is evident that glycosylation plays an important role in CRC carcinogenesis and progression, though further studies are necessary to fully elucidate the biological mechanisms behind these changes.

1.6   Altered glycoprotein expression in colorectal cancer

Following the review in the above section, it is clear in the literature that glycan structures are altered in CRC. Given that the majority of proteins are likely to be glycosylated [9], it is therefore not surprising that numerous glycoproteins are also altered in this disease. Here, the major classes of altered glycoproteins associated with CRC – mucins, integrins and cadherins – and how their glycosylation and overall expression is changed in CRC will be discussed below.

1.6.1 Mucins

Mucins are a class of high molecular weight proteins that are known to be extensively *O*-glycosylated [181, 182]. Currently, there are 22 known members of this family. Mucins can be subdivided into three classes: secretory gel-forming mucins, membrane-bound mucins and small-soluble mucins [183]. Mucins, particularly mucin-2 (MUC2), are known to form a substantial component of the mucus layer in the colon [184, 185]. The mucus layer plays a crucial role in the healthy function of the colon, by providing a protective barrier against harmful exogenous micro-organisms [184, 186] or chemical agents, mechanical shear [185], and chemical signals against bacterial infection [187]. The glycosylation and expression of mucins are known to be altered during colon cancer [102].

In CRC, the *O*-glycosylation of the mucin protein, MUC2, has been relatively well-studied using LC-MS/MS [102, 122, 177]. These investigations have revealed a diverse suite of glycan structures with more than 100 MUC2 *O*-glycans in normal tissue [122] and only about 40 observed in tumours [102]. While the majority of *O*-glycans were based on the core 3 structure, smaller amounts of core 2 and core 4 structures were also detected [102, 122]. Most importantly, the glycosylation pattern of MUC2 is known to be altered in CRC. In fact, the increased expression of sialyl-Tn and a core 3 sialyl-Lewis[X] hexasaccharide [NeuAcα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3(NeuAcα2-6)GalNAc] was found on MUC2 isolated from colon tumours.

However, not only is the glycosylation of MUC2 altered, there is evidence to suggest that the expression of the MUC2 protein is changed in CRC. MUC2 appears to be present at lower levels during CRC progression, with MUC2 undetected by anti-MUC2 antibodies in about half of CRC tumours while strong expression was detected in normal colonic mucosa [79]. This loss of MUC2 was significantly associated with disease recurrence (P<0.031) and tumour localisation (P<0.048). This trend also appears to be predictive of patient outcome with MUC2-positive tumours correlated with longer disease-free survival (P=0.059) [79]. In a separate study, the suppression of MUC2 expression by RNA interference also enhanced the proliferation, invasion and migration of cells of the mucinous CRC cell line, LS174T [87]. In addition, the growth rate of tumours of MUC2-suppressed LS174T cells was also significantly higher than control cells (P<0.05) [87]. Taken together, these findings indicate that down-regulation of MUC2 plays a role in CRC progression by augmenting the invasive and proliferative capabilities of colon cancer cells.

Whilst the expression and glycosylation pattern of MUC2 is most well-studied in CRC, other mucins are also known to be expressed in the normal or cancerous colon such as mucin-1 (MUC1), mucin-3 (MUC3), mucin-4 (MUC4), mucin-5B (MUC5B), mucin-6 (MUC6) and mucin-13 (MUC13) [188, 189] to name a few. With the exception of MUC3, MUC4 and MUC13, all these mucins are known to be potential carriers of cancer-associated glycan antigens such as sialyl-Tn, T (core 1), sialyl Lewis[A] and sialyl Lewis[X] antigens [190-194], indicating alterations in mucin protein expression may play a critical role in altering glycan expression.

Moreover, changes in mucin expression are known to impact the metastatic potential of CRC cells. Using high and low mucin variants of LS174T cell line, it was observed that the high-mucin variant were more highly tumorigenic than the low-mucin cell line when injected in athymic nude mice [88]. Furthermore, high-mucin variants also produced more lymph node or hepatic metastases [90]. Overall, there is mounting evidence to suggest that mucin expression and its glycosylation are altered during colorectal cancer and these changes may enhance the malignant behaviour of colonic tumours.

1.6.2   Integrins

Integrins are a class of heterodimeric transmembrane receptors that are important mediators of tumour metastasis through their modulation of cell properties such as cell adhesion, migration and signal transduction [195]. Currently, there are at least 25 distinct heterodimers formed by combination of 19 α-subunits and 8 β-subunits [196]. Similar to the mucins, dysregulation of protein expression and glycosylation has been associated with increased CRC malignancy.

Examples of the most well-studied integrins include αvβ3, αvβ6, α5β1 and α4β1 amongst the 25 heterodimers. These have been found altered in expression during cancer progression in a variety of different cancers such as melanoma, ovarian, cervical and colon cancer [195]. In colon cancer, there is evidence to suggest that the induction of integrin αvβ6 promotes cell migration and activation of autocrine TGF-β production [197]. Most importantly, integrin β6 is a potential prognostic indicator of poor survival [197]. Immunohistochemical analysis of 488 colorectal carcinomas revealed a decrease in patient survival for patients with β6 expression [197]. A similar trend is also found in patients with cervical cancer [198]. Therefore, while increased β6 expression is not CRC-specific, the similar resulting phenotype of reduced patient survival suggests that increased integrin β6 expression is a notable contributor towards cancer progression in various cancers, including CRC.

In terms of integrin glycosylation, altered sialylation and core-fucosylation have been linked to changes in integrin function and cell metastatic properties. For sialylation, a 4-fold increase was observed in the α2-6 sialylation of β1 integrin ($P < 0.02$) in adenocarcinoma colon tumours compared to pair-matched normal tissues [199]. In this same study, this increased α2-6 sialylation of β1 integrin increased the cell adhesion and migration of colon cancer cells [199]. Conversely, the opposite can also occur when α2-6-sialylation of β1 integrin is decreased. For example, in another study, the reduction of α2-6-sialylation of β1 integrin has been reported to result in decreased cell adhesion, migration and invasion [200]. Taken together, this suggests that alterations in the α2-6 sialylation of β1 integrin in CRC are implicated in β1 integrin-mediated cell adhesion and migration. Interestingly, differential α2-6 sialylation itself does not alter cell invasion in cancer cells lacking in β1 integrin expression [200], indicating that β1 integrin expression itself, as well as its altered glycosylation, both play a crucial role in cancer progression.

On the other hand, core-fucosylation has been found to be essential for integrin-mediated cell signalling using fucosyltransferase 8-positive (Fut8$^{+/+}$) and negative (Fut8$^{-/-}$) mouse embryonic fibroblasts. In this study, reduced α3β1 integrin-mediated cell migration was found in Fut8$^{-/-}$ cells [201]. Moreover, integrin-stimulated cell signalling was also shown to be decreased in Fut8$^{-/-}$ cells by the decreased phosphorylation of FAK, a protein-tyrosine kinase known to be involved in integrin-mediated signalling [201]. However, this was reversed via reintroduction of Fut8. Taken together, these results indicate the core-fucosylation and Fut8 expression are important in α3β1 integrin function. Collectively, both expression of integrin and its glycosylation changes appear to be critical factors for its function. When dysregulated, this may result in altered integrin-mediated signalling and further downstream impact on cancer malignancy.

1.6.3 Cadherins

Cadherins are a large family of glycoproteins consisting of more than 100 members with key roles in calcium-dependent cell adhesion [202], as well as cell signalling [202, 203], sorting [204] and recognition [175]. Structurally, members of this protein family are characterised by an outer extracellular domain responsible for cell-cell interactions, a transmembrane domain and a cytoplasmic domain which is frequently linked to the cytoskeleton [175]. They can be divided into subfamilies such as Type I, Type II, desmosomal and seven-pass transmembrane cadherins. They are frequently located in intercellular adhesion sites (junctions) and well-known examples include adherens junctions, desmosomes, intercalated discs and synaptic junctions [175]. Classically, they are complexed with cytoplasmic proteins called catenins and disruption of intact cadherin-catenin complexes may result in increased tumour aggressiveness, invasion and metastasis [205]. In CRC, reduced or loss of E-cadherin expression is generally associated with more advanced and aggressive tumour behaviour and is regarded as an unfavourable prognostic indicator of reduced overall survival [206, 207].

E-cadherin is known to be glycosylated with core-fucose [155, 208], bisecting and branched structures [209, 210] in various cell lines and tissues. The relationship between E-cadherin expression, bisecting and branching *N*-glycans has been elucidated mainly using cell lines. Interestingly, both E-cadherin expression and its glycosylation were altered depending on culture conditions. E-cadherin was characterised by a higher expression of bisecting *N*-glycans, as detected by E-PHA lectin, in dense culture compared to sparse culture [210]. E-cadherin overall expression was also greater in dense culture. A similar trend was also found for the core-fucosylation of E-cadherin with core-fucosylation increased in dense culture [155]. These lines of evidence lend support to the notion that E-cadherin expression and glycosylation may be altered in a cell density-dependent manner during cell culturing.

Even so, the expression of E-cadherin and the glycosyltransferases responsible for bisecting (GlcNAcT-III, product of MGAT3 gene) and branching structures (GlcNAcT-V, product of MGAT5 gene) have been shown

to be related. Restoration of E-cadherin expression in a previously E-cadherin deficient cell line (MDA-MB-231) resulted in significant increase in GlcNAcT-III activity [209]. However, knockdown of GlcNAcT-III did not alter overall E-cadherin expression, but resulted in membrane delocalisation of E-cadherin leading to cytoplasmic accumulation [209]. Moreover in the same study, GlcNAcT-III and GlcNAcT-V were shown to competitively modify E-cadherin *N*-glycans in both cell lines and gastric carcinoma tissues. Taken together, this indicates the presence of bidirectional crosstalk between E-cadherin expression and glycosyltransferase regulation of GlcNAcT-III and GlcNAcT-V. It also indicates that cell surface glycoproteins may influence the expression of glycosyltransferases which in turn can further modify the glycosylation of the cell surface glycoprotein.

As with other glycoproteins mentioned thus far, the perturbation of E-cadherin glycosylation appears to alter cellular properties and functions, such as cell morphology and cell signalling which may have implication for tumour malignancy. For example, an increase of bisecting *N*-glycans on E-cadherin prevents alteration of cell morphology in MGAT3-transfected colon cancer cells after epithelial growth factor (EGF) treatment [211]. In contrast, in mock-transfected cells, cell morphology is changed after only 15 minutes of EGF stimulation with looser cell-cell contact. Moreover, in the same study, cell signalling was altered with a decrease in downstream tyrosine phosphorylation of β-catenin in MGAT3 transfectants [211]. Thus, the addition of bisecting residues to E-cadherin decreases tyrosine phosphorylation of β-catenin, and this down-regulation may have implications in suppression of tumour progression by modulating important properties such as motility [206]. Collectively, these findings indicate that glycosylation of E-cadherin plays a vital role in modulating cell properties and cancer-related cell signalling, which may ultimately impact on cancer behaviour. Overall, there is accumulating evidence which implicates the expression of glycoproteins (such as mucins, integrins and cadherins) and their glycosylation in CRC progression.

1.7   Glycans and glycoproteins as potential biomarkers in CRC

1.7.1   Current biomarkers of CRC have limited sensitivity and specificity

There are several types of biomarkers: diagnostic, prognostic, and predictive [76]. Biomarkers can also be used for monitoring cancer progression or recurrence. A brief explanation of these types of biomarkers is shown in Table 10.

Table 10. Common types of biomarkers and their purpose(s)

Biomarker type	Purpose
Diagnostic	To differentiate the presence of cancer from other non-malignant conditions such as benign tumours or inflammatory diseases.
Prognostic	To give indication of how disease may develop.
Predictive	To give indication of probable treatment effect.
Disease monitoring/surveillance	To give indication of disease progress or recurrence.

Biomarkers can originate from different biofluids (e.g. serum, urine, saliva or faecal samples) or tissues. The two main considerations for evaluating the clinical utility of a biomarker are sensitivity and specificity. Sensitivity is the ability to correctly identify all patients with the disease, or the true positive rate [212]. Specificity is the ability to correctly identify patients without the disease, or the true negative rate [212].

Although numerous candidate biomarkers have been found, at present the only FDA approved biomarker for CRC is carcinoembryonic antigen (CEA), which can be measured in the serum [213]. Currently, it is recommended by the European Society for Medical Oncology [214] and the American Society of Clinical Oncology [215] for use in monitoring response to therapy and postoperative surveillance of disease recurrence. However, this biomarker has limited specificity and sensitivity in CRC diagnosis. For example, at a cut-off of 2.5 ng/mL, there is a specificity of 87% and sensitivity of 36% for stage 1 and 2 patients [213]. Although sensitivity rises for more advanced stages, early detection of CRC is preferable since patient survival rates are higher [216]. Therefore, this biomarker has not been recommended for use in population-wide CRC screening.

Aside from CEA, blood from the faecal occult blood test (FOBT) is another frequently employed screening tool [163]. FOBT detects the existence of CRC by presence of blood in stools, which may have been released from tumours. However, this test also suffers from the problem of low clinical specificity and specificity, since blood may also be released from other premalignant conditions such as polyps and adenomas, or present in certain foodstuffs. Thus, test-positive patients still require confirmation of results by endoscopy [214]. Overall, current biomarkers and screening tests have limited specificity and sensitivity and there remains an urgent need for biomarkers with improved specificity and sensitivity, particularly diagnostic biomarkers for early detection of CRC. To this end, changes in glycosylation may serve as a potential resource for novel biomarkers.

1.7.2 Biomarker discovery of glycans and glycosylation-related changes using mass spectrometry

Although it is clear that altered glycosylation is commonly observed in colorectal cancer, there has been relatively limited application of mass spectrometry for glycan biomarker discovery. One major reason which

may explain this is the time-consuming nature of the data analysis, which remains largely manual and therefore, greatly reduces throughput (discussed in section 1.4.2). Only a very small number of studies have investigated differences in glycan profiles of proteins in CRC. Using LC-MS/MS, differences in *O*-glycan profiles of MUC2 mucin isolated from patient-matched carcinomas and tissue (n=3) from resection margins were reported by Robbe-Masselot *et. al.* [102]. The most notable finding of this study was the increased expression of core 3 sialyl-Le^X hexasaccharide, NeuAcα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3(NeuAcα2-6)GalNAc, and sialyl-Tn in the tumours, previously mentioned in section 1.5. However, the size of patient cohort was relatively small with only 3 patients analysed in this study [102]. For protein *N*-glycans, profiles of from 13 pair-matched control and tumour colon tissues were recently reported by Balog *et. al.* [38]. The proteins from the tissues were isolated using chloroform-methanol precipitation and also analysed by LC-MS/MS. Bisecting *N*-glycans were shown to be decreased in the tumours, while sulfated, paucimannosidic and glycans containing sialyl Lewis type epitope were increased. Both studies identified altered glycan structures not previously found using other targeted techniques. In the latter study, many of these were bisected *N*-glycans [38].

Given the relatively restricted use of LC-MS/MS for glycan profiling in CRC, there is still a large scope for its application to investigate glycan changes associated in CRC, especially in regards to membrane proteins (instead of mucins or on the whole cell proteome as performed previously). Additionally, the identification of novel glycan changes using mass spectrometry not previously discovered using other techniques demonstrates the unparalleled sensitivity and high glycan structural determination capacity of LC-MS/MS, which is a clear advantage in biomarker discovery. As well, there may be additional glycan changes that remain untapped as potential biomarker candidates in CRC. These potential glycan marker candidates can be used singly or in combination with other biomarkers, which may improve their sensitivity and specificity. Consequently, a major aim of this study is to identify the profile of both membrane glycan and glycoprotein biomarker(s) of CRC using LC-MS/MS. In light of the discrepancies in reported glycan alterations, especially between cell line and tissues, the use of colon cancer cell line models for the study of membrane glycosylation in colorectal cancer and in particular, glycan biomarker discovery will also be evaluated.

1.8 Overall aim and scope of this study

Broad aim of this study: To evaluate colon cancer cell lines as representative models of disease for the study of cell surface glycans and glycoproteins in the context of CRC biomarker discovery.
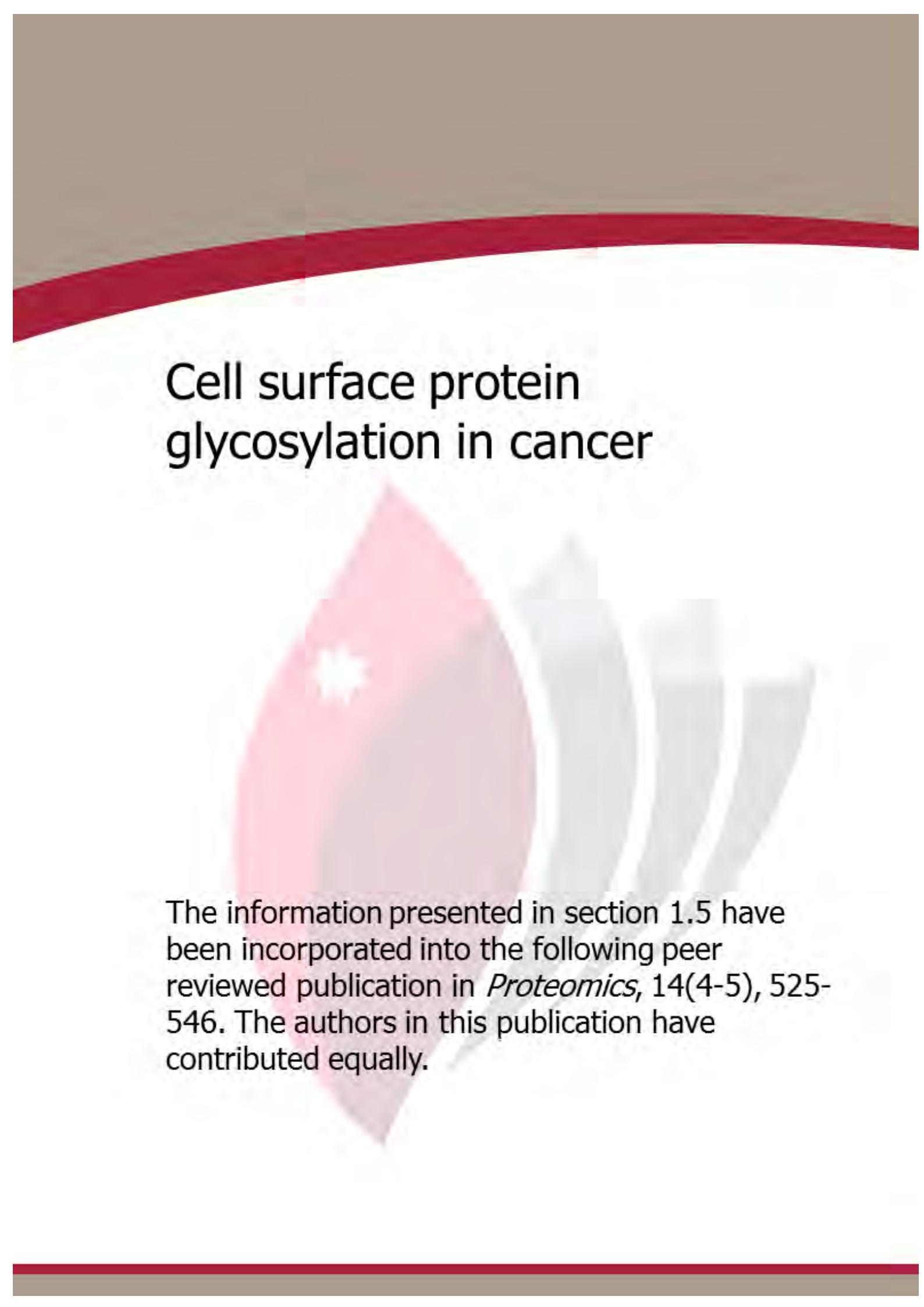
Aims

1. **To compare membrane protein glycosylation profiles of commonly used colorectal cancer cell lines and CRC tumours using mass spectrometry**

Out of the three broad systems of scientific investigation (*in vitro/ex vivo/in vivo*), cell lines (*in vitro* models) are a common scientific tool and laboratory substitute for tumour tissue. However, some inconsistencies previously reported in the literature between the protein glycosylation of cancer cell lines and tissues, have highlighted the potential of glycosylation differences between these investigation systems. Importantly, the degree that cell lines reflect that of parent tumours is unknown for the study of glycan alterations in cancer and these differences may represent a substantial impediment towards true cancer biomarker discovery. Therefore, a major aim of this thesis is to compare the glycosylation profiles of five colorectal cancer cell lines with cancerous colorectal tissue to evaluate whether cell lines closely reflect tumour glycosylation status.

2. **To compare the membrane glycosylation and glycoprotein profiles from pair-matched normal and tumour tissue using mass spectrometry towards biomarker discovery**

With the knowledge gained from the first aim, the membrane protein glycosylation profiles of patient-matched normal and tumour colonic tissues will be compared to identify novel glycan and/or glycoprotein candidate biomarkers. Both glycan and glycoproteins were profiled in detail to enhance the identification of new biomarker combinations with the potential of becoming diagnostic biomarkers with higher sensitivity and specificity than those currently available in CRC.

Cell surface protein glycosylation in cancer

Pages 58-79 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages:

Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., & Packer, N. H. (2014). Cell surface protein glycosylation in cancer. *Proteomics, 14*(4-5), 525-546. https://doi.org/10.1002/pmic.201300387

Chapter 2

Comprehensive glycomics comparison between colon cancer cell cultures and tumours: implications for biomarker studies

The work carried out in this chapter fulfils Aim 1 as detailed in the Introduction.
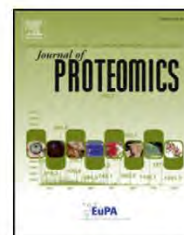
Comprehensive glycomics comparison between colon cancer cell cultures and tumours: Implications for biomarker studies

Jenny H.L. Chik[a], Jerry Zhou[b], Edward S.X. Moh[a], Richard Christopherson[b], Stephen J. Clarke[c], Mark P. Molloy[a,d], Nicolle H. Packer[a,*]

[a]Department of Chemistry and Biomolecular Sciences, Faculty of Science, Macquarie University, Sydney, Australia
[b]School of Molecular Bioscience, University of Sydney, Sydney, Australia
[c]Department of Medicine, Royal North Shore Hospital, University of Sydney, Australia
[d]Australian Proteome Analysis Facility, Macquarie University, Sydney, Australia

ARTICLE INFO

ABSTRACT

Altered glycosylation is commonly observed in colorectal cancer. *In vitro* models are frequently used to study this cancer but little is known about the differences that may exist between these model cell systems and tumour tissue. We have compared the membrane protein glycosylation of five colorectal cancer cell lines (SW1116, SW480, SW620, SW837, LS174T) with epithelial cells from colorectal tumours using liquid chromatography tandem mass spectrometry. Remarkably, there were five abundant O-glycans in the tumour cells that were undetected in the low-mucin producing cell lines, although two were found in the mucinous LS174T cells. The O-glycans included the well-known glycan cancer marker, sialyl-Tn, which has been associated with mucins. Using qRT-PCR, sialyl-Tn expression was found to be associated with an increase in α2,6-sialyltransferase gene (ST6GALNAC1) and a decrease in core 1 synthase gene (C1GALT1) in LS174T cells. The expression of a subset of mucins (MUC2, MUC6, MUC5B) was also correlated with sialyl-Tn expression in LS174T cells. Overall, the membrane protein glycosylation of the model cell lines was found to differ from each other and from the epithelial cells of tumour tissue. These findings should be noted in the design of biomarker discovery experiments particularly when cell surface targets are being investigated.

Biological significance
The extent of protein glycosylation differences between *in vitro* cell lines and *ex vivo* tumours in colorectal cancer research is unknown. Our study expands current knowledge by characterising the membrane protein glycosylation profiles of five different colorectal cancer cell lines and of epithelial cells derived from resected colorectal cancer tumour

tissue, using liquid chromatography tandem mass spectrometry. The detailed structural differences found in both N- and O-linked glycan structures on the membrane glycoproteins were determined and correlated with the mRNA expression of the relevant proteins in the cell lines. The glycosylation differences found between cultured cancer cell lines and epithelial cells from tumour tissue have important implications for glycan biomarker discovery.

1. Introduction

Colorectal cancer (CRC) is a major type of malignancy worldwide, being the 3rd most common cancer in males and second most common in females [1]. It has a high incidence and mortality with an estimate of over one million new cases and half a million deaths each year [1]. Prognosis is poor for the many patients that are diagnosed at late stages where metastasis has occurred [2]. The commonality of this malignancy, high mortality rate and reduction in 5-year survival upon late diagnosis is consistent with the improved need for biomarkers to enable earlier diagnosis or better prediction/monitoring of therapeutic treatment.

One potential and under-explored source of disease markers is protein glycosylation [3–6]. Protein glycosylation is the enzymatic construction and addition of sugar chains (glycans) to proteins. There are two main types of protein glycosylation; N- and O-glycosylation [7]. Briefly, N-glycosylation occurs on an asparagine residue that is part of a tripeptidic consensus sequence of asparagine–X–serine/threonine, where X denotes any amino acid except proline [8]. All N-glycans share a common structural motif known as the chitobiose core consisting of three mannose and two N-acetylglucosamine (GlcNAc) residues. N-glycans can be separated into three structural categories: high mannose, hybrid and complex based on the type of carbohydrate units which extend from the chitobiose core. On the other hand, O-glycans [9,10] are covalently bonded to a serine or threonine residue and form generally shorter, more varied structures compared with N-glycans. These are based on a number of glycan structural 'cores' of which core 1 to 4 are more well-studied in mammals [10].

Both N- and O-glycans have been shown to differ in their abundance levels during cancer progression [11,12]. In CRC, examples include the well-known marker, CA 19-9, that has been characterised as specifically comprising the sialyl-Lewis A epitope [13,14] and the glycoproteins CA 50 and CA 242, which have not yet been fully structurally characterised [15]. Other carbohydrate cancer associated alterations such as both sialylated and non-sialylated Lewis antigens [16], increased sialylation [17] and decreased bisecting N-glycans [12] have been described. Amongst the most well-known pan-carcinoma marker is the over-expression of the sialyl-Tn antigen (NeuAcα2-6GalNAc). This O-glycan was observed by immunohistochemistry of colorectal tumours, using anti-sialyl-Tn monoclonal antibodies, to be expressed in a majority (83.4%) of colon cancer cases whilst it is not usually observed in normal mucosa [18].

Many of these glycosylation changes were initially observed using anti-glycan antibodies or lectins, but with the advancement in mass spectrometry techniques investigators have begun to analyse glycans using mass spectrometry (MS). For O-glycans on proteins, glycan MS profiling has been restricted to colonic mucins of interest [19], such as MUC2 [20,21]. One notable finding was the increased expression of a core 3 hexasaccharide containing the sialyl-Lewis X epitope (NeuAcα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3(NeuAcα2-6)GalNAc) on tumour compared to normal mucins, which may be a potential marker of malignant transformation [20]. The N-glycosylation of proteins from matched normal mucosa and colorectal tumours has been reported recently [12]. This study indicated that bisecting N-linked structures are reduced in normal mucosa compared with matched tumour tissue. On the other hand, sialyl Lewis-type, paucimannosidic and sulphated N-glycans were observed to increase in the tumour tissue [12].

Although many studies into glycosylation changes in CRC have been performed with ex vivo tissue, such as the analysis of mucus [20], blood [22], or tissues [12,23], many other investigative approaches utilise in vivo animal models [24,25] or in vitro established cell lines [26,27]. From a biomarker perspective one must be cautious with the results obtained using model systems as the molecular aspects of these "isolated" systems may differ compared with the in vivo setting. Advantages of cell lines include homogeneity and the fact that sample availability is not limited, allowing extensive analyses over long periods. However, alterations to glycome/proteome may occur due to two-dimensional culture on plastic plates and accumulation of mutations from extended subculturing [28,29]. Moreover, cell lines may differ for various characteristics, for example, the colon cancer cell lines LS174T and HM7 produce mucin in culture, whilst others such as SW480 and SW1417 were reported as non-mucinous [11].

We are particularly interested to understand how well the cell surface glycome compares between cultured colon cancer cells and CRC tumours. Therefore, we affinity isolated epithelial cells from CRC tumours and utilised liquid chromatography tandem mass spectrometry (LC–MS/MS) to compare the N- and O-glycan profiles from cell membrane proteins of four colorectal cancer cell lines (SW1116, SW480, SW620, SW837) with colorectal tumours. We also examined the O-glycans attached to mucins produced by the LS174T mucinous cell line [19] and compared these to the glycans of colorectal cancer tissue.

2. Materials and methods

The workflow of the glycomic analysis of the membrane proteins from cultured and epithelial tissue cells is summarised in Fig. 1.

2.1. Cell lines and culturing conditions

Five colorectal cancer (CRC) cell lines were investigated in this study and their characterisation has been tabulated (Suppl. Table 1). SW1116, SW480, SW620 and SW837 were cultured in
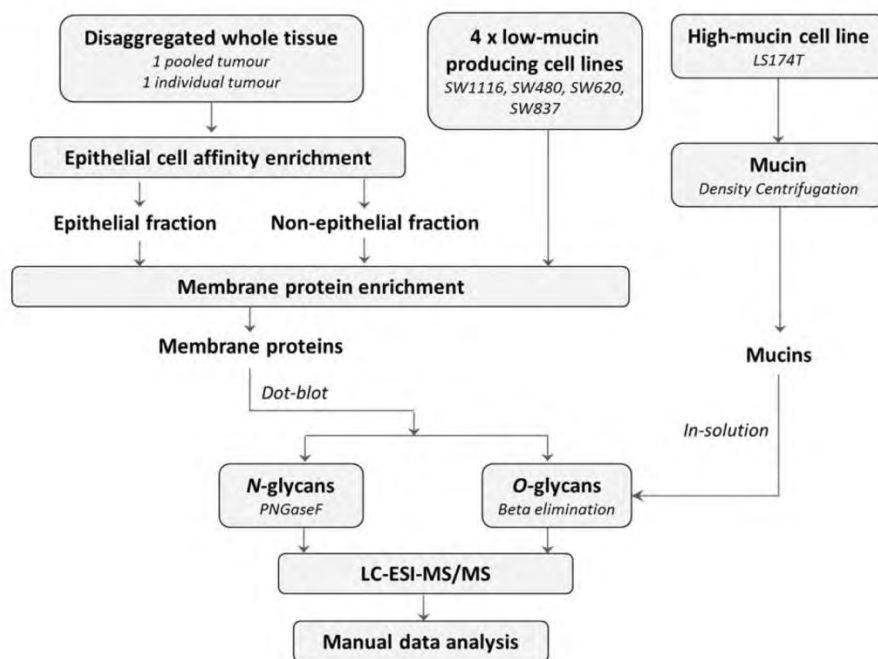
Fig. 1 – Glycan analysis workflow for colorectal tumour tissues and cell lines.

DMEM containing 10 mM HEPES. LS174T was grown in RPMI 1640 containing 2 mM L-glutamine. All cell lines were cultured with the addition of 10% foetal bovine serum and antibiotic–antimycotic cocktail (100 μg/mL streptomycin, 100 U/mL penicillin, 0.25 μg/mL amphotericin B, Invitrogen) in cell culture dishes (Corning, 150 × 25 mm, Sigma Aldrich). All cultures were grown to confluency at 37 °C with 5% CO_2 in a humidified incubator.

2.2. Tissue samples

Written and informed consent was provided by all patients under an ethics application approved by the Sydney South West Area Health Service (protocol number: X08-0614). In total, five tumour samples collected as part of surgical resection (detailed in [30]) were used and their clinicopathological characteristics are shown in Suppl. Table 2. Pathological examination of resected specimens followed the Australian Clinico-Pathological Staging (ACPS) protocol for CRC [31], which is compatible with other staging systems such as the TNM (tumours/nodes/metastases, [32]). All samples used were classified as stage B1 according to ACPS which correspond with the T3N0M0 stage in the TNM classification system.

Samples were processed according to Zhou et al. [30]. Immediately following the surgery, patient tissue samples were collected and stored in Hanks' balanced salt solution (HBSS; Sigma-Aldrich) at 4 °C for no longer than 6 h to maintain cell viability. The samples were cut into 2 mm strips and incubated with occasional gentle mixing for 60 min at 37 °C with an equal volume of RPMI 1640 medium containing 2% (v/v) collagenase type 4 (Worthington, USA) and 0.2% (w/v) deoxyribonuclease I from bovine pancreas (DNAse I; Sigma-

Aldrich). Alternatively, 1% (v/v) collagenase type 4 could be used, with the incubation time increased to 80 min. The semi-digested tissue was then forced gently through a fine wire mesh strainer using the plunger from a 20 mL syringe; cells were washed through with HBSS. The resulting cell suspension was sequentially passed through 200 μm and 50 μm Filcon filters (BD Biosciences) to remove cell aggregates. Cells were stored frozen at −80 °C in heat-inactivated foetal calf serum with 10% dimethyl sulphoxide (DMSO). Thawed cell suspensions were treated with 0.1% (w/v) DNAse I. Cell viability was determined using trypan blue exclusion (Sigma-Aldrich). Disaggregated cells were stored in DMSO at −80 °C until further use, to maintain their viability. In order to obtain adequate material for enrichment of tissue epithelial cells, tumours from three patients were pooled for one of the enrichment samples (patients 2–4, referred to as tumour 2, Supp. Table 2). In total, two independent enrichments were performed (tumour 1 and tumour 2).

2.3. Enrichment of epithelial cell fraction from tissue using immunomagnetic beads

Tumours were thawed at a 37 °C water bath and immediately diluted in HBSS–HEPEs containing 2% AB serum (HBSS with 9.8 g/L HEPES, 2% (v/v) AB serum, Sigma Aldrich). Cells were pelleted and counted. Immunomagnetic beads conjugated to an antibody against epithelial molecular marker EpCAM (Dynabead CELLection Epithelial Enrich, Invitrogen) were washed twice in HBSS–HEPEs with 2% AB serum (180 μL beads/10^7 viable cells). Beads and cells were mixed and incubated for 1 h at 4 °C with gentle agitation. The bead-cell suspension was washed twice with PBS (phosphate buffered saline; 2.7 mM potassium chloride, 1.5 mM potassium phosphate monobasic, 136.7 mM sodium

chloride, 8.1 mM sodium phosphate dibasic) to remove traces of buffer. Unbound cells in PBS ('non-epithelial fraction') and bead-bound cells ('epithelial fraction') were pelleted and stored at −20 °C until use.

2.4. Western Blotting of tissue fractions

Epithelial enrichment was validated by performing a Western Blot on epithelial and non-epithelial fractions of tumour tissue. An orthogonal epithelial marker Keratin 18 (Cell Signalling Technologies) and a representative non-epithelial marker Vimentin (Cell Signalling Technologies) for fibroblasts/mesenchymal cells were used to indicate epithelial cell partitioning. The epithelial and non-epithelial fractions from the tumour sample were resuspended in ice-cold lysis buffer (50 mM Tris–HCl, 100 mM sodium chloride, 1 mM EDTA) containing protease inhibitors (Roche Diagnostics, Germany) and lysed mechanically using Omni Polytron (Omni International). Lysed cells were clarified by centrifugation (2000 g at 4 °C, 20 min). Protein quantification was performed by BCA assay (Thermo Scientific).

To precipitate proteins, 9 volumes of ice-cold acetone was added to non-epithelial and epithelial fractions (25 μg in each fraction) and precipitated overnight at −20 °C. Precipitated protein was pelleted, air-dried, solubilised in up to 10% SDS (v/v), mixed with loading buffer (Invitrogen) and reduced in 70 mM DTT at 70 °C for 10 min. Sample was loaded onto a 4–12% (w/v) acrylamide Bis–Tris gel (Invitrogen) and electrophoresis performed for 50 min at 200 V with MOP buffer (Invitrogen). The gel-separated proteins were wet-transferred onto nitrocellulose membrane (Bio-Rad, 0.45 μm) at 30 V for 90 min. Membrane was washed briefly in TBS (50 mM Tris, 150 mM sodium chloride, pH 7.4) and non-specific binding was blocked by incubation of membrane at room temperature in 5% (w/v) skim milk in TBS for 1 h.

Membrane was washed with TBST (TBS with 0.1% (v/v) Tween 20) and incubated overnight in 4 °C with vimentin and keratin 18 antibodies at 1:1000 and 1:2000 dilutions in 5% skim milk (w/v) in TBST respectively. Subsequently, the membrane was washed 3 times and incubated with secondary antibodies, anti-mouse IRDye 800 (LiCor) and anti-rabbit IRDye 680 (LiCor), diluted 1:10,000 and 1:15,000 in 5% skim milk TBST (w/v). Membrane was protected from light during the 50-minute incubation at room temperature and then imaged wet using Odyssey Imager (LiCor).

2.5. Membrane protein enrichment

Membrane protein enrichment was required for glyco-profiling to enrich for glycosylated proteins. This was performed by Triton X114 phase partitioning using essentially the same method as detailed elsewhere [33,34]. Two independent enrichments from each cell line were performed. Cells were washed four times with PBS and scraped off the culture dish in PBS containing protease inhibitor cocktail (Roche Diagnostics, Germany) using a commercial cell scraper. Cells were counted, pelleted and frozen at −80 °C if not used immediately.

Both cell lines (10^7 cells) and epithelial-enriched tumour cells were lysed on ice using Omni Polytron in lysis buffer

(50 mM Tris–HCl, 100 mM sodium chloride, 1 mM EDTA, protease inhibitor cocktail, Roche, pH 7.4). Samples were centrifuged at 2000 g for 20 min at 4 °C to remove unlysed cells. Supernatant was diluted in buffer to final concentration of 0.1 M sodium carbonate and ultracentrifuged at 120,000 g for 80 min at 4 °C. The pellet was resuspended in buffer (20 mM Tris–HCl, 100 mM sodium chloride) with 1% Triton X114 (v/v) [33], incubated on ice for 10 min, then at 37 °C for 20 min, followed by centrifugation at 1000 g for 3 min to partition sample into aqueous (upper) and detergent (lower; membrane protein) layers. Upper layer was removed and the proteins in the detergent layer were acetone-precipitated overnight by addition of at least nine volumes of ice-cold acetone and incubation at −20 °C.

2.6. Mucin preparation using isopycnic density gradient centrifugation

Similar to the mucins prepared according to Kilcoyne et al. [35], confluent LS174T cells were washed with PBS three times. 4 M guanidine hydrochloride (Amresco) added to the culture plates to lyse cells and cell lysates were collected using a commercial cell scraper. The cell extract was reduced by addition of freshly prepared DTT (10 mM) and incubation at 37 °C for 5 h with occasional mixing, following by alkylation with IAA (25 mM, Sigma Aldrich) and incubation overnight at room temperature in the dark. Cesium chloride (CsCl, Biochemicals) was added and sample density adjusted to 1.4 g/mL. Samples were centrifuged on a CsCl density gradient at 36,000 rpm for 72 h and mucin fractions sedimenting at density 1.3–1.4 g/mL [35] were dialysed with deionised water using a dialysis membrane (Selby Anax), lyophilised and weighed. 10 μg of purified mucins was used for glycan analysis and two independent mucin isolations were performed.

2.7. N- and O-glycan release

The procedure for the release of N- and O-glycans was basically the same as detailed previously [36]. Briefly, N-glycans from cell membrane proteins of four cell lines (SW1116, SW480, SW620, SW837) and enriched epithelial cells from tumour samples were analysed. Proteins were immobilised by dot-blotting onto a pre-activated PVDF membrane (0.2 μm, Bio-Rad). Protein spots were cut and placed wells of a 96-well plate containing 1% PVPP (v/v) in 50% methanol (v/v), then washed thoroughly with water and allowed to incubate overnight at 37 °C with 5 μL of PNGase F (0.5 U/μL, Roche) with 10 μL of water. Released N-glycans were incubated with 100 mM ammonium acetate (pH 5.0) at room temperature for 1 h, dried, and reduced by 20 μL of 1 M sodium borohydride in 50 mM potassium hydroxide (KOH) by incubation for 2 h at 50 °C. Samples were acidified by 2 μL of glacial acetic acid and desalted using cation exchange resin (Dowex AG 50 W X8), followed by three additions of 100 μL methanol, dried to completeness after each addition.

O-glycans from membrane proteins were released directly from the same dot-blots after N-glycan release, whilst mucin O-glycans from LS174T were released in-solution; using beta-elimination via addition of 0.5 M sodium borohydride in 50 mM KOH and incubation at 50 °C for 16 h. Samples

were acidified, desalted and dried as above. Released LS174T mucin O-glycans were further purified by tips using porous graphitised carbon (Extract-clean carbon SPE cartridge, Grace) manually packed onto C_{18} reverse phase stage tips (Thermo Fisher Scientific) and dried. Reduced N- and O-glycans were taken up in water for analysis by LC–MS/MS.

2.8. Analysis by liquid chromatography tandem mass spectrometry

Released N- and O-glycans were separated using a commercial porous graphitised carbon column (Thermo Hypercarb, 5 μm, 100 mm × 180 μm, pore size 250 Å, Thermo Scientific) with a flow rate of 2 μL/min using an Agilent 1100 or 1260 series HPLC (Agilent Technologies Inc. USA) [36]. Briefly, N-glycan separation was performed using an 85 min gradient of 0–45% (v/v) acetonitrile in 10 mM ammonium bicarbonate (NH_4HCO_3). O-glycans were separated using a 45 min gradient of 0–90% (v/v) acetonitrile in 10 mM NH_4HCO_3. Eluates were introduced directly into an Agilent 6330 electrospray source (Agilent Technologies Inc., USA). Mass spectrometry was performed in negative ion mode using either Agilent LCD/MSD Trap XCT Ultra (Agilent Technologies Inc., USA) or Bruker HCT Ultra (Bruker Daltonics, Germany). No in-source dissociation was performed. The voltage of the capillary outlet was set at 3 kV and temperature of the transfer capillary was maintained at 300 °C. The following scan events were performed: full MS scans were between m/z 100–2200 with a maximum accumulation time of 200 ms and data-dependent MS/MS scan of the top two most abundant ions (for N-glycans) and top three most abundant ions (for O-glycans).

2.9. Glycan structural determination and relative quantification

Nomenclature from the Consortium of Functional Glycomics/Essentials of Glycobiology has been used to represent glycan structures. Theoretical glycan composition was calculated from total ion mass using GlycoMod (http://web.expasy.org/glycomod/) with a mass tolerance of ±0.5 Da. Additional data analysis was performed in Data Analysis 4.0 (Bruker Daltonics, Germany). MS/MS spectra for each structure were manually inspected to determine structure details based on structural feature ions [37,38], and a knowledge of the known biosynthetic pathways [8,9] with reference to the annotated MS2 spectra available in the public glycan database UniCarb-DB [39] (http://unicarb-db.org). Examples of feature ions used to support proposed structures include m/z 350/368 [37] which indicates core-fucosylation, the m/z 510 ion (Gal-[Fuc]GlcNAc), that indicates outer-arm fucosylation and the D-221 ion which indicates the presence of bisecting GlcNAc [37,38,40]. Where there is potential variability in glycosidic sequence the depiction of the structures (Suppl. Tables 3 and 4) is shown with bracketed monosaccharides in the 'Probable Structure' column and the presence of isomeric structures is indicated in the Tables as separate entries. Depiction of bond stereochemistry and specific linkages (for example by linkage angles) are not intended in the glycan representations. Each glycan structure was quantified relative to the total content by integration of the extracted ion chromatogram peak area (EIC). Prior to integration, the peak

was smoothed using Gaussian algorithm (1 cycle and 1 pnts). For comparison of glycan abundances across samples, area-under-the-curves (AUCs) of each glycan structure were normalised to total AUC (AUCs of all glycan structures in a sample summed) and expressed as a percentage. The relative abundance of all glycan masses and their proposed structures are summarised in Suppl. Table 3 (N-glycans) and 4 (O-glycans). Glycan structures were considered differentially expressed by using a 2-fold cut-off.

2.10. RNA extraction and quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR)

Two independent RNA extractions were performed for each cell line. RNA was extracted from cell pellets using Trizol reagent (Life Technologies). Crude RNA was further purified using RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA concentration and purity were estimated using A260/280 ratio (>1.94 for all samples). Additionally, RNA was treated with Turbo DNA-Free kit (Life Technologies) according to the vendor's instructions to minimise genomic contamination.

2 μg of RNA was reverse transcribed using SuperScript VILO cDNA synthesis kit as per vendor's instructions (Life Technologies) with OligoDT primers (Life Technologies). cDNA was subsequently diluted 1 in 20 and used in a 5 μL PCR reaction containing 1 μL of diluted template, 2.5 μL GoTAQ qPCR Master Mix (Promega), and 1.5 μL of primers (2.5 μM of forward and reverse primers). The primer sequences used for glycosyltransferases (ST6GALNAC1, C1GALT1, MGAT3), mucin genes (MUC1, MUC2, MUC3A, MUC4, MUC5B, MUC6, MUC13) and reference gene (β-actin) are shown in Suppl. Table 5. All primers were designed in-house by Primer-BLAST software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) except for β-actin primers, which were previously used in another study [41].

PCR detection of gene expression was performed by Mastercycler Realplex (Eppendorf International, Germany). PCR reactions were performed in triplicate for each sample. For each primer pair, no template controls, no reverse-transcriptase controls and melt-curve analysis were included to ensure specific gene amplification with minimal primer dimers. The cycling conditions were as follows: initial denaturation at 95 °C for 2 min, followed by 40 amplification cycles of 95 °C for 15 s, 55 °C for 15 s and 68 °C for 20 s. Data was reported as $2^{-\Delta CT}$ values [42]. For mucin mRNA, $2^{-\Delta CT}$ values were also expressed as percentage of total mucin mRNA.

2.11. Data visualisation and statistical analysis

Generation of heatmaps was performed by R statistical software (v. 3.0.1) after log-transformation of normalised AUC values. Statistical comparisons of mRNA expression were performed using GraphPad Prism 6. Values are displayed as means ± SD (standard deviation) with a minimum of five data points used for each calculation. Statistical significance was calculated using Student's t-test or one-way ANOVA followed by Tukey's multiple comparison test. P-values <0.05 were considered significant.

3. Results

3.1. Enrichment of tumour epithelial cells

Epithelial cells from single cell suspensions of CRC tumour tissues were positively selected by immunoaffinity capture using the epithelial marker protein, EpCAM [43]. The epithelial cell enrichment was confirmed by Western Blotting (Fig. 2A). Epithelial marker keratin 18 (46 kDa) [44] was observed in the EpCAM-positive (epithelial) fraction and was absent from the remaining fraction. Conversely, Vimentin (57 kDa), a mesenchymal marker [45] was not detected in the EpCAM-enriched fraction. These findings are consistent with selective enrichment of epithelial cells from CRC tumours into an EpCAM-positive epithelial cell fraction.

3.2. Comparison of glycan profiles of tumour isolated epithelial and non-epithelial cell fractions

Two independent enrichments of epithelial and non-epithelial cells from CRC tumours demonstrated that the N-linked glycan profiles of non-epithelial and epithelial fractions were similar (Suppl. Fig. 1). On the other hand, although there were major O-glycans (Fig. 2B) expressed in both epithelial and non-epithelial fraction of tumours in relatively high abundance, such as the trisaccharide $(Hex)_1(HexNAc)_1(NeuAc)_1$ with m/z 675.3, and the tetrasaccharide $(Hex)_1(HexNAc)_1(NeuAc)_2$ with m/z 966.3, we also observed the presence of two glycans, disaccharide m/z 513 $(HexNAc)_1(NeuAc)_1$, and trisaccharide with m/z 716, $(HexNAc)_2(NeuAc)_1$, which were highly elevated in abundance in the tumour epithelial cell fraction. This indicates that there are O-glycans associated with tumours, which are enriched using EpCAM-immunoaffinity capture from the mixed population of

tissue cells. For a bona fide glycan comparison between CRC epithelial cell lines and tumours, these enriched epithelial cell fractions from the tumour tissue were compared with the cultured cells from different patient origins (Suppl. Table 1). We are unaware of previous reports adopting this approach that overcomes the problem of cell-type heterogeneity in tumour tissue for the study of protein glycosylation.

3.3. Comparison of membrane protein glycan profiles between cell lines and enriched epithelial tumour cells

The N- and O-glycoprofiles from five colorectal cancer cell lines were compared with enriched epithelial cells of tumour tissues. Overall, 173 N-glycan and 43 O-glycan structures were identified and their MS2 annotated (to the extent depicted by the glycan representations, Suppl. Tables 3 and 4). The annotated MS2 of all structures is shown in Suppl. Fig. 5 (for N-glycans) and Suppl. Fig. 6 (For O-glycans). The MS2 peak lists have also been deposited in the public glycan database, UniCarb-DB (http://unicarb-db.org).

3.3.1. Analysis of N-glycans

173 N-glycans were identified, verified by MS2 annotation and quantified by AUC integration (Suppl. Table 3) for the membrane proteins of four CRC cell lines (SW1116, SW480, SW620, SW837) and two tumours. An additional 6 low abundance N-glycans were confirmed compositionally and quantified but were excluded from further analysis since they represented less than 1% of the glycome for the majority of samples (Suppl. Table 3).

Overall, only minor differences between the four cell lines and the tumour samples were observed in the N-glycans on the membrane proteins when average MS (summed across total elution time) was compared (Fig. 3A, Suppl. Fig. 3). High-mannose glycans were highly abundant features in both
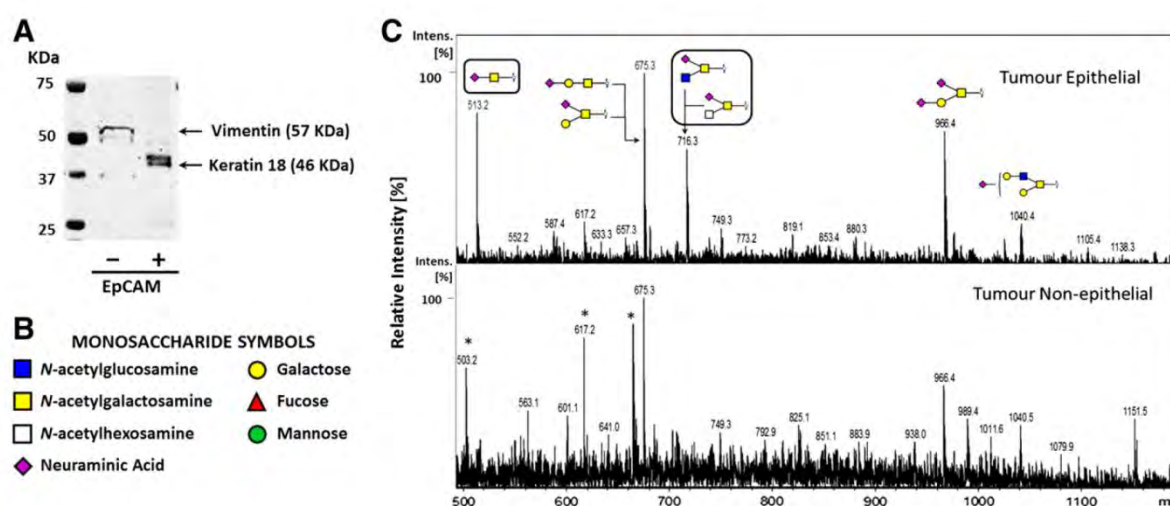
Fig. 2 – EpCAM immunoaffinity enrichment of epithelial cells from tumours. A) Western Blot of epithelial (EpCAM-positive) and non-epithelial (EpCAM-negative) fractions of a tumour sample processed using EpCAM-conjugated dynabeads. Epithelial marker keratin 18 and mesenchymal/fibroblast marker vimentin indicate cell-type partitioning. B) Monosaccharide symbols used for the depiction of glycans according to the Consortium of Functional Glycomics nomenclature. C) Average mass spectra of epithelial and non-epithelial fractions from the same tumour. Major glycan differences have been boxed. Asterisked peaks do not correspond to glycans as indicated by MS2.
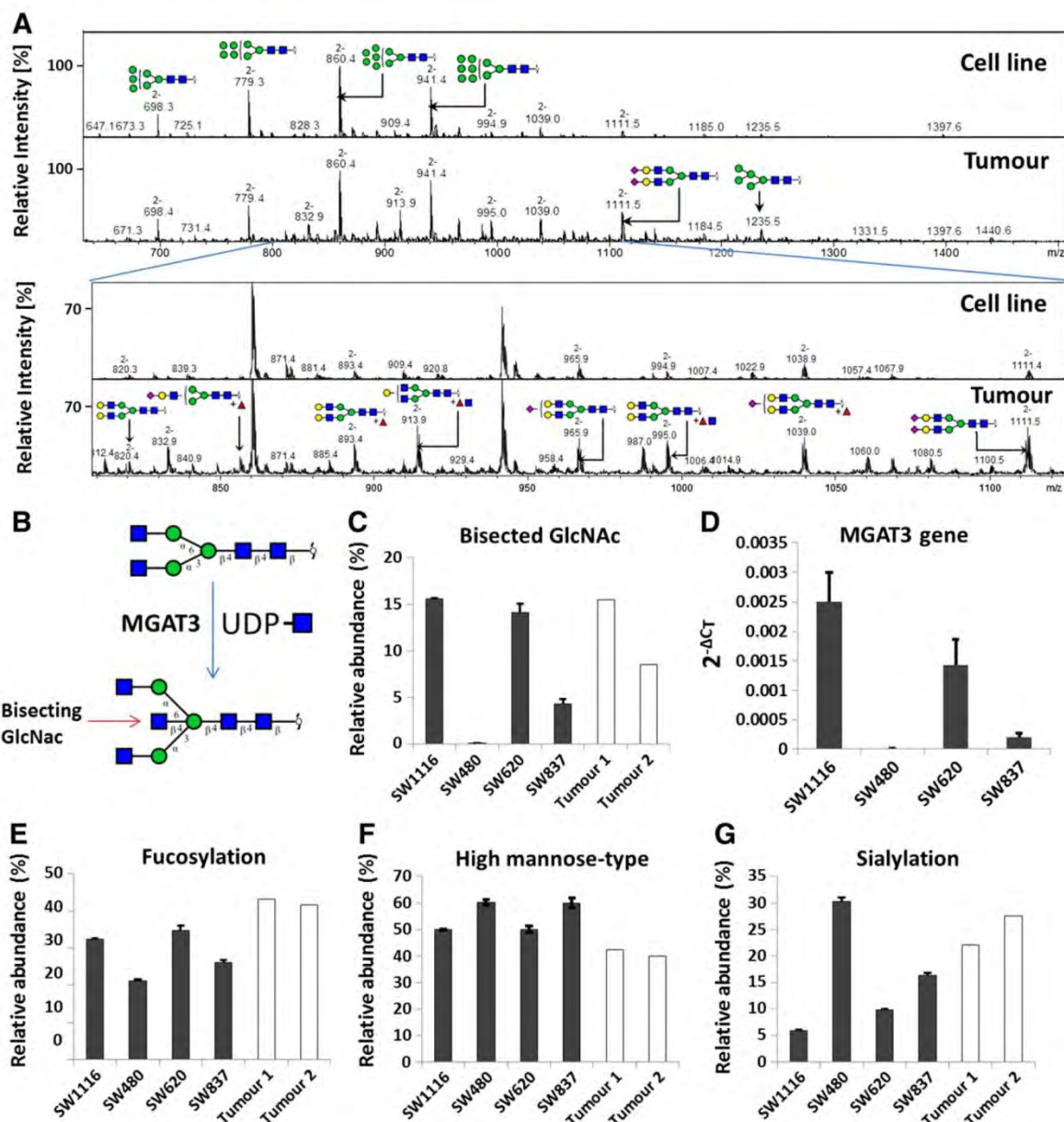
Fig. 3 – Differences in N-glycan profiles and N-glycan classes between cell lines and tumours. A) Typical N-glycan profile (summed across the total elution time) of low mucin-producing colorectal cancer cell lines and tumours showing high-mannose glycans as most abundant in both cell lines and tumours. Relatively small differences in other complex/hybrid glycans were observed (examples shown in zoomed in portion of average MS — see text for details). B) Schematic showing the involvement of N-acetylglucosaminyltransferase (encoded by MGAT3 gene) in the synthesis of bisecting N-glycans. C, E–G) Relative abundance of different glycan types analysed by LC–MS/MS. Error bars of cell lines represent maximum and minimum of duplicate glycan analysis. Tumours (n = 2) are shown as separate biological replicates. C) Changes in bisecting N-glycans between colorectal cancer cell lines (black bars) and tumours (white bars). D) mRNA expression of MGAT3 gene in cell lines by qRT-PCR. Data represented as mean ± SD of at least five data points from two triplicate PCRs. E–G) Changes in other N-glycan classes between colorectal cancer cell lines (black bars) and tumours (white bars).

cell lines and tumours (Fig. 3A and F) whilst more varied levels of expression were observed for the hybrid/complex glycans. In total, there were 14 glycans which were prominent on the membrane proteins (>3% of total N-glycans in at least one sample). Of these, 6 were high-mannose glycans (with two isomers for m/z 941) and the remaining 8

were hybrid/complex. The 8 hybrid/complex structures were at lower abundance and are shown in the zoomed portion of Fig. 3A (Structures 28, 37, 56, 65, 81, 94, 105, 130 in Suppl. Table 3).

When the hybrid/complex glycans were grouped in different glycan types, namely bisecting, fucosylated, sialylated,
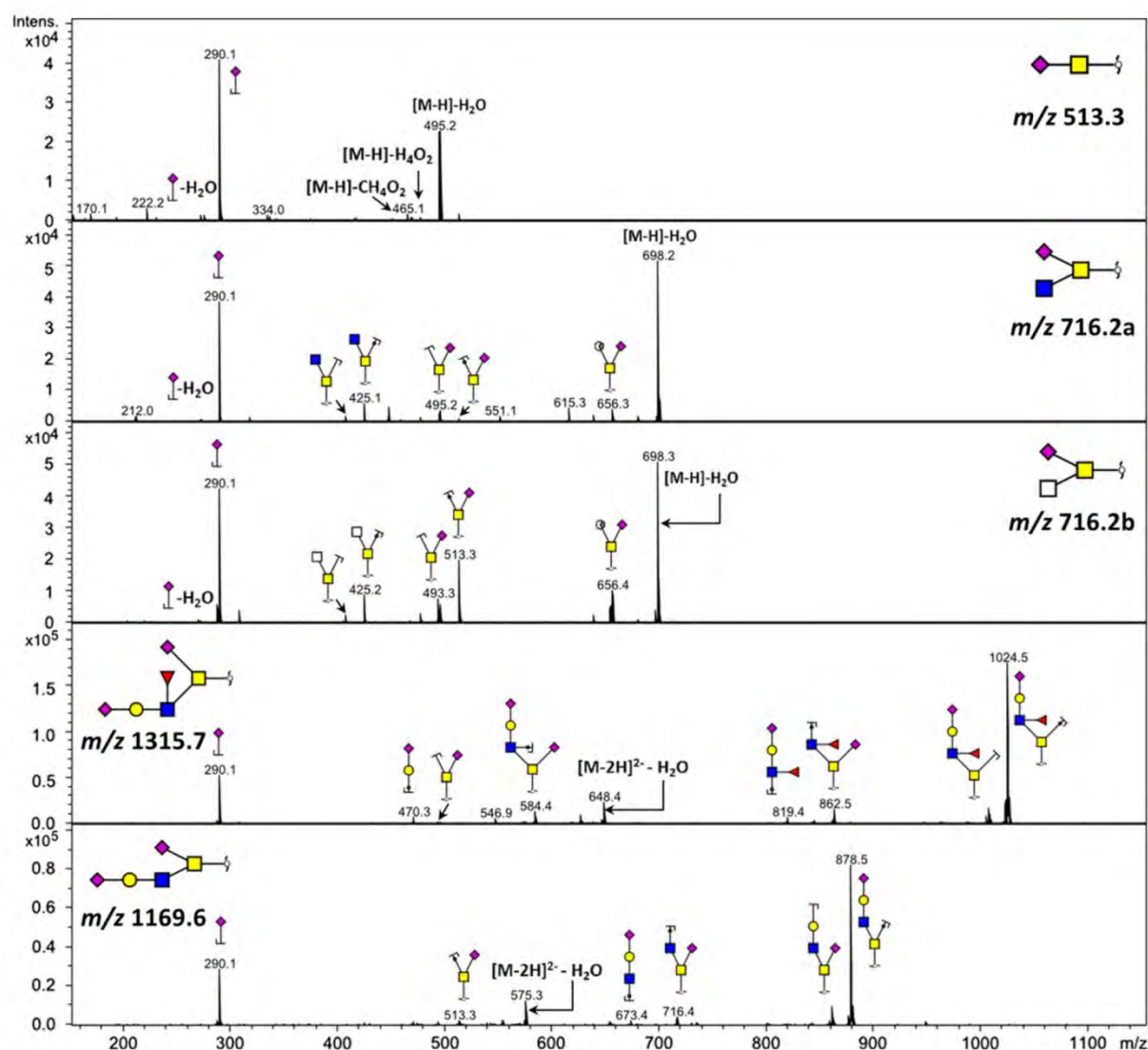
Fig. 4 – Annotation of representative MS2 spectra for the five most abundant O-glycan structures detected in tumours but absent in the low mucin-producing colorectal cancer cell lines (SW1116, SW480, SW620, SW837). Only m/z 513.3 and m/z 716.2b were observed in LS174T O-glycans isolated from mucins. The first isomer of m/z 716.2 (m/z 716.2a) is annotated according to glycan data available from Unicarb-DB. As core 5/6/7 cannot be excluded, the second isomer is depicted with a N-acetyl hexosamine residue at the non-reducing terminus. Glycan nomenclature was derived from GlycoWorkbench.

paucimannosidic and Lewis-type structures, some notable differences could be observed. For some glycan classes examined, the abundance (expressed as percentage of total N-glycans) was either relatively small (paucimannosidic, <3.1%, and Lewis-type, <2.9%, data not shown) or differential abundance was not generally observed based on a 2-fold change (fucosylation, Fig. 3E) between the different cell lines or the tissues. In contrast, total N-linked sialylation (Fig. 3G) and bisected N-glycans (Fig. 3C) displayed significant differential abundance (>2-fold) between the cell lines. There were also notable inverse differences for sialylated and bisecting N-glycans between the cell lines, particularly in SW480 with bisecting glycans in very low in abundance (<0.1%) with a corresponding very high total sialylation (30%) (Fig. 3C, G). In

contrast bisecting GlcNAc was detected at high abundance in other cell lines, for example, 15.7% for SW1116 and 14.2% in SW620 (Fig. 3C) with a consequent low sialylation. In addition, there was high variation in the expression of bisecting (0.1%–15.7%, Fig. 3C) and sialylated N-glycans (5.9%–29.8%, Fig. 3G) across the cell lines, which was also observed to a smaller degree in the tumours (bisecting N-glycans, 8.5–15.6%; sialylated N-glycans, 22.0%–27.5%, Fig. 3C, G). This variability is likely a reflection of biological differences given the small variance observed in the glycan analysis of duplicate growths of cell lines (biological replicates) for these N-glycan classes. Therefore, our observations indicate marked biological variation in expression of sialylated and bisected N-glycans between cell lines and tumours.

3.3.2. Analysis of O-glycans

For the analysis of O-glycosylation in colon cancer, it was considered that since mucins are the predominant O-glycan carrying proteins in the colon [21,41,46], a known mucinous cell line should be included in the analysis. Specifically for this purpose the mucins produced by the cell line LS147T were
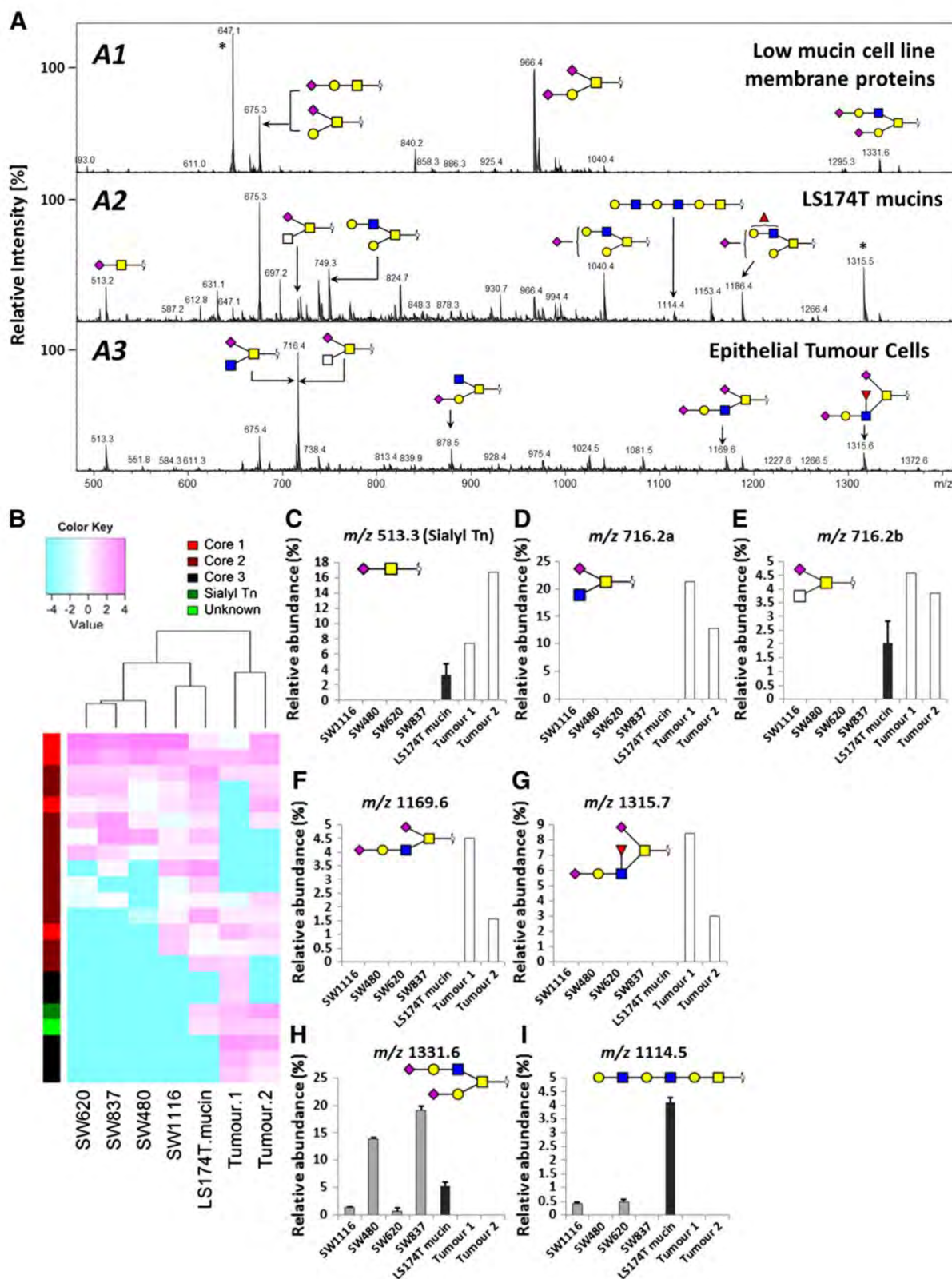
isolated and compared with the O-glycosylation found in the tumour tissue.

Representative average mass spectra of O-glycans observed in the cell lines and tumours are shown in Fig. 5A. MS2 annotation and relative quantification by area-under-the-curve integration were performed on all 43 O-glycan structures (Suppl. Table 4). In total, there were 22 abundant structures (>3% in at least one sample) where the expression varied amongst cell lines and tumours. Varied abundance can also be detected for other glycans of relatively low abundance (0.3–2%) although these consisted mostly of O-glycans detected at low abundance in one tumour exclusively, and therefore these have been excluded for comparison of individual glycans. When the quantities of the 22 most abundant glycan structures were represented in a heatmap and unsupervised clustering was performed, there was a clear separation of tumour and cell lines (Fig. 5B).

Some striking observations were made when comparing the membrane protein O-glycosylation of cultured cell lines (SW1116, SW480, SW620, SW837) with epithelial cells isolated from CRC tumours and mucins purified from LS174T cultured cells (Fig. 5A). Only 4 glycans were observed solely in cell lines (for example, m/z 1114.5, 0.4–4.3%, Fig. 5I; and m/z 1331, 1.1%–19.8%, Fig. 5H; glycans 30 and 40, Suppl. Fig. 6) but there were 5 O-glycans that were highly expressed in epithelial cells of tumours (with 2 also expressed on LS174T cell mucins) whilst they were not able to be detected at all in low-mucin producing cell lines (SW1116, SW480, SW620, SW837, Fig. 5C–G). The MS2 used to validate these structures is shown in Fig. 4. These five O-glycan structures corresponded to sialyl-Tn $(HexNAc)_1(NeuAc)_1$ with m/z 513 (7.6%–16.8% in tumours, 3.4% in LS174T; Fig. 5C), two isomers of trisaccharide $(HexNAc)_2(NeuAc)_1$ with m/z 716 (12.9–21.9% for isomer 1, 3.9–4.7% for isomer 2 in tumours, 2.0% isomer 2 only in LS174T; Fig. 5D, E), $(Hex)_1(HexNAc)_2(NeuAc)_2$ with m/z 1169 (1.6%–4.6% in tumours, Fig. 5F) and hexasaccharide $(Hex)_1(HexNAc)_2(Deoxyhexose)_1(NeuAc)_2$ with m/z 1315 (3.0–8.6% in tumours; Fig. 5G).

The $\alpha 2,6$-linkage of sialic acid on sialyl-Tn in the tumour cells was confirmed with exoglycosidase digestion (Suppl. Fig. 2) where separate aliquots from the same pooled tumour sample were digested with either an $\alpha 2,3/6$-sialidase (ABS) and an $\alpha 2,3$-sialidase (NAN1). According to known glycan synthetic pathways, the presence of ion m/z 513 in NAN1-treated and its absence in ABS-digested sample suggest that the sialic acid on sialyl-Tn is $\alpha 2,6$-linked. Although the two isomers of 716 (a & b) have similar MS2 fragmentation (Fig. 4), there is a difference in elution time of approximately two minutes (Supp Fig. 4), which is comparable to an earlier study [21]. The first eluting isomer of m/z 716 (Fig. 5D) is assigned as a core 3 O-glycan since it is consistent with the elution order previously established in another structural glycan investigation under similar experimental conditions [47]. The second eluting isomer of m/z 716 was also annotated based on MS2 fragmentation data (Fig. 4) and prior reports of this structure [20,21,48]. For this isomer, the fragment m/z 513, and accompanying ion of 18 Da lower, strongly indicates that the sialic acid is attached to the terminal reducing end GalNAc. The fragment m/z 493 may be explained by cross-ring cleavage ($^{2,5}X$ ion) of sialic acid accompanied by the loss of water (m/z 511–18).

Overall, out of the 5 major O-glycans differentially expressed between tumours and cell lines, 3 were core 3 O-glycan structures (Fig. 5D, F, G). Since these three sialylated core 3 glycan structures were exclusively found in the tumour tissue epithelial cells, the abundance of the different O-linked core types was further explored between cell lines and tumours (Fig. 6B–E). The core 3 structures were highly abundant only in tumours (19.0%–42.1%, Fig. 6D) but below detection in all cell lines. In contrast, core 1 glycans were generally more highly abundant in low mucin cell lines (SW1116, SW480, SW620) than on LS174T mucins or tumours (Fig. 6B). Core 2 O-glycan type varied between cell lines and tumours, with the greatest abundance of core 2 structures detected on LS174T mucins (Fig. 6C). Core 4 was detected in only one tumour and at low abundance in one cell line (SW837, Fig. 6E).

The 5 abundant O-glycan structures found in the tumour tissue (Fig. 5C–G) have been previously observed in mucins of human colonic tumour biopsies [20]. The results of our investigation are consistent with these observations. These glycans were not apparent in any of the low mucin-producing cell lines in our study. Interestingly, of these 5 O-glycans, two sialylated structures (sialyl-Tn and a trisaccharide with m/z 716) were also detected in the LS174T cell line mucin O-glycans. The observation of sialyl-Tn in LS174T is in agreement with previous studies [49,50]. However, this is the first known report of expression of the trisaccharide, $(HexNAc)_2(NeuAc)_1$ on LS174T mucins.

Fig. 5 – Differences in the O-glycan profiles of colorectal cancer cultured cells and colorectal tumours. Representative average mass spectra from major O-glycan membrane proteins are shown in A1–A3: A1) low mucin-producing cell lines (SW1116, SW480, SW620, SW837); A2) mucins from the mucinous LS174T cell line; and A3) epithelial cells enriched from tumours. Sialyl-Tn (m/z 513.3) was not detected in membrane protein O-glycans from low-mucin cell lines but was observed in the mucin O-glycans from LS174T. Asterisked peaks were not characterised as O-glycans as indicated by MS2. B) Unsupervised clustering of the relative amounts (% of total area) of the most-abundant glycans (>3% of total O-glycans, n = 22) depicted as a heatmap.
C–I) Major O-glycan differences between membrane proteins of low-mucin producing cell lines (grey bars), LS174T mucins (black bars) and tumours (white bars). The relative abundance of O-glycans was quantified as a percentage of total O-glycans. Error bars of cell lines represent maximum and minimum of duplicate analysis from two independent cell lysates. Tumour tissues (n = 2) are shown as separate biological samples. Five O-glycans were detected in tumours but were undetected in low mucin-producing cell lines (C–G) including sialyl-Tn (C). Remarkably, mucins from LS174T also contained sialyl-Tn (C) as well as one isomer of glycan m/z 716.2 (E). Four O-glycans were detected in cell lines exclusively (e.g. H–I).

3.4. Glycosyltransferase and mucin mRNA expression by qRT-PCR

To validate the changes in glycan structures observed in the different CRC cell line models, the mRNA expression of the enzymes involved in the synthesis of the differently expressed glycans was determined using qRT-PCR (Suppl. Table 6). For the N-glycans expressed on the membrane proteins of CRC cell lines (SW1116, SW480, SW620, SW837), the abundance of bisected N-glycans (Fig. 3C) was in agreement with the expression of the

MGAT3 gene (Fig. 3D). This was expected as the MGAT3 gene product catalyses the addition of the N-acetylglucosamine (GlcNAc) necessary for the formation of bisecting N-glycans [51]. It also provides evidence that MGAT3 expression is an important component in the expression of these bisecting structures in the different colon cancer cell lines.

For the O-glycans the most dominant observation was the lack of expression of the well-known sialyl-Tn cancer marker in the low mucin-producing CRC cell lines compared with the mucinous LS174T cells. To investigate the glycosylation
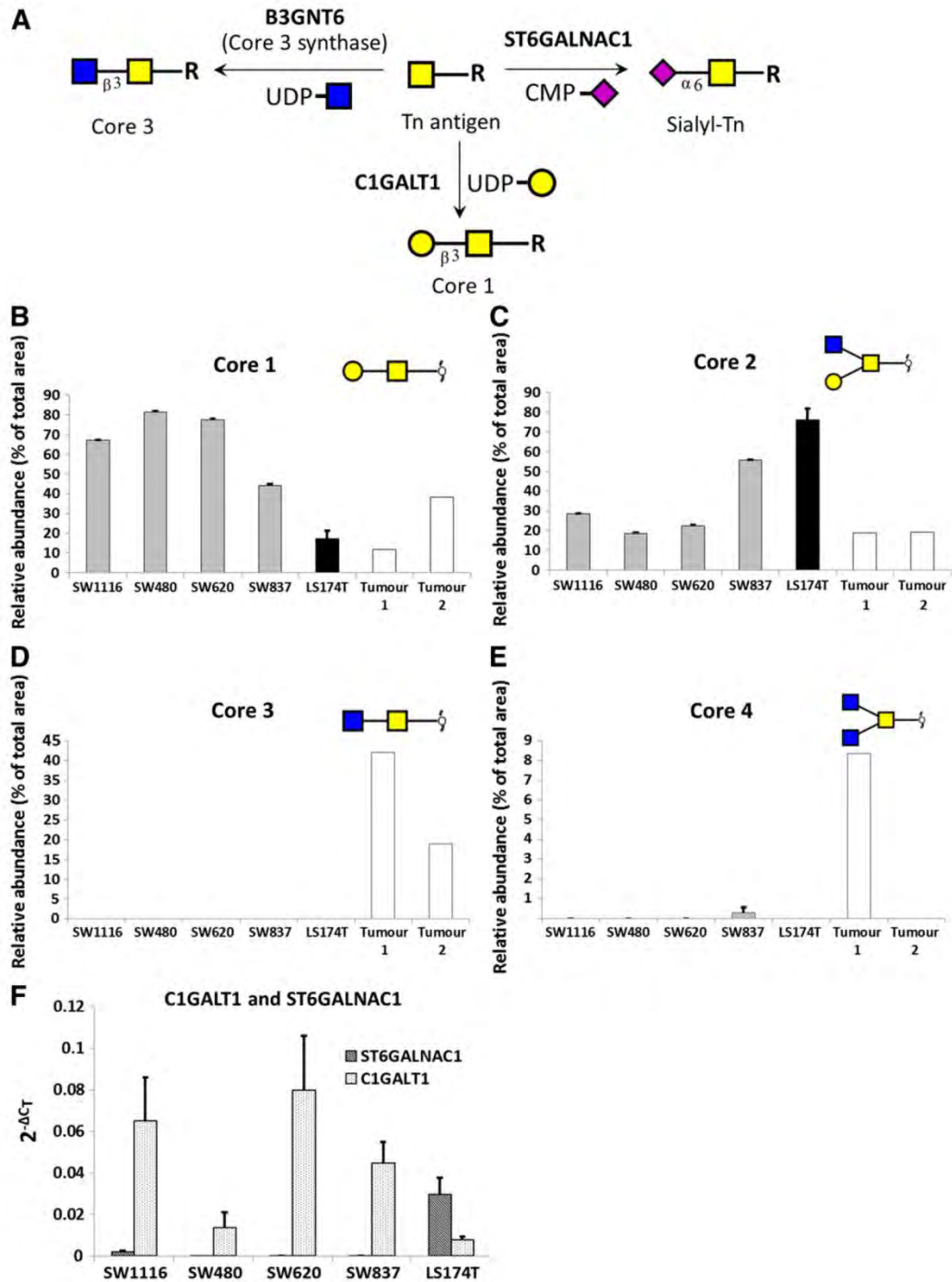
pathways that may be regulating the expression of sialyl-Tn in the mucinous cell lines but not in the low mucin-producing cell lines, the mRNA expression of the enzymes involved in the synthesis of sialyl-Tn was determined. Three glycosyltransferase enzymes (products of the ST6GALNAC1, C1GALT1 and B3GNT6 genes) compete for the Tn antigen substrate and thereby regulate sialyl-Tn versus core 1 and core 3 O-glycan biosynthesis (Fig. 6A). The expression of these genes was compared to the abundance of the sialyl-Tn structure determined by mass spectrometry in the five cell lines. As would be expected from the lack of core 3 O-glycan expression in the cell lines (Fig. 6D), B3GNT6 (the core 3 synthase gene) was not detected in any cell line (data not shown). However, we observed a strong correlation between the expression of sialyl-Tn (Fig. 4) and ST6GALNAC1 (Fig. 5F), the sialyltransferase responsible for addition of sialic acid in a α2,6-linkage to the Tn antigen. The ST6GALNAC1 gene was shown to be almost exclusively expressed and was significantly greater than C1GALT1 expression (P < 0.001) in LS174T cells compared to all other low mucin-producing cell lines, thus correlating with the sialyl-Tn expression found on their mucins.

There was also good correlation between a higher abundance of core 1 O-glycans and the expression of C1GALT1 (Fig. 5F), in the low mucin-producing cell lines compared to the mucinous cell line LS147T (Fig. 5B), although it was noted that C1GALT1 expression in SW480 correlated less well than expected based on the glycan mass profiles. Nonetheless, the C1GALT1 mRNA expression was significantly greater (P < 0.001) than ST6GALNAC1 mRNA expression in all the low-mucin producing CRC cell lines (SW1116, SW480, SW620, SW837, Fig. 6F) indicating the preferred expression of core 1 over sialyl-Tn O-glycans in these colorectal cancer cell models.

The fact that the sialyl-Tn structure and the corresponding ST6GALNAC1 gene were only expressed in the mucinous LS174T cell line, suggests that mucins are one of the major carrier proteins for sialyl-Tn. This is further supported by the observation of sialyl-Tn by mass spectrometry on mucins isolated from human colonic biopsies [20,21]; and the co-localisation of anti-mucin and anti-Sialyl Tn antibodies in cancers of the colon, liver, breast and ovary [52,53]. Therefore, the mRNA expression of mucins was also investigated in the cell lines to determine whether mRNA expression of known sialyl-Tn carriers such as MUC1, MUC2 and MUC6 [53] is altered between the different cell lines. A panel of 7 mucins

previously detected in LS174T cells, either as protein or mRNA, was used (MUC1, MUC2, MUC3, MUC4, MUC5B, MUC6, MUC13 [54–57]) and their mRNA expression determined (Suppl. Table 6). The expression of these mucins was compared between the four low mucin-producing cell lines and the mucinous LS174T cells (Fig. 7). There was significantly higher expression of MUC2, MUC6 and MUC5B in LS174T cells with almost none expressed in the four other colon cancer cell lines (P < 0.0001) (Fig. 7A). The level of other mucins varied considerably between the cell lines. For example, MUC 13 is virtually the only mucin expressed in SW1116 (0.476 [99%], Fig. 7A, B2) and SW837 (0.501 [97%], Fig. 7A, B5) compared to LS174T (0.074 [30%], Fig. 7A, B1); whilst MUC1 was most highly expressed in SW620 (0.017 [50%], Fig. 7A, B4) and SW837 (0.013 [2.5%], Fig. 7A, B5) compared to other cell lines.

The mRNA expression data therefore correlate well with the observed glycosylation profiles in the CRC cell lines, showing a strong correlation between the genes expressed and the resultant differing glycan structures observed across the cultured cells. In addition, the only cell line that expresses the well-described O-glycan cancer mucin marker (LS174T) uniquely expresses several unique mucins not expressed in the other cell lines.

4. Discussion and conclusions

This is the first report of a glycomic analysis of cell membrane proteins from a series of model CRC cell lines and epithelial enriched cell populations from CRC tumour tissue. Because CRC tumours contain a variety of cell types, including immune and stromal cells [58], we first enriched for epithelial cells using EpCAM immunoaffinity prior to glycomic characterisation. We considered that this was essential as it allowed for a true comparison with epithelial colon cancer cells grown in culture.

A major finding from the comparison of glycan profiles of membrane proteins from four epithelial CRC cell lines (SW1116, SW480, SW620, SW837) and epithelial cells from tumours was the absence of five high abundance protein O-glycans in these cell lines, including the well-known pan-carcinoma glycan marker commonly expressed in colorectal tumours, sialyl-Tn antigen [23,59,60] but only rarely in normal mucosa [18,61,62]. Sialyl-Tn antigen has been previously reported on the mucin, MUC2, isolated from colon cancer

Fig. 6 – Differences in the relative abundance of O-glycan cores between cell lines and tissues. A) Initial steps of O-glycan biosynthesis involving Tn antigen. There is known competition between glycosyltransferases encoded by ST6GALNAC1, C1GALT1 and B3GNT6 genes for Tn antigen during the formation of sialyl Tn, core 1 and core 3. B–D) Relative abundances of O-glycan cores between cell lines and tissues. Cores of O-glycans were derived from MS2 annotation (Supp. Table 4) and common knowledge of glycan biosynthetic pathways. Their relative abundance was quantified as a percentage of total O-glycans. Error bars of cell lines represent maximum and minimum of biological duplicates performed from two separate cell lysates. Tumour tissues (n = 2) were plotted as separate biological replicates. Grey bars, low mucin cell lines; black bars, LS174T; white bars, tumours. B) Core 1 structures were generally more abundant in low mucin-producing cell lines (SW1116, SW480, SW620, SW837) than mucinous LS174T cells. C) Abundance of core 2 varied between cell lines and tumours. LS174T mucins had the highest number of core 2 structures. D) Core 3 structures were highly abundant in tumours but were below detection in all cell lines. E) Core 4 structures were abundant in one tumour only and generally low in abundance or undetected in other samples. F) mRNA expression of glycosyltransferase genes, C1GALT1 and ST6GALNAC1. Both these genes are known to compete for Tn antigen. The difference between the expression of ST6GALNAC1 and C1GALT1 in each cell line was highly statistically significant (P < 0.001) as determined by Student's t-test.
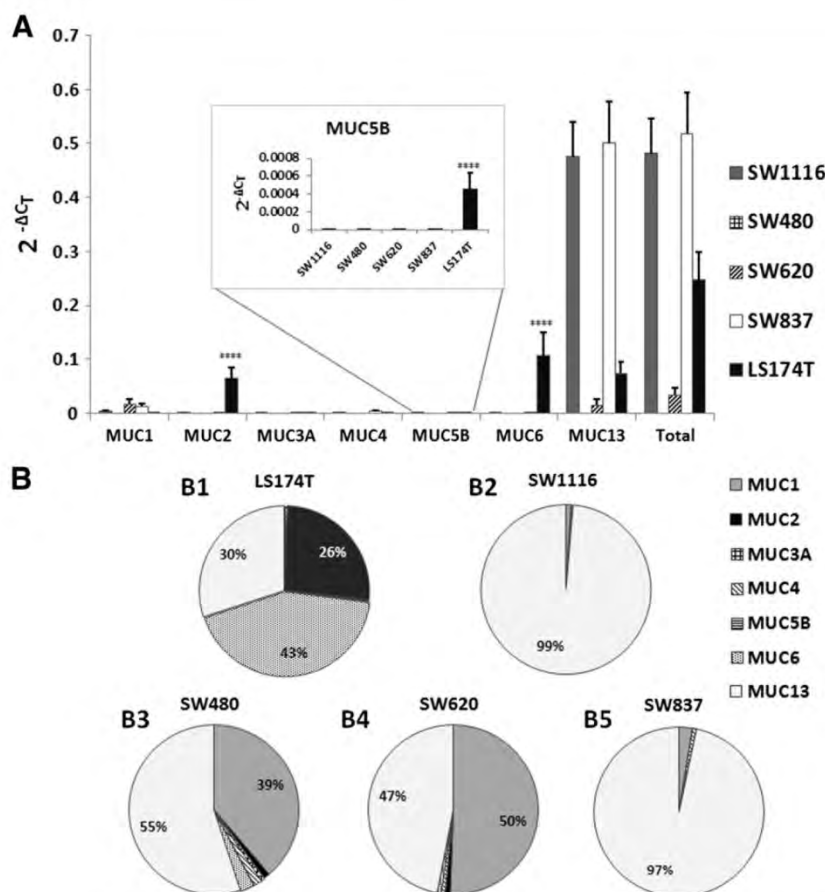
Fig. 7 – Relative quantification of mucin mRNA expression using qRT-PCR. The expression of 7 mucin genes previously detected in LS174T was determined (MUC1, MUC2, MUC3A, MUC4, MUC5B, MUC6, MUC13). A) Differences in the expression of profiled mucin genes between CRC cell lines. Error bars represent mean ± SD of at least five data points from two triplicate PCRs. The expression of MUC2, MUC5B and MUC6 was significantly higher in LS174T compared with the other low mucin-producing cell lines as determined using ANOVA (P < 0.001). B1–B5) Proportional differences in mucin mRNA expression between the different colorectal cancer cell lines.

biopsies [20,21]. Interestingly, sialyl-Tn was found in our study to be expressed on the mucin fraction isolated from the LS174T cell line which are known to be mucinous cells [11].

The results of our study complement a previous study which investigated sialyl-Tn expression in a panel of CRC cell lines, including SW480, LS174T and SW1116 [11]. In that work, strong staining by anti-sialyl-Tn antibodies was observed on LS174T cells and no staining was observed for SW480 [11] supporting our observation. However, moderate antibody staining was also observed for SW1116 cells although sialyl-Tn was not detected in SW1116 in our mass spectrometric approach. SW1116 is known as a weak mucin producer [11] and the high expression of MUC13 mRNA in this cell line was noted in our study. However, MUC13 has not been previously reported as a carrier of sialyl-Tn. Therefore, insufficient production by SW1116 of this mucin carrying sialyl Tn for MS analysis cannot be excluded.

Another glycan of interest seen in the tumour tissue which was also not seen in cell lines was the core 3 disialyl-Lewis X hexsaccharide, $(Hex)_1(HexNAc)_2(Deoxyhexose)_1(NeuAc)_2$. This has also been previously reported as a potential marker of

malignant transformation in the colon by LC–MS/MS [20]. Remarkably, this glycan was not detected across all cell lines, including LS174T. Whilst there are no known reports of the other three tumour abundant O-glycans being correlated with cancer, the lack of detection of sialyl-Tn and the core 3 disialyl-Lewis X hexsaccharide indicates that potential O-glycan biomarkers are not readily expressed in some commonly used CRC cell line models.

This observed difference in the expression of Sialyl-Tn and core 3 O-glycans could be regulated by many factors such as the mRNA turnover and the availability of substrate, nucleotide sugar transporters and relevant glycosyltransferases. Therefore, differential gene expression of competing glycosyltransferases in the Tn biosynthetic pathway, such as B3GNT6, ST6GALNAC1 and C1GALT1 was investigated. ST6GALNAC1 is one of six known members of the N-acetylgalactosamine (GalNAc) α2,6-sialyltransferase family [63] which is responsible for addition of α2,6-linked sialic acid to GalNAc. The ST6GalNac1 enzyme is reported to act preferentially on the substrate Tn antigen for formation of sialyl-Tn antigen [64,65] which was validated by the high expression of the ST6GALNAC1 mRNA

observed in sialyl-Tn expressing LS174T cells compared to other cell lines where sialyl-Tn was not detected. Interestingly, SW1116 had a low expression of ST6GALNAC1 although sialyl-Tn was not detected, and this may explain the previous report where sialyl-Tn was detected by an anti-sialyl-Tn antibody in this cell line [11].

Other enzymes known to act on Tn antigen are C1GALT1 (core 1 synthase) and B3GNT6 (core 3 synthase). Reduced core 1 synthase activity in a LS174T variant has been previously shown to result in the synthesis of only Tn and sialyl-Tn oligosaccharides [66]. In this study, an inverse relationship was observed where an increased formation of sialyl-Tn in LS174T cells correlated with higher expression of ST6GALNAC1 and lower expression of C1GALT1 genes, indicating the competition of C1GALT1 and ST6GALNAC1 gene products for the same substrate (Tn antigen). The inverse relationship was also observed in the four other cell lines in which there was no Sialyl Tn. Correspondingly, these cell lines also had an increase in core 1 structures which generally correlated with an increase in expression of C1GALT1.

However, although there was general concordance between mRNA expression of C1GALT1 and the abundance of core 1 structures, the expression of C1GALT1 was lower than expected in SW480. The reason for this is unknown although disparity between enzyme activity and mRNA expression has been previously reported, including instances where enzyme activity greatly exceeded mRNA expression [67]. One explanation to reconcile the discrepancy is a difference in co-expression of C1GALT1 chaperone, Cosmc, between SW480 and the other cell lines. Cosmc is known to be essential for core 1 synthase activity and somatic mutations of C1GALT1 are reported to result in Tn-related disorder such as Tn syndrome [68]. A recent study has shown that overexpression of Cosmc increases core 1 expression and interestingly, SW480 was reported to express Cosmc in the same study [69]. Although this remains to be further investigated, overexpression of Cosmc is a potential compensatory mechanism for the abundant expression of core 1 glycans in SW480, despite its relatively low C1GALT1 mRNA expression.

In addition to the enzyme products of the genes ST6GALNAC1 and C1GALT1, core 3 synthase (core 3 β1-3 N-acetylglucosaminyltransferase gene B3GNT6) also competes for the Tn antigen as a substrate [70]. There was no expression of B3GNT6 gene in any of the cell lines which correlated completely with the absence of core 3 structures. This finding is supported by an earlier study, where six colorectal cancer cell lines, including SW1116, displayed no core 3 synthase activity [71]. Conversely, the abundant expression of core 3 structures in the tumour tissues in this study is also consistent with the activity of the core 3 synthase previously detected in colonic tumours [71]. Therefore, it is likely that elevated core 3 synthase gene (B3GNT6) expression contributed to the higher abundance of core 3 structures in the tumours. Overall, the expression of sialyl-Tn in the colon cancer tumours may be explained by the regulation of glycosyltransferases competing for the Tn antigen (gene products of ST6GALNAC1, B3GNT6 and C1GALT1), a trend that has previously been reported [72,73].

Sialyl-Tn and the other four highly abundant tumour O-glycans not detected on the majority of cell lines studied have been reported as being expressed on mucins. Earlier, all five glycans were reported to be carried by the mucin MUC2 isolated from colon cancer biopsies using LC–MS/MS [20,21]. Moreover, sialyl-Tn and other mucins (MUC1 [52,65], MUC2 [74], MUC6 [53] and MUC5B [75]) were observed to be co-localised with sialyl-Tn in antibody-based studies [52,53] further supporting the notion that these mucins are sialyl-Tn carriers. In our study, we observed stark differences in mucin expression between mucinous LS174T cells and the other less mucinous cell lines (SW1116, SW480, SW620, SW837). MUC2, MUC6 and MUC5B were highest in expression in the LS174T cell line whilst MUC13 and/or MUC 1 were the major mucins expressed in the other cell lines. Overall, these findings demonstrate an association between sialyl-Tn expression and the mucins, MUC2, MUC6 and MUC5B in LS174T, which is consistent with these mucins being sialyl-Tn carriers.

With the exception of MUC1, which is a well-known sialyl-Tn carrier [52,53,65], the status of other mucins expressed in colon cancer cells (MUC3, MUC4, MUC13) as sialyl-Tn carrier proteins is currently unknown. However, there is notable variability in mucin expression in these cell lines, particularly in the expression of MUC13. MUC13 is a recently discovered transmembrane mucin with little known about its glycosylation, although N- and O-glycosylation has been indicated by PNGase F treatment and lectins [76]. Unexpectedly, MUC13 mRNA was extremely highly expressed in SW1116 and SW837, which has not been previously reported. Here, high MUC13 expression did not correlate with the presence of sialyl-Tn or the relevant glycosyltransferases genes, ST6GALNAC1 and C1GALT1 (Figs. 5C, 6F, 7A). To account for this observation, either the mature MUC 13 glycoprotein has not been properly expressed from the transcribed mRNA due to protein misfolding and degradation [77,78] or low mRNA stability [79]. Alternatively, MUC13 is expressed but does not carry sialyl Tn, which is supported by the absence of sialyl-Tn and the greatly reduced expression of ST6GALNAC1 in SW1116 and SW837.

On a different note, cell line differences in protein N-glycans were less marked than for the O-glycans. The most interesting observation was the alteration in the abundance of bisecting N-glycans. This glycan type is characterised by the presence of GlcNAc attached to the mannose of the trimannosyl core and this glycosidic linkage is catalysed by the glycosyltransferase encoded by the MGAT3 gene [70]. In this study, the abundance of bisecting N-glycans was well correlated with the MGAT3 gene expression in the cell lines, which indicates that mRNA expression of MGAT3 is a required factor in the likely complex regulation of bisecting GlcNAc synthesis that includes enzyme expression and the correct substrates and transporters. Since SW620 is a more metastatic cell line derived from the same patient as SW480, the low expression of bisecting GlcNAc on SW480 compared to SW620 cells therefore suggest bisecting glycans may play a role in colorectal metastatic progression [80].

In summary, comparison of N- and O-glycan profiles obtained by LC–MS/MS from various colorectal cancer cell lines showed remarkable differences in the glycans expressed on the cell membrane proteins. Amongst the most striking differences included the expression of well-known glycan marker sialyl-Tn, which is upregulated in CRC epithelial tumour cells and in the mucinous LS147T cell line. The strong association between the mRNA of ST6GALNAC1/C1GALT1 and sialyl-Tn structural abundance suggests that mRNA

expression of pertinent glycosyltransferase genes is an important component of sialyl-Tn biosynthesis. Moreover, sialyl-Tn carrying mucins such as MUC2 and MUC6 are co-expressed in LS174T, the only cell line which expressed sialyl-Tn. In addition, the higher similarity between the O-glycosylation of LS174T and tumour tissues suggests that LS147T is the most useful *in vitro* model for gaining insights into O-glycan expression and regulation seen in tumours. However, not all tumour-abundant O-glycans were detected even in these cells. Our data also highlights that differences in protein glycosylation exist between different cancer cell line models and the epithelial cells from tumour tissue, and that this needs to be carefully considered, especially when designing glycan biomarker studies.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.05.002.

REFERENCES

[1] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61(2):69–90.
[2] Siegel R, Naishadham D, Jemal A. Cancer statistics. CA Cancer J Clin 2012;62(1):10–29.
[3] Brooks SA, Carter TM, Royle L, Harvey DJ, Fry SA, Kinch C, et al. Altered glycosylation of proteins in cancer: what is the potential for new anti-tumour strategies. Anticancer Agents Med Chem 2008;8(1):2–21.
[4] Kim YJ, Varki A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. Glycoconj J 1997;14(5):569–76.
[5] Meany DL, Chan DW. Aberrant glycosylation associated with enzymes as cancer biomarkers. Clin Proteomics 2011;8(1):7.
[6] Ono M, Hakomori S. Glycosylation defining cancer cell motility and invasiveness. Glycoconj J 2004;20(1):71–8.
[7] Varki A, Sharon N. Historical background and overview. In: Varki A, et al, editors. Essentials of Glycobiology; 2009 [Cold Spring Harbor (NY)].
[8] Stanley P, Schachter H, Taniguchi N. N-glycans. In: Varki A, et al, editors. Essentials of Glycobiology; 2009 [Cold Spring Harbor (NY)].
[9] Brockhausen I, Schachter H, Stanley P. O-GalNAc glycans. In: Varki A, et al, editors. Essentials of Glycobiology; 2009 [Cold Spring Harbor (NY)].
[10] Brockhausen I. Pathways of O-glycan biosynthesis in cancer cells. Biochim Biophys Acta 1999;1473(1):67–95.
[11] Dahiya R, Itzkowitz SH, Byrd JC, Kim YS. Mucin oligosaccharide biosynthesis in human colonic cancerous tissues and cell lines. Cancer 1992;70(6):1467–76.
[12] Balog CIA, Stavenhagen K, Fung WLJ, Koeleman CA, McDonnell LA, Verhoeven A, et al. N-glycosylation of colorectal cancer tissues. Mol Cell Proteomics 2012;11(9):571–85.
[13] Kannagi R. Carbohydrate antigen sialyl Lewis A—its pathophysiological significance and induction mechanism in cancer progression. Chang Gung Med J 2007;30(3):189–209.
[14] Magnani JL, Nilsson B, Brockhaus M, Zopf D, Steplewski Z, Koprowski H, et al. A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-fucopentaose II. J Biol Chem 1982;257(23):14365–9.
[15] Hundt S, Haug U, Brenner H. Blood markers for early detection of colorectal cancer: a systematic review. Cancer Epidemiol Biomarkers Prev 2007;16(10):1935–53.
[16] Nakamori S, Kameyama M, Imaoka S, Furukawa H, Ishikawa O, Sasaki Y, et al. Involvement of carbohydrate antigen sialyl Lewis X in colorectal cancer metastasis. Dis Colon Rectum 1997;40(4):420–31.
[17] Bresalier RS, Ho SB, Schoeppner HL, Kim YS, Sleisenger MH, Brodt P, et al. Enhanced sialylation of mucin-associated carbohydrate structures in human colon cancer metastasis. Gastroenterology 1996;110(5):1354–67.
[18] Julien S, Videira PA, Delannoy P. Sialyl-Tn in cancer: (how) did we miss the target? Biomol 2012;2(4):435–66.
[19] Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. EMBO Rep 2006;7(6):599–604.
[20] Robbe-Masselot C, Herrmann A, Maes E, Carlstedt I, Michalski JC, Capon C. Expression of a core 3 disialyl-Le(x) hexasaccharide in human colorectal cancers: a potential marker of malignant transformation in colon. J Proteome Res 2009;8(2):702–11.
[21] Larsson JM, Karlsson H, Sjovall H, Hansson GC. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. Glycobiology 2009;19(7):756–66.
[22] Akamine S, Nakagoe T, Sawai T, Tsuji T, Tanaka K, Hidaka S, et al. Differences in prognosis of colorectal cancer patients based on the expression of sialyl Lewisa, sialyl Lewisx and sialyl Tn antigens in serum and tumor tissue. Anticancer Res 2004;24(4):2541–6.
[23] Itzkowitz SH, Bloom EJ, Kokal WA, Modin G, Hakomori S-I, Kim YS. Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients. Cancer 1990;66(9):1960–6.
[24] Moriwaki K, Noda K, Furukawa Y, Ohshima K, Uchiyama A, Nakagawa T, et al. Deficiency of GMDS leads to escape from NK cell-mediated tumor surveillance through modulation of TRAIL signaling. Gastroenterology 2009;137(1):188–98.
[25] Morgenthaler J, Kemmner W, Brossmer R. Sialic acid dependent cell adhesion to collagen IV correlates with in vivo tumorigenicity of the human colon carcinoma sublines HCT116, HCT116a and HCT116b. Biochem Biophys Res Commun 1990;171(2):860–6.
[26] Vercoutter-Edouart A-S, Slomianny M-C, Dekeyzer-Beseme O, Haeuw J-F, Michalski J-C. Glycoproteomics and glycomics investigation of membrane N-glycosylproteins from human colon carcinoma cells. Proteomics 2008;8(16):3236–56.
[27] Terada M, Khoo KH, Inoue R, Chen CI, Yamada K, Sakaguchi H, et al. Characterization of oligosaccharide ligands expressed on SW1116 cells recognized by mannan-binding

protein. A highly fucosylated polylactosamine type N-glycan. J Biol Chem 2005;280(12):10897–913.

[28] Moyret C, Madsen MW, Cooke J, Briand P, Theillet C. Gradual selection of a cellular clone presenting a mutation at codon 179 of the p53 gene during establishment of the immortalized human breast epithelial cell line HMT-3522. Exp Cell Res 1994;215(2):380–5.

[29] Hughes P, Marshall D, Reid Y, Parkes H, Gelber C. The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? Biotechniques 2007;43(5):575 [577–8, 581–2 passim].

[30] Zhou J, Belov L, Huang PY, Shin J-S, Solomon MJ, Chapuis PH, et al. Surface antigen profiling of colorectal cancer using antibody microarrays with fluorescence multiplexing. J Immunol Methods 2010;355(1–2):40–51.

[31] Davis NC, Newland RC. Terminology and classification of colorectal adenocarcinoma: the Australian clinico-pathological staging system. Aust N Z J Surg 1983;53(3):211–21.

[32] Fielding LP, Arsenault PA, Chapuis PH, Dent O, Gathright B, Hardcastle JD, et al. Clinicopathological staging for colorectal cancer: an International Documentation System (IDS) and an International Comprehensive Anatomical Terminology (ICAT). J Gastroenterol Hepatol 1991;6(4):325–44.

[33] Lee A, Kolarich D, Haynes PA, Jensen PH, Baker MS, Packer NH. Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. J Proteome Res 2009;8(2):770–81.

[34] Chick JM, Haynes PA, Molloy MP, Bjellqvist B, Baker MS, Len AC. Characterization of the rat liver membrane proteome using peptide immobilized pH gradient isoelectric focusing. J Proteome Res 2008;7(3):1036–45.

[35] Kilcoyne M, Gerlach JQ, Gough R, Gallagher ME, Kane M, Carrington SD, et al. Construction of a natural mucin microarray and interrogation for biologically relevant glyco-epitopes. Anal Chem 2012;84(7):3330–8.

[36] Jensen PH, Karlsson NG, Kolarich D, Packer NH. Structural analysis of N- and O-glycans released from glycoproteins. Nat Protoc 2012;7(7):1299–310.

[37] Everest-Dass A, Abrahams J, Kolarich D, Packer N, Campbell M. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. J Am Soc Mass Spectrom 2013;24(6):895–906.

[38] Harvey DJ, Royle L, Radcliffe CM, Rudd PM, Dwek RA. Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. Anal Biochem 2008;376(1):44–60.

[39] Hayes CA, Karlsson NG, Struwe WB, Lisacek F, Rudd PM, Packer NH, et al. UniCarb-DB: a database resource for glycomic discovery. Bioinformatics 2011;27(9):1343–4.

[40] Harvey DJ. Fragmentation of negative ions from carbohydrates: part 3. Fragmentation of hybrid and complex N-linked glycans. J Am Soc Mass Spectrom 2005;16(5):647–59.

[41] Lin P, Hu SW, Chang TH. Correlation between gene expression of aryl hydrocarbon receptor (AhR), hydrocarbon receptor nuclear translocator (Arnt), cytochromes P4501A1 (CYP1A1) and 1B1 (CYP1B1), and inducibility of CYP1A1 and CYP1B1 in human lymphocytes. Toxicol Sci 2003;71(1):20–6.

[42] Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 2008;3(6):1101–8.

[43] Momburg F, Moldenhauer G, Hammerling GJ, Moller P. Immunohistochemical study of the expression of a Mr 34,000 human epithelium-specific surface glycoprotein in normal and malignant tissues. Cancer Res 1987;47(11):2883–91.

[44] Moll R, Franke WW, Schiller DL, Geiger B, Krepler R. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell 1982;31(1):11–24.

[45] Franke WW, Schmid E, Osborn M, Weber K. Different intermediate-sized filaments distinguished by immunofluorescence microscopy. Proc Natl Acad Sci 1978;75(10):5034–8.

[46] Holmen Larsson JM, Thomsson KA, Rodriguez-Pineiro AM, Karlsson H, Hansson GC. Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. Am J Physiol Gastrointest Liver Physiol 2013;305(5):G357–63.

[47] Everest-Dass AV, Jin D, Thaysen-Andersen M, Nevalainen H, Kolarich D, Packer NH. Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by *Candida albicans*. Glycobiology 2012;22(11):1465–79.

[48] Robbe C, Capon C, Coddeville B, Michalski JC. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. Biochem J 2004;384(Pt 2):307–16.

[49] Ogata S, Chen A, Itzkowitz SH. Use of model cell lines to study the biosynthesis and biological role of cancer-associated sialosyl-Tn antigen. Cancer Res 1994;54(15):4036–44.

[50] Axelsson MAB, Karlsson NG, Steel DM, Ouwendijk J, Nilsson T, Hansson GC. Neutralization of pH in the Golgi apparatus causes redistribution of glycosyltransferases and changes in the O-glycosylation of mucins. Glycobiology 2001;11(8):633–44.

[51] Priatel JJ, Sarkar M, Schachter H, Marth JD. Isolation, characterization and inactivation of the mouse Mgat3 gene: the bisecting N-acetylglucosamine in asparagine-linked oligosaccharides appears dispensable for viability and reproduction. Glycobiology 1997;7(1):45–56.

[52] Pinto R, Carvalho AS, Conze T, Magalhaes A, Picco G, Burchell JM, et al. Identification of new cancer biomarkers based on aberrant mucin glycoforms by in situ proximity ligation. J Cell Mol Med 2012;16(7):1474–84.

[53] Sasaki M, Yamato T, Nakanuma Y. Expression of sialyl-Tn, Tn and T antigens in primary liver cancer. Pathol Int 1999;49(4):325–31.

[54] Rodriguez-Pineiro AM, Post S, Johansson ME, Thomsson KA, Nesvizhskii AI, Hansson GC. Proteomic study of the mucin granulae in an intestinal goblet cell model. J Proteome Res 2012;11(3):1879–90.

[55] Becker S, Oelschlaeger TA, Wullaert A, Pasparakis M, Wehkamp J, Stange EF, et al. Bacteria regulate intestinal epithelial cell differentiation factors both *in vitro* and *in vivo*. PLoS One 2013;8(2):e55620.

[56] Bu XD, Li N, Tian XQ, Huang PL. Caco-2 and LS174T cell lines provide different models for studying mucin expression in colon cancer. Tissue Cell 2011;43(3):201–6.

[57] van Klinken BJ, Oussoren E, Weenink JJ, Strous GJ, Buller HA, Dekker J, et al. The human intestinal cell lines Caco-2 and LS174T as models to study cell-type specific mucin expression. Glycoconj J 1996;13(5):757–68.

[58] Jankova L, Chan C, Fung CL, Song X, Kwun SY, Cowley MJ, et al. Proteomic comparison of colorectal tumours and non-neoplastic mucosa from paired patient samples using iTRAQ mass spectrometry. Mol Biosyst 2011;7(11):2997–3005.

[59] Itzkowitz SH, Yuan M, Montgomery CK, Kjeldsen T, Takahashi HK, Bigbee WL, et al. Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. Cancer Res 1989;49(1):197–204.

[60] Imada T, Rino Y, Hatori S, Takahashi M, Amano T, Kondo J, et al. Sialyl Tn antigen expression is associated with the prognosis of patients with advanced colorectal cancer. Hepatogastroenterology 1999;46(25):208–14.

[61] Ogata S, Koganty R, Reddish M, Michael Longenecker B, Chen A, Perez C, Itzkowitz S. Different modes of sialyl-Tn expression during malignant transformation of human colonic mucosa. Glycoconj J 1998;15(1):29–35.

[62] Thor A, Ohuchi N, Szpak CA, Johnston WW, Schlom J. Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. Cancer Res 1986;46(6):3118–24.

[63] Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi MA, Samyn-Petit B, Julien S, Delannoy P. The human sialyltransferase family. Biochimie 2001;83(8):727–37.

[64] Marcos NT, Pinho S, Grandela C, Cruz A, Samyn-Petit B, Harduin-Lepers A, et al. Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated Sialyl-Tn antigen. Cancer Res 2004;64(19):7050–7.

[65] Sewell R, Bäckström M, Dalziel M, Gschmeissner S, Karlsson H, Noll T, et al. The ST6GalNAc-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated Sialyl-Tn O-glycan in human breast cancer. J Biol Chem 2006;281(6):3586–94.

[66] Brockhausen I, Yang J, Dickinson N, Ogata S, Itzkowitz SH. Enzymatic basis for sialyl-Tn expression in human colon cancer cells. Glycoconj J 1998;15(6):595–603.

[67] Glanemann C, Loos A, Gorret N, Willis LB, O'Brien XM, Lessard PA, et al. Disparity between changes in mRNA abundance and enzyme activity in *Corynebacterium glutamicum*: implications for DNA microarray analysis. Appl Microbiol Biotechnol 2003;61(1):61–8.

[68] Ju T, Cummings RD. Protein glycosylation: chaperone mutation in Tn syndrome. Nature 2005;437(7063):1252.

[69] Huang J, Che M-I, Lin N-Y, Hung J-S, Huang Y-T, Lin W-C, et al. The molecular chaperone Cosmc enhances malignant behaviors of colon cancer cells via activation of Akt and ERK. Mol Carcinog 2014;53(Suppl 1):E62–71.

[70] Iwai T, Inaba N, Naundorf A, Zhang Y, Gotoh M, Iwasaki H, et al. Molecular cloning and characterization of a novel UDP-GlcNAc:GalNAc-peptide β1,3-N-acetylglucosaminyltransferase (β3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans. J Biol Chem 2002;277(15):12802–9.

[71] Vavasseur F, Yang JM, Dole K, Paulsen H, Brockhausen I. Synthesis of O-glycan core 3: characterization of UDP-GlcNAc: GalNAc-R beta 3-N-acetyl-glucosaminyltransferase activity from colonic mucosal tissues and lack of the activity in human cancer cell lines. Glycobiology 1995;5(3):351–7.

[72] Barrow H, Tam B, Duckworth CA, Rhodes JM, Yu LG. Suppression of core 1 Gal-transferase is associated with reduction of TF and reciprocal increase of Tn, sialyl-Tn and Core 3 glycans in human colon cancer cells. PLoS One 2013;8(3):e59792.

[73] Iwai T, Kudo T, Kawamoto R, Kubota T, Togayachi A, Hiruma T, et al. Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. Proc Natl Acad Sci U S A 2005;102(12):4572–7.

[74] Conze T, Carvalho AS, Landegren U, Almeida R, Reis CA, David L, et al. MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. Glycobiology 2010;20(2):199–206.

[75] Thomsson KA, Prakobphol A, Leffler H, Reddy MS, Levine MJ, Fisher SJ, et al. The salivary mucin MG1 (MUC5B) carries a repertoire of unique oligosaccharides that is large and diverse. Glycobiology 2002;12(1):1–14.

[76] Williams SJ, Wreschner DH, Tran M, Eyre HJ, Sutherland GR, McGuckin MA. Muc13, a novel human cell surface mucin expressed by epithelial and hemopoietic cells. J Biol Chem 2001;276(21):18327–36.

[77] Olzmann JA, Kopito RR, Christianson JC. The mammalian endoplasmic reticulum-associated degradation system. Cold Spring Harb Perspect Biol 2013;5(9).

[78] Haynes CM, Titus EA, Cooper AA. Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death. Mol Cell 2004;15(5):767–76.

[79] Geisberg, Joseph V., Z. Moqtaderi, X. Fan, F. Ozsolak, and K. Struhl, Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell. 156(4): p. 812–824.

[80] Flatmark K, Maelandsmo GM, Martinsen M, Rasmussen H, Fodstad O. Twelve colorectal cancer cell lines exhibit highly variable growth and metastatic capacities in an orthotopic model in nude mice. Eur J Cancer 2004;40(10):1593–8.

2.2 Supplemental information

Supplementary Tables

Supplementary Table 1. Clinicopathological characterisation of colorectal cancer cell lines investigated. Numbered references refer to references as listed in Chapter 6 of this thesis.

Cell line	Derived from	Age	Duke's Stage	Reference
SW1116	Large intestine, colon, adenocarcinoma	73	A	[217]
SW480	Moderately differentiated adenocarcinoma of descending colon	50	B	[217]
SW620	Lymph-node metastasis from colon adenocarcinoma	51	C	[217]
SW837	Large intestine rectum, adenocarcinoma	53	C	[217]
LS174T	Moderately well-differentiated colon adenocarcinoma	58	Unknown	[218]

Supplementary Table 2. Clinicopathological characterisation of patient adenocarcinoma tumours. Patients were classified using the Australian clinicopathological staging system (ACP), which is compatible with TNM (Tumour/Node/Metastasis) staging [53]. The first tumour sample corresponded to patient number one. The second tumour sample used for glycoprofiling was a pooled sample containing tumours (patient number 2, 3 and 4) from three separate Stage B1 individuals (asterisked). Sample number five was used for validation of enrichment by Western Blot. T3, size of tumour; N0, no evidence of lymph node spread; M0, no evidence of metastases.

Patient	Gender	Age	Blood type	ACP staging	TNM staging	Differentiation	Tumour type
1	F	76	O-	B1	T3 N0 M0	Moderate	Rectal adenocarcinoma
2*	F	84	A+	B1	T3 N0 M0	Moderate	Colon adenocarcinoma
3*	F	56	O+	B1	T3 N0 M0	Moderate	Rectal adenocarcinoma
4*	M	57	O+	B1	T3 N0 M0	Moderate	Colorectal adenocarcinoma
5	F	74	O+	B1	T3 N0 M0	Moderate to poor	Colorectal adenocarcinoma

Supplementary Table 3 (over page). *N*-glycan characterisation of cell lines and epithelial-enriched tumour tissues and their normalised areas under the curve. Observed *N*-glycan ions have been reduced. Replicates of cell lines denoted with letters, e.g. SW1116A (replicate 1), SW1116B (replicate 2). Structural elucidation was based annotation of spectral fragments using common knowledge of biosynthetic pathways. The presence of additional structural isomers cannot be excluded. n.d., not detected; T.b.c., to be confirmed.

#	[M – H]⁻	[M – 2H]²⁻	[M – H]⁻ calc	Δ[M – H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
	Observed ions (Obs)		Theoretical (Thr)	Deviation (Thr - Obs)												
1	895.5	n.d.	895.5	0.159	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1$		0.14	0.15	0.77	0.60	0.38	0.38	0.14	0.21	0.41	1.56
2	911.4	n.d.	911.4	0.064	$(Hex)_3 (HexNAc)_2$		0.28	0.33	0.51	0.62	0.40	0.32	0.26	0.41	n.d.	0.55
3	1057.4	n.d.	1057.4	0.006	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1$		1.18	0.97	1.67	1.36	0.78	0.84	0.68	0.73	0.38	0.95
4	1235.6	n.d.	1235.6	0.158	$(Hex)_2 + (Man)_3 (GlcNAc)_2$		0.86	0.87	0.58	0.40	0.73	0.48	1.06	1.51	1.95	0.87
5	n.d.	698.3	1397.6	0.105	$(Hex)_3 + (Man)_3 (GlcNAc)_2$		4.59	4.89	4.32	3.28	6.20	6.91	2.36	3.04	3.10	3.57
6	n.d.	698.3	1397.6	0.105	$(Hex)_3 + (Man)_3 (GlcNAc)_2$		0.41	0.28	0.35	0.34	0.61	0.41	n.d.	n.d.	0.71	0.25
7	n.d.	710.8	1422.6	0.074	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.11	0.07	0.21	0.17	0.07	0.10	n.d.	n.d.	n.d.	n.d.
8	n.d.	718.8	1438.6	0.079	$(Hex)_2 (HexNAc)_1 + (Man)_3 (GlcNAc)_2$		0.23	0.23	0.16	0.21	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
9	n.d.	731.3	1463.6	0.047	$(HexNAc)_2 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.05	0.03	0.05	0.05	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
10	n.d.	731.3	1463.6	0.047	$(HexNAc)_2 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.16	0.12	0.46	0.54	n.d.	n.d.	0.10	0.14	n.d.	n.d.
11	n.d.	731.3	1463.6	0.047	$(HexNAc)_2 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.12	0.11	0.34	0.31	0.29	0.24	n.d.	n.d.	0.60	0.92
12	n.d.	739.3	1479.6	0.052	$(Hex)_1 (HexNAc)_2 + (Man)_3 (GlcNAc)_2$		0.09	0.07	0.05	0.07	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
13	n.d.	739.3	1479.6	0.052	$(Hex)_1 (HexNAc)_2 + (Man)_3 (GlcNAc)_2$		0.31	0.28	0.10	0.08	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
14	n.d.	759.9	1520.8	0.226	$(HexNAc)_3 + (Man)_3 (GlcNAc)_2$		0.49	0.39	0.53	0.53	0.09	0.05	n.d.	n.d.	0.29	0.70
15	n.d.	779.4	1559.8	0.253	$(Hex)_4 + (Man)_3 (GlcNAc)_2$		7.66	9.62	5.21	7.22	10.35	13.95	5.83	8.11	4.33	6.69
16	n.d.	779.4	1559.8	0.253	$(Hex)_4 + (Man)_3 (GlcNAc)_2$		1.82	2.03	1.68	1.45	2.56	2.41	1.35	1.07	1.85	1.90
17	n.d.	783.5	1568	0.436	$(Hex)_1 (HexNAc)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		0.23	0.17	0.15	0.14	0.31	0.30	0.34	0.44	1.23	2.40
18	n.d.	783.5	1568	0.436	$(Hex)_1 (HexNAc)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		n.d.	n.d.	0.02	n.d.	0.24	0.16	0.08	0.11	n.d.	0.21

#	[M – H]⁻	[M – 2H]²⁻	[M – H]⁻ calc	Δ[M – H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
	Observed ions (Obs)		Theoretical (Thr)	Deviation (Thr - Obs)												
19	n.d.	783.5	1568	0.436	$(Hex)_1 (HexNAc)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	0.02	0.03	0.05	0.09	0.08	n.d.	n.d.	n.d.	n.d.
20	n.d.	791.8	1584.6	0.021	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.12	0.19	0.04	0.03	0.36	0.33	0.07	0.05	0.21	0.17
21	n.d.	791.9	1584.8	0.221	$(Hex)_2 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.14	0.18	0.14	0.14	0.12	0.11	n.d.	0.27
22	n.d.	799.8	1600.6	0.026	$(Hex)_3 (HexNAc)_1 + (Man)_3(GlcNAc)_2$		0.72	0.72	0.73	0.74	1.30	0.84	0.55	0.39	0.47	1.00
23	n.d.	812.3	1625.6	-0.005	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.06	0.06	n.d.	n.d.
24	n.d.	812.3	1625.6	-0.005	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.14	0.09	0.24	0.22	n.d.	n.d.	n.d.	n.d.	n.d.	0.14
25	n.d.	812.3	1625.6	-0.005	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.14	0.17	0.18	0.13	n.d.	n.d.	n.d.	n.d.	n.d.	0.09
26	n.d.	812.3	1625.6	-0.005	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.43	0.21	0.99	1.10	n.d.	n.d.	n.d.	n.d.	0.78	0.76
27	n.d.	820.4	1641.8	0.199	$(Hex)_2 (HexNAc)_2 + (Man)_3(GlcNAc)_2$		1.87	1.78	1.31	1.37	0.13	0.08	n.d.	n.d.	1.16	n.d.
28	n.d.	820.4	1641.8	0.199	$(Hex)_2 (HexNAc)_2 + (Man)_3(GlcNAc)_2$		3.17	3.83	2.01	1.95	0.39	0.27	n.d.	n.d.	0.29	n.d.
29	n.d.	820.4	1641.8	0.199	$(Hex)_2 (HexNAc)_2 + (Man)_3(GlcNAc)_2$		0.10	n.d.	0.23	0.30	0.60	0.47	0.35	0.28	0.49	1.30
30	n.d.	832.8	1666.6	-0.031	$(HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.60	0.53	0.92	1.08	0.22	0.16	n.d.	n.d.	2.05	2.53
31	n.d.	832.8	1666.6	-0.031	$(HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.19	0.17	0.32	0.37	0.38	0.46	0.15	0.23	0.47	0.39
32	n.d.	832.8	1666.6	-0.031	$(HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.03	0.01	0.12	0.13	0.13	0.10	n.d.	n.d.	n.d.	0.26
33	n.d.	832.8	1666.6	-0.031	$(HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.07	0.06	0.07	0.11	0.10	0.11	n.d.	n.d.	n.d.	n.d.
34	n.d.	840.9	1682.8	0.173	$(Hex)_1 (HexNAc)_3 + (Man)_3(GlcNAc)_2$		1.07	1.09	1.97	1.64	0.31	0.25	n.d.	n.d.	0.69	n.d.
35	n.d.	856.3	1713.6	-0.021	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.05	0.03	0.27	0.31	0.36	0.39	0.10	0.15	n.d.	0.08
36	n.d.	856.3	1713.6	-0.021	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.51	n.d.

#	[M − H]⁻	[M − 2H]²⁻	[M − H]⁻calc	Δ[M − H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
37	n.d.	856.4	1713.8	0.178	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.29	0.24	0.64	0.66	0.42	0.39	0.87	0.85	1.37	3.32
38	n.d.	856.4	1713.8	0.178	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.03	0.03	0.15	0.15	0.13	0.18	0.13	0.15	n.d.	0.34
39	n.d.	860.3	1721.6	0	$(Hex)_5 + (Man)_1(GlcNAc)_2$		0.08	0.10	0.10	0.03	0.06	0.10	0.07	0.07	13.54	n.d.
40	n.d.	860.3	1721.6	0	$(Hex)_5 + (Man)_1(GlcNAc)_2$		14.74	15.08	10.40	11.39	16.91	16.73	10.20	12.14	n.d.	8.56
41	n.d.	860.4	1721.8	0.2	$(Hex)_5 + (Man)_3(GlcNAc)_3$		0.58	0.26	0.48	0.33	0.77	0.42	1.88	1.56	n.d.	0.35
42	n.d.	864.4	1729.8	0.183	$(Hex)_2 (HexNAc)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.57	0.55	0.62	0.53	0.91	1.02	2.38	2.18	1.02	2.39
43	n.d.	864.4	1729.8	0.183	$(Hex)_2 (HexNAc)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.11	0.13	0.54	0.72	0.34	0.38	n.d.	0.16
44	n.d.	864.9	1730.8	0.163	$(Hex)_2 (HexNAc)_1 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.47	0.29	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
45	n.d.	872.9	1746.8	0.168	$(Hex)_3 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		2.48	2.87	0.34	0.25	0.75	0.60	0.14	0.33	0.73	0.74
46	n.d.	872.9	1746.8	0.168	$(Hex)_3 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_1(GlcNAc)_2$		0.08	0.08	0.19	0.12	1.71	1.91	n.d.	n.d.	n.d.	n.d.
47	n.d.	872.9	1746.8	0.168	$(Hex)_3 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_1(GlcNAc)_2$		0.13	0.15	0.32	0.40	0.17	0.16	0.25	0.13	n.d.	0.28
48	n.d.	885.4	1771.8	0.136	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.08	0.16	0.09	0.08	0.13	0.09	n.d.	n.d.	0.24	n.d.
49	n.d.	885.4	1771.8	0.136	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.39	0.29	0.21	0.25	0.20	0.12	0.04	0.09	0.47	0.69
50	n.d.	885.4	1771.8	0.136	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.17	0.10	n.d.	n.d.	n.d.	n.d.	0.30	0.24	n.d.	n.d.
51	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.13	0.16	n.d.	0.03	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
52	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.09	0.09	n.d.	0.01	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
53	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.29	0.35	0.68	0.66	0.04	0.01	0.09	0.17	0.43	0.34
54	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		1.97	2.02	2.47	2.11	n.d.	n.d.	0.09			

#	[M–H]⁻	[M–2H]²⁻	Theoretical (Thr) [M–H]⁻calc	Deviation (Thr–Obs) Δ[M–H]⁻	Composition	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
55	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	n.d.	n.d.	n.d.	n.d.	0.44	0.42	n.d.	n.d.	n.d.	n.d.
56	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	0.16	0.08	0.73	0.63	1.05	0.82	0.73	0.64	2.12	3.78
57	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	n.d.	0.09	0.22	0.26	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
58	n.d.	893.4	1787.8	0.141	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	0.03	0.03	0.12	0.10	0.06	0.22	n.d.	n.d.	n.d.	0.27
59	n.d.	901.4	1803.8	0.147	$(Hex)_3 (HexNAc)_2 + (Man)_3(GlcNAc)_2$	1.14	0.74	0.71	0.70	0.16	0.12	n.d.	n.d.	n.d.	n.d.
60	n.d.	901.4	1803.8	0.147	$(Hex)_3 (HexNAc)_2 + (Man)_3(GlcNAc)_2$	0.25	0.28	0.17	0.13	0.03	n.d.	n.d.	n.d.	n.d.	n.d.
61	n.d.	901.4	1803.8	0.147	$(Hex)_4 (HexNAc)_2 + (Man)_3(GlcNAc)_2$	0.72	0.75	0.34	0.32	0.18	0.12	n.d.	n.d.	n.d.	n.d.
62	n.d.	901.4	1803.8	0.147	$(Hex)_4 (HexNAc)_2 + (Man)_3(GlcNAc)_2$	2.23	2.67	0.83	1.12	0.18	0.15	n.d.	n.d.	n.d.	n.d.
63	n.d.	913.9	1828.8	0.115	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	0.96	0.69	0.22	0.17	0.05	0.03	n.d.	n.d.	0.39	0.48
64	n.d.	913.9	1828.8	0.115	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	0.66	0.44	1.26	1.14	0.10	0.09	n.d.	n.d.	n.d.	0.61
65	n.d.	913.9	1828.8	0.115	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	0.98	1.00	3.55	3.20	0.34	0.13	n.d.	n.d.	4.18	0.96
66	n.d.	921.9	1844.8	0.12	$(Hex)_2 (HexNAc)_3 + (Man)_3(GlcNAc)_2$	1.15	0.96	1.50	1.59	0.66	0.45	n.d.	n.d.	0.42	n.d.
67	n.d.	934.3	1869.6	-0.111	$(HexNAc)_4 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	n.d.	n.d.	0.17	0.17	0.02	0.02	n.d.	n.d.	n.d.	n.d.
68	n.d.	934.3	1869.6	-0.111	$(HexNAc)_4 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	n.d.	0.13	0.31	0.35	0.11	0.19	0.08	0.19	n.d.	n.d.
69	n.d.	937.3	1875.6	-0.074	$(Hex)_2 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$	0.18	0.25	0.50	0.51	0.51	0.58	1.15	1.12	0.61	0.60
70	n.d.	937.3	1875.6	-0.074	$(Hex)_2 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$	n.d.	n.d.	0.12	0.08	0.20	0.21	0.28	0.26	n.d.	0.24
71	n.d.	941.4	1883.8	0.147	$(Hex)_6 + (Man)_3(GlcNAc)_1$	n.d.	0.09	0.10	0.09	0.06	0.06	0.12	0.12	n.d.	n.d.
72	n.d.	941.4	1883.8	0.147	$(Hex)_6 + (Man)_3(GlcNAc)_2$	16.41	13.97	23.38	24.68	17.17	18.19	32.62	31.06	12.17	15.98

#	$[M-H]^-$	$[M-2H]^{2-}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
73	n.d.	941.4	1883.8	0.147	$(Hex)_6 + (Man)_3(GlcNAc)_2$		2.70	2.35	2.41	2.10	2.69	2.09	3.67	2.79	2.53	1.68
74	n.d.	945.4	1891.8	0.131	$(Hex)_3 (HexNAc)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		1.07	1.61	1.43	1.16	1.83	2.12	2.88	3.11	1.23	1.75
75	n.d.	945.4	1891.8	0.131	$(Hex)_3 (HexNAc)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.33	0.37	1.68	1.19	0.78	0.66	n.d.	0.21
76	n.d.	945.9	1892.8	0.11	$(Hex)_3 (HexNAc)_1 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.56	0.63	n.d.	n.d.	1.05	0.98	n.d.	n.d.	n.d.	n.d.
77	n.d.	953.8	1908.6	-0.084	$(Hex)_4 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	0.06	0.07	0.06	0.05	n.d.	n.d.
78	n.d.	953.8	1908.6	-0.084	$(Hex)_4 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	0.23	0.14	n.d.	n.d.	n.d.	n.d.
79	n.d.	957.8	1916.6	-0.1	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.07	0.06	n.d.	0.14	0.10	0.14	n.d.	0.40
80	n.d.	957.8	1916.6	-0.1	$(Hex)_1 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.05	0.09	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
81	n.d.	965.9	1932.8	0.104	$(Hex)_2 (HexNAc)_2 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.34	0.30	0.95	1.00	3.38	2.75	2.15	3.20
82	n.d.	965.9	1932.8	0.104	$(Hex)_2 (HexNAc)_2 (NeuAc)_1 + (Man)_3(GlcNAc)_3$		n.d.	n.d.	0.06	0.10	0.72	0.45	0.61	0.65	0.35	0.89
83	n.d.	966.5	1934	0.284	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.30	0.39	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
84	n.d.	966.5	1934	0.284	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.16	n.d.
85	n.d.	966.5	1934	0.284	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	0.76	1.14	n.d.	n.d.	n.d.	n.d.
86	n.d.	974.4	1949.8	0.089	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.70	0.75	0.03	0.02	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
87	n.d.	974.4	1949.8	0.089	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.26	0.24	0.01	0.01	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
88	n.d.	974.4	1949.8	0.089	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.22	0.26	0.41	0.31	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
89	n.d.	974.4	1949.8	0.089	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.51	0.68	0.58	0.47	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
90	n.d.	974.4	1949.8	0.089	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.30	0.32	0.31	0.25	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

#	[M−H]⁻	[M−2H]²⁻	[M−H]⁻ calc	Δ[M−H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
91	n.d.	986.9	1974.8	0.057	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_2 + (Man)_3 (GlcNAc)_2$		0.98	1.30	0.61	0.49	0.10	0.09	n.d.	n.d.	1.23	0.77
92	n.d.	986.9	1974.8	0.057	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_2 + (Man)_3 (GlcNAc)_2$		0.08	n.d.	n.d.	0.21	0.22	0.11	n.d.	n.d.	0.61	0.69
93	n.d.	994.9	1990.8	0.062	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.92	0.62	0.11	0.04	0.03	0.02	n.d.	n.d.	0.15	0.33
94	n.d.	994.9	1990.8	0.062	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		1.52	1.38	3.79	2.89	0.76	0.77	n.d.	0.13	2.52	0.51
95	n.d.	994.9	1990.8	0.062	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.15	0.27	n.d.	n.d.	0.37	0.32	n.d.	n.d.	0.57	n.d.
96	n.d.	994.9	1990.8	0.062	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.05	0.05	0.37	0.38	n.d.	n.d.	0.04	0.05	0.22	0.11
97	n.d.	1002.7	2006.4	-0.332	$(Hex)_3 (HexNAc)_3 + (Man)_3 (GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.09	0.14	0.72	n.d.
98	n.d.	1006.4	2013.8	0.127	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_2 (Sulph)_1 + (Man)_3 (GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
99	n.d.	1018.4	2037.8	0.073	$(Hex)_3 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		0.18	0.20	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.00
100	n.d.	1018.4	2037.8	0.073	$(Hex)_3 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		0.78	0.68	0.29	0.25	0.41	0.46	0.67	0.44	0.28	n.d.
101	n.d.	1022.4	2045.8	0.094	$(Hex)_7 + (Man)_3 (GlcNAc)_2$		1.74	1.35	0.82	1.42	1.91	1.41	4.31	3.02	0.43	1.10
102	n.d.	1026.4	2053.8	0.078	$(Hex)_4 (HexNAc)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		0.12	n.d.	n.d.	0.09	0.29	0.17	0.39	0.38	n.d.	n.d.
103	n.d.	1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		0.11	0.06	0.02	n.d.	0.11	0.14	0.61	0.43	0.29	0.68
104	n.d.	1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		n.d.	n.d.	0.05	0.06	0.27	0.25	n.d.	n.d.	n.d.	n.d.
105	n.d.	1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		0.10	0.07	0.47	0.61	1.17	0.86	3.19	3.78	1.27	2.37
106	n.d.	1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		n.d.	n.d.	0.05	0.06	0.17	0.18	0.23	0.10	n.d.	n.d.
107	n.d.	1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		n.d.	n.d.	0.19	0.11	0.47	0.37	0.48	0.42	0.88	1.76
108	n.d.	1039.5	2080	0.226	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_3 + (Man)_3 (GlcNAc)_2$		0.47	0.43	0.20	0.13	0.74	0.69	n.d.	n.d.	1.46	1.29

#	[M − H]⁻	[M − 2H]²⁻	[M − H]⁻ calc	Δ[M − H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
109	n.d.	1039.5	2080	0.226	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.17	0.13	0.67	0.58	n.d.	n.d.	n.d.	0.61
110	n.d.	1039.5	2080	0.226	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.15	0.14	0.72	0.72	1.61	1.62	n.d.	n.d.	n.d.	2.17
111	n.d.	1039.5	2080	0.226	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.07	0.05	0.07	0.04	0.83	0.69	n.d.	n.d.	n.d.	2.15
112	n.d.	1047.4	2095.8	0.031	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.62	0.65	0.08	0.18	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
113	n.d.	1055.5	2112	0.236	$(Hex)_4 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.28	0.38	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
114	n.d.	1055.5	2112	0.236	$(Hex)_4 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.07	n.d.	0.11	0.13	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
115	n.d.	1055.5	2112	0.236	$(Hex)_4 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.09	0.07	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
116	n.d.	1059.4	2119.8	0.019	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.39	0.34	n.d.	n.d.	n.d.	n.d.	n.d.	0.46
117	n.d.	1059.4	2119.8	0.019	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.20	0.11	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
118	n.d.	1059.4	2119.8	0.019	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.08	0.16	n.d.	n.d.
119	n.d.	1060	2121	0.199	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$	T.b.c.	n.d.	0.29	n.d.	n.d.	0.29	n.d.	0.27	0.24	1.51	n.d.
120	n.d.	1068	2137	0.204	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.74	1.03	0.01	0.02	0.01	0.01	n.d.	n.d.	0.16	0.28
121	n.d.	1068	2137	0.204	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		2.67	2.24	0.74	0.46	0.18	0.12	n.d.	n.d.	1.19	0.45
122	n.d.	1068	2137	0.204	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.39	0.20	0.72	0.66	0.84	0.63	n.d.	n.d.	n.d.	n.d.
123	n.d.	1075.9	2152.8	0.009	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.32	0.25	0.08	0.08	0.36	0.22	0.42	0.38
124	n.d.	1075.9	2152.8	0.009	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.08	0.05	0.07	0.10	n.d.	n.d.	n.d.	0.25
125	n.d.	1075.9	2152.8	0.009	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.06	0.05	n.d.	0.12	n.d.	n.d.	n.d.	0.19
126	n.d.	1080.6	2162.2	0.372	$(HexNAc)_4 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.14	0.24	0.12	0.14	0.31	0.37	0.44	0.45	2.55	0.43

#	[M − H]⁻	[M − 2H]²⁻	[M − H]⁻ calc	Δ[M − H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
127	n.d.	1096.5	2194	0.183	$(Hex)_2(HexNAc)_4(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.48	0.43	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
128	n.d.	1096.5	2194	0.183	$(Hex)_2(HexNAc)_4(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.68	0.90	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
129	n.d.	1103.4	2207.8	0.041	$(Hex)_8 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.40	0.45	n.d.	n.d.
130	n.d.	1111.5	2224	0.209	$(Hex)_2(HexNAc)_2(NeuAc)_2 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.07	0.09	0.49	0.61	2.40	1.14	3.25	2.33
131	n.d.	1111.5	2224	0.209	$(Hex)_2(HexNAc)_2(NeuAc)_2 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.01	0.02	0.41	0.41	0.15	0.12	n.d.	n.d.
132	n.d.	1111.9	2224.8	-0.011	$(Hex)_2(HexNAc)_2(Deoxyhexose)_2(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.21	0.15	0.09	0.17	0.35	0.31	1.39	1.34	1.52	1.85
133	n.d.	1111.9	2224.8	-0.011	$(Hex)_2(HexNAc)_2(Deoxyhexose)_2(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.01	0.02	0.03	0.02	0.53	0.49	0.13	0.12	n.d.	n.d.
134	n.d.	1111.9	2224.8	-0.011	$(Hex)_2(HexNAc)_2(Deoxyhexose)_2(NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	0.01	0.09	0.10	0.58	0.59	0.06	0.04	n.d.	n.d.
135	n.d.	1111.9	2224.8	-0.011	$(Hex)_2(HexNAc)_2(Deoxyhexose)_2(NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	0.00	0.02	0.01	0.07	0.03	n.d.	n.d.	n.d.	n.d.
136	n.d.	1111.9	2224.8	-0.011	$(Hex)_2(HexNAc)_2(Deoxyhexose)_2(NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	0.00	0.00	0.01	0.11	0.12	n.d.	n.d.	n.d.	n.d.
137	n.d.	1132.4	2265.8	-0.037	$(Hex)_1(HexNAc)_1(Deoxyhexose)_2(NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	n.d.	n.d.	0.20	0.20	0.48	0.57	0.45	n.d.
138	n.d.	1140.4	2281.8	-0.032	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.33	0.27	0.81	0.77	0.16	0.09	n.d.	n.d.	0.63	0.32
139	n.d.	1140.4	2281.8	-0.032	$(Hex)_2(HexNAc)_3(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.14	0.12	0.03	0.03	0.02	n.d.	n.d.	n.d.	n.d.	0.16
140	n.d.	1140.4	2281.8	-0.032	$(Hex)_2(HexNAc)_3(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.33	0.24	0.81	0.77	0.16	0.09	n.d.	n.d.	0.41	0.33
141	n.d.	1140.4	2281.8	-0.032	$(Hex)_2(HexNAc)_3(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.31	0.27	0.24	0.33	n.d.	n.d.	n.d.	n.d.
142	n.d.	1140.5	2282	0.167	$(Hex)_2(HexNAc)_3(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		n.d.	n.d.	0.31	0.27	0.24	0.33	n.d.	n.d.	n.d.	n.d.
143	n.d.	1141	2283	0.146	$(Hex)_2(HexNAc)_3(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		2.31	3.20	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.80	0.41
144	n.d.	1141	2283	0.146	$(Hex)_2(HexNAc)_3(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		1.47	1.31	n.d.	n.d.	0.56	0.32	n.d.	n.d.	0.78	0.81

#	[M − H]⁻	[M − 2H]²⁻	[M − H]⁻ calc	Δ[M − H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
145	n.d.	1141	2283	0.146	(Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_3$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	0.17	0.23	n.d.	n.d.	0.71	n.d.
146	n.d.	1148.4	2297.8	-0.027	(Hex)$_2$ (HexNAc)$_3$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.25	0.23	n.d.	n.d.
147	n.d.	1148.4	2297.8	-0.027	(Hex)$_3$ (HexNAc)$_3$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.07	0.07	n.d.	n.d.
148	n.d.	1149.1	2299.2	0.351	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	0.06	n.d.	0.29	0.15	n.d.	n.d.	n.d.	n.d.
149	n.d.	1149.1	2299.2	0.351	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	0.16	0.16	n.d.	n.d.	n.d.	n.d.
150	n.d.	1149.1	2299.2	0.351	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$	structure + 2x	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.07	n.d.	n.d.	n.d.
151	n.d.	1152.9	2306.8	-0.064	(HexNAc)$_4$ (Deoxyhexose)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.13	0.23	n.d.	n.d.
152	n.d.	1177.5	2356	0.13	(Hex)$_3$ (HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure +	0.19	n.d.	0.74	0.64	n.d.	n.d.	n.d.	n.d.	0.12	n.d.
153	n.d.	1184.5	2370	0.151	(Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	0.08	0.10	0.23	0.17	1.20	1.30	1.12	0.62
154	n.d.	1184.5	2370	0.151	(Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	0.16	0.09	0.44	0.44	1.67	1.94	0.26	n.d.
155	n.d.	1184.5	2370	0.151	(Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_2$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	0.39	0.42	0.21	0.23	0.53	0.40
156	n.d.	1205.6	2412.2	0.304	(Hex)$_1$ (HexNAc)$_3$ (Deoxyhexose)$_3$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	T.b.c.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.17	0.23	n.d.	n.d.
157	n.d.	1205.6	2412.2	0.304	(Hex)$_1$ (HexNAc)$_3$ (Deoxyhexose)$_3$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	T.b.c.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
158	n.d.	1213.6	2428.2	0.309	(Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure +	0.97	1.28	0.32	0.29	n.d.	n.d.	n.d.	n.d.	1.44	n.d.
159	n.d.	1221.5	2444	0.114	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	0.13	0.12	n.d.	n.d.	0.82	1.02	0.96	n.d.
160	n.d.	1221.5	2444	0.114	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	0.04	0.04	n.d.	n.d.	0.22	0.10	0.26	n.d.
161	n.d.	1221.5	2444	0.114	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	0.09	n.d.	0.28	0.25	n.d.	n.d.
162	n.d.	1222	2445	0.093	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_3$ + (Man)$_3$(GlcNAc)$_2$	structure	n.d.	n.d.	n.d.	n.d.	0.12	0.12	n.d.	n.d.	n.d.	n.d.

Table (numbered rows 163–179). Columns: Observed ions (Obs) — [M–H]⁻, [M–2H]²⁻; Theoretical (Thr) — [M–H]⁻calc; Deviation (Thr–Obs) — Δ[M–H]⁻; Composition; Probable Structure; and percentage columns for SW1116 A, SW1116 B, SW620A, SW620B, SW837A, SW837B, SW480A, SW480B, Tumour 1, Tumour 2.

#	[M–H]⁻	[M–2H]²⁻	[M–H]⁻calc	Δ[M–H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620B %	SW837A %	SW837B %	SW480A %	SW480B %	Tumour 1 %	Tumour 2 %
163	n.d.	1222	2445	0.093	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_3 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	0.16	0.14	n.d.	n.d.	0.30	n.d.
164	n.d.	1225	2451	0.173	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_3 (NeuAc)_1 (Sulph)_1 + (Man)_3 (GlcNAc)_2$	T.b.c.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.60	n.d.
165	n.d.	1257.5	2516	0.093	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_2 (NeuAc)_2 + (Man)_3 (GlcNAc)_2$	T.b.c.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.39	n.d.
166	n.d.	1258.5	2518	0.077	$(Hex)_1 (HexNAc)_4 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	0.10	0.13	0.07	0.14	0.10	0.15	n.d.	n.d.
167	n.d.	1258.5	2518	0.077	$(Hex)_4 (HexNAc)_4 (Deoxyhexose)_4 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	0.05	0.03	n.d.	n.d.	0.24	n.d.
168	n.d.	1286	2573	0.071	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	0.04	n.d.	0.17	0.17	n.d.	n.d.	n.d.	n.d.
169	n.d.	1294.1	2589.2	0.276	$(Hex)_3 (HexNAc)_3 (NeuAc)_2 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.11	0.13	0.52	n.d.
170	n.d.	1330.6	2662.2	0.235	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_3 (NeuAc)_2 + (Man)_3 (GlcNAc)_2$	(structure, 3x)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
171	n.d.	1331	2663	0.04	$(Hex)_4 (HexNAc)_4 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.12	0.08	n.d.	n.d.
172	n.d.	1331	2663	0.04	$(Hex)_4 (HexNAc)_4 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.03	0.01	n.d.	n.d.
173	n.d.	1367	2735	0.018	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_2 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.14	0.18	n.d.	n.d.
174	n.d.	1367	2735	0.018	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_2 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.33	0.34	n.d.	n.d.
175	n.d.	1404	2809	-0.017	$(Hex)_4 (HexNAc)_4 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	0.02	0.05	n.d.	n.d.	0.21	0.14	n.d.	n.d.
176	n.d.	1404	2809	-0.017	$(Hex)_4 (HexNAc)_4 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	0.04	0.05	n.d.	n.d.	0.11	0.08	n.d.	n.d.
177	n.d.	1440	2881	-0.038	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_2 (NeuAc)_2 + (Man)_3 (GlcNAc)_2$	T.b.c.	n.d.	n.d.	n.d.	n.d.	0.08	0.09	0.12	0.07	1.46	n.d.
178	n.d.	1512.6	3026.2	0.123	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_3 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.21	0.17	n.d.	n.d.
179	n.d.	1549.7	3100.4	0.286	$(Hex)_4 (HexNAc)_4 (Deoxyhexose)_1 (NeuAc)_2 + (Man)_3 (GlcNAc)_2$	(structure)	n.d.	n.d.	n.d.	0.01	n.d.	n.d.	0.49	0.42	n.d.	n.d.

Supplementary Table 4 (over page). *O*-glycan characterisation of cell lines and epithelial-enriched tumour tissues with their normalised areas under the curve. All observed ions have been reduced. Replicates of cell lines denoted with letters, e.g. SW1116A (replicate 1), SW1116B (replicate 2). Structural elucidation was based annotation of spectral fragments using common knowledge of biosynthetic pathways. Structural diagrams not intended to specific linkage positions for example, there is no distinction between the three-branch and six-branch of core 4. The presence of additional structural isomers cannot be excluded. n.d., not detected.

#	[M−H]⁻	[M−2H]²⁻	[M−H]⁻ calc	Δ[M−H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620 B %	SW837A %	SW837 B %	SW480A %	SW480 B %	LS174T mucin A %	LS174T mucin B %	Tumour 1 %	Tumour 2 %
	Observed ions (Obs)		Theoretical (Thr)	Deviation (Obs-Thr)														
1	513.3	n.d.	513.3	0.106	(HexNAc)$_1$ (NeuAc)$_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.00	4.69	7.57	16.75
2	587.3	n.d.	587.3	0.069	(Hex)$_1$ (HexNAc)$_2$		0.39	0.88	0.60	n.d.	1.07	0.65	n.d.	n.d.	n.d.	2.33	0.31	3.30
3	587.3	n.d.	587.3	0.069	(Hex)$_1$ (HexNAc)$_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.10	1.04
4	675.3	n.d.	675.3	0.053	(Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_1$		1.92	2.19	2.54	n.d.	2.05	2.07	0.03	1.02	2.48	6.68	n.d.	10.00
5	675.3	n.d.	675.3	0.053	(Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_1$		11.96	10.92	22.66	29.19	9.19	12.72	25.70	21.92	4.65	8.25	7.50	13.58
6	716.2	n.d.	716.2	-0.072	(HexNAc)$_2$ (NeuAc)$_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	21.86	12.87
7	716.2	n.d.	716.2	-0.072	(HexNAc)$_2$ (NeuAc)$_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.25	2.81	4.67	3.86
8	716.2	n.d.	716.2	-0.072	(HexNAc)$_2$ (NeuAc)$_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.31	n.d.
9	749.2	n.d.	749.2	-0.082	(Hex)$_2$ (HexNAc)$_2$		1.48	1.56	6.95	4.73	5.56	6.02	0.60	0.58	2.43	12.93	n.d.	3.60
10	829.3	n.d.	829.3	0.06	(Hex)$_2$ (HexNAc)$_2$ (Sulph)$_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.35	n.d.	0.33	n.d.
11	878.4	n.d.	878.4	0.074	(Hex)$_1$ (HexNAc)$_2$ (NeuAc)$_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.96	n.d.
12	878.4	n.d.	878.4	0.074	(Hex)$_1$ (HexNAc)$_2$ (NeuAc)$_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.72	5.64	3.28	n.d.
13	895.4	n.d.	895.4	0.059	(Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$		1.80	2.42	n.d.	n.d.	2.61	1.92	n.d.	n.d.	0.55	3.47	n.d.	n.d.
14	n.d.	913	1827	0.331	(Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_2$ (NeuAc)$_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.19	n.d.	1.21	n.d.
15	966.4	n.d.	966.4	0.058	(Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_2$		51.60	51.92	50.91	48.24	32.87	28.77	55.36	58.05	1.19	2.36	0.54	14.84

#	Observed ions (Obs) [M – H]⁻	[M – 2H]²⁻	Theoretical (Thr) [M – H]⁻calc	Deviation (Obs-Thr) Δ[M – H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620 B %	SW837A %	SW837 B %	SW480A %	SW480 B %	LS174T mucin A %	LS174T mucin B %	Tumour 1 %	Tumour 2 %
16	966.4	n.d.	966.4	0.058	$(Hex)_1 (HexNAc)_1 (NeuAc)_2$		0.76	1.67	0.45	n.d.	0.63	n.d.	n.d.	0.67	n.d.	n.d.	n.d.	n.d.
17	975.4	n.d.	975.4	0.102	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (Sulph)_1$		4.38	4.67	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.47	1.86	7.50	3.21
18	975.4	n.d.	975.4	0.102	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (Sulph)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.87
19	975.4	n.d.	975.4	0.102	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (Sulph)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.06	n.d.
20	1016.4	n.d.	1016.4	0.075	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_1 (Sulph)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.21	n.d.
21	1024.5	n.d.	1024.5	0.116	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.30	1.56
22	1024.5	n.d.	1024.5	0.116	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	3.58	n.d.
23	1040.5	519.7	1040.5	0.121	$(Hex)_2 (HexNAc)_2 (NeuAc)_1$		4.21	4.36	4.31	3.30	3.23	3.65	2.40	1.83	14.51	21.55	2.79	2.43
24	1040.5	519.7	1040.5	0.121	$(Hex)_2 (HexNAc)_2 (NeuAc)_1$		0.71	n.d.	4.21	5.67	19.16	21.27	1.82	0.98	3.81	n.d.	n.d.	1.88
25	1040.5	519.7	1040.5	0.121	$(Hex)_2 (HexNAc)_2 (NeuAc)_1$		1.60	1.63	6.80	6.80	2.69	2.98	0.44	0.69	3.33	n.d.	n.d.	n.d.
26	1081.5	n.d.	1081.5	0.095	$(Hex)_1 (HexNAc)_3 (NeuAc)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.46	n.d.
27	1081.5	n.d.	1081.5	0.095	$(Hex)_1 (HexNAc)_3 (NeuAc)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.62	n.d.
28	1081.5	n.d.	1081.5	0.095	$(Hex)_1 (HexNAc)_3 (NeuAc)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	0.57	n.d.	n.d.	n.d.	n.d.	2.10	n.d.
29	n.d.	551.8	1104.6	0.259	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 (Sulph)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.53	n.d.
30	1114.5	n.d.	1114.5	0.084	$(Hex)_3 (HexNAc)_3$		0.39	0.46	0.57	0.38	n.d.	n.d.	n.d.	n.d.	4.28	3.91	n.d.	n.d.

#	Observed [M−H]⁻	[M−2H]²⁻	Theoretical [M−H]⁻calc	Deviation Δ[M−H]⁻	Composition	Probable Structure	SW1116 A %	SW1116 B %	SW620A %	SW620 B %	SW837A %	SW837 B %	SW480A %	SW480 B %	LS174T mucin A %	LS174T mucin B %	Tumour 1 %	Tumour 2 %
31	1120.5	n.d.	1120.5	0.164	$(Hex)_2(HexNAc)_2(NeuAc)_1(Sulph)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.24	n.d.
32	1169.6	584.3	1169.6	0.179	$(Hex)_1(HexNAc)_2(NeuAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.15	n.d.	12.73	4.60	1.55
33	1186.4	n.d.	1186.4	−0.036	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_1$		1.55	0.92	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	10.83	n.d.	1.64	1.55
34	1186.4	n.d.	1186.4	−0.036	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_1$		2.06	2.99	n.d.	0.36	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.20	1.43
35	1227.7	n.d.	1227.7	0.237	$(Hex)_1(HexNAc)_3(Deoxyhexose)_1(NeuAc)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.93	1.12	n.d.
36	1260.5	n.d.	1260.5	0.026	$(Hex)_3(HexNAc)_3(Deoxyhexose)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.88	0.68	n.d.	n.d.
37	1266.4	632.7	1266.4	0.006	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_1(Sulph)_1$		4.67	4.29	n.d.	n.d.	19.77	18.44	13.65	14.12	0.49	n.d.	1.12	1.68
38	1266.4	632.7	1266.4	0.006	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_1(Sulph)_1$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.51	n.d.	0.53	n.d.
39	1315.7	657.4	1315.7	0.221	$(Hex)_1(HexNAc)_2(Deoxyhexose)_1(NeuAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.62	8.64	2.99
40	1331.6	665.3	1331.6	0.126	$(Hex)_2(HexNAc)_2(NeuAc)_2$		1.56	1.10	n.d.	1.34	n.d.	n.d.	n.d.	n.d.	6.05	n.d.	n.d.	n.d.
41	1372.6	685.9	1372.6	0.099	$(Hex)_1(HexNAc)_3(NeuAc)_2$		n.d.	n.d.	n.d.	n.d.	1.18	0.93	n.d.	n.d.	n.d.	3.56	4.26	n.d.
42	1477.6	738.3	1477.6	0.068	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_2$		8.95	8.04	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	36.22	n.d.	n.d.	n.d.
43	n.d.	840	1681	0.388	$(Hex)_2(HexNAc)_3(Deoxyhexose)_1(NeuAc)_2$		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.85	n.d.

Supplementary Table 5. Primers used for qRT-PCR (in Chapter 2).

Gene	Orientation	Sequence (5' - 3')
C1GALT1	Forward	AAGCAGGGCTACATGAGTGG
	Reverse	GCATCTCCCCAGTGCTAAGT
ST6GALNAC1	Forward	CAGAGGCACAATCATGGAAG
	Reverse	GCTGACTTTTGGGAATGAGC
MGAT3	Forward	CGGGACCTGAACTACATCCG
	Reverse	GTCCAGCAGGTAGTGGAACC
MUC1	Forward	TTTCCAGCCCGGGATACCTA
	Reverse	TAGGGGCTACGATCGGTACT
MUC2	Forward	GTTCATCAAGCTGGCTCCCT
	Reverse	TTCCAGCTGTTCCCGAAGTC
MUC3	Forward	TGGCTGAGCAACAACTCTGT
	Reverse	GGGAGGCTACTGCTTGGTTT
MUC4	Forward	TCACCTCAACAGGCTCAACA
	Reverse	GTCATCATCTGCGTGAGGGT
MUC5B	Forward	GGCTGTTTCAGCACACACTG
	Reverse	CTTGTAGGTGGCCTCGTTGT
MUC6	Forward	TGACAGCCACCTCATGGTTC
	Reverse	AGCAAACTTGTGGGGTTCCA
MUC13	Forward	TAATCACCGCTTCATCTCCA
	Reverse	TGTTTAGGGTGCTGGTCTCC
β-actin	Forward	TCATGAAGTGTGACGTGGACATC
	Reverse	CAGGAGGAGCAATGATCTTGATCT

Supplementary Table 6. mRNA expression of glycosyltransferase and mucin genes assayed by qRT-PCR. Data are shown as mean (2-Δ^{CT}) ± SD of a minimum of five data points from two triplicate PCRs.

Genes	SW1116	SW480	SW620	SW837	LS174T
MUC1	4.7E-03 ± 3.3E-04	4.3E-04 ± 3.1E-04	1.7E-02 ± 8.7E-03	1.3E-02 ± 4.2E-03	1.5E-03 ± 6.1E-04
MUC2	5.4E-05 ± 1.5E-05	1.2E-05 ± 1.2E-05	3.2E-04 ± 3.0E-04	4.6E-05 ± 2.5E-05	6.4E-02 ± 2.1E-02
MUC3A	1.3E-03 ± 7.0E-04	7.6E-06 ± 1.2E-05	2.6E-05 ± 1.1E-05	4.7E-05 ± 6.1E-05	1.3E-04 ± 7.4E-05
MUC4	3.0E-04 ± 1.5E-04	1.8E-05 ± 1.8E-05	2.5E-04 ± 2.8E-04	3.7E-03 ± 1.7E-03	2.7E-04 ± 1.1E-04
MUC5B	1.3E-06 ± 8.5E-07	4.7E-07 ± 1.7E-07	1.5E-06 ± 1.5E-06	3.0E-06 ± 1.9E-06	4.6E-04 ± 1.7E-04
MUC6	1.8E-05 ± 1.6E-05	3.0E-05 ± 2.8E-05	2.4E-04 ± 8.1E-05	8.1E-05 ± 5.5E-05	1.1E-01 ± 4.4E-02
MUC13	4.8E-01 ± 6.5E-02	6.1E-04 ± 4.0E-04	1.6E-02 ± 1.0E-02	5.0E-01 ± 7.6E-02	7.4E-02 ± 2.1E-02
ST6GALNAC1	1.8E-03 ± 8.6E-04	1.6E-05 ± 9.6E-06	1.2E-04 ± 1.2E-04	2.1E-04 ± 5.1E-05	3.0E-02 ± 8.2E-03
C1GALT1	6.5E-02 ± 2.1E-02	1.4E-02 ± 7.6E-03	8.0E-02 ± 2.6E-02	4.5E-02 ± 1.0E-02	7.7E-03 ± 1.6E-03
MGAT3	2.5E-03 ± 4.9E-04	7.8E-06 ± 9.0E-06	1.4E-03 ± 4.4E-04	2.0E-04 ± 7.3E-05	1.0E-05 ± 3.9E-06

Supplementary Figure 1. Average MS of *N*-glycan profiles of EpCAM-positive fractions and EpCAM-negative fractions of two independent enrichments for epithelial cells from colorectal tumours.
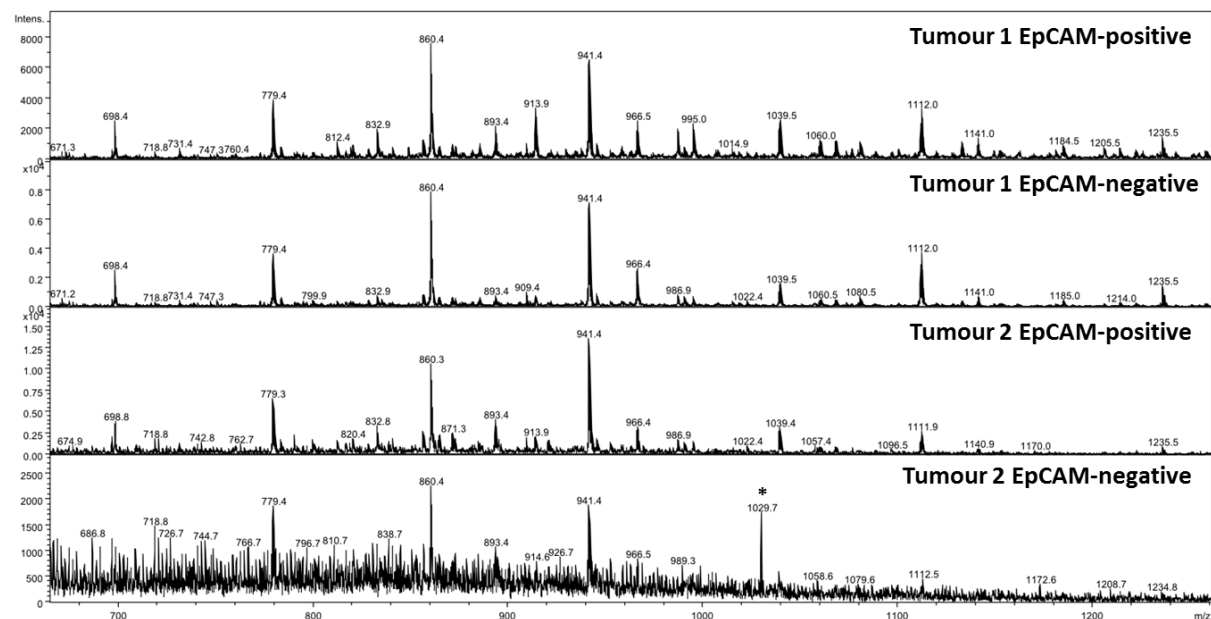
Supplementary Figure 2. Confirmation of α2,6-sialic acid linkage of sialyl-Tn (*m/z* 513.2) using exoglycosidase digestion of pooled glycan sample. A pooled glycan sample was divided evenly into three aliquots. One aliquot contained buffer only (control) while the others were treated with either ABS or NAN1 exoglycosidase enzymes. ABS is a sialidase which cleaves α2,3/6 linked sialic acids. NAN1 is a sialidase which cleaves α2,3-linked sialic acids. The extracted ion chromatogram of *m/z* 513.2 shown is consistent with an α2,6-linked sialic acid. The 1 min difference in retention time is within usual variability observed during daily operation of the porous graphitised carbon column.
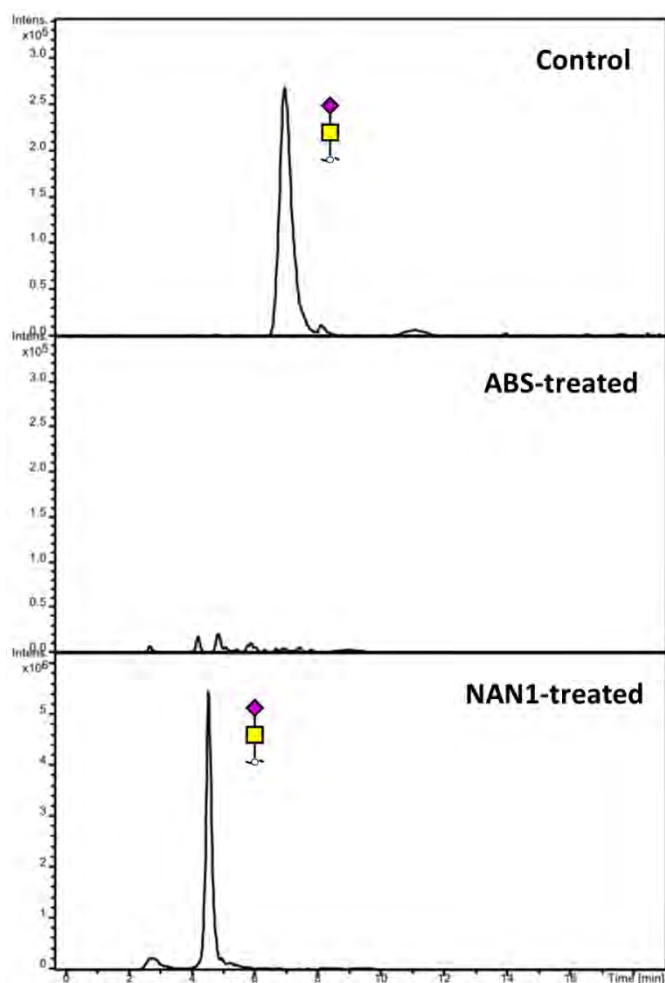
Supplementary Figure 3. Representative average MS of *N*-glycan profiles from cell lines summed across total elution time. Analysis of cell lines was performed in biological duplicate.
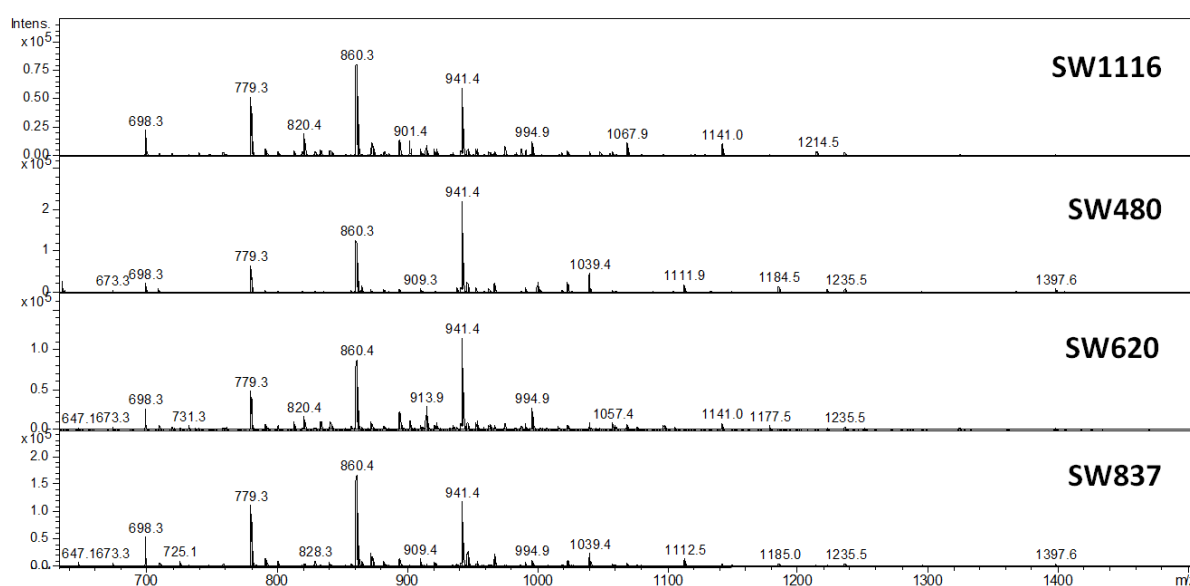
Supplementary Figure 4. Representative extracted ion chromatogram of *m/z* 716.2. Two peaks were clearly observed which indicated the presence of two major isomers. There was also a difference in elution time of approximately two mins, which is comparable with a previous study [122].
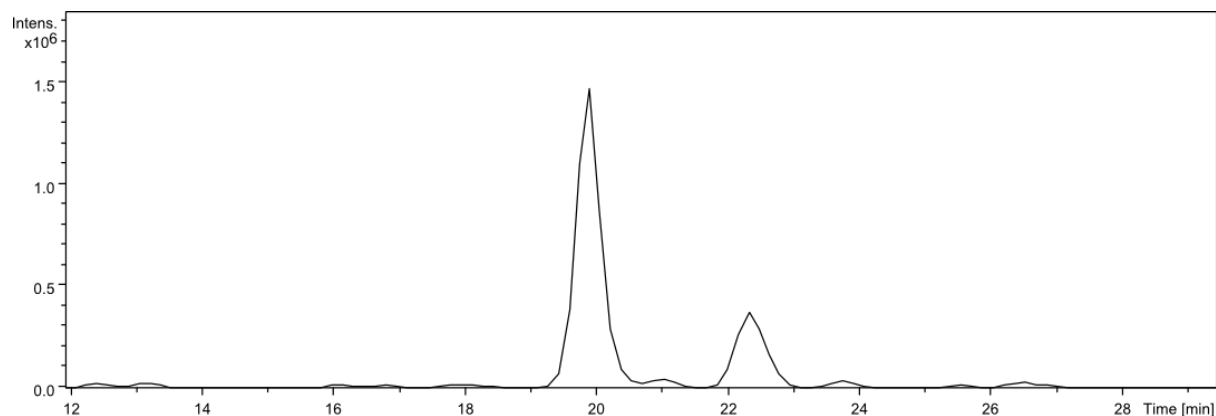
Supplementary Figure 5. Annotated MS/MS for *N*-glycans. *N*-glycans have been numbered according to Supplementary Table 3. The 6 structures (number 119, 156, 157, 164, 165, 177) not confirmed by MS/MS are not shown. Glycan schemes were derived from GlycoWorkbench. Annotation was based on the presence of structural feature ions and common knowledge of known glycan synthetic pathways. Additional structural isomers cannot be excluded. Structural schemes not intended to show specific linkages (linkages not shown by linkage angles). Where ambiguity is indicated in parent structure, for example in the assignment of fucose, depiction of specific linkages and isomers are not intended for corresponding glycan fragments.

Glycan 1

Parent ion: *m/z* 895.5⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁
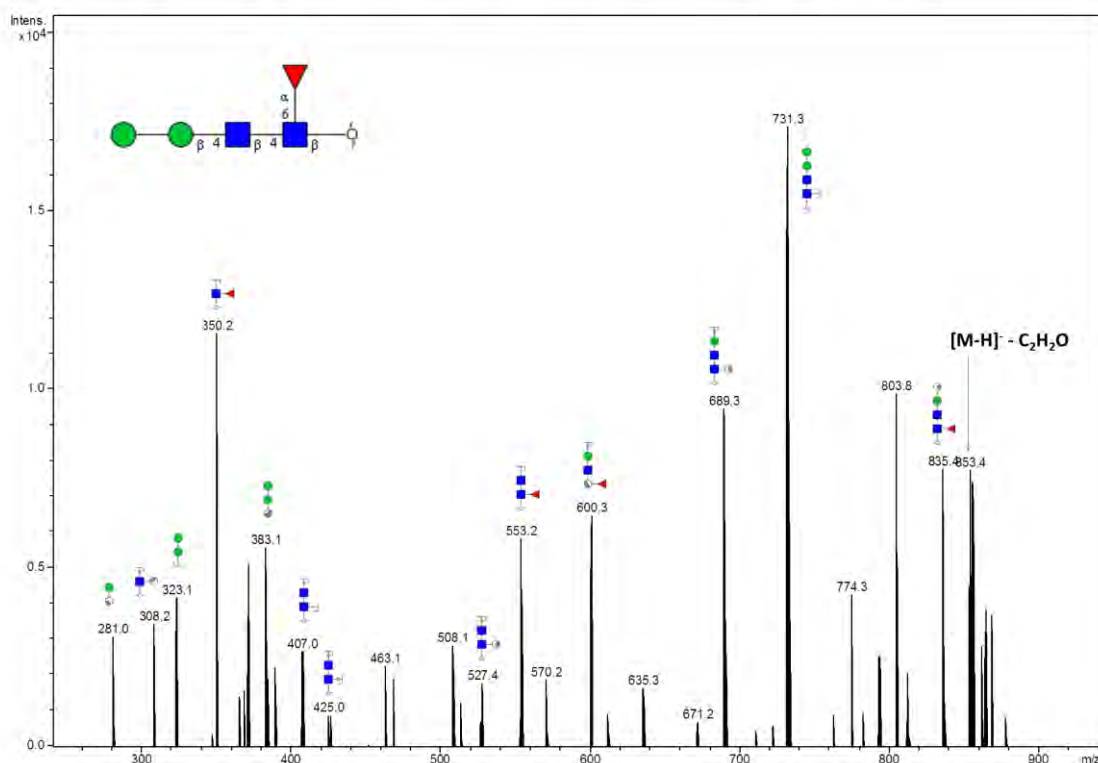
Glycan 2

Parent ion: *m/z* 911.4⁻

Composition: $(Hex)_3 (HexNAc)_2$
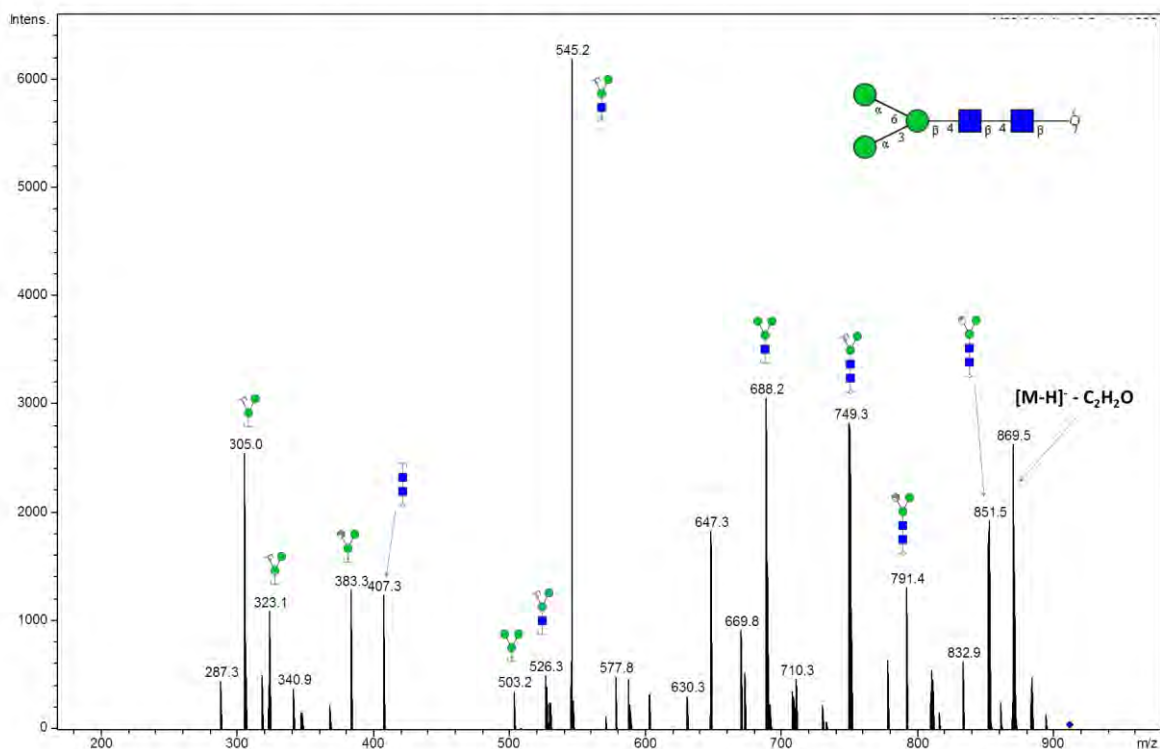
Glycan 3

Parent ion: *m/z* 1057.4⁻

Composition: $(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1$
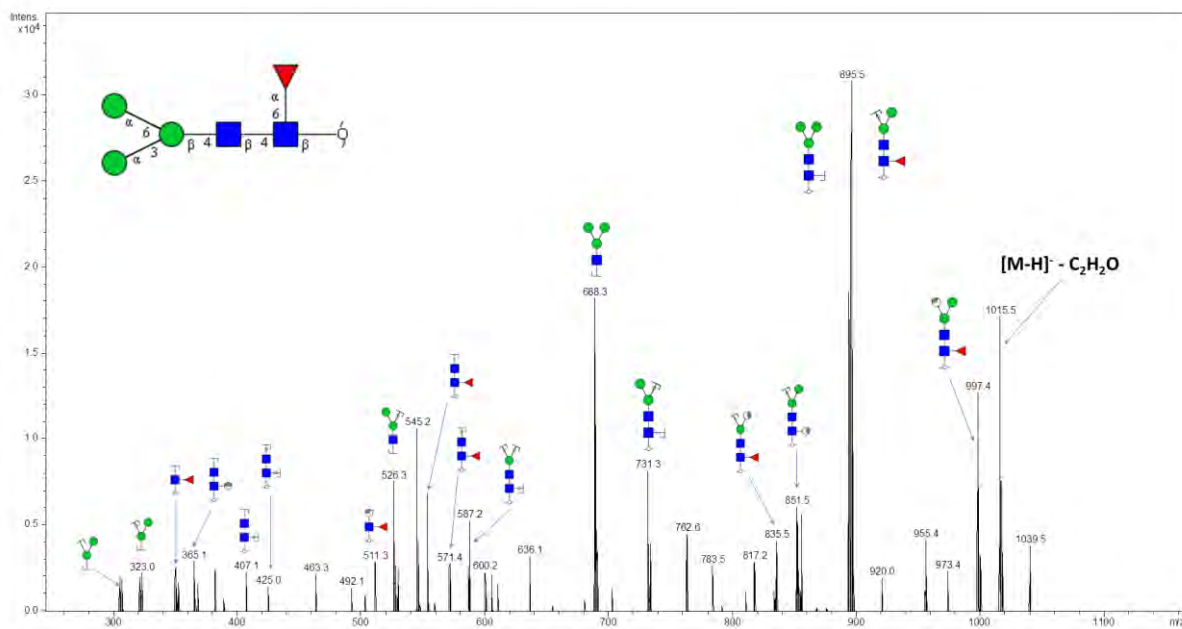
Glycan 4

Parent ion: *m/z* 1235.6⁻

Composition: $(Hex)_2 + (Man)_3(GlcNAc)_2$
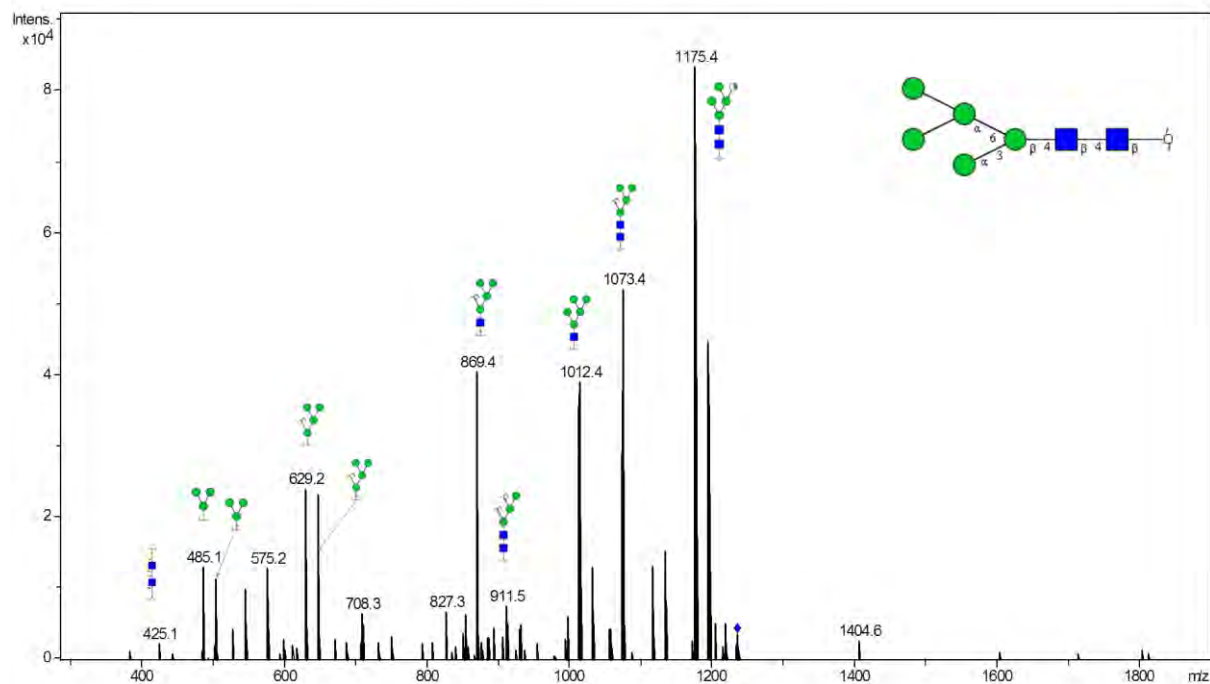
Glycan 5

Parent ion: *m/z* 698.3²⁻

Composition: $(Hex)_3 + (Man)_3(GlcNAc)_2$

Notes: No distinction intended between six-branch and three-branch in fragmentation scheme.
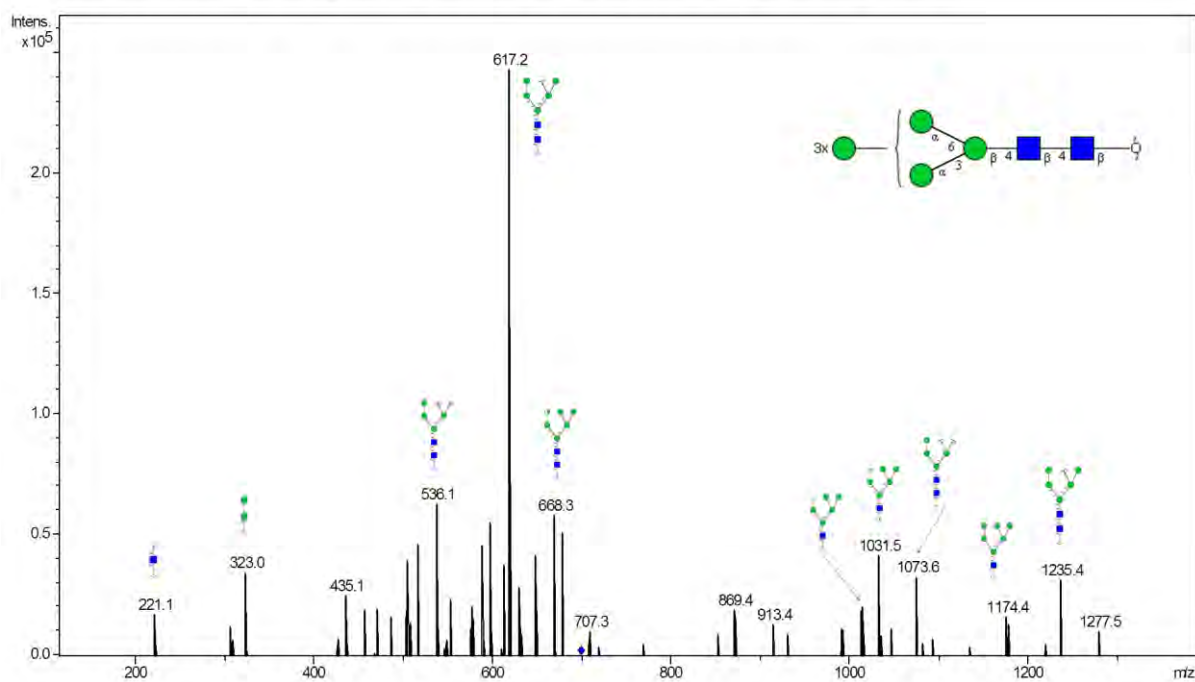
Glycan 6

Parent ion: *m/z* 698.3^{2-}

Composition: (Hex)$_3$ + (Man)$_3$(GlcNAc)$_2$
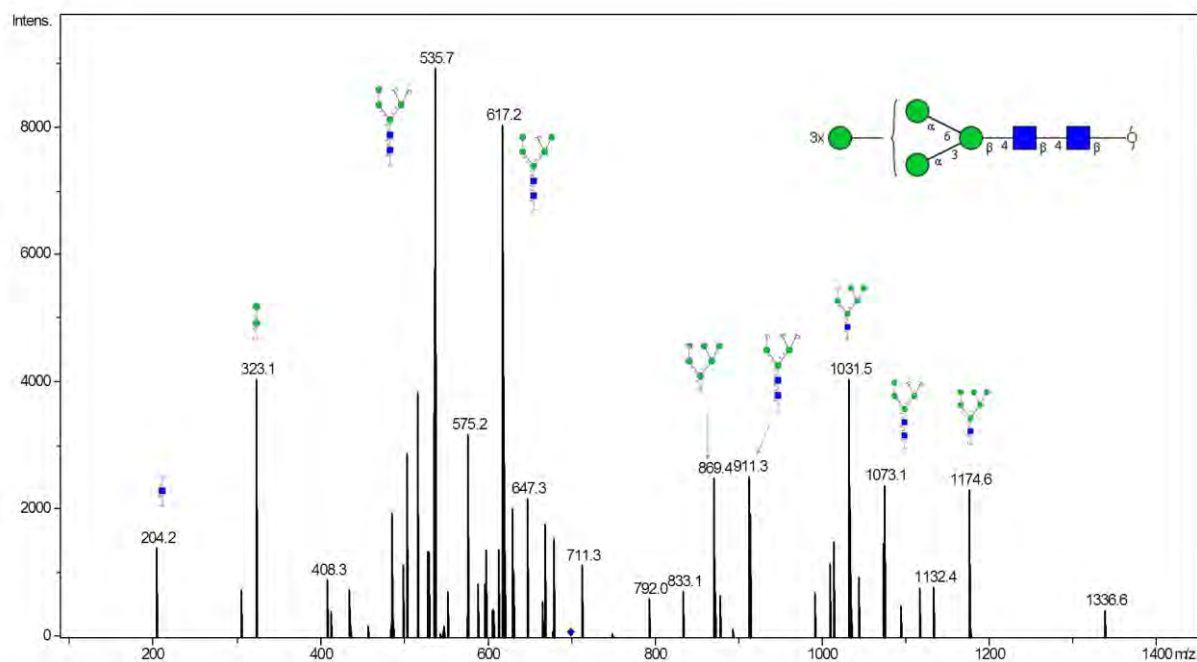
Glycan 7

Parent ion: *m/z* 710.8^{2-}

Composition: (Hex)$_1$ (HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

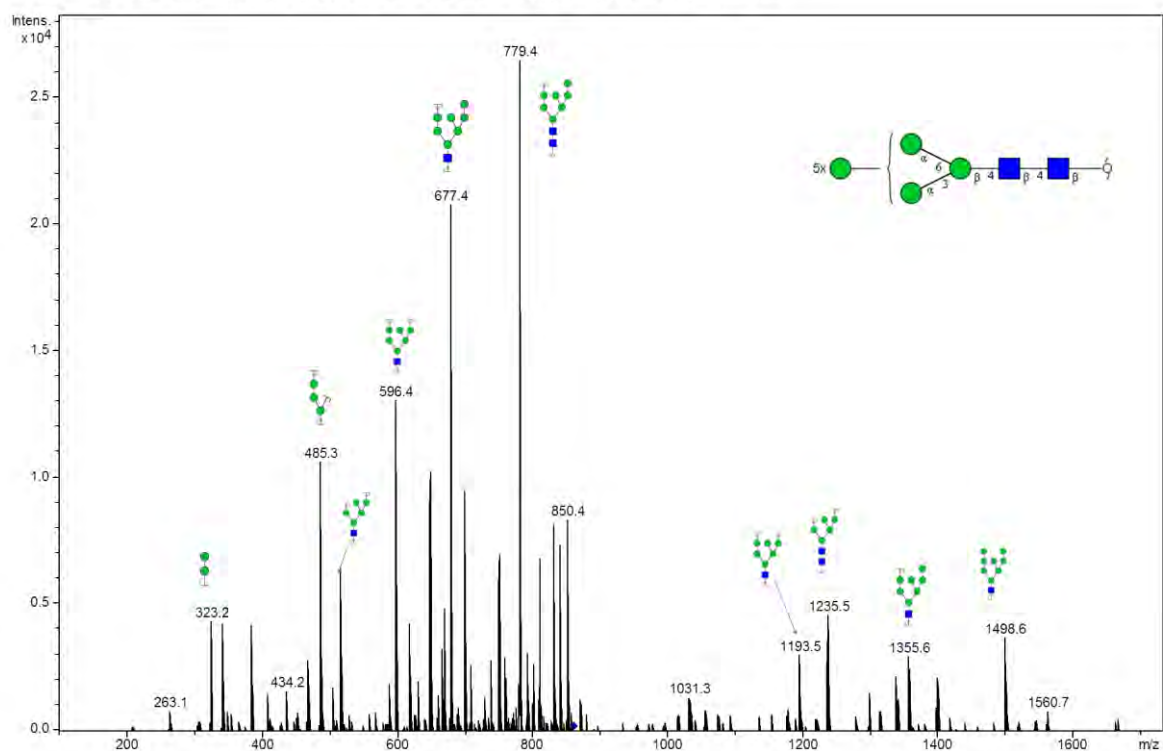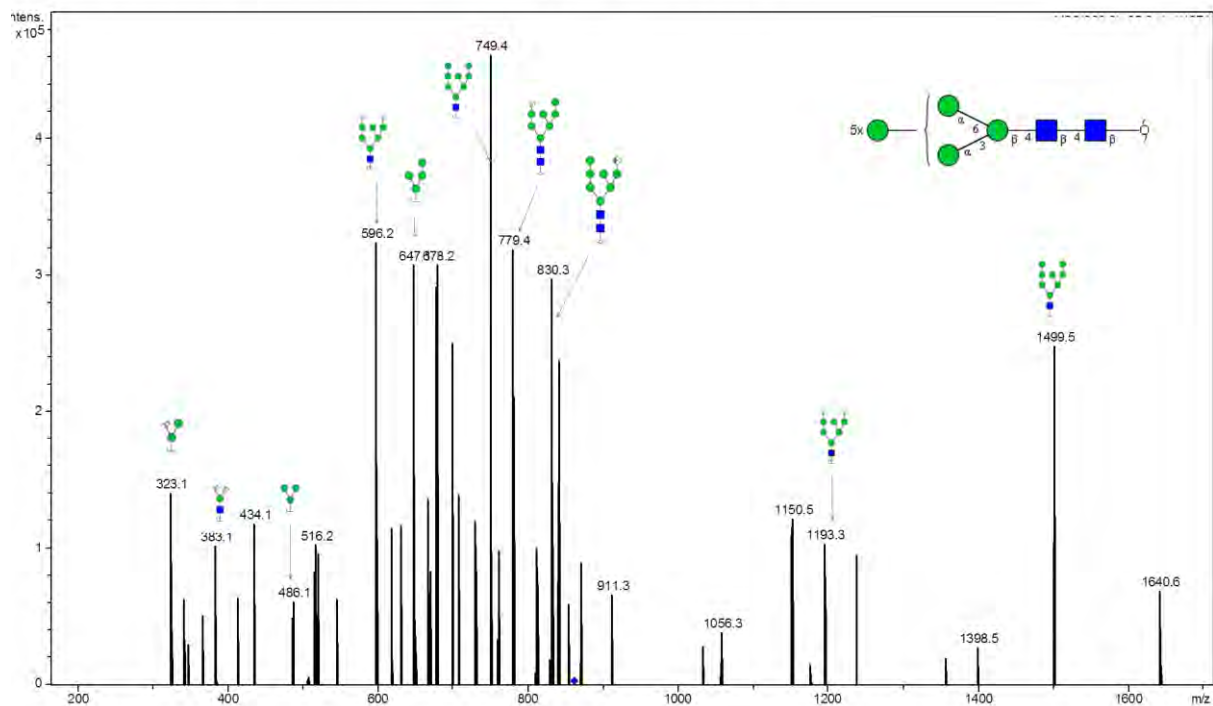Notes: No distinction intended between 3-arm and 6-arm in fragmentation scheme.
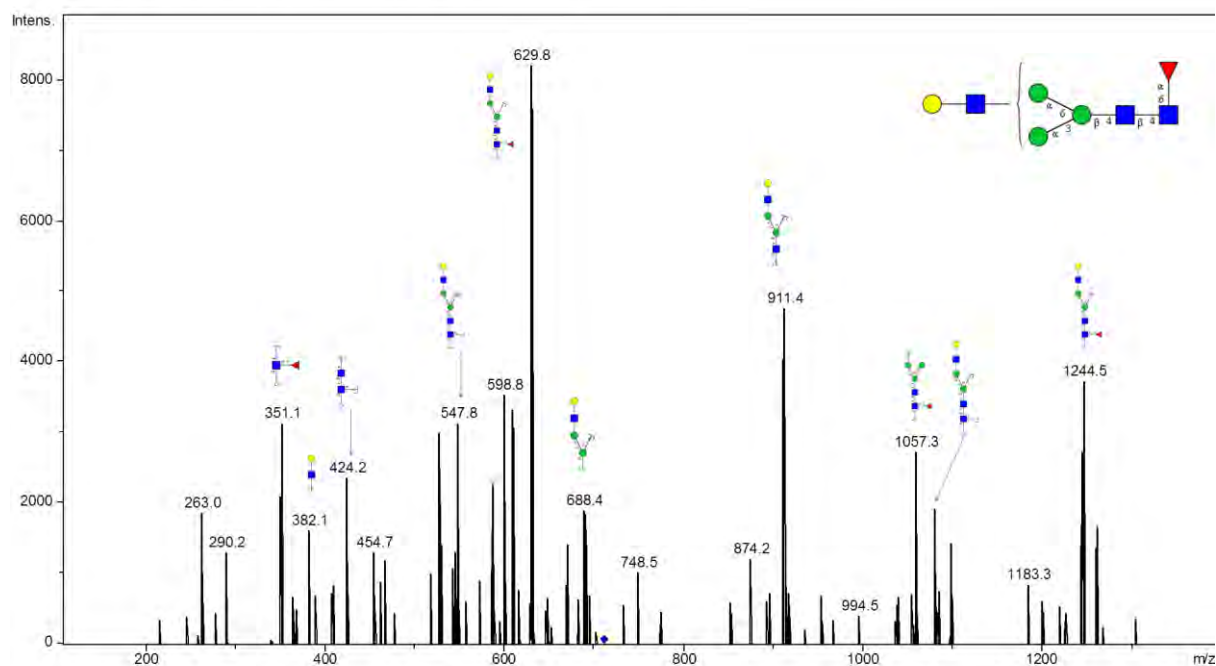
Glycan 8

Parent ion: *m/z* 718.8^{2-}

Composition: (Hex)$_2$ (HexNAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
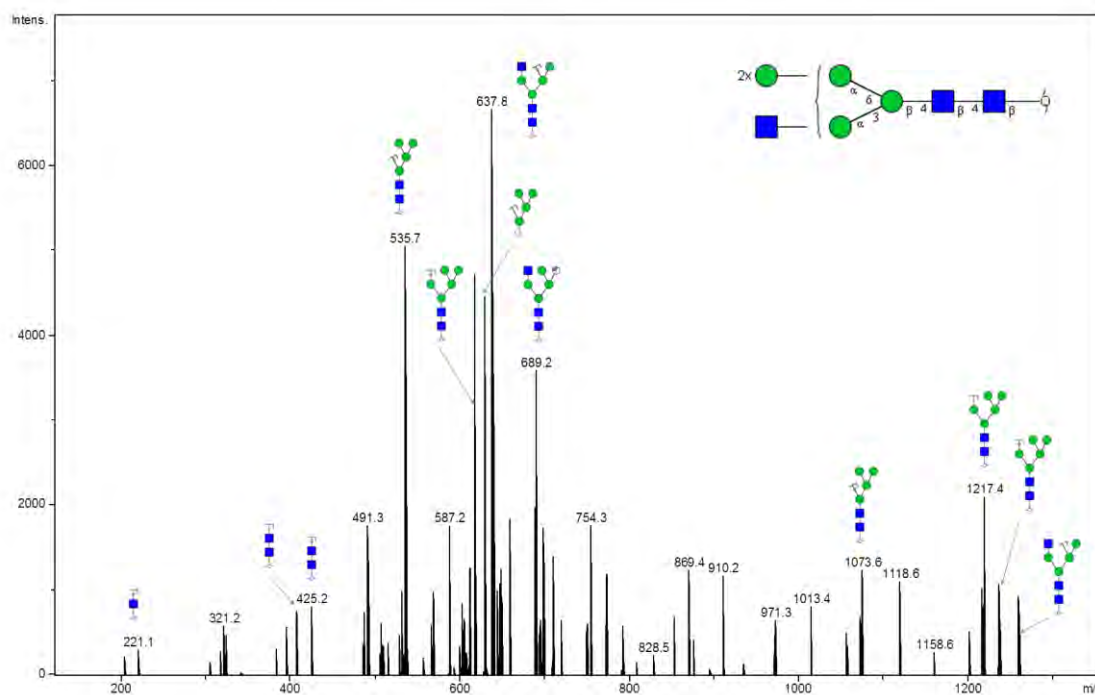
Glycan 9

Parent ion: *m/z* 731.3^{2-}

Composition: (HexNAc)$_2$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
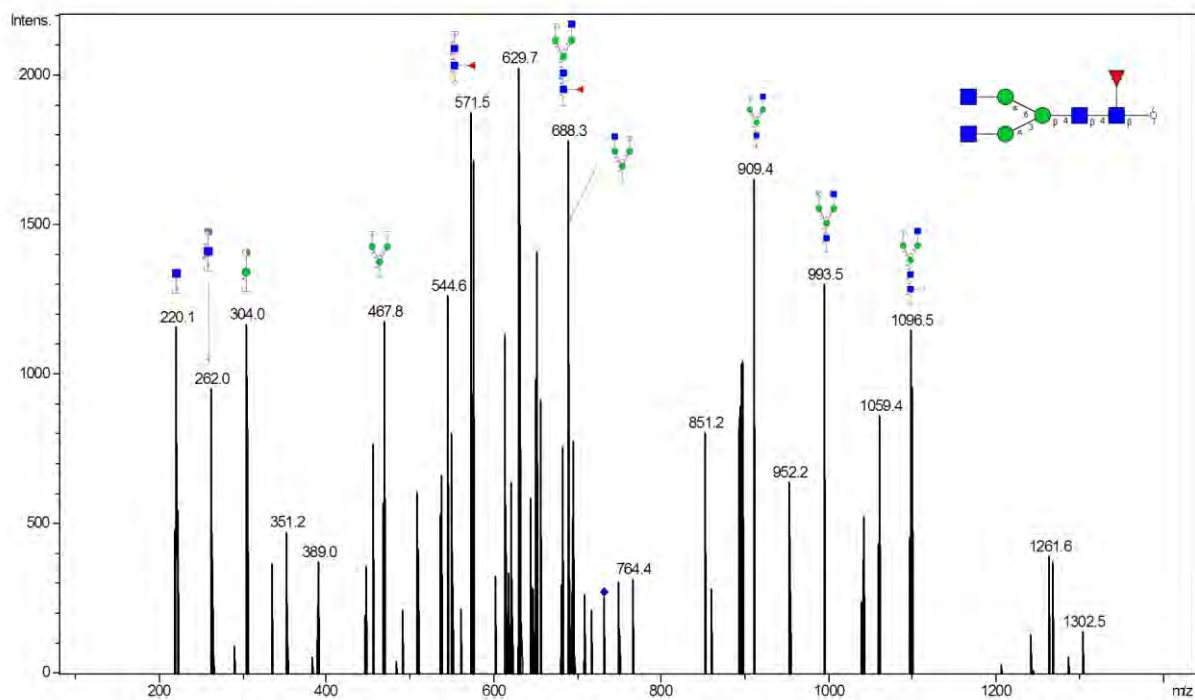
Glycan 10

Parent ion: *m/z* 731.3²⁻

Composition: $(HexNAc)_2$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$
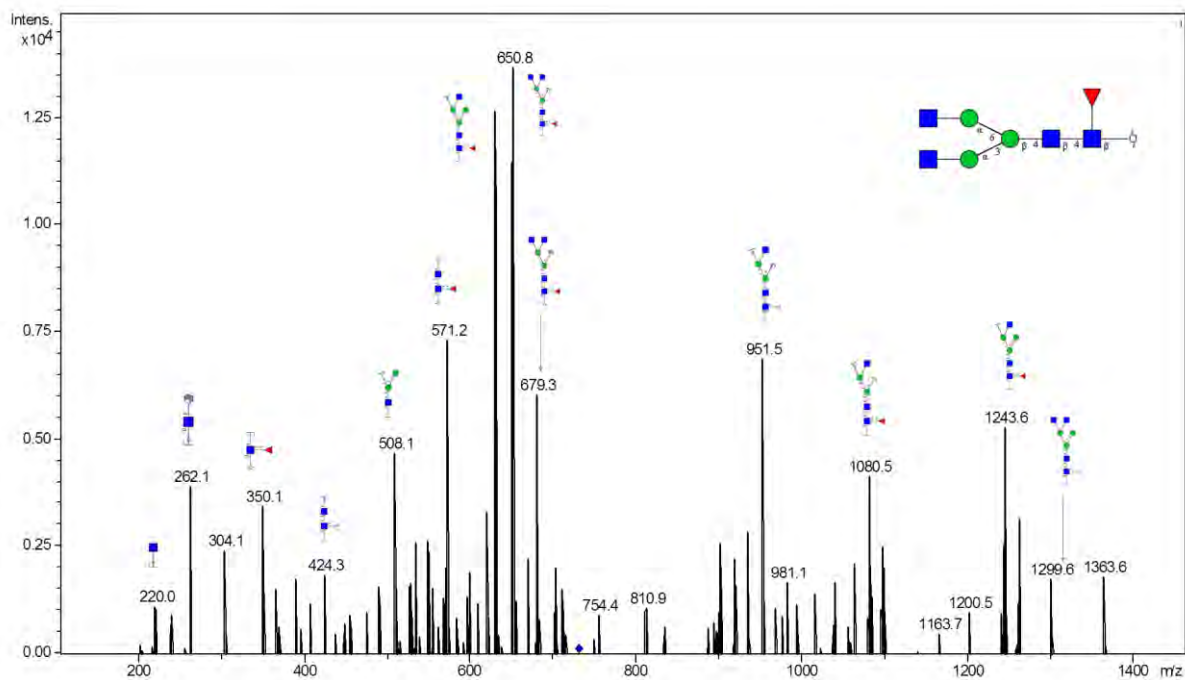
Glycan 11

Parent ion: *m/z* 731.3²⁻

Composition: $(HexNAc)_2$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$
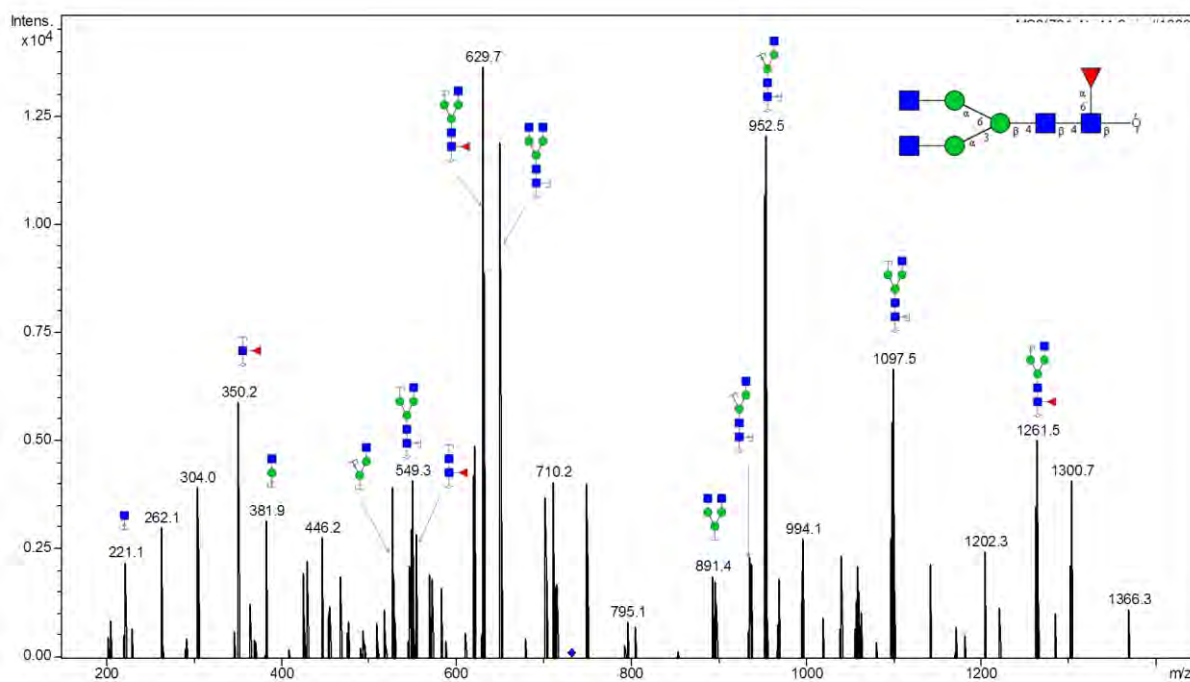
Glycan 12

Parent ion: *m/z* 739.3^{2-}

Composition: (Hex)$_1$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm and 6-arm in glycan fragment scheme.
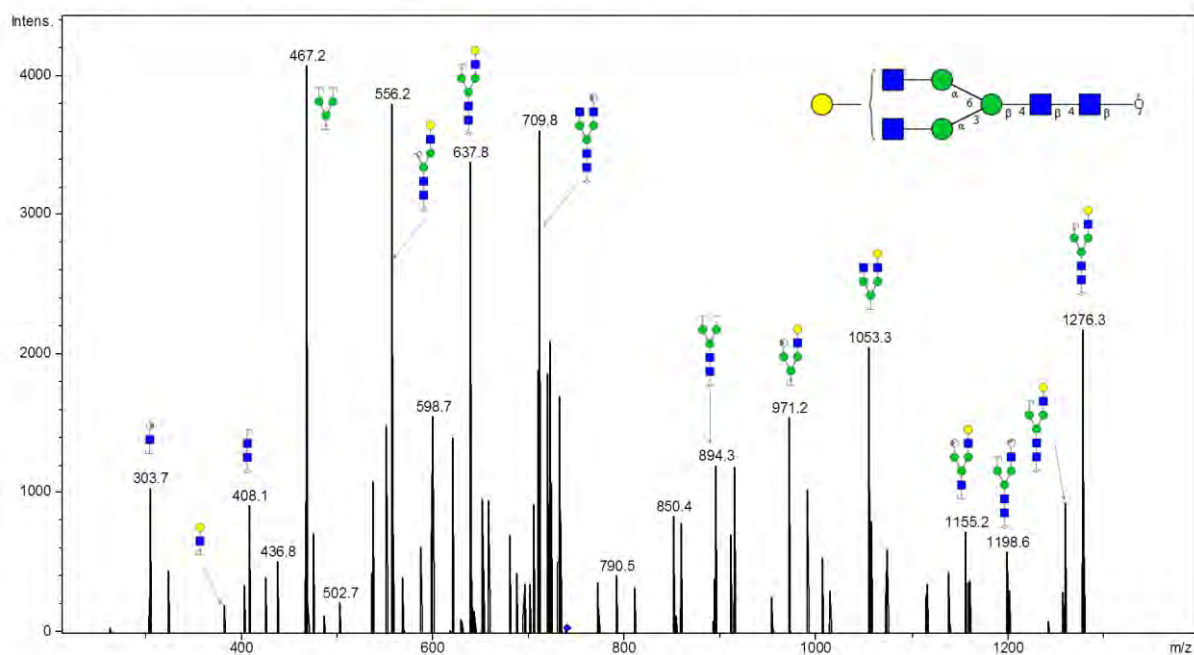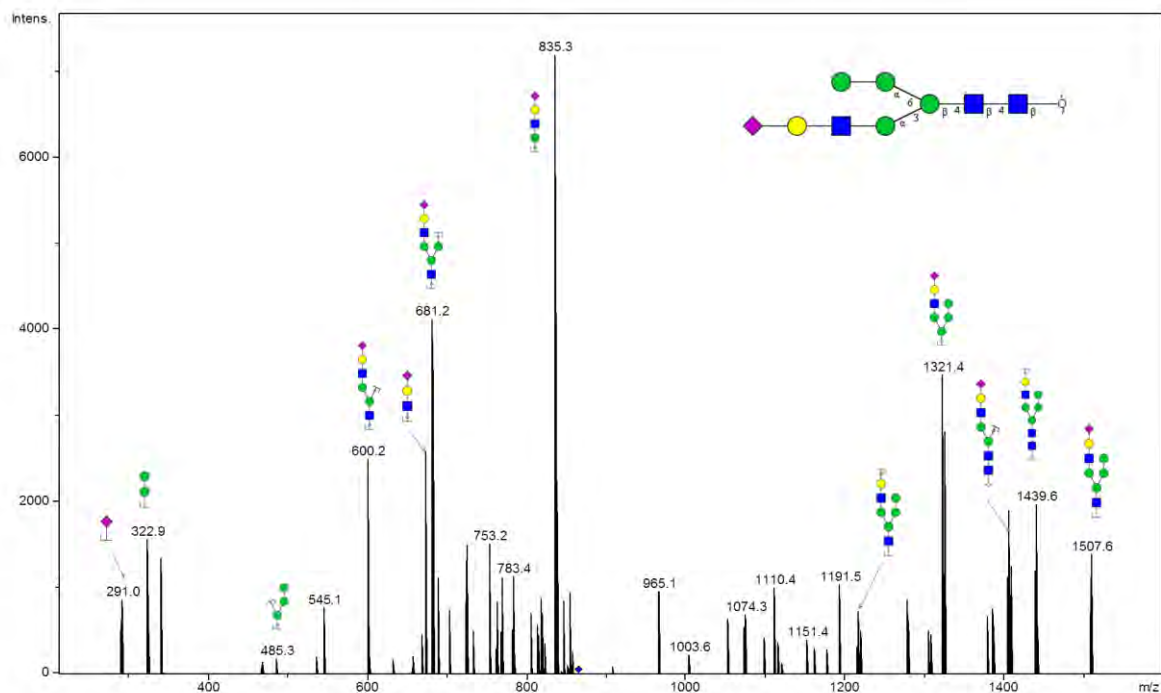

Glycan 13

Parent ion: *m/z* 739.3^{2-}

Composition: (Hex)$_1$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$
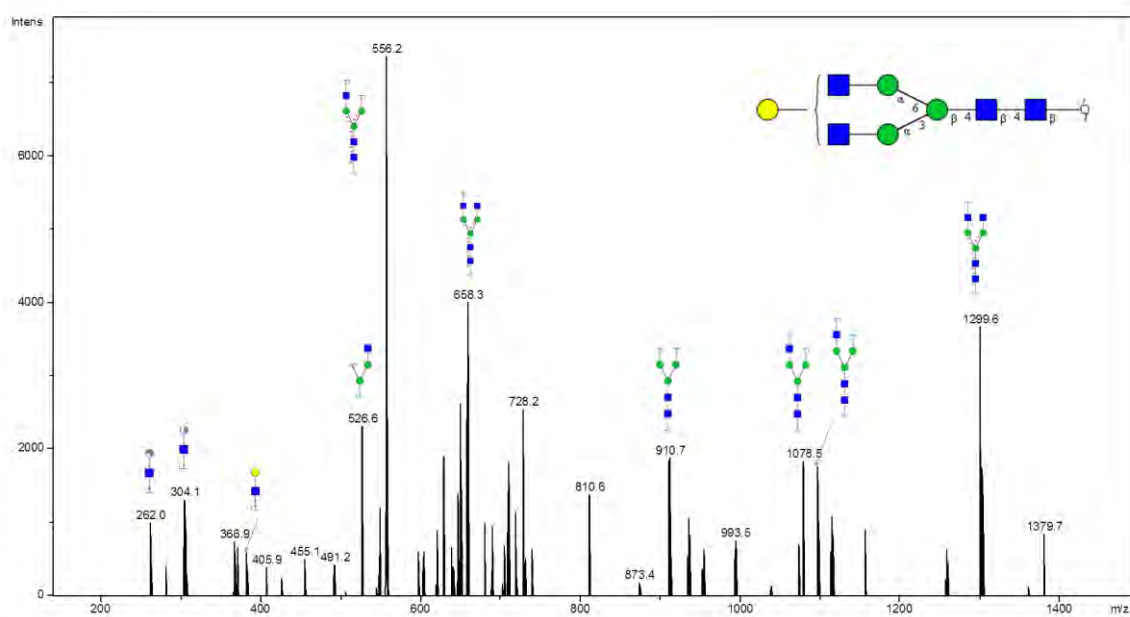
Glycan 14

Parent ion: *m/z* 759.9^{2-}

Composition: (HexNAc)$_3$ + (Man)$_3$(GlcNAc)$_2$
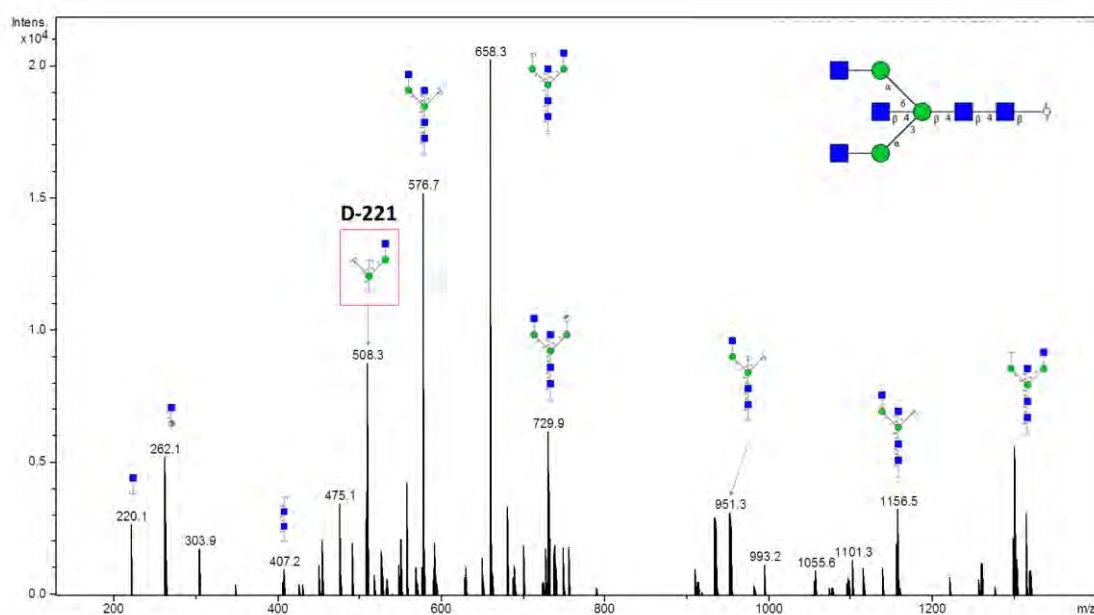
Glycan 15

Parent ion: *m/z* 779.3^{2-}

Composition: (Hex)$_4$ + (Man)$_3$(GlcNAc)$_2$

Notes: Depiction of specific branching not intended in glycan fragment scheme.
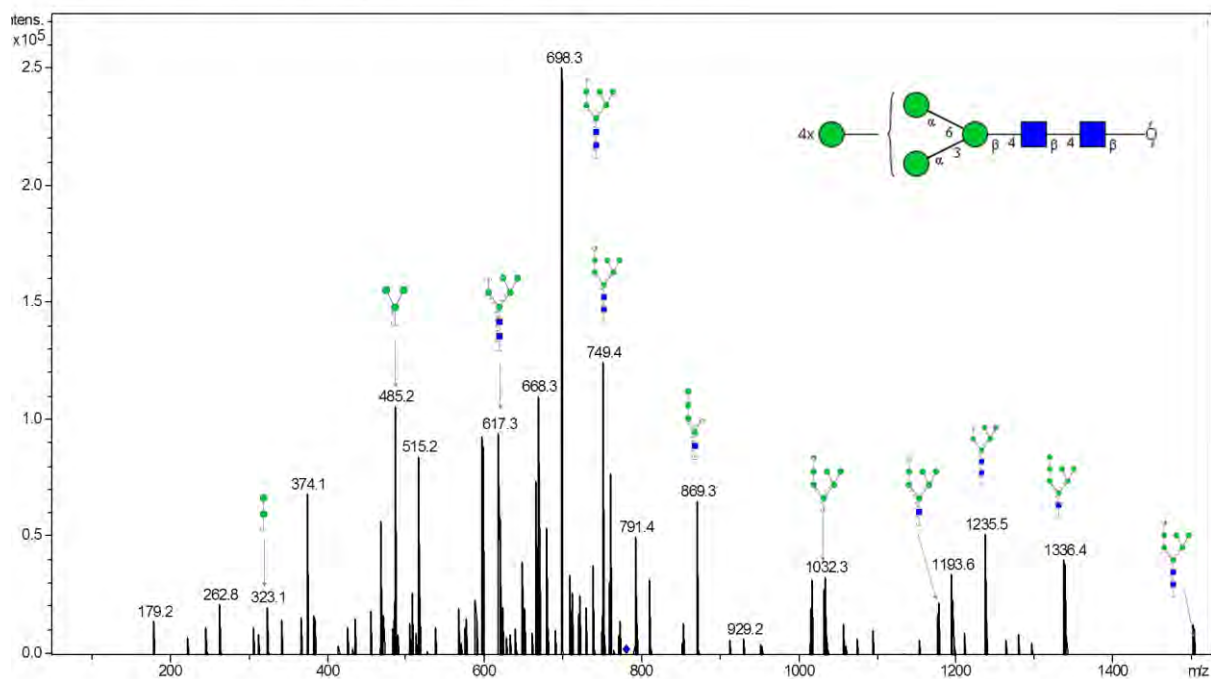
Glycan 16

Parent ion: *m/z* 779.3^{2-}

Composition: (Hex)$_4$ + (Man)$_3$(GlcNAc)$_2$

Notes: Depiction of specific branching not intended in glycan fragment scheme.
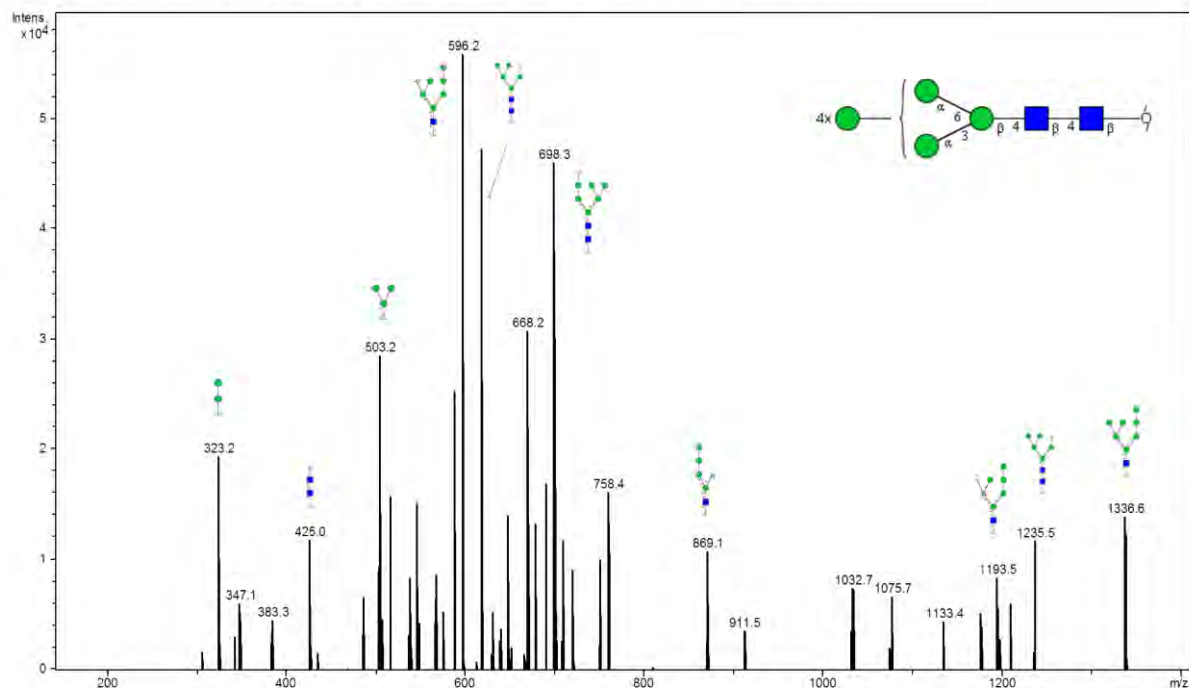
Glycan 17

Parent ion: *m/z* 783.5^{2-}

Composition: (Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm and 6-arm in glyan fragment scheme
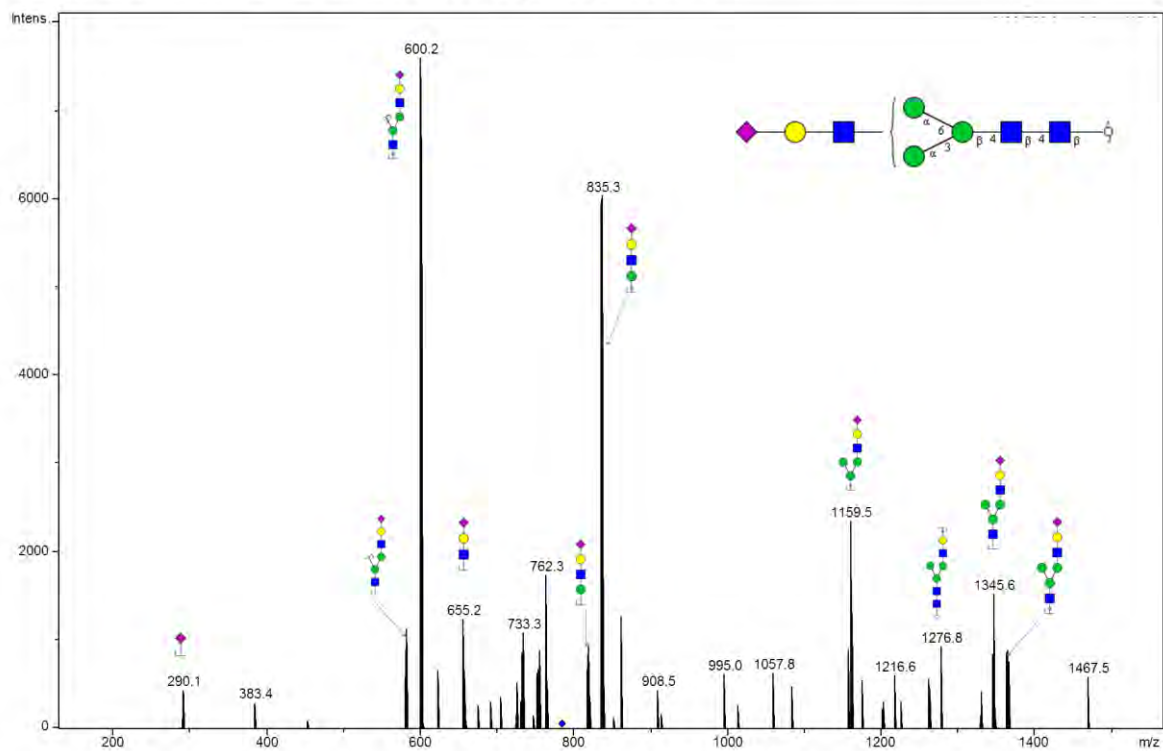
Glycan 18

Parent ion: *m/z* 783.5²⁻

Composition: (Hex)₁ (HexNAc)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glyan fragment scheme
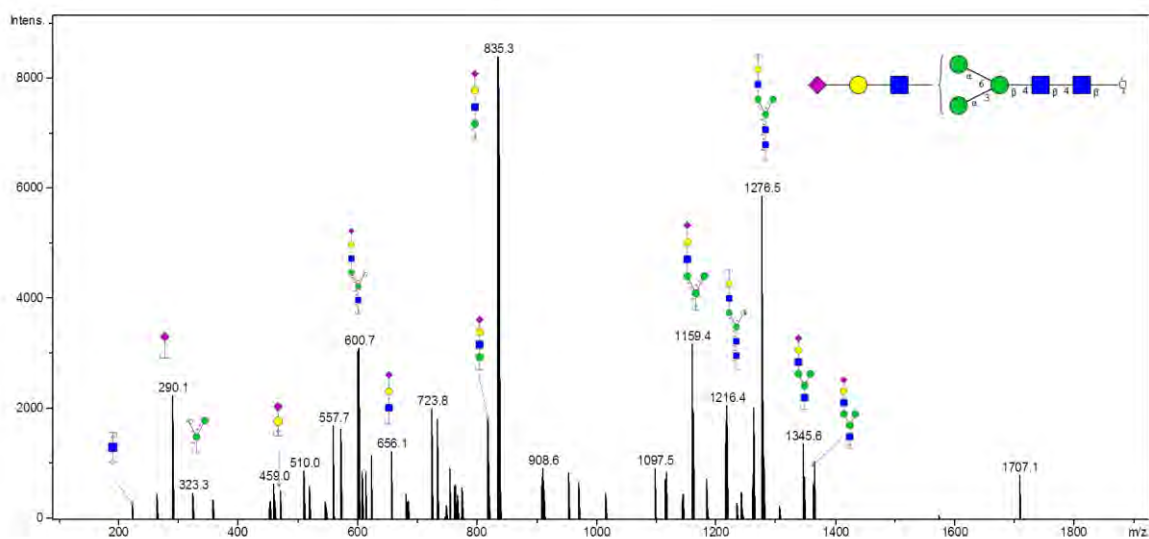
Glycan 19

Parent ion: *m/z* 783.5²⁻

Composition: (Hex)₁ (HexNAc)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glyan fragment scheme
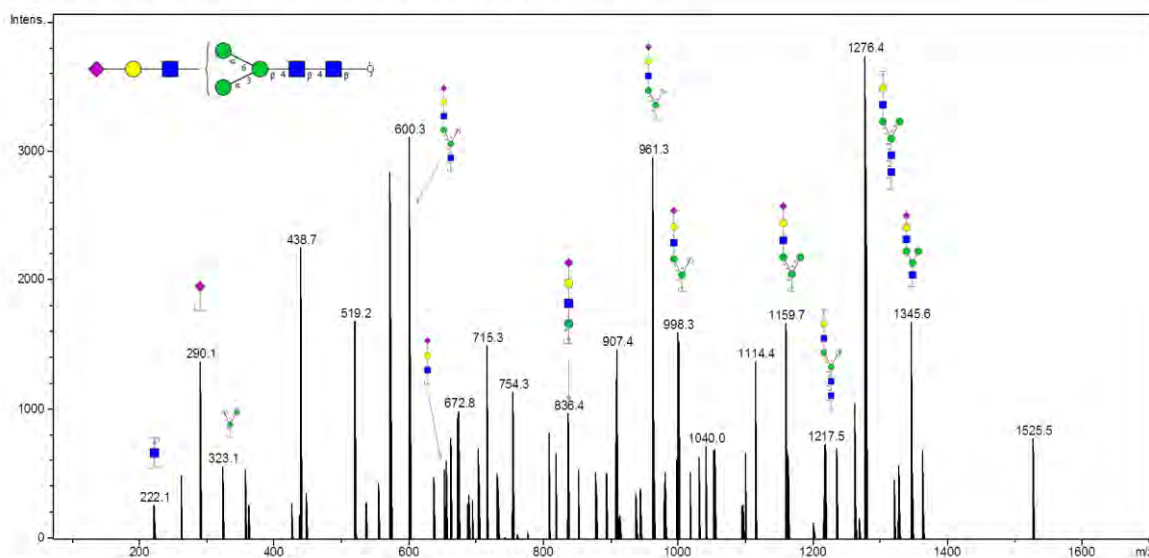
Glycan 20

Parent ion: *m/z* 791.8^{2-}

Composition: (Hex)$_2$ (HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Specific linkage for fucose not intended in glycan scheme, i.e. fucose may be attached to galactose or GlcNAc.
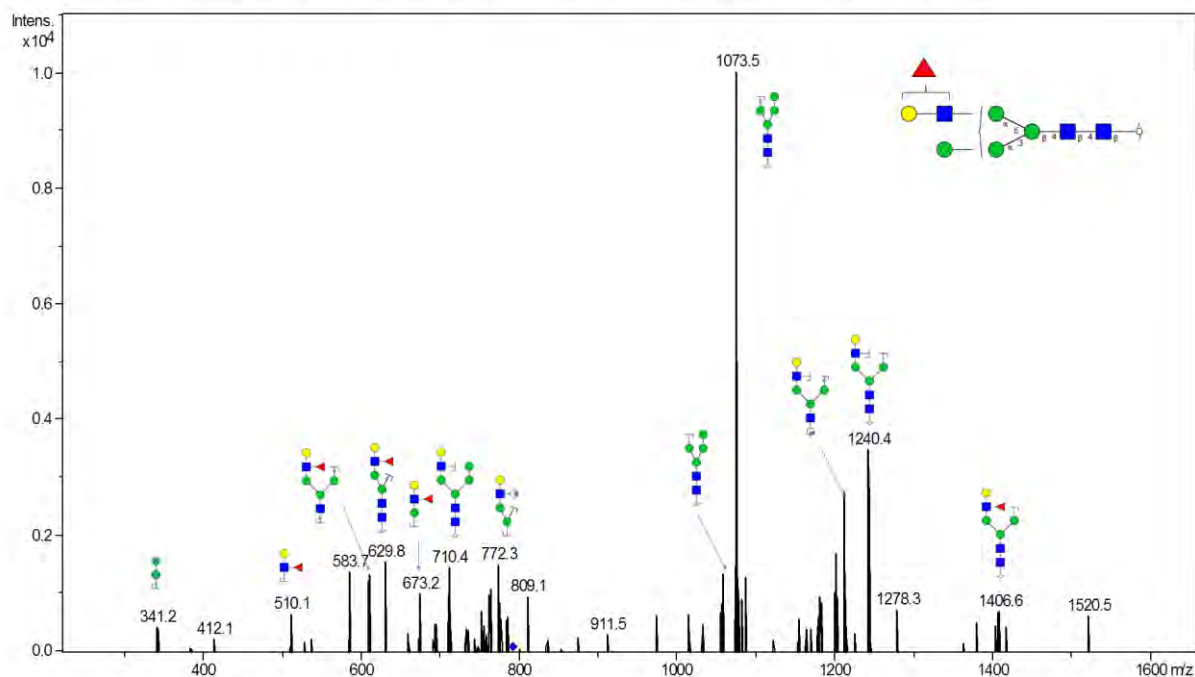
Glycan 21

Parent ion: *m/z* 791.9^{2-}

Composition: (Hex)$_2$ (HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
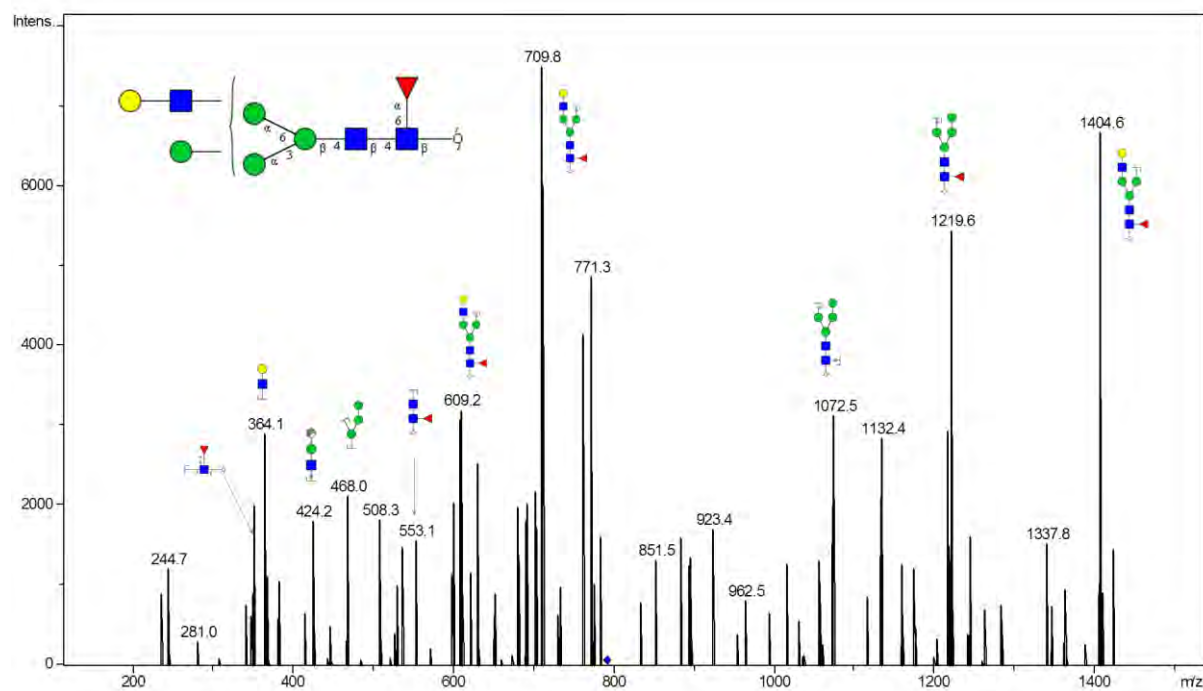
Glycan 22

Parent ion: *m/z* 799.8^{2-}

Composition: (Hex)$_3$ (HexNAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
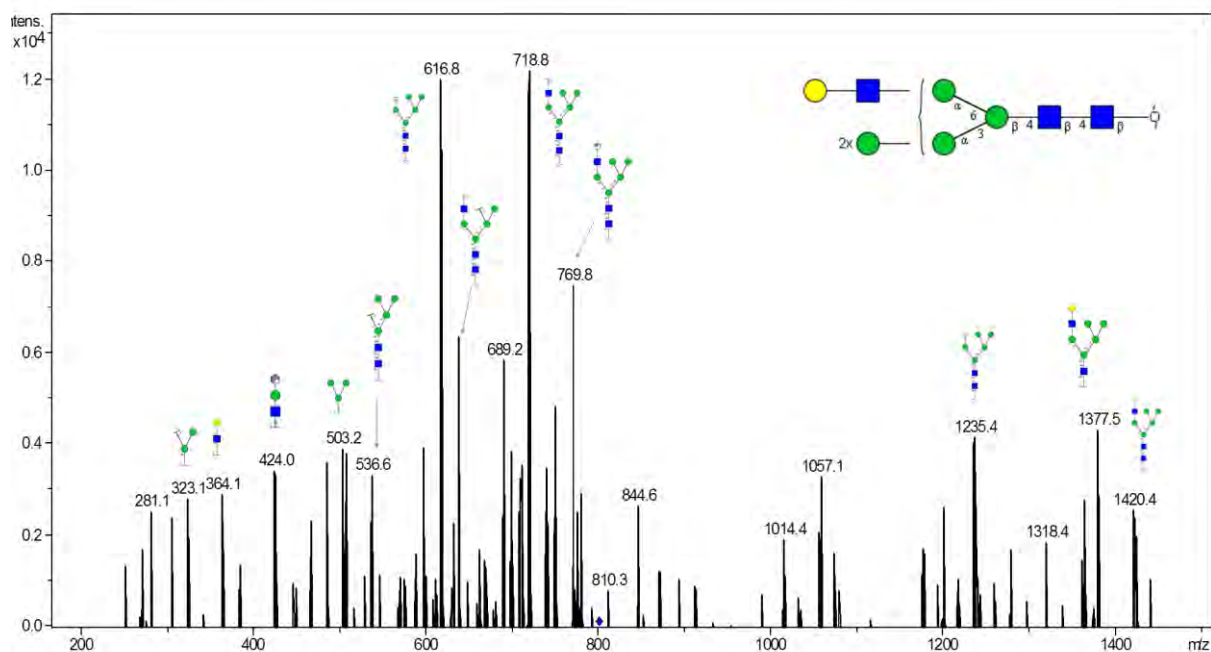
Glycan 23

Parent ion: *m/z* 812.4^{2-}

Composition: (Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
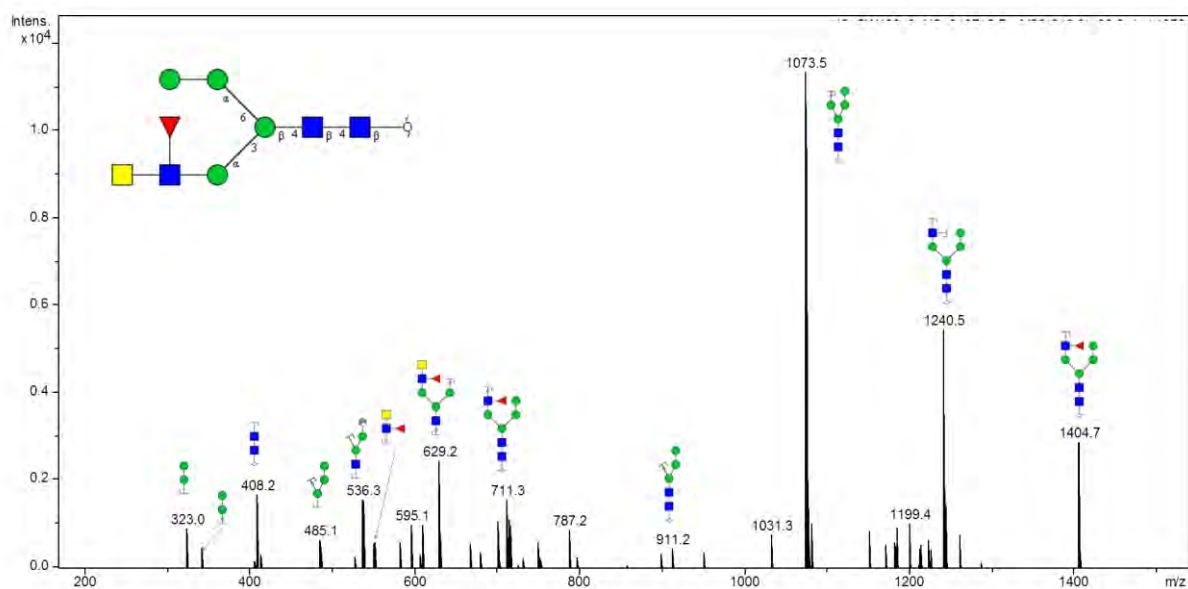
Glycan 24

Parent ion: *m/z* 812.4²⁻

Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
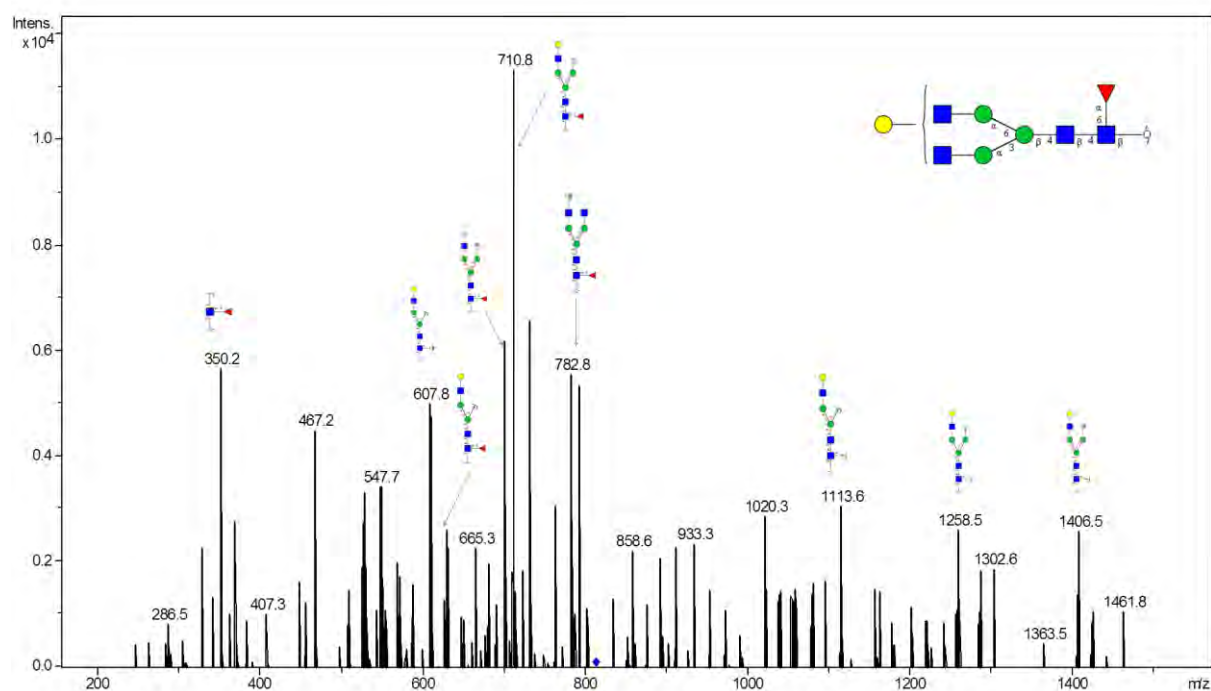
Glycan 25

Parent ion: *m/z* 812.4²⁻

Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
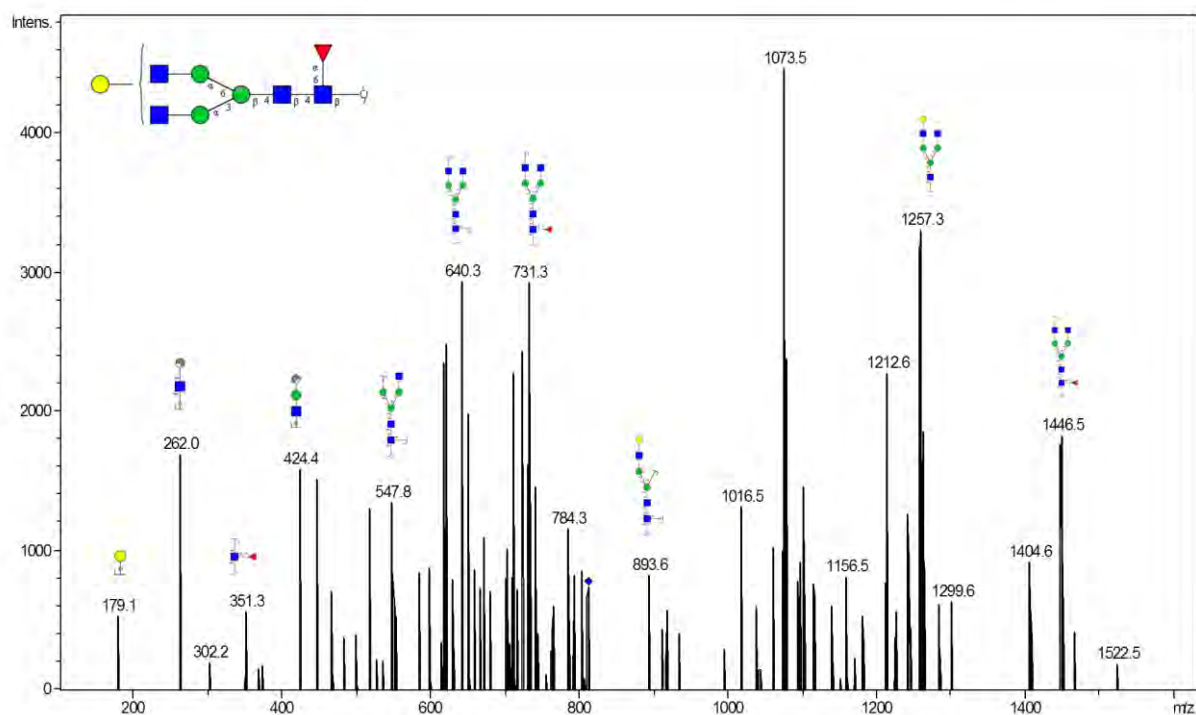
Glycan 26

Parent ion: *m/z* 812.4^{2-}

Composition: (Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
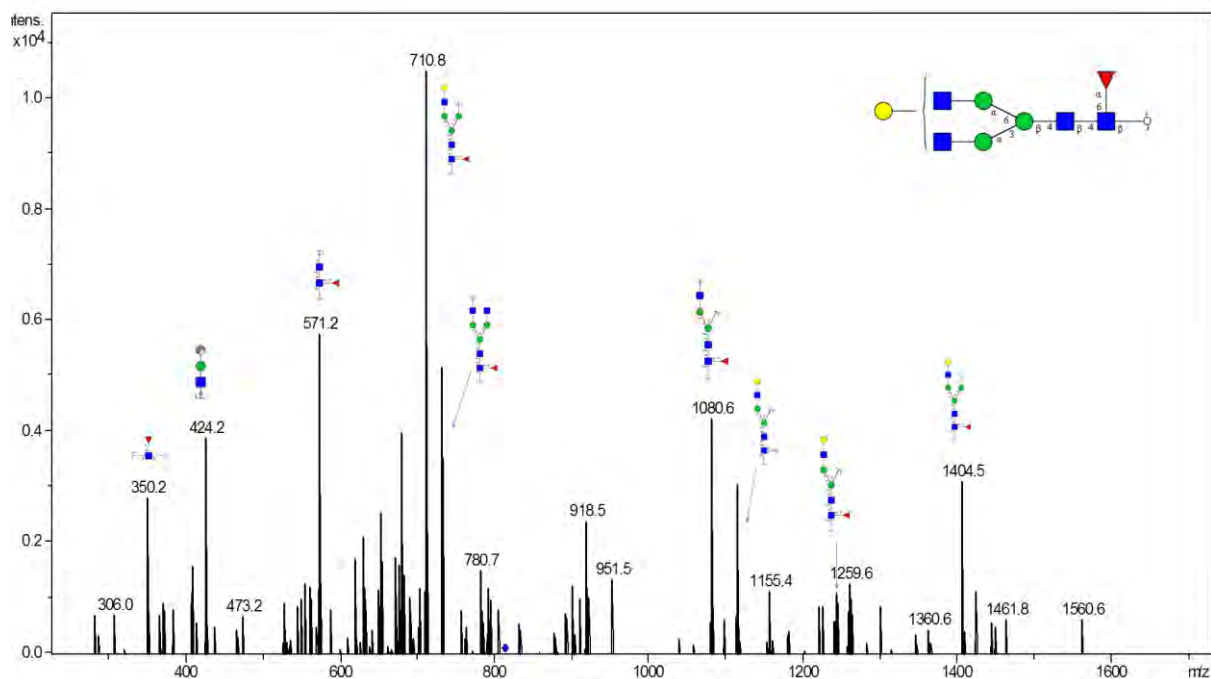
Glycan 27

Parent ion: *m/z* 820.3^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$
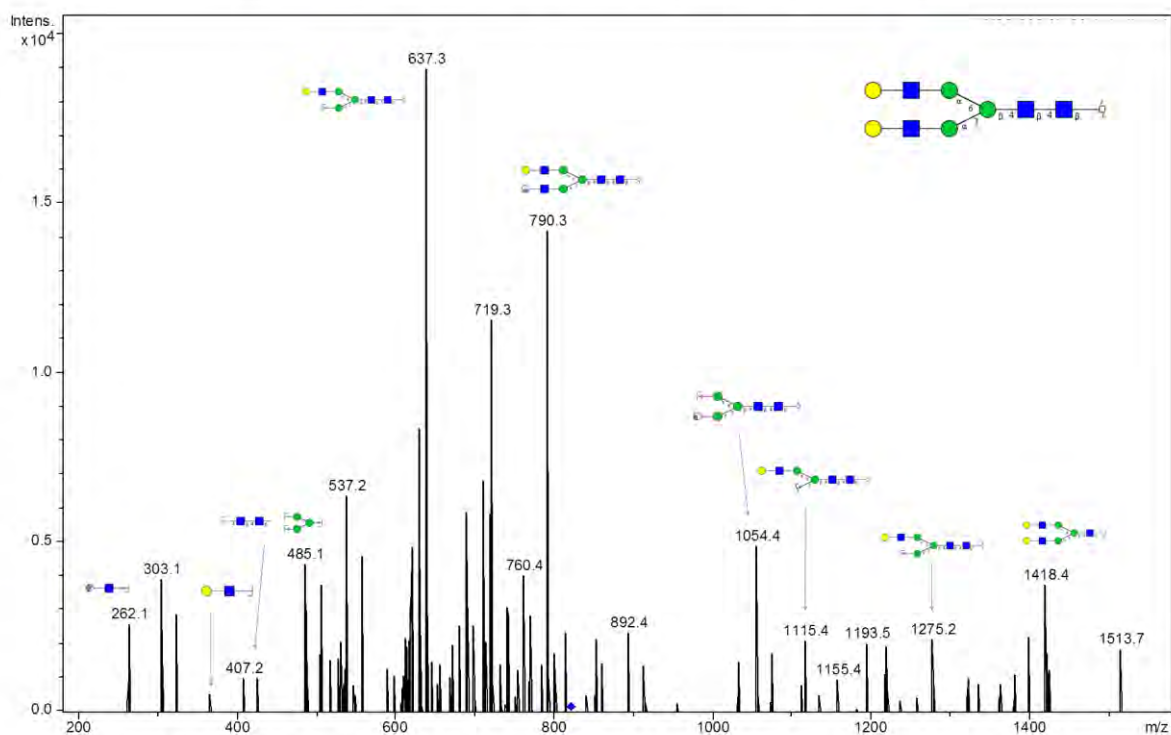
Glycan 28

Parent ion: *m/z* 820.3^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

[M-2H]$^{2-}$ - C$_2$H$_2$O
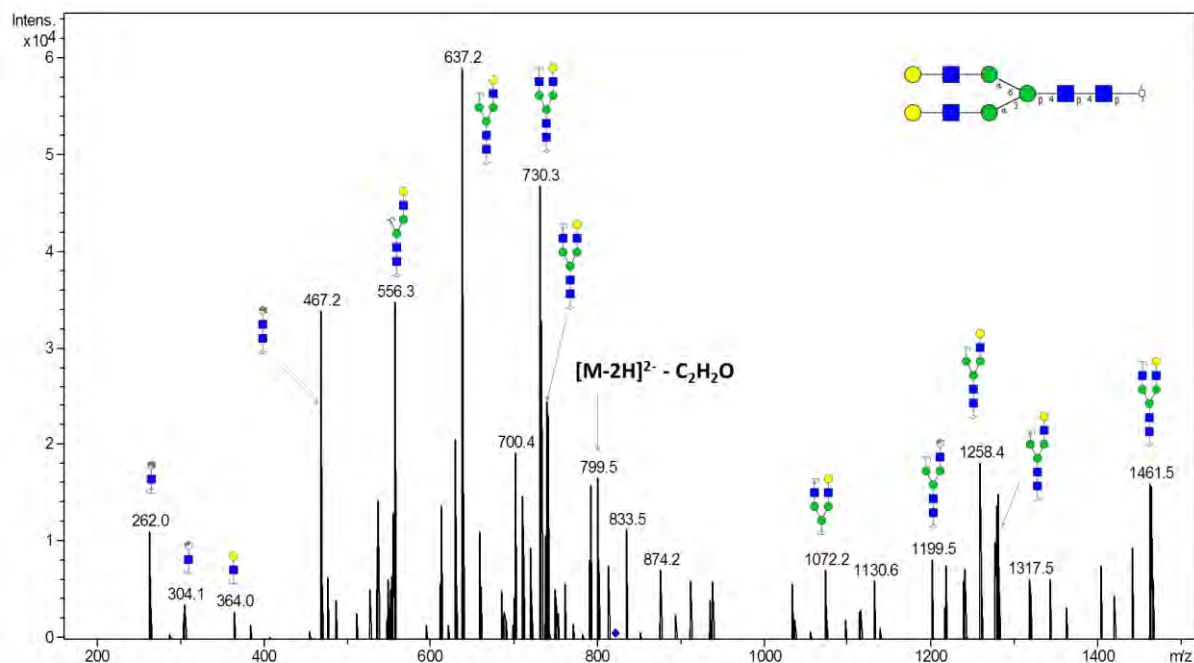
Glycan 29

Parent ion: *m/z* 820.3^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$
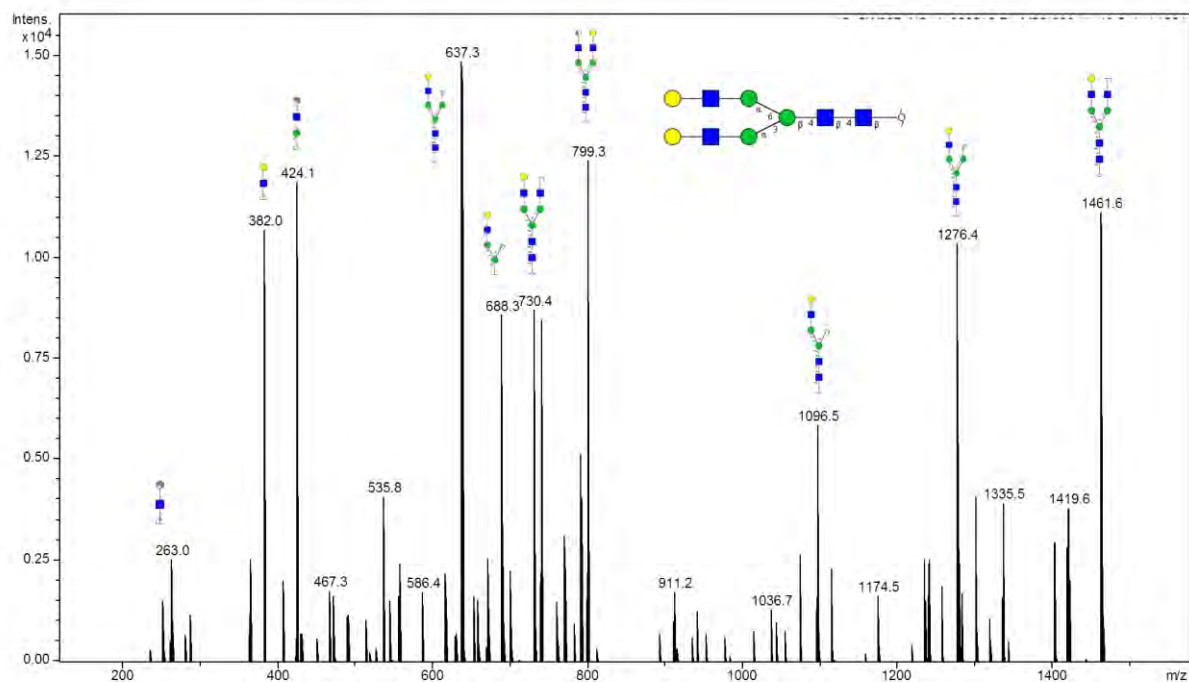
Glycan 30

Parent ion: *m/z* 832.9^{2-}

Composition: (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
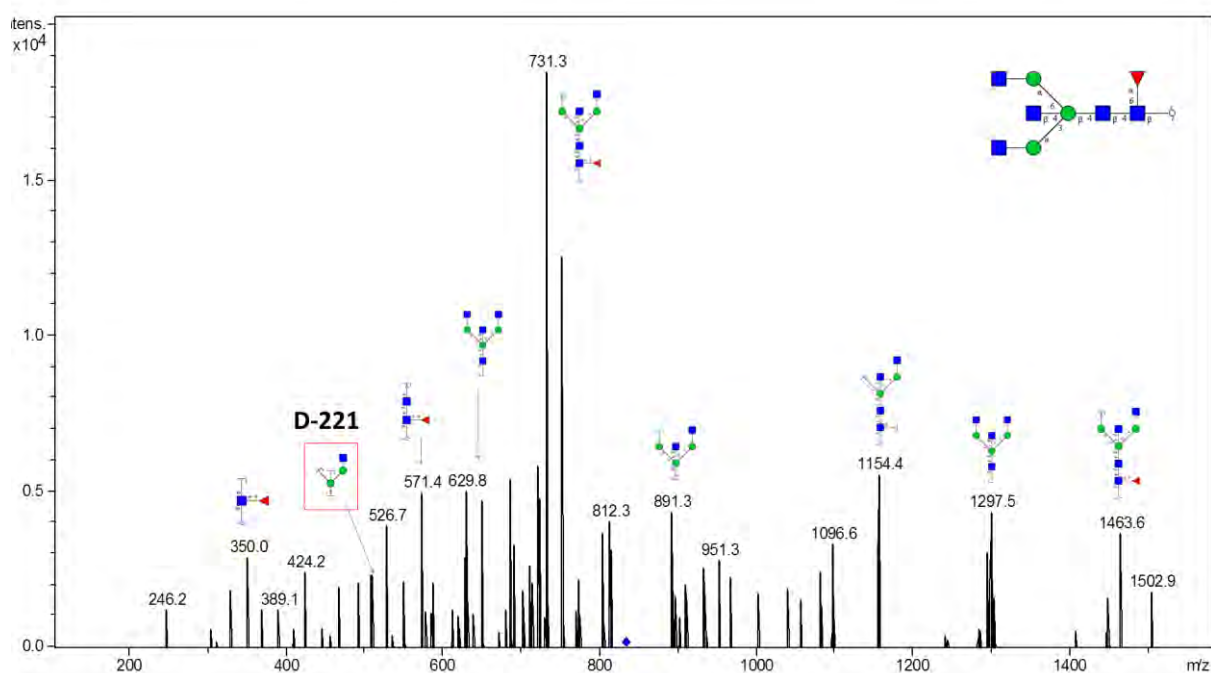
Glycan 31

Parent ion: *m/z* 832.9^{2-}

Composition: (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arn and 6-arm in glycan fragmentation scheme.
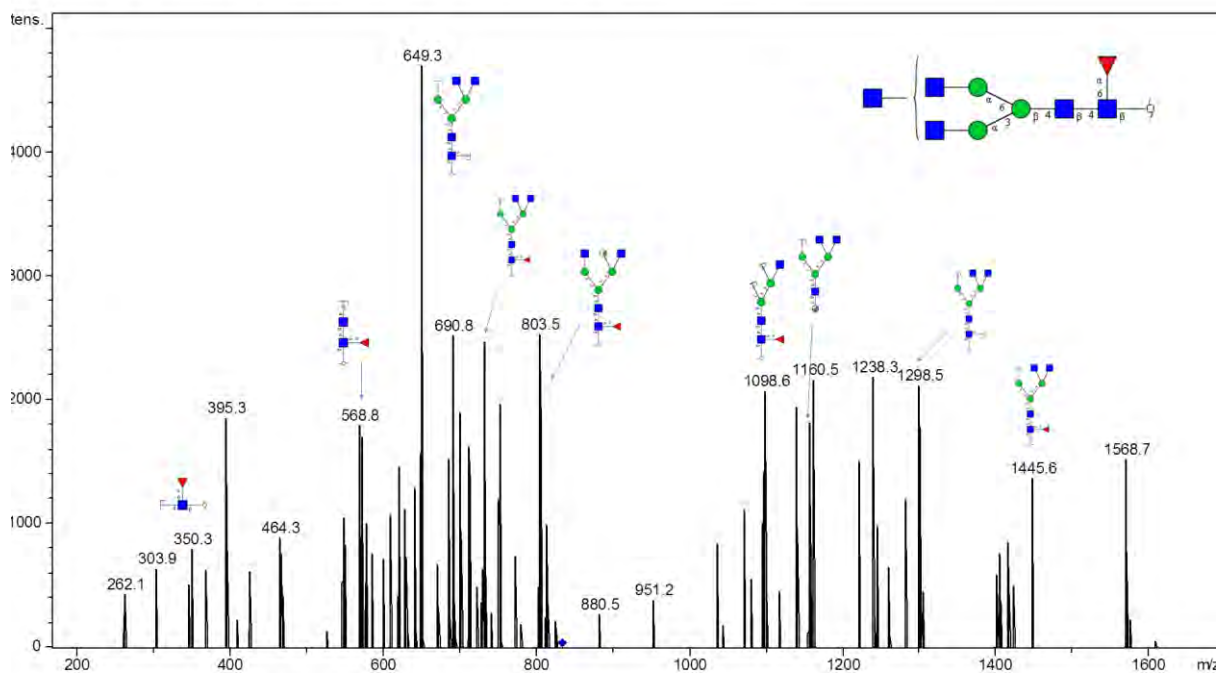
Glycan 32

Parent ion: *m/z* 832.9^{2-}

Composition: (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arn and 6-arm in glycan fragmentation scheme.
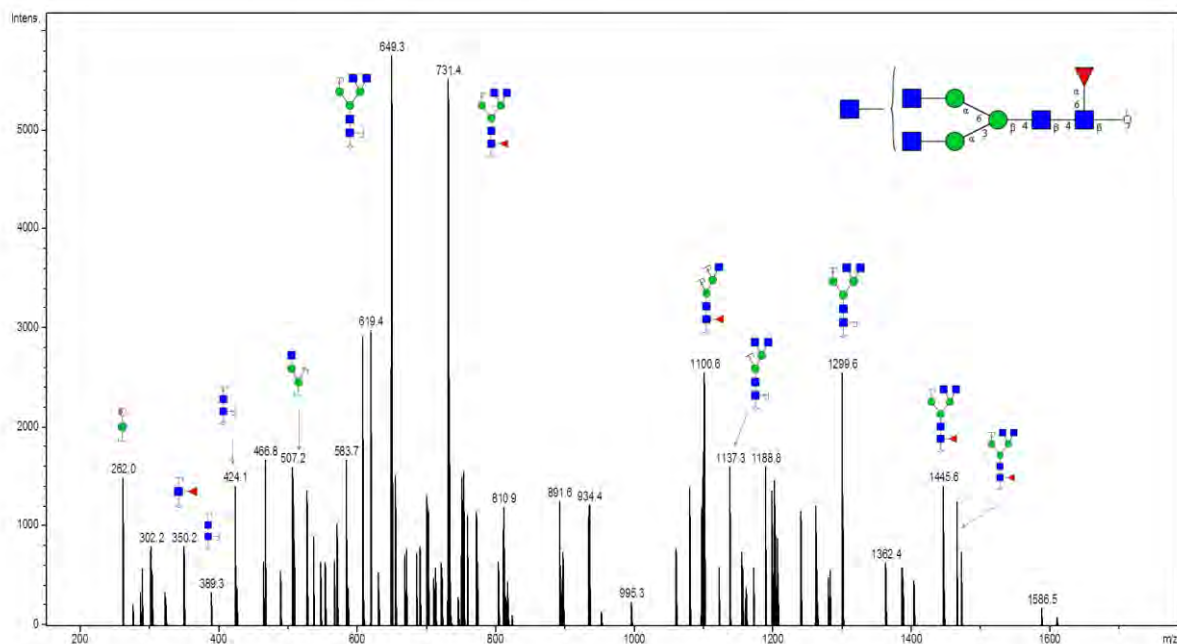
Glycan 33

Parent ion: *m/z* 832.9^{2-}

Composition: (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arn and 6-arm in glycan fragmentation scheme.
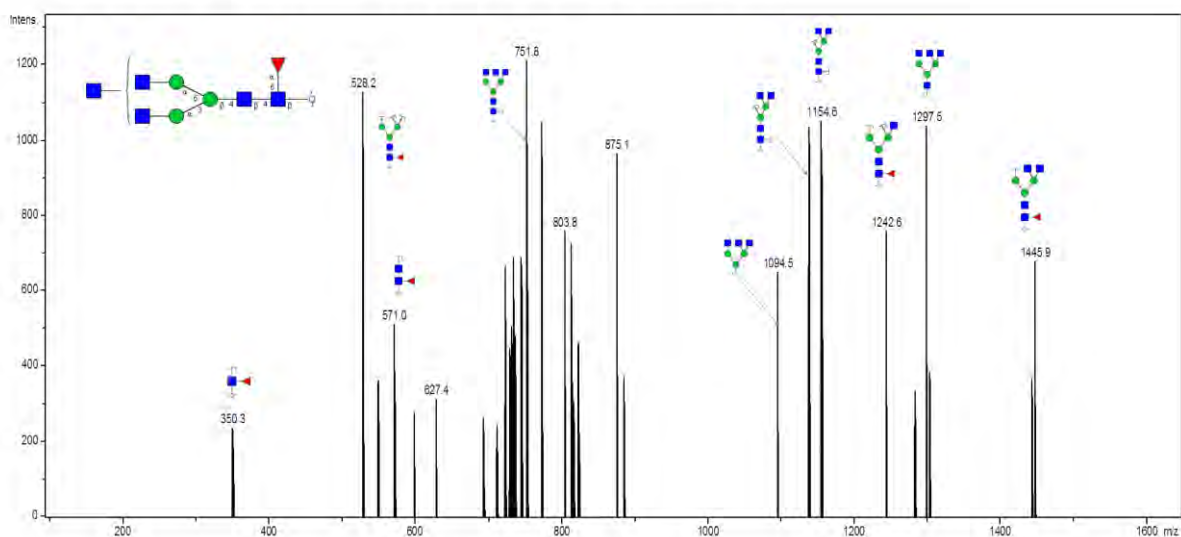
Glycan 34

Parent ion: *m/z* 840.9²⁻

Composition: (Hex)₁ (HexNAc)₃ + (Man)₃(GlcNAc)₂
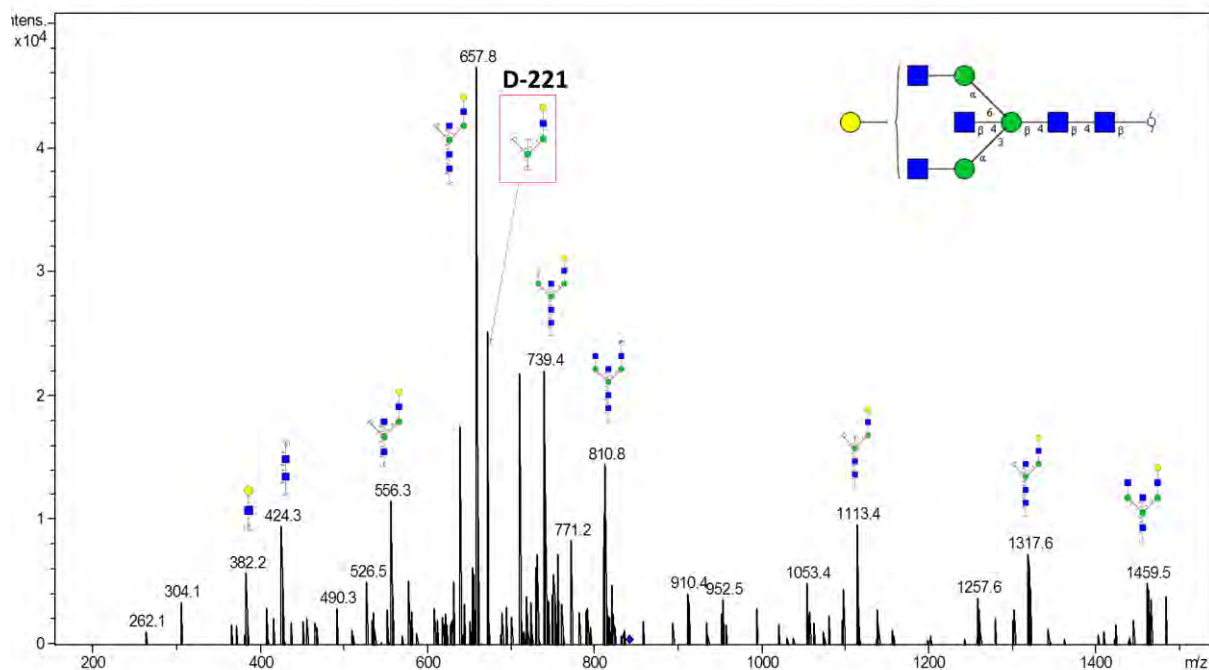
Glycan 35

Parent ion: *m/z* 856.3²⁻

Composition: (Hex)₁ (HexNAc)₁ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂
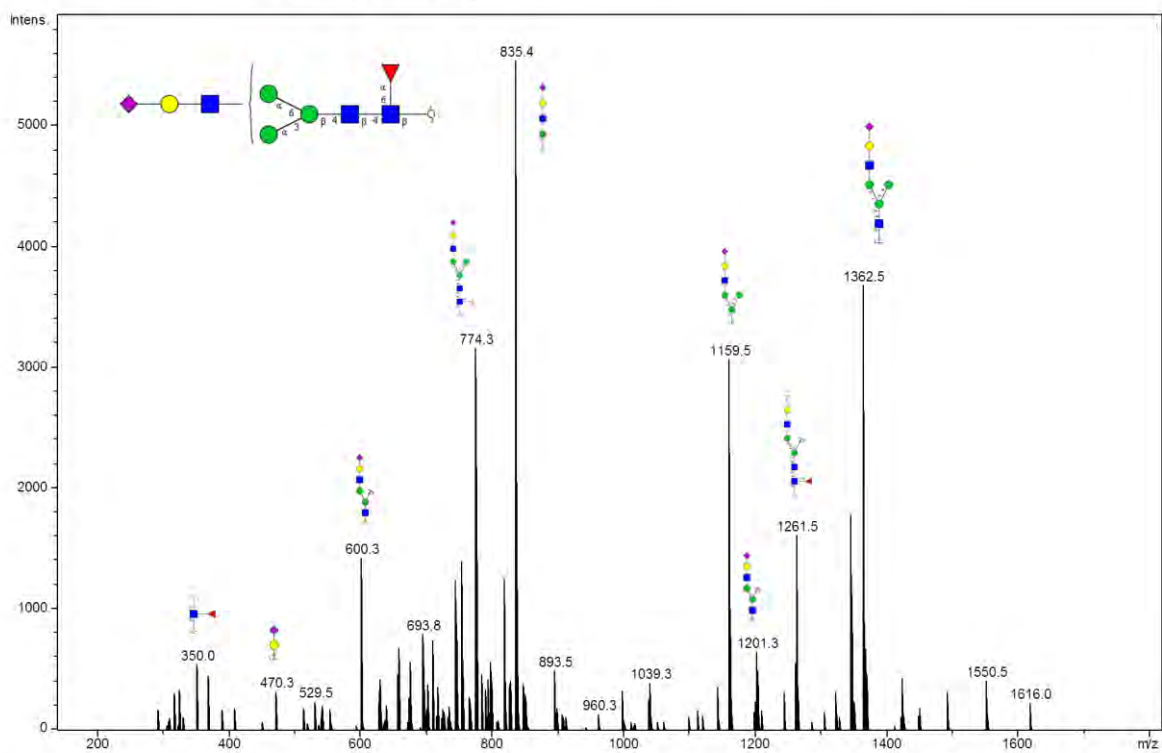
Glycan 36

Parent ion: *m/z* 856.3²⁻
Composition: (Hex)$_1$ (HexNAc)$_1$ (Deoxyhexose)$_1$ (NeuAc)$_1$ +
(Man)$_3$(GlcNAc)$_2$

Notes: Specific linkage of fucose not intended in glycan fragmentation scheme. I.e. fucose may be attached to outer-arm or chitobiose core.
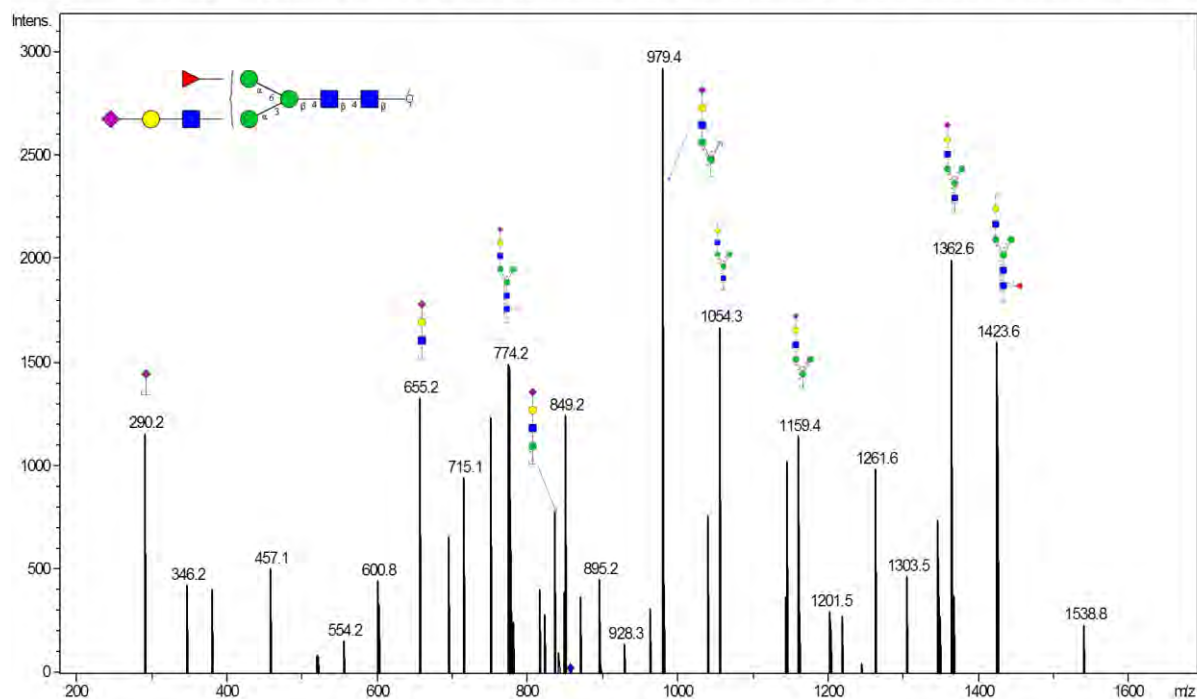
Glycan 37

Parent ion: *m/z* 856.3²⁻
Composition: (Hex)$_1$ (HexNAc)$_1$ (Deoxyhexose)$_1$ (NeuAc)$_1$ +
(Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
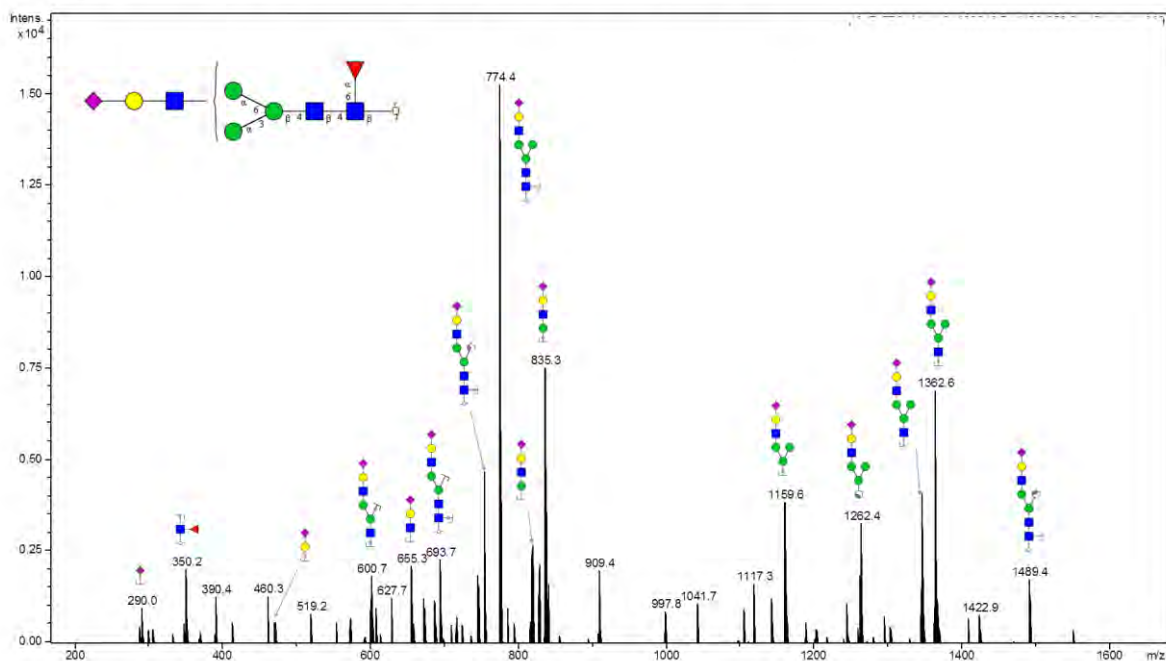
Glycan 38

Parent ion: *m/z* 856.3²⁻
Composition: (Hex)₁ (HexNAc)₁ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂
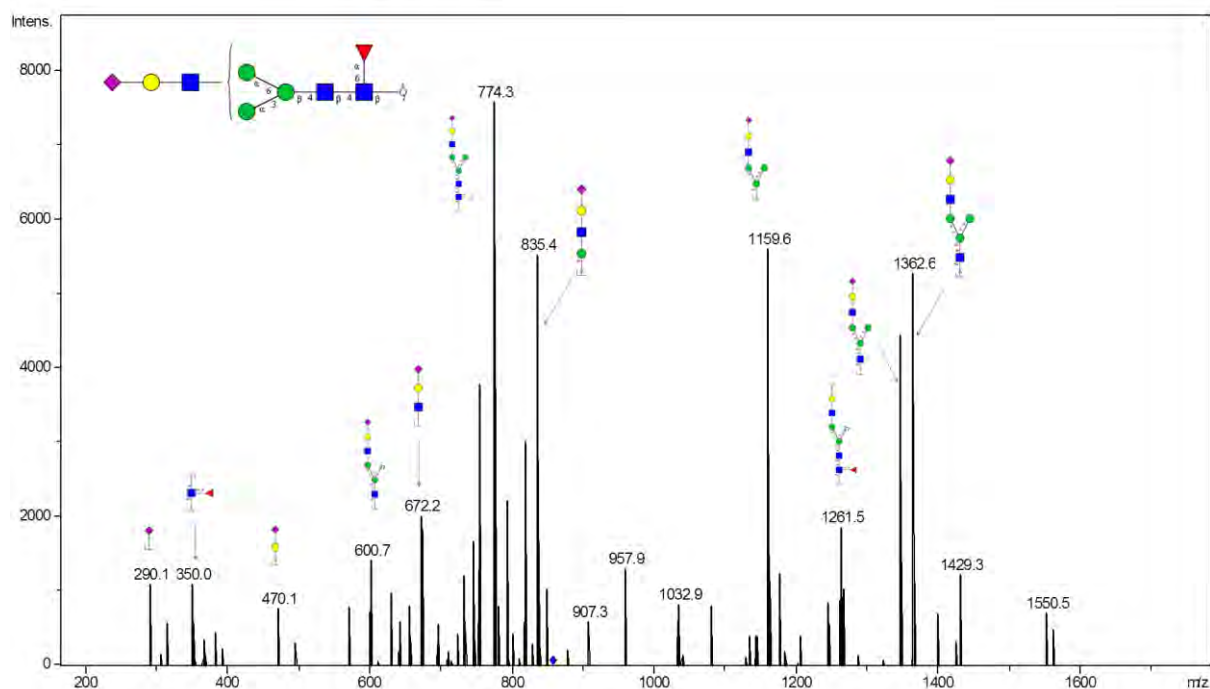
Glycan 39

Parent ion: *m/z* 860.3²⁻
Composition: (Hex)₅ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan fragmentation scheme.

Glycan 40

Parent ion: *m/z* 860.3²⁻

Composition: (Hex)₅ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan fragmentation scheme.

Glycan 41

Parent ion: *m/z* 860.4²⁻

Composition: (Hex)₅ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arn and 6-arm in glycan fragmentation scheme.

138
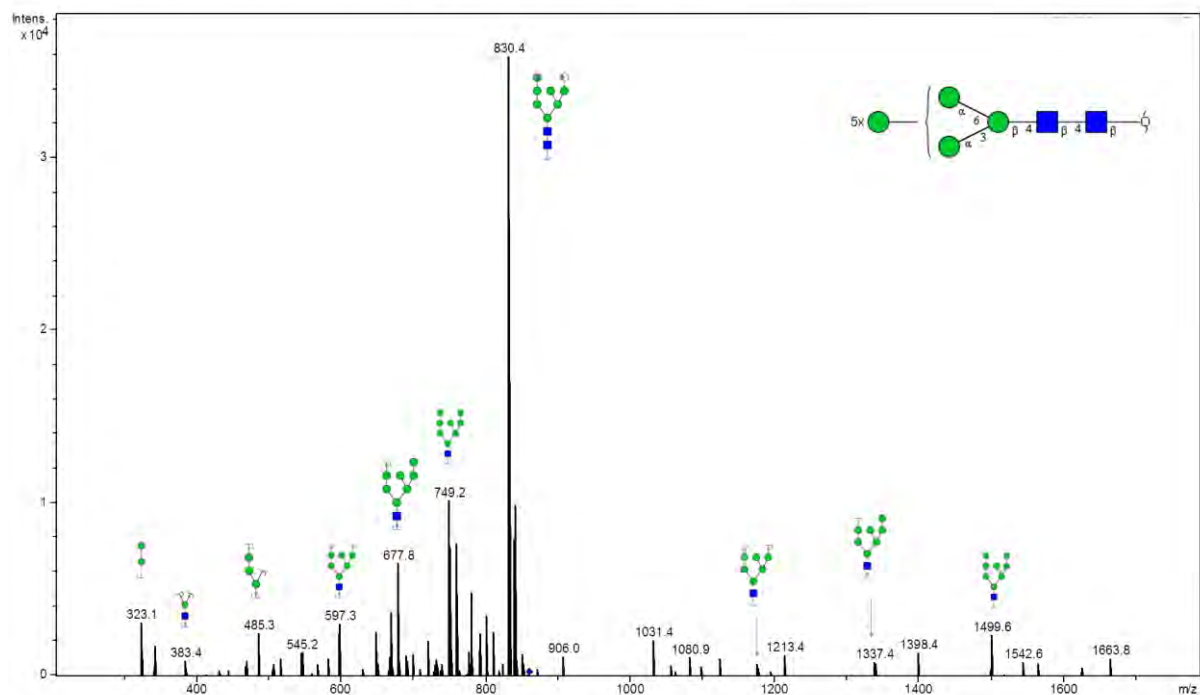
Glycan 42

Parent ion: *m/z* 864.4²⁻

Composition: (Hex)₂ (HexNAc)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂
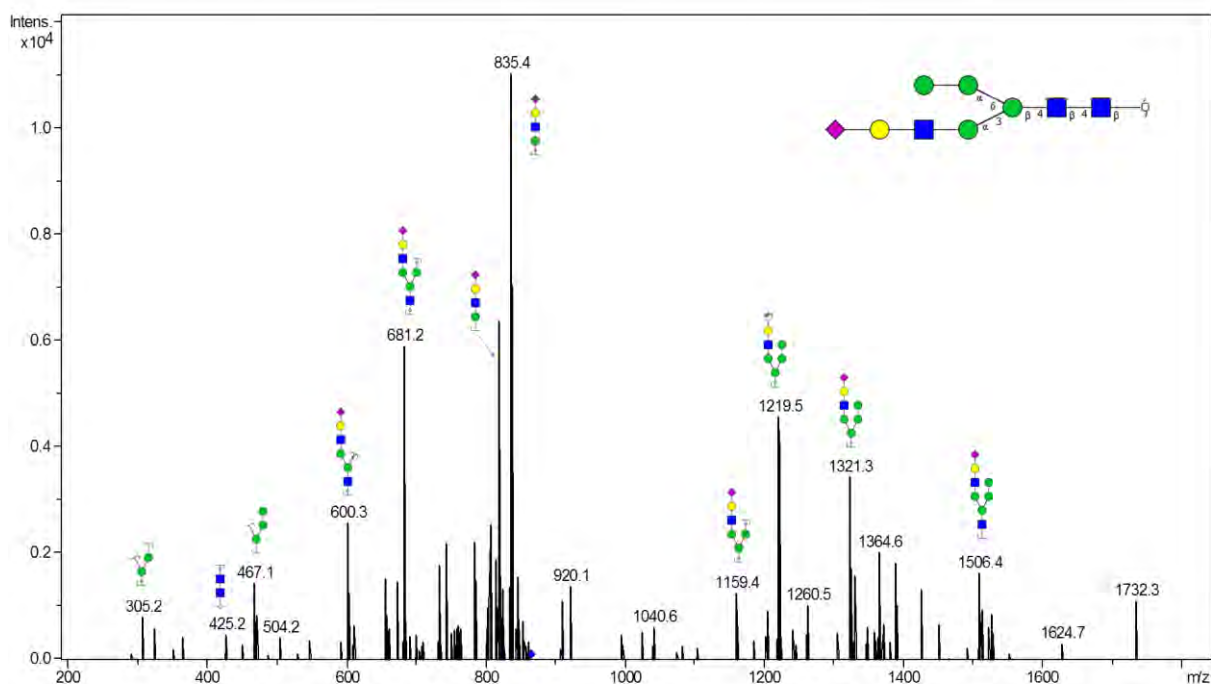
Glycan 43

Parent ion: *m/z* 864.4²⁻

Composition: (Hex)₂ (HexNAc)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Glycan 44

Parent ion: *m/z* 864.9²⁻

Composition: (Hex)₂ (HexNAc)₁ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: Specific linkage of outer-arm fucose not intended in glycan fragmentation scheme.
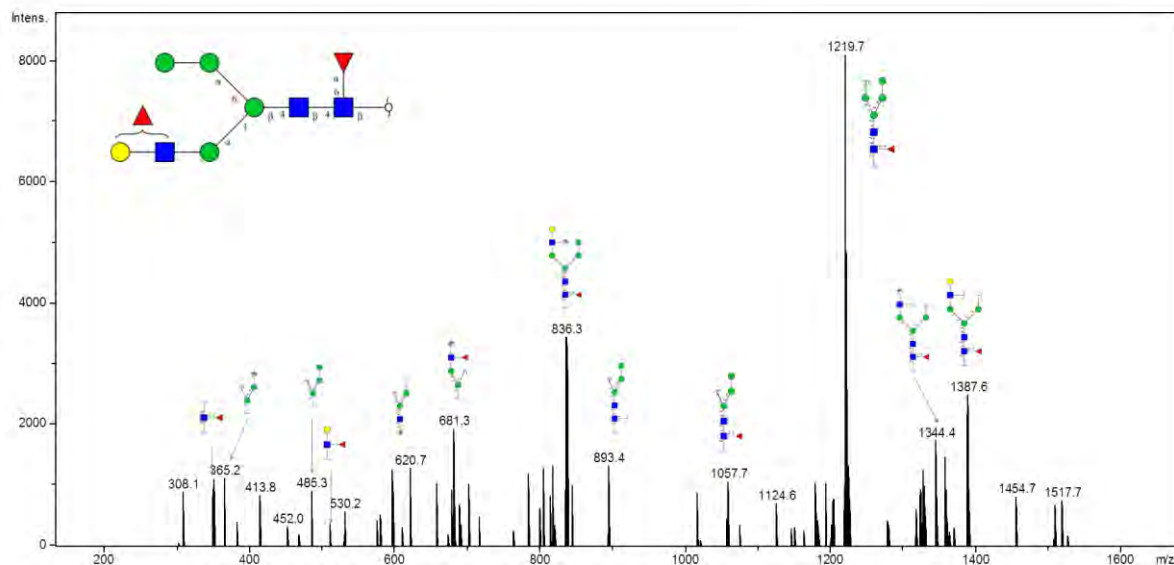
Glycan 45

Parent ion: *m/z* 872.9²⁻

Composition: (Hex)₃ (HexNAc)₁ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Specific linkage of outer-arm fucose not intended in glycan fragmentation scheme.
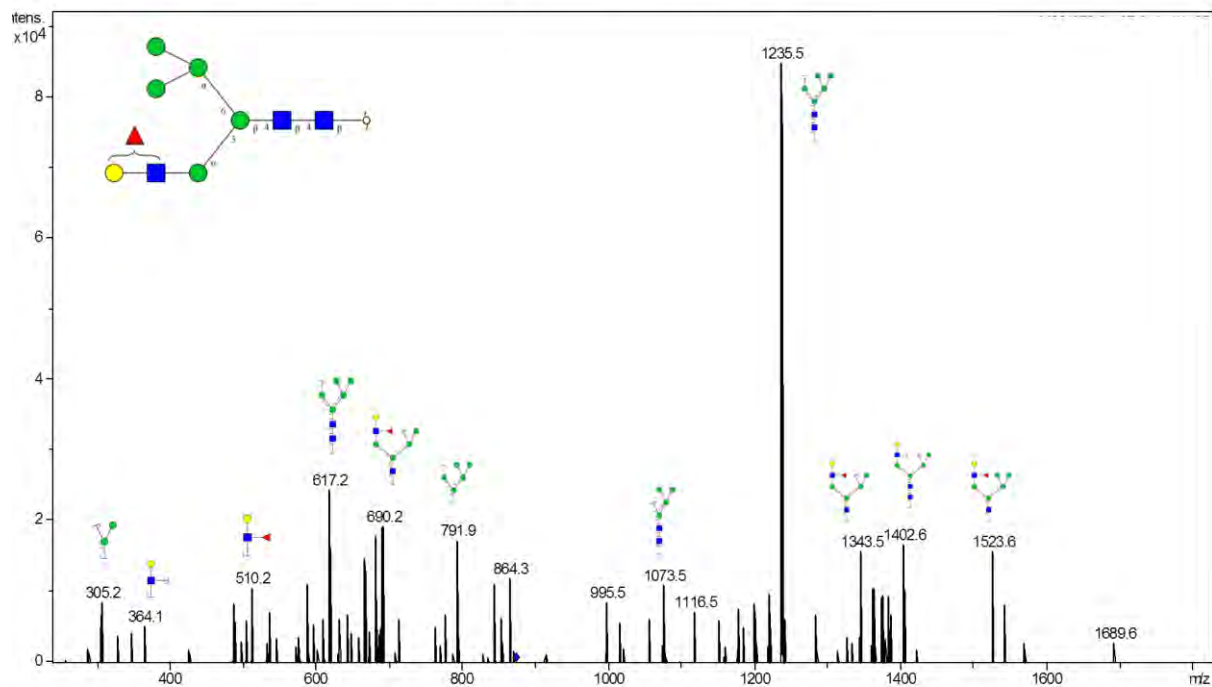
Glycan 46

Parent ion: *m/z* 872.9^{2-}
Composition: (Hex)$_3$ (HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
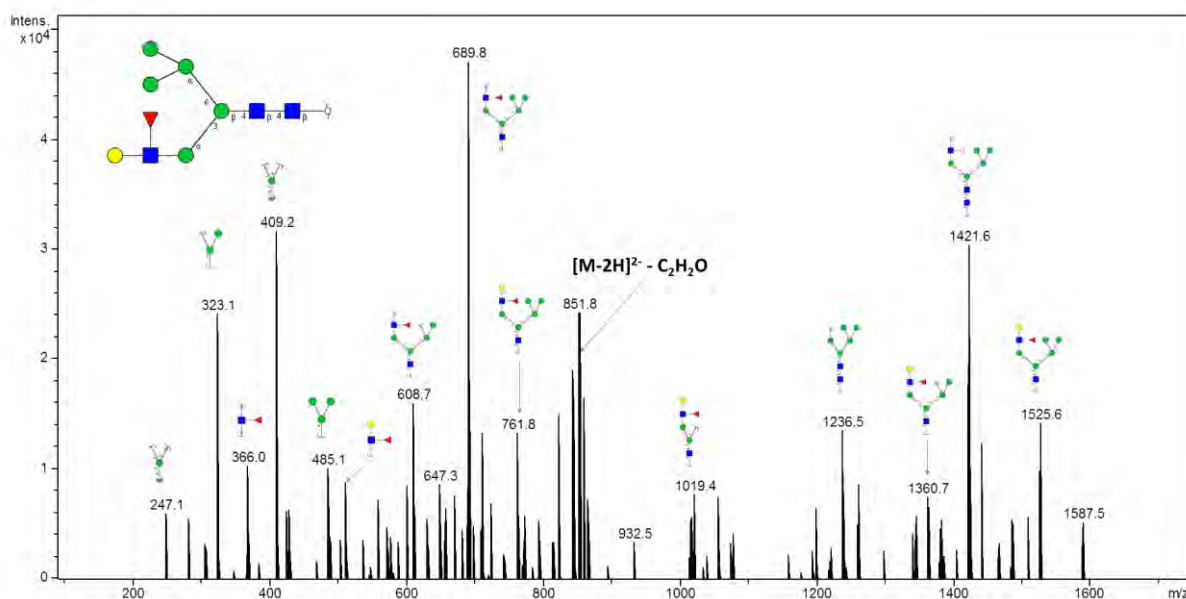
Glycan 47

Parent ion: *m/z* 872.9^{2-}
Composition: (Hex)$_3$ (HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
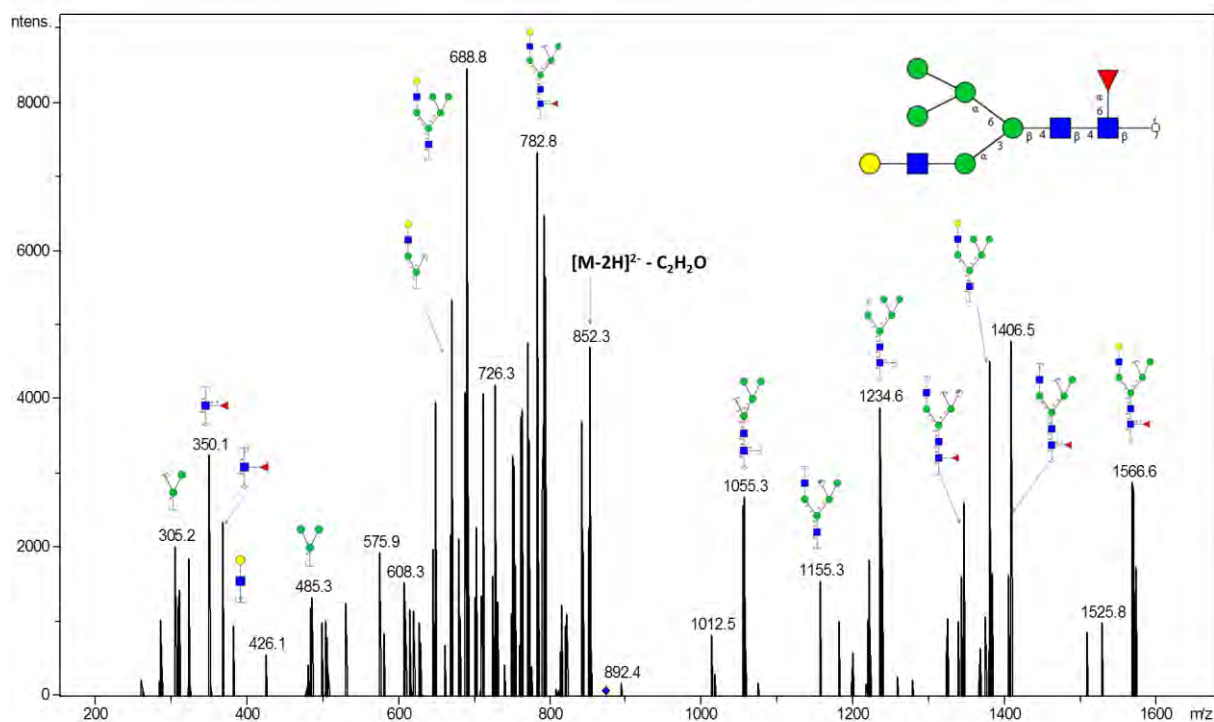
Glycan 48

Parent ion: *m/z* 885.4^{2-}

Composition: (Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$
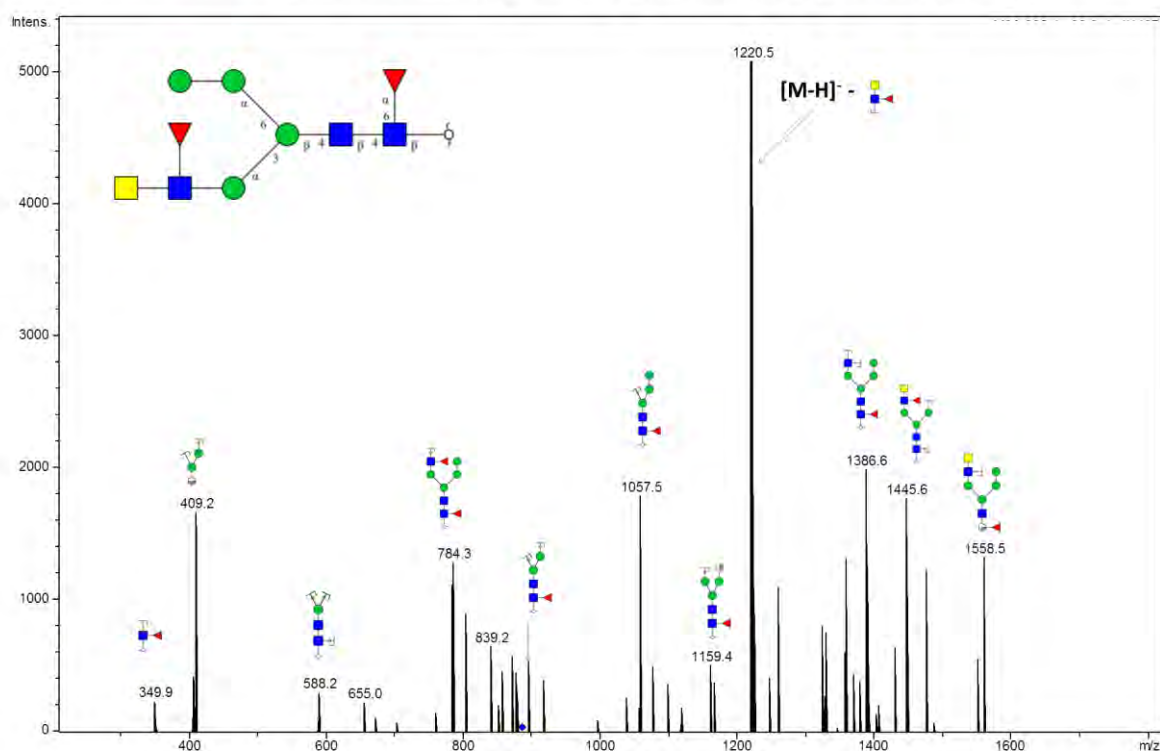
Glycan 49

Parent ion: *m/z* 885.4^{2-}

Composition: (Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Specific linkage of outer-arm fucose not intended in glycan fragmentation scheme.
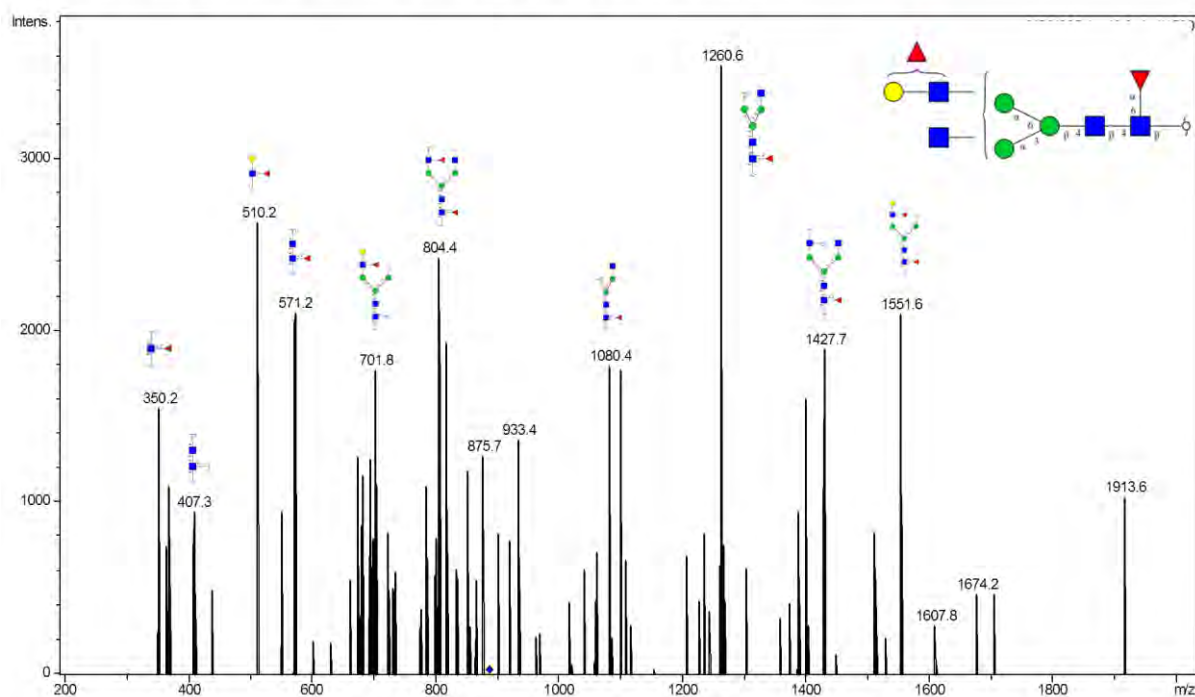
Glycan 50

Parent ion: *m/z* 885.4²⁻

Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragmentation scheme.
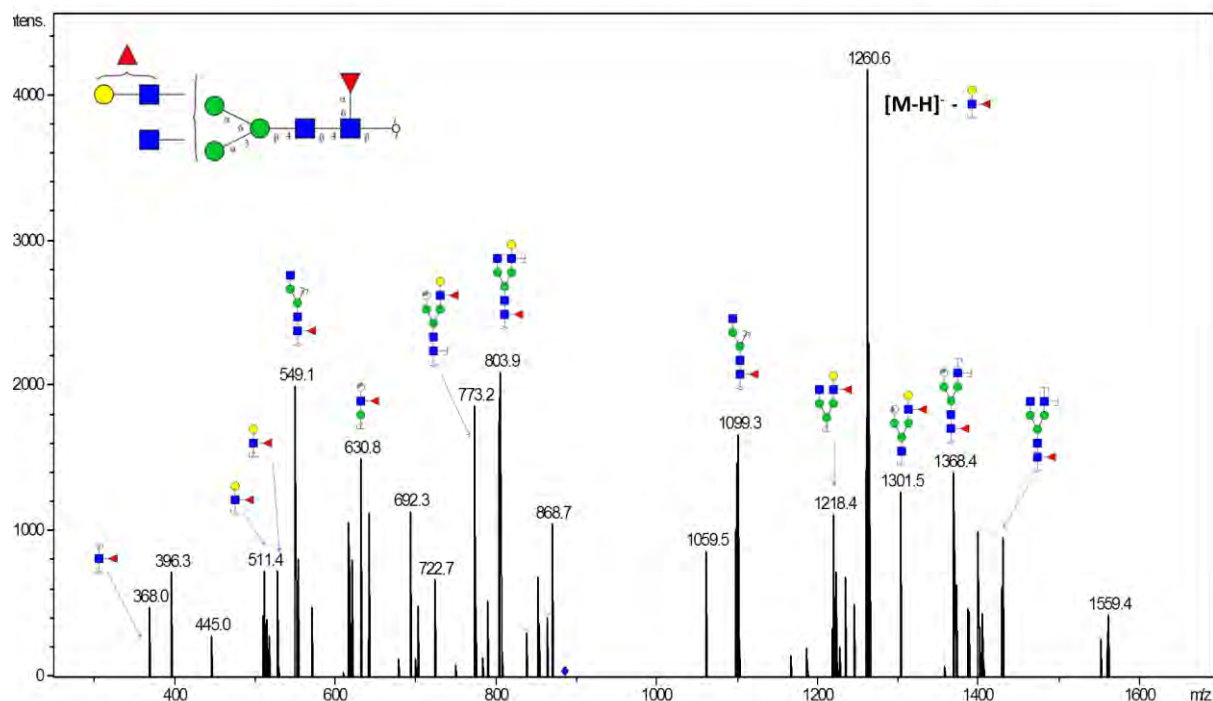
Glycan 51

Parent ion: *m/z* 893.4²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan fragment scheme.
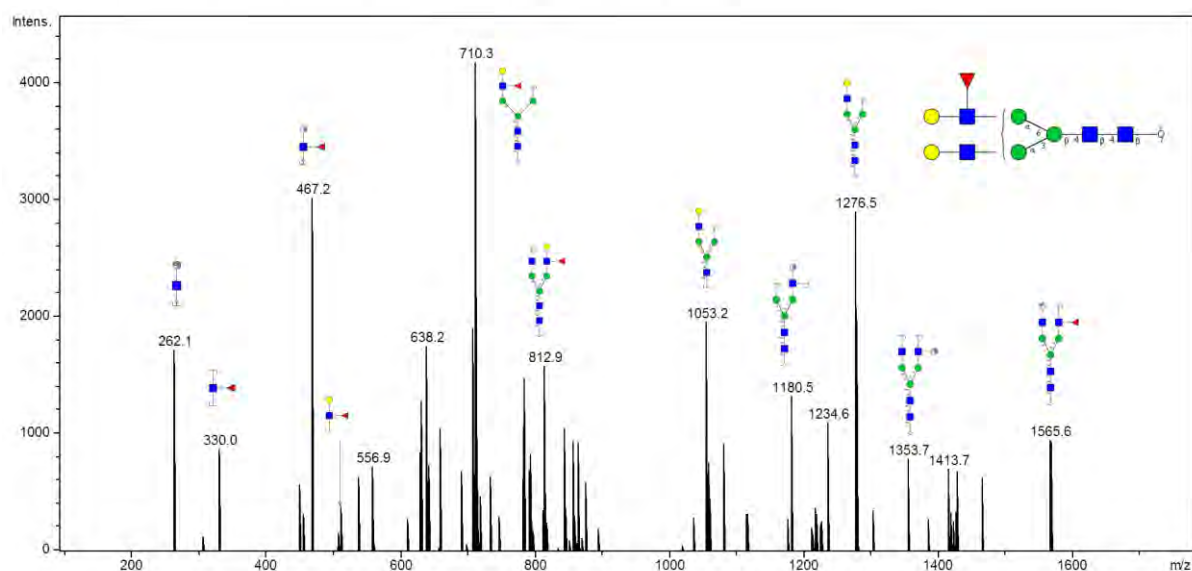
Glycan 52

Parent ion: *m/z* 893.4²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan fragment scheme.
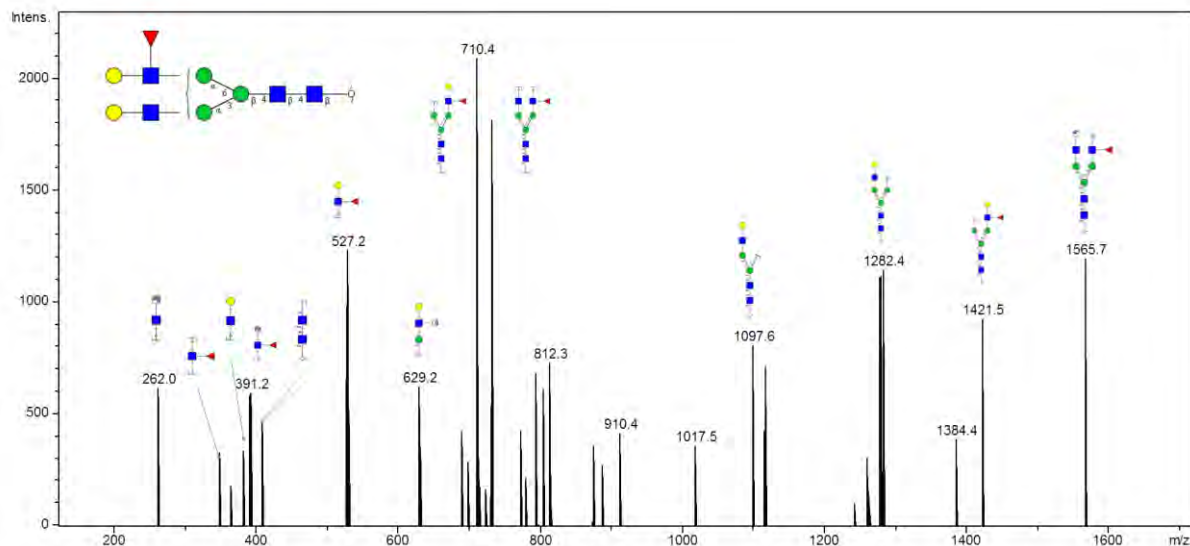
Glycan 53

Parent ion: *m/z* 893.4²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
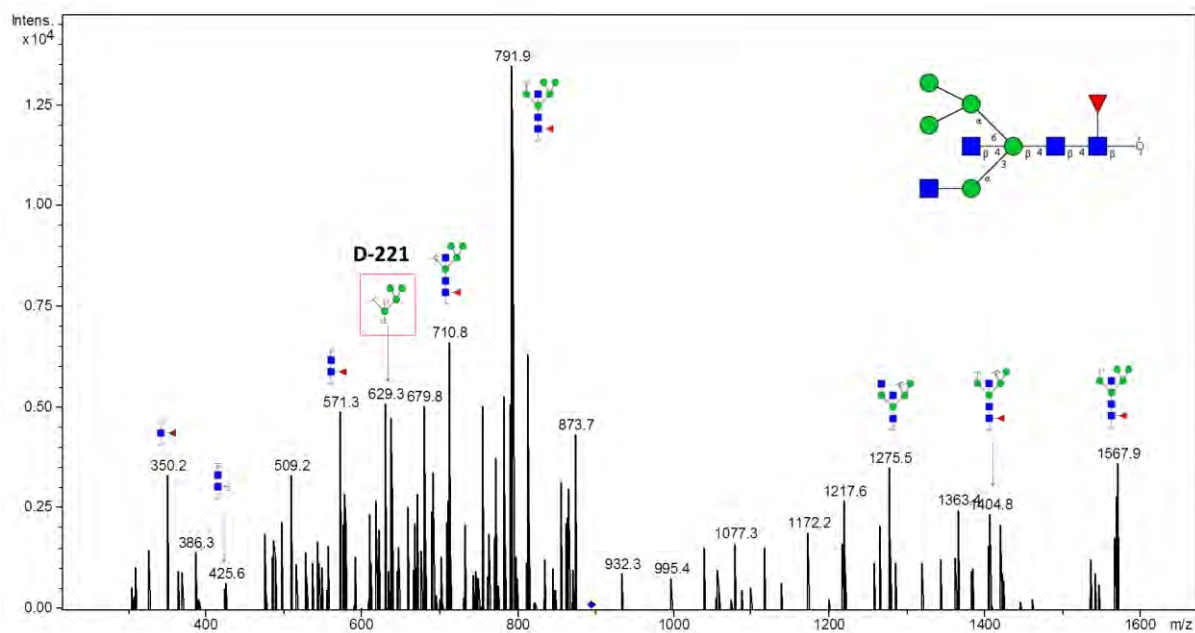
Glycan 54

Parent ion: *m/z* 893.4^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
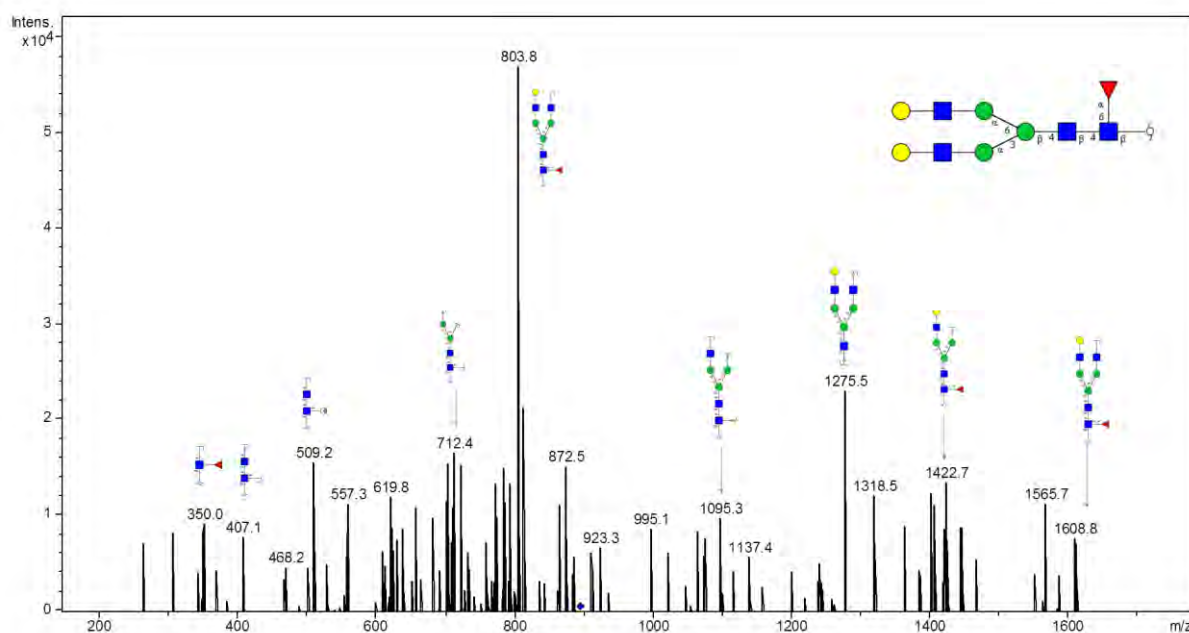
Glycan 55

Parent ion: *m/z* 893.4^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm and 6-arm in glycan fragment scheme.
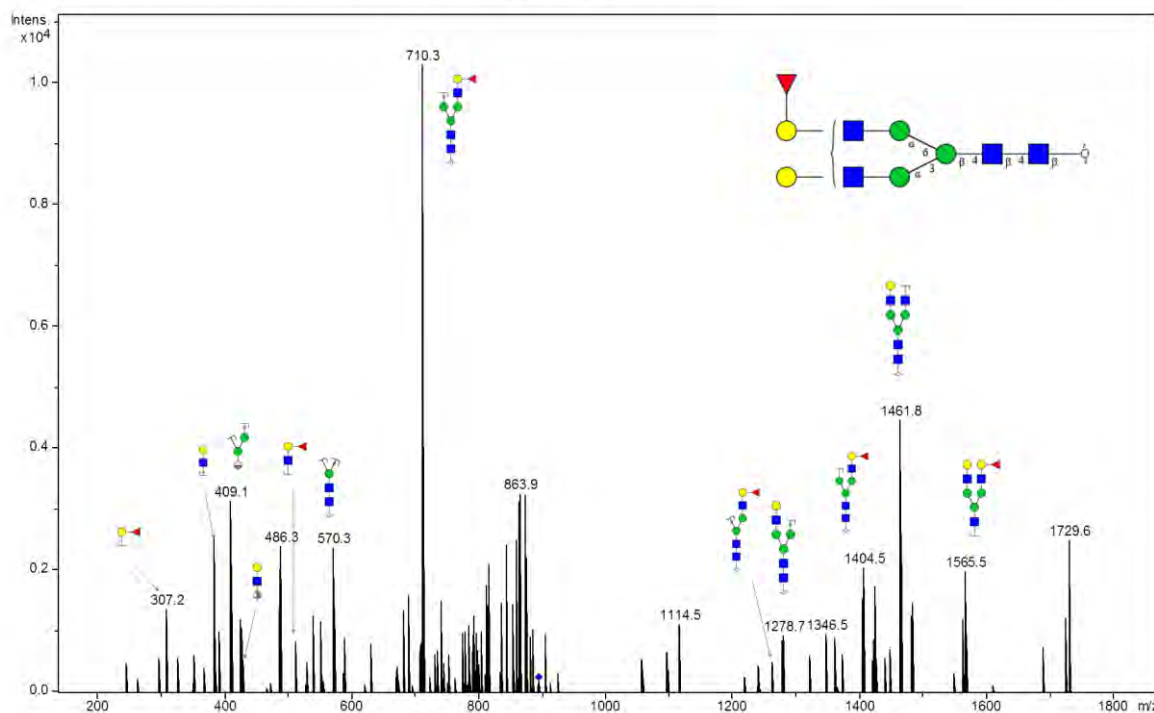
Glycan 56

Parent ion: *m/z* 893.4²⁻

Composition: $(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$
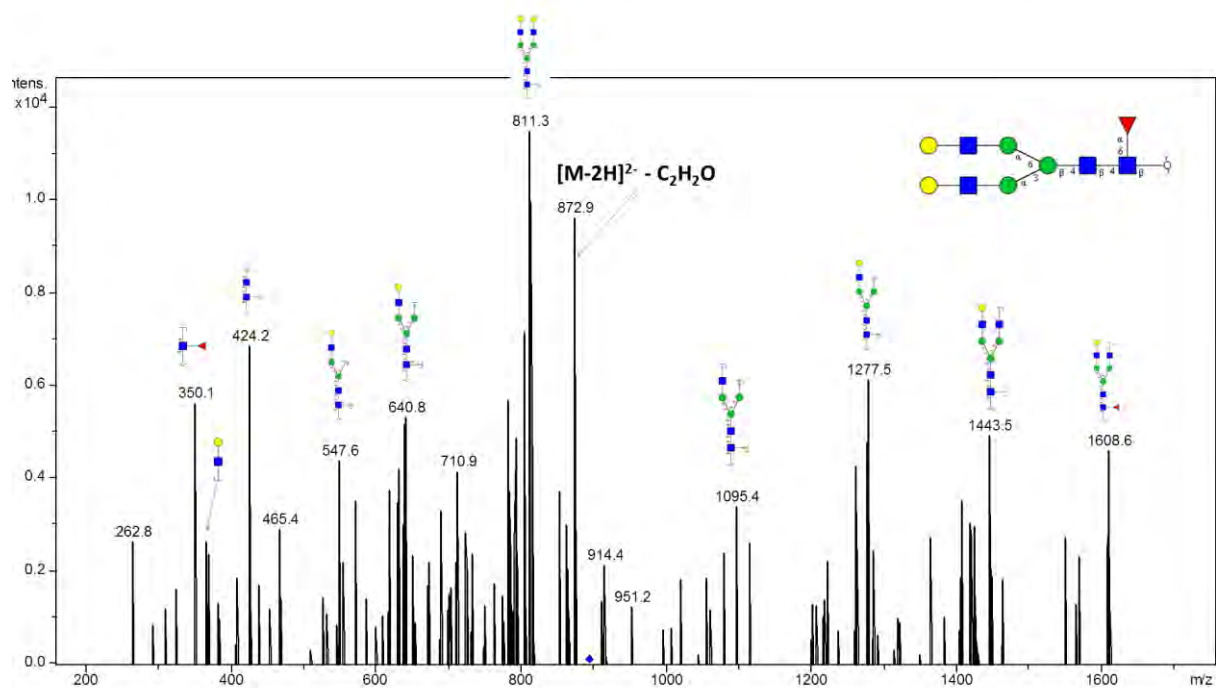
Glycan 57

Parent ion: *m/z* 893.4²⁻

Composition: $(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$
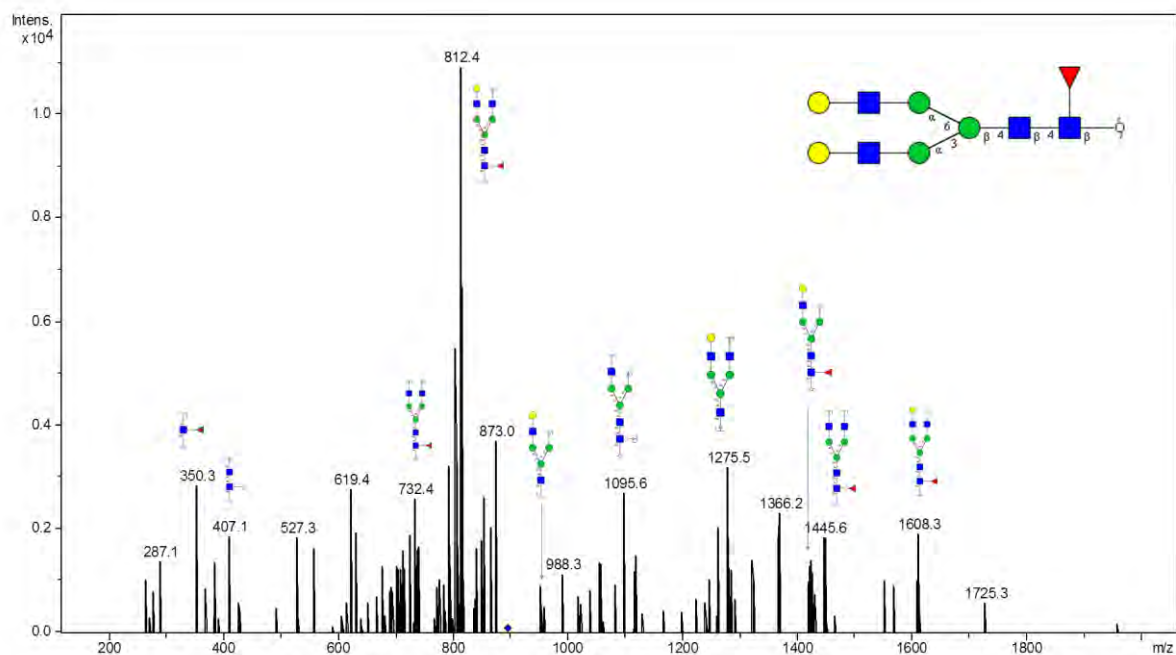
Glycan 58

Parent ion: *m/z* 893.4^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
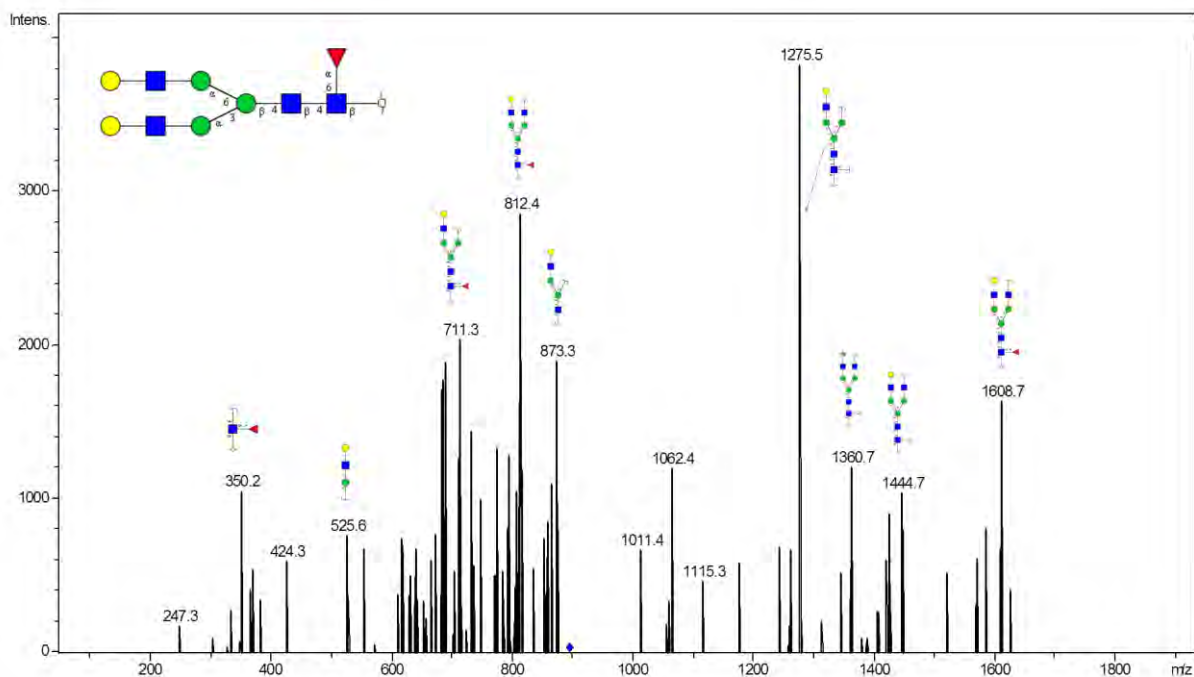
Glycan 59

Parent ion: *m/z* 901.4^{2-}

Composition: (Hex)$_3$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$
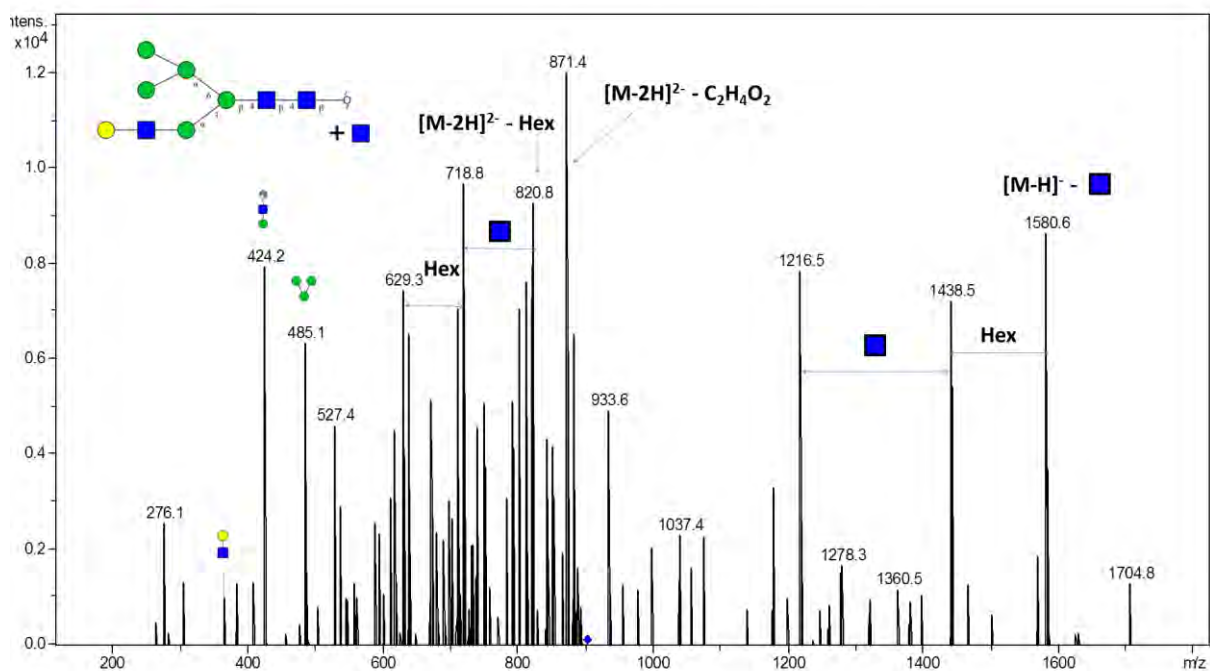
Notes: Hex, Hexose

Glycan 60

Parent ion: *m/z* 901.4^{2-}

Composition: (Hex)$_3$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Hex, Hexose; Fuc, Fucose

Intens.

871.6 [M-2H]$^{2-}$ - C$_2$H$_2$O

[M-2H]$^{2-}$ - Hex

616.8

535.8

718.7

665.3

383.1

789.3

1149.3

1438.5

1558.3

1661.9

467.0 - 18

1002.1

947.9

1504.5

1747.7

1215.5

1606.7

1702.7
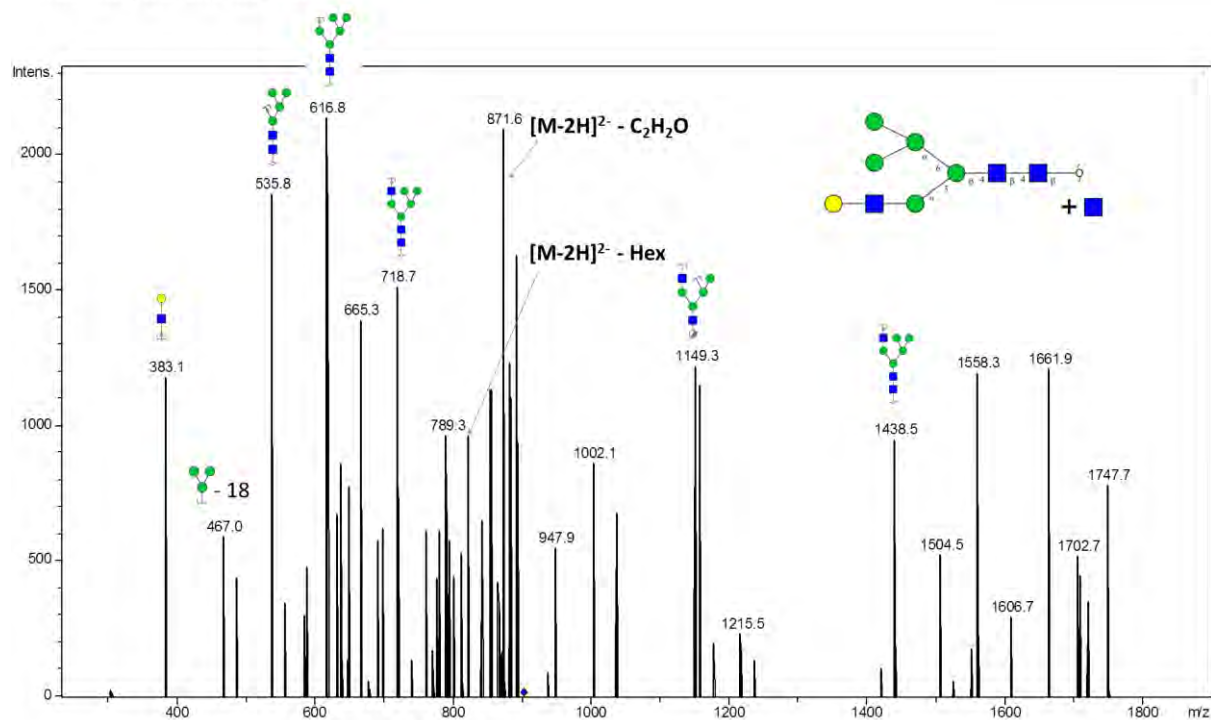
Glycan 61

Parent ion: *m/z* 901.4^{2-}

Composition: (Hex)$_3$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Hex, Hexose; Fuc, Fucose

itens.
x10^4

820.3 [M-2H]$^{2-}$ - Hex

718.3

[M-2H]$^{2-}$ - C$_2$H$_4$O$_2$

871.3

467.1 - H$_2$O

424.2

637.3

1054.6

[M-H]$^-$ -

[M-H]$^-$ - Hex

304.1

550.1

1169.9

1277.6 Hex 1419.7

1623.6

1702.8

952.3

1317.7
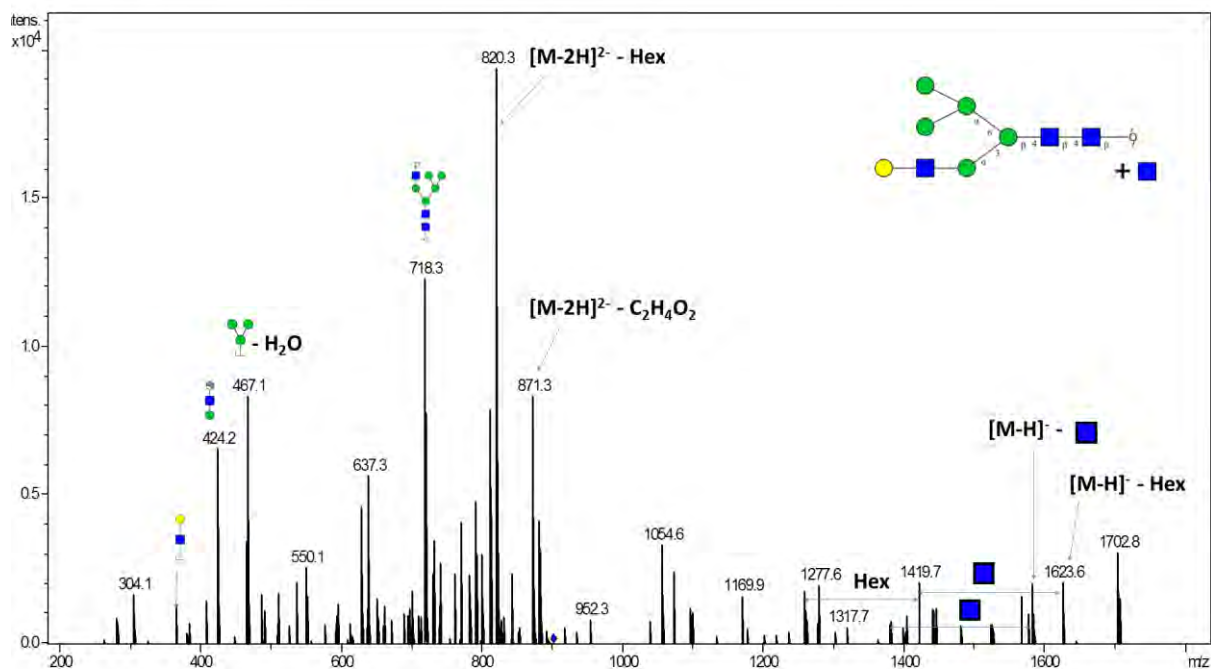
Glycan 62

Parent ion: *m/z* 901.4^{2-}

Composition: (Hex)$_3$ (HexNAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Hex, Hexose; Fuc, Fucose

Labeled peaks and annotations: 262.2, 323.1, 485.1, 535.7, 629.3, 718.8, 811.3 [M-2H]$^{2-}$ - Hex, [M-2H]$^{2-}$ - C$_2$H$_4$O$_2$ 871.4, 936.2, 1072.3, 1235.5, 1401.3, 1475.8, 1580.5, 1624.8, [M-H]$^-$ - Hex, [M-H]$^-$ - ■, Hex
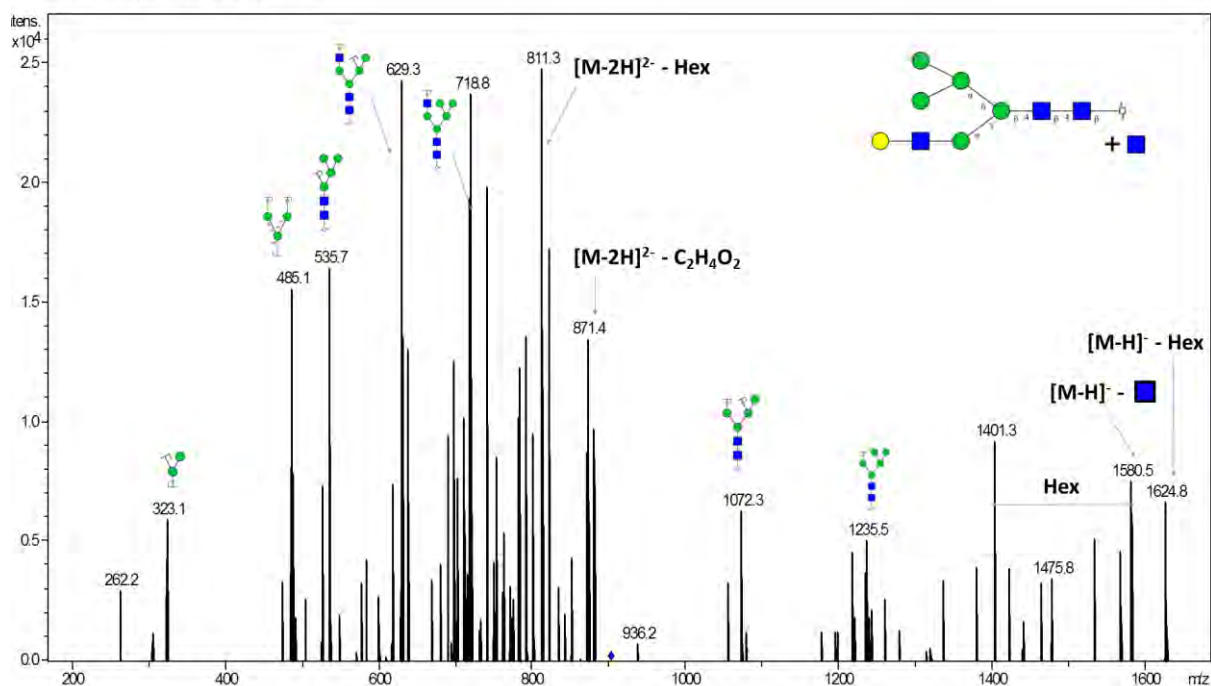
Glycan 63

Parent ion: *m/z* 913.9^{2-}

Composition: (Hex)$_1$ (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction between 3-arm/6-arm intended in glycan fragment scheme. Hex, Hexose; Fuc, Fucose

Labeled peaks and annotations: 262.3, 364.1, 425.2, 491.2, 527.2, 571.8, 611.2, 658.3, 730.7, 816.3 D - 221, 884.3, 1037.5, 1115.5, 1260.4, 1317.6, 1455.8, 1608.5, [M-H]$^-$ - ■, Fuc, Hex
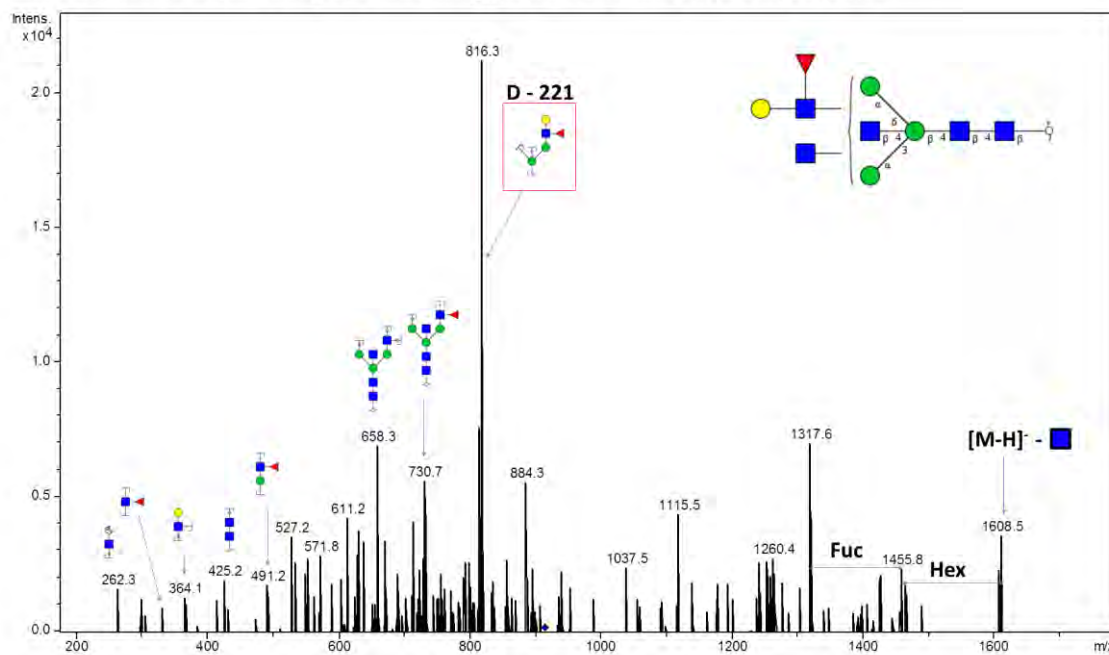
Glycan 64

Parent ion: *m/z* 913.9²⁻

Composition: (Hex)₁ (HexNAc)₃ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Hex, Hexose; Fuc, Fucose
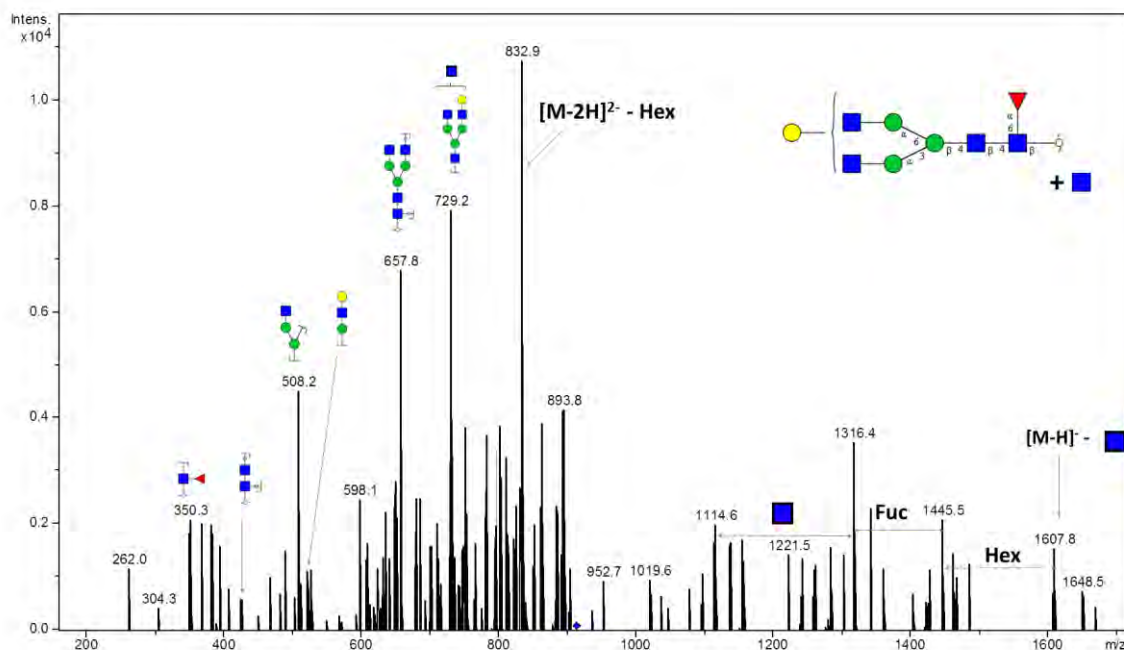
Glycan 65

Parent ion: *m/z* 913.9²⁻

Composition: (Hex)₁ (HexNAc)₃ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan fragment scheme.
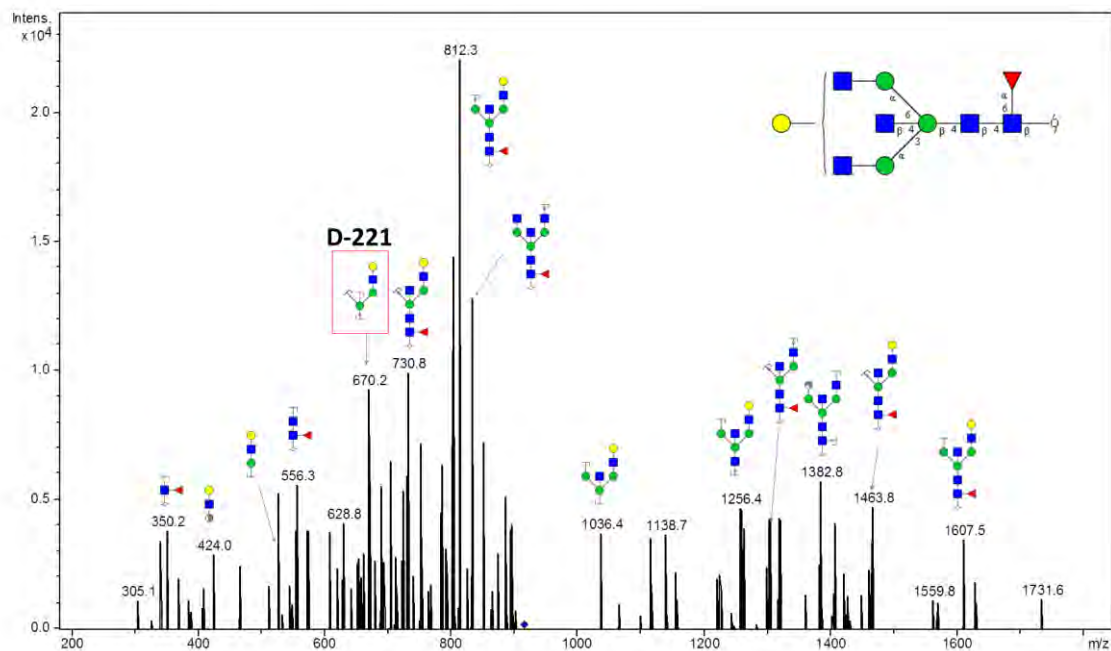
Glycan 66

Parent ion: *m/z* 921.9²⁻
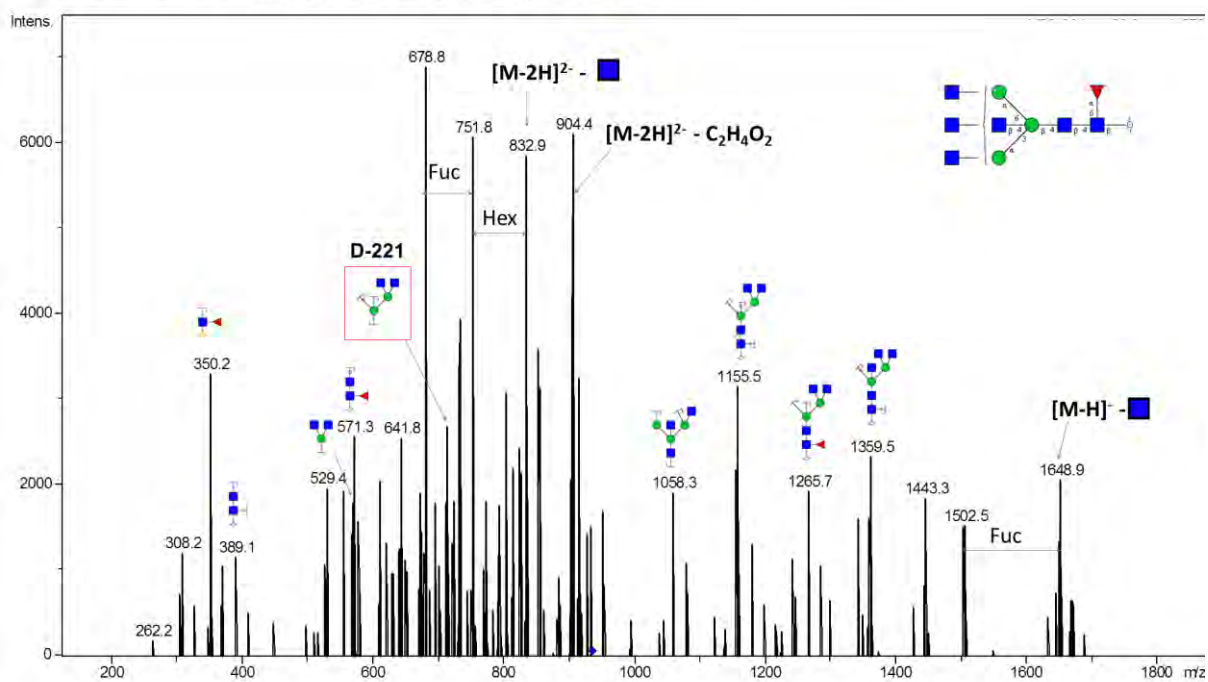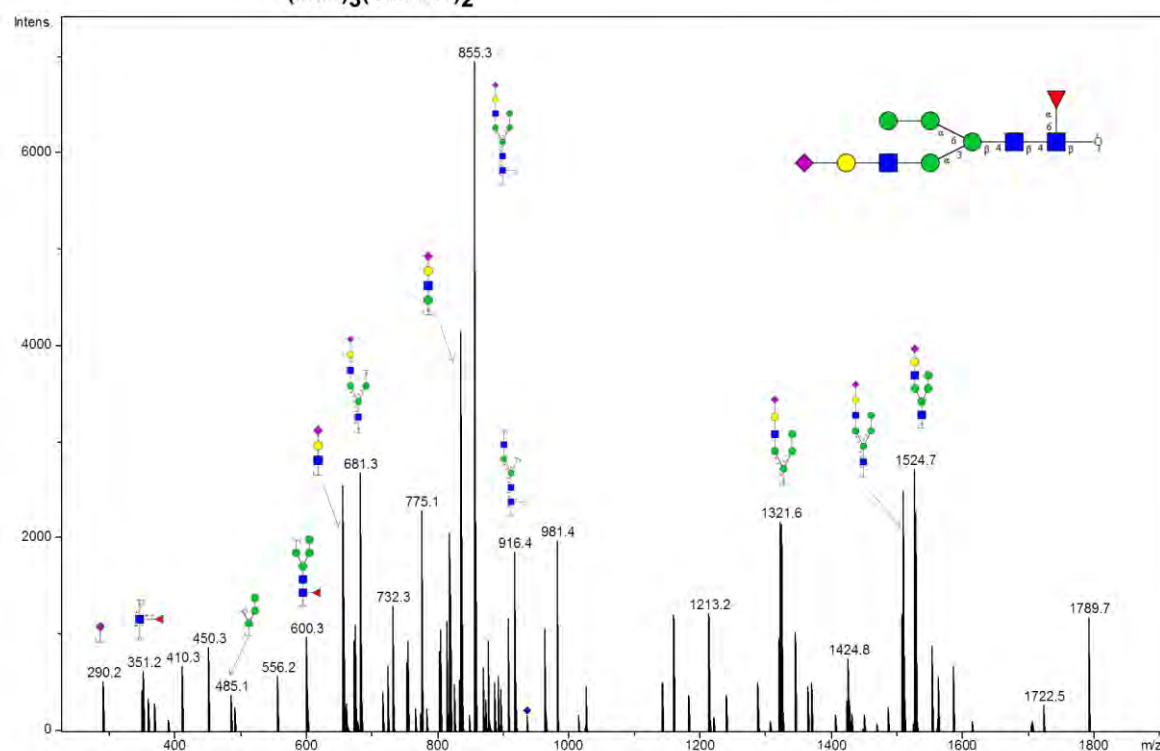
Composition: (Hex)₂ (HexNAc)₃ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan fragment scheme.
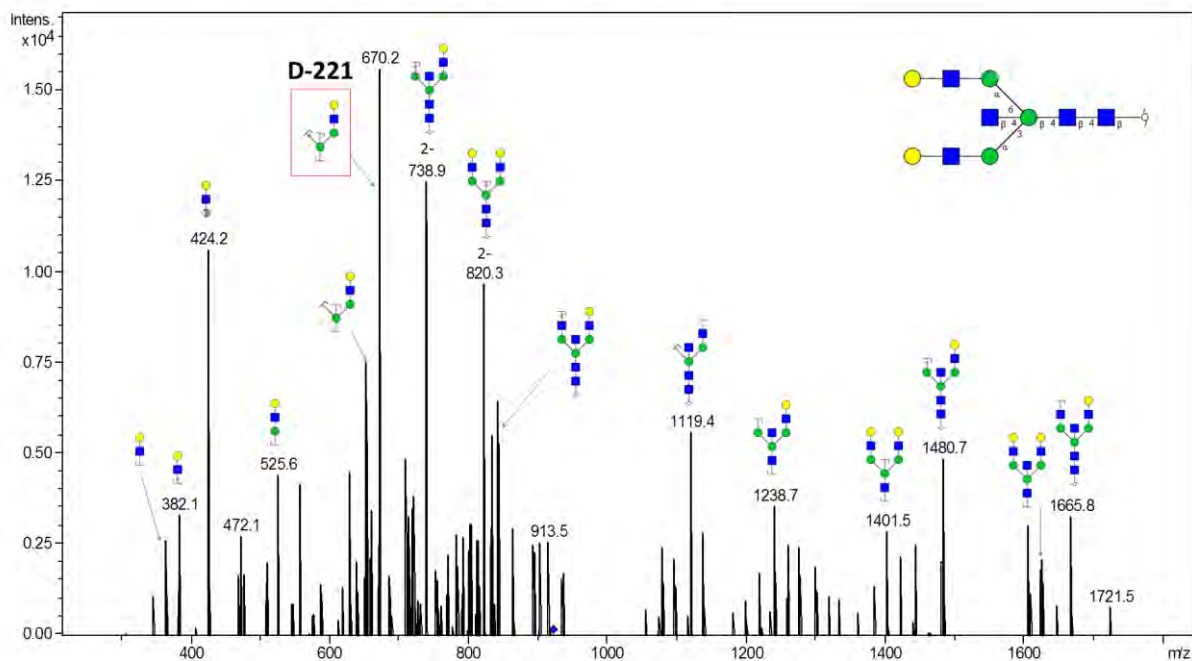
Glycan 67

Parent ion: *m/z* 934.3²⁻

Composition: (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended for 3-arm/6-arm in glycan fragment scheme.

Glycan 68

Parent ion: *m/z* 934.3²⁻

Composition: (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
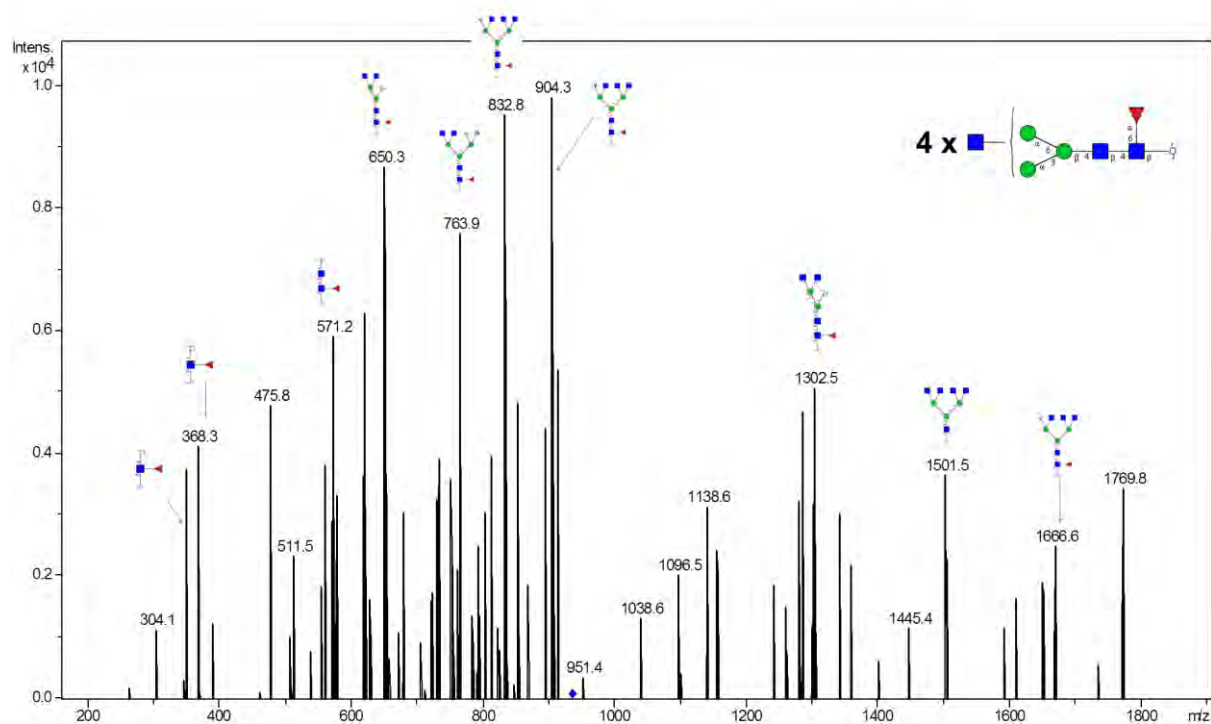
Glycan 69

Parent ion: *m/z* 937.3²⁻

Composition: (Hex)₂ (HexNAc)₁ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Glycan 70

Parent ion: *m/z* 937.3^{2-}
Composition: (Hex)$_2$ (HexNAc)$_1$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
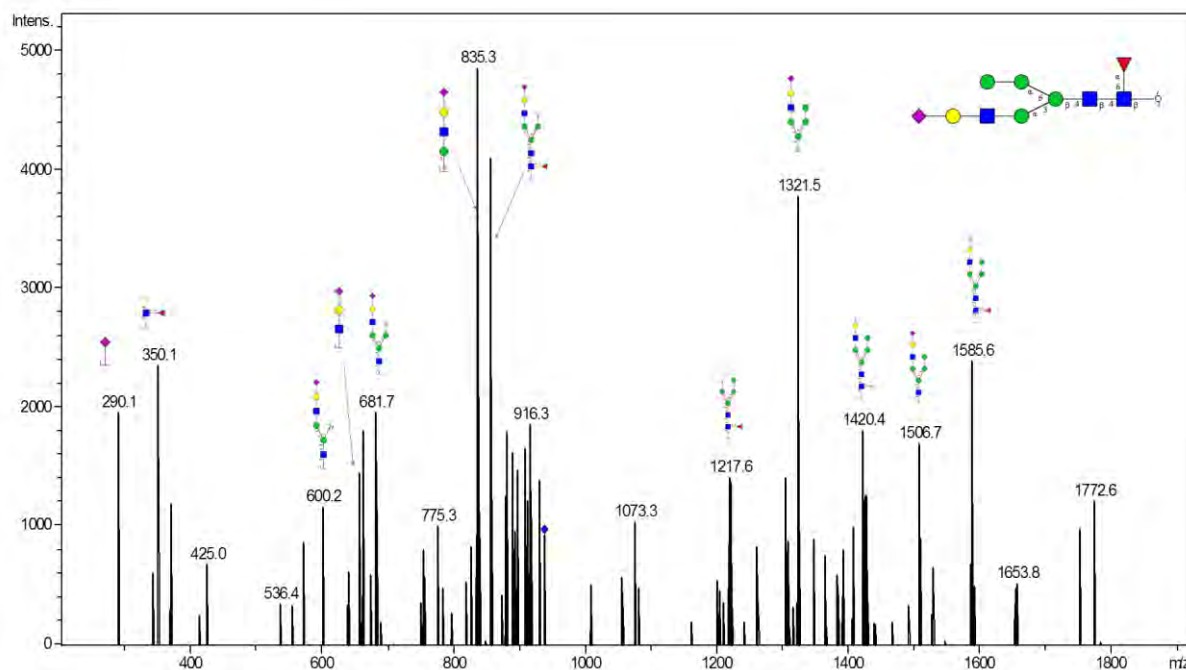
Glycan 71

Parent ion: *m/z* 941.4^{2-}
Composition: (Hex)$_6$ + (Man)$_3$(GlcNAc)$_2$
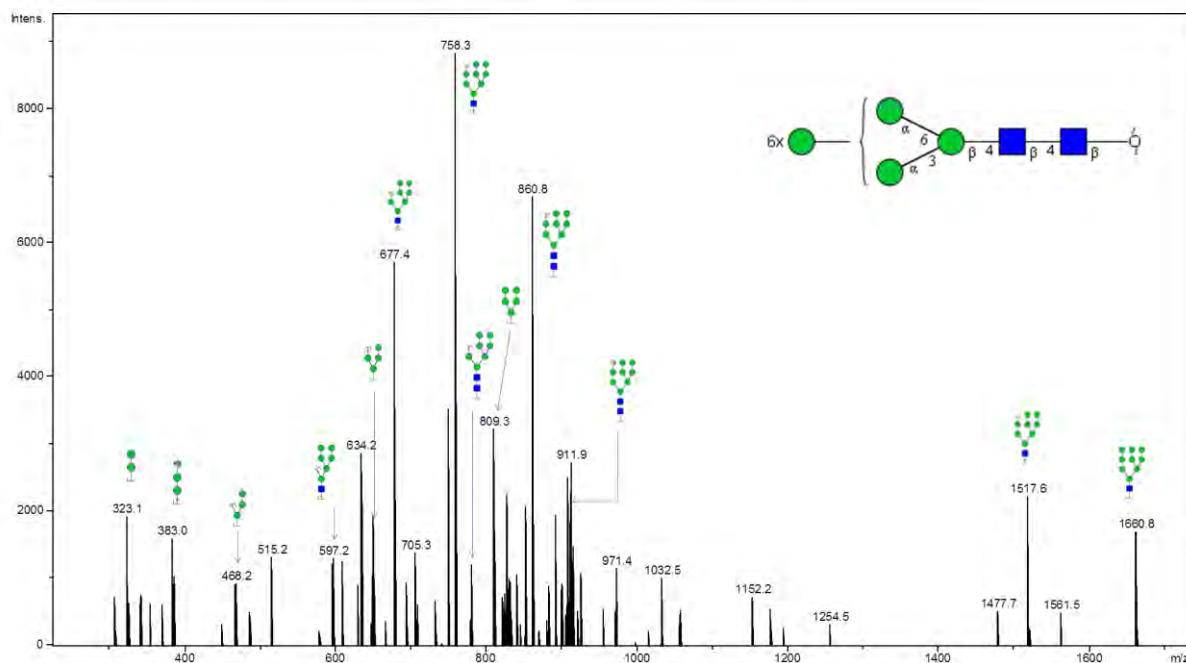
Glycan 72

Parent ion: *m/z* 941.4²⁻
Composition: (Hex)₆ + (Man)₃(GlcNAc)₂
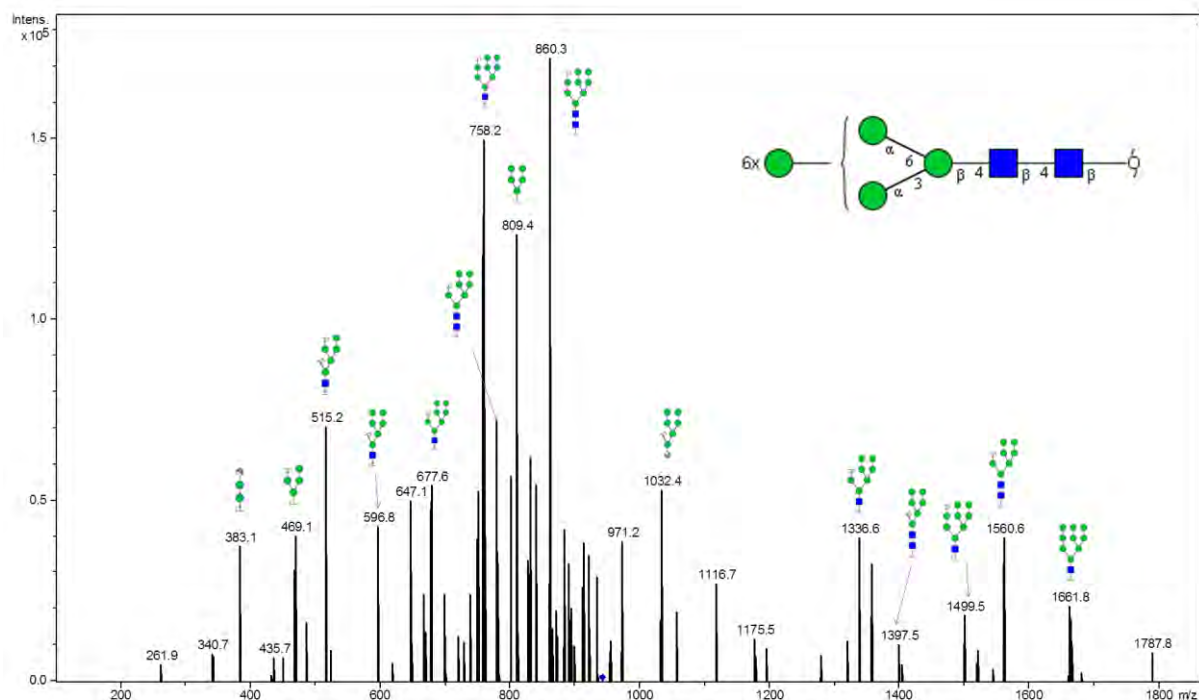
Glycan 73

Parent ion: *m/z* 941.4²⁻
Composition: (Hex)₆ + (Man)₃(GlcNAc)₂
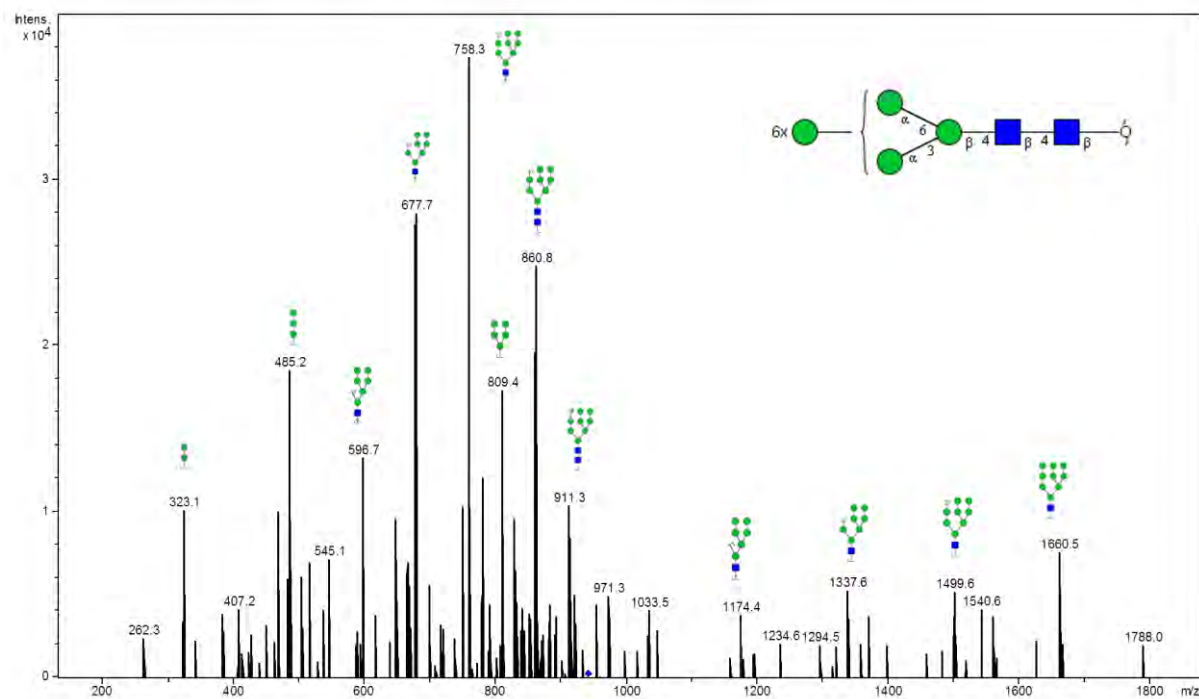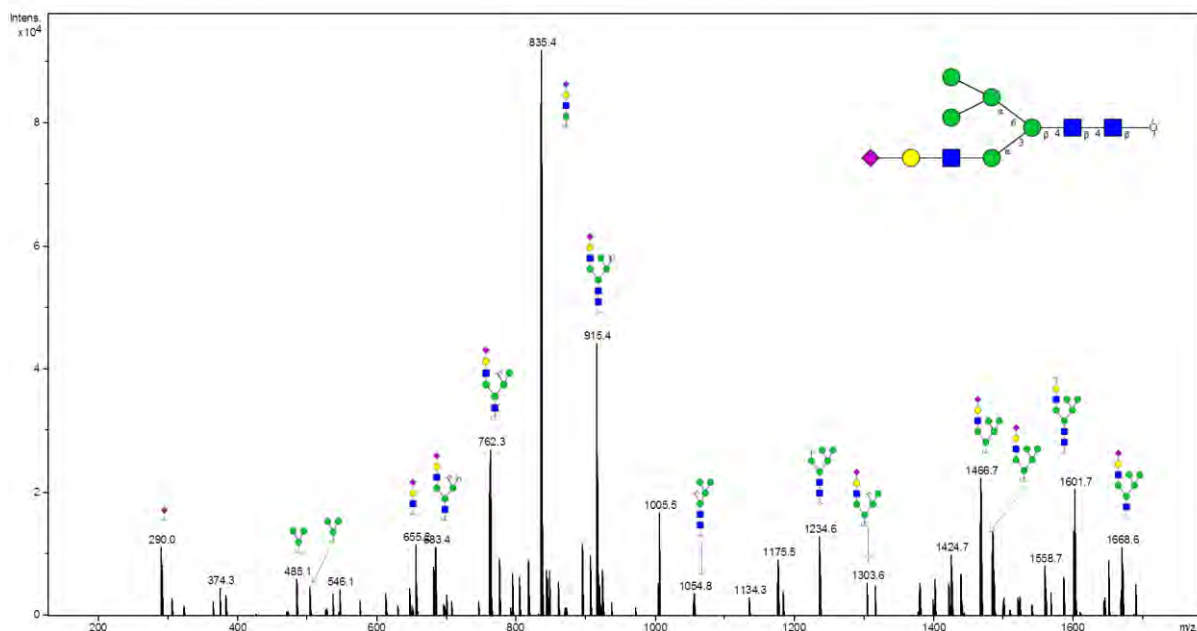
Glycan 74

Parent ion: *m/z* 945.4[2-]
Composition: (Hex)₃ (HexNAc)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂
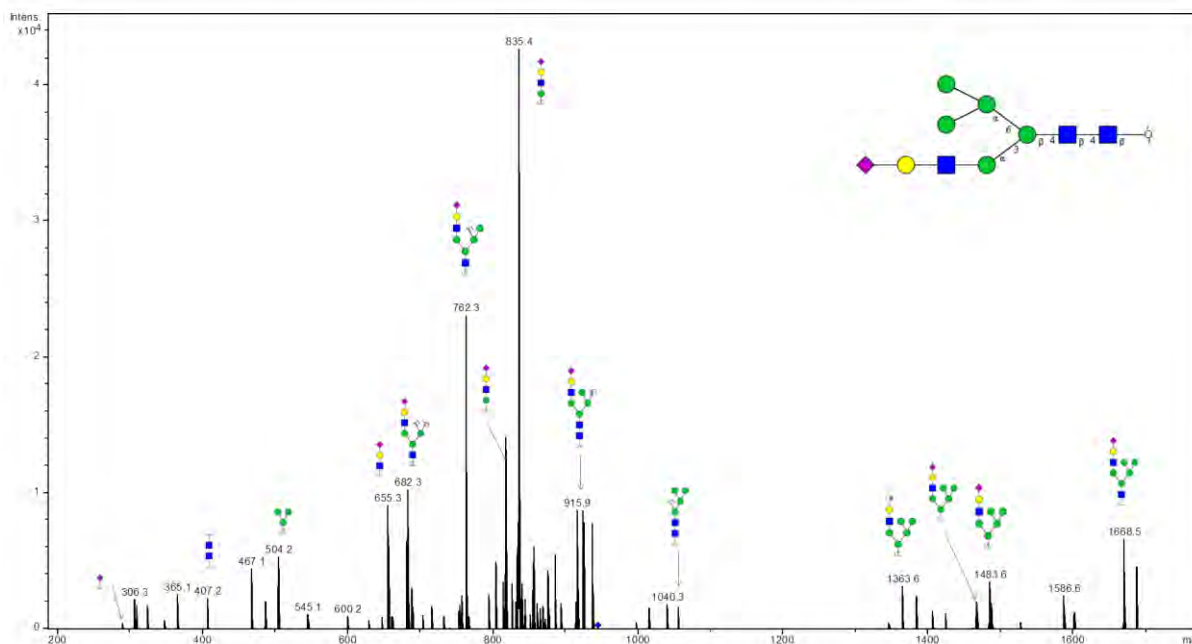
Glycan 75

Parent ion: *m/z* 945.4[2-]
Composition: (Hex)₃ (HexNAc)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Glycan 76

Parent ion: *m/z* 945.9^{2-}
Composition: (Hex)$_3$ (HexNAc)$_1$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$
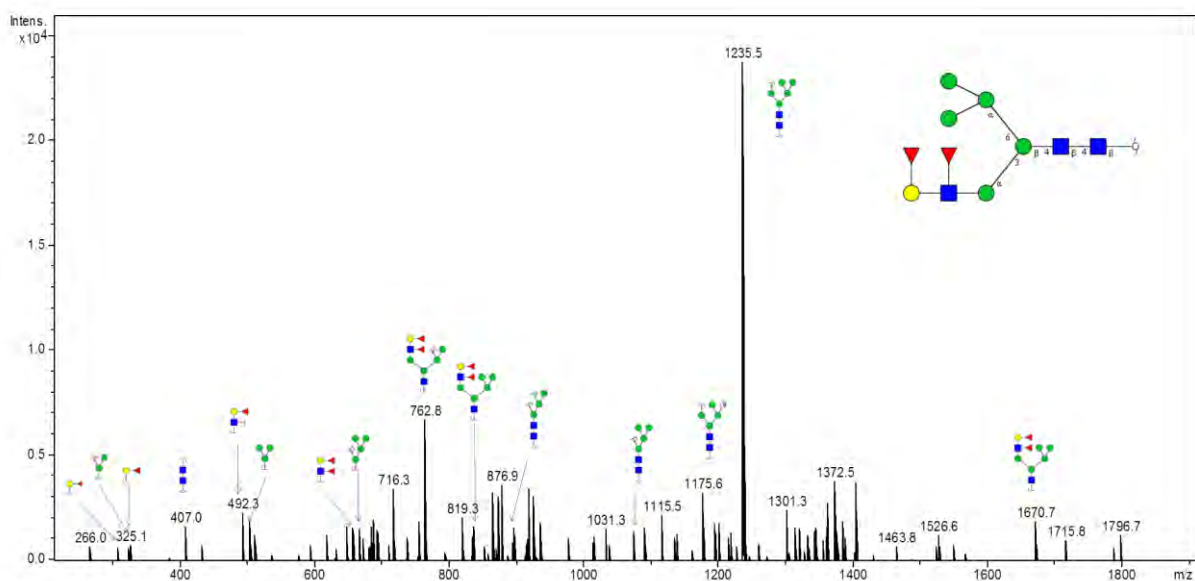
Glycan 77

Parent ion: *m/z* 953.8^{2-}
Composition: (Hex)$_4$ (HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
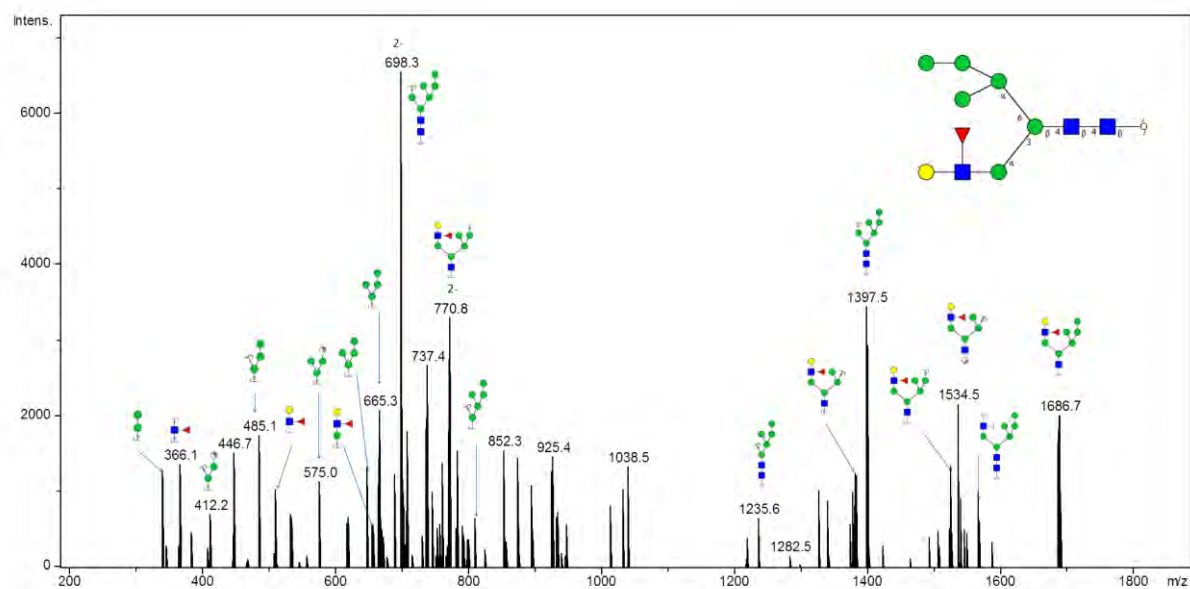
Glycan 78

Parent ion: *m/z* 953.8^{2-}
Composition: (Hex)$_4$ (HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Specific linkage of outer-arm fucose not intended in fragment scheme. (i.e. fucose can be on Gal or GlcNac)
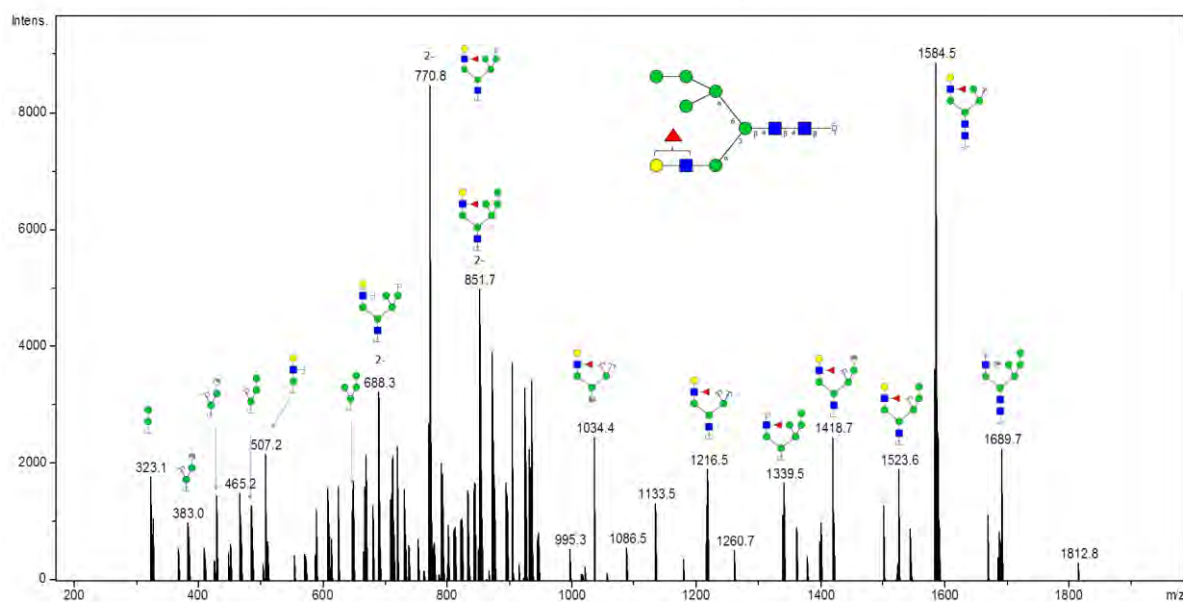
Glycan 79

Parent ion: *m/z* 957.8^{2-}
Composition: (Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended in between 3-arm and 6-arm in glycan fragment scheme
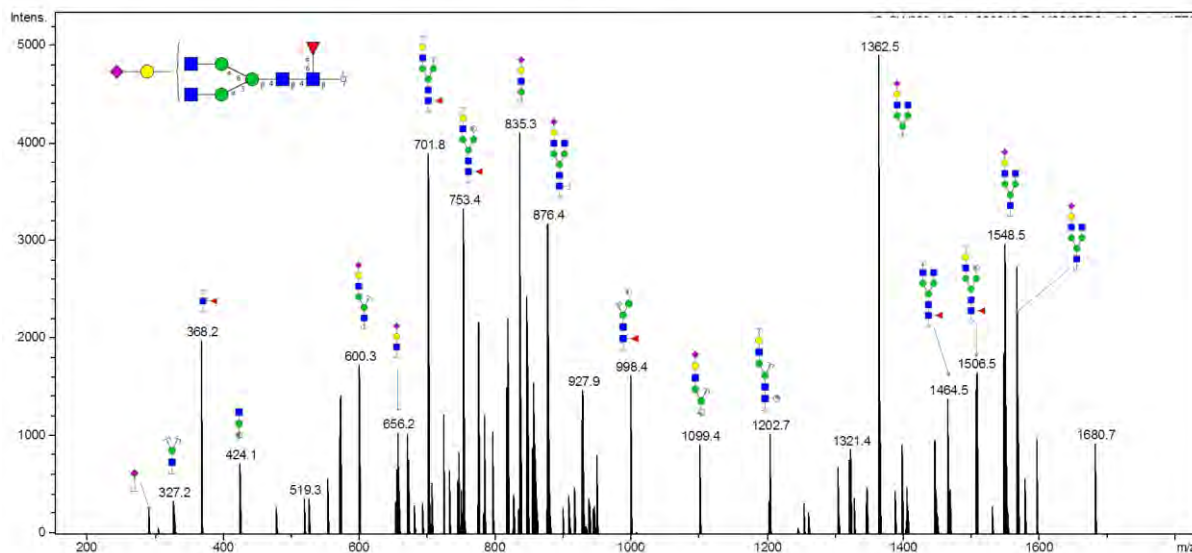
Glycan 80

Parent ion: *m/z* 957.8^{2-}
Composition: (Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended in between 3-arm and 6-arm in glycan fragment scheme

Glycan 81

Parent ion: *m/z* 965.9^{2-}
Composition: (Hex)$_2$ (HexNAc)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended in between 3-arm and 6-arm in glycan fragment scheme

158

Glycan 82

Parent ion: *m/z* 965.9²⁻
Composition: (Hex)₂ (HexNAc)₂ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended in between 3-arm and 6-arm in glycan fragment scheme
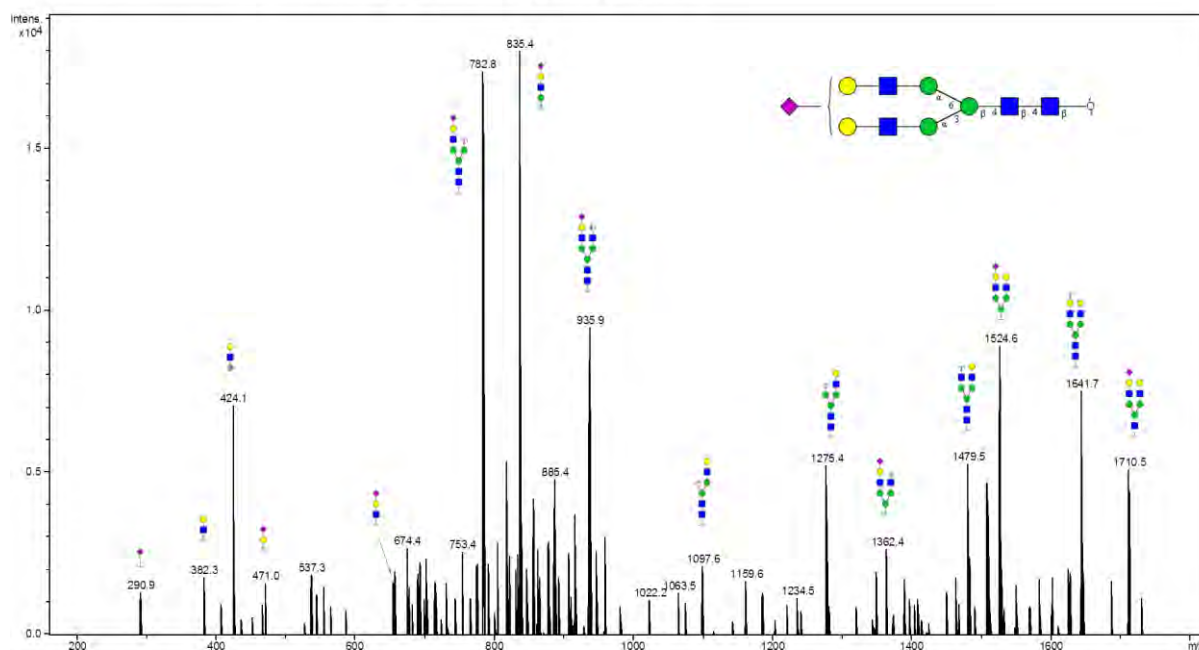
Glycan 83

Parent ion: *m/z* 966.5²⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: No distinction intended in between 3-arm and 6-arm in glycan fragment scheme

159
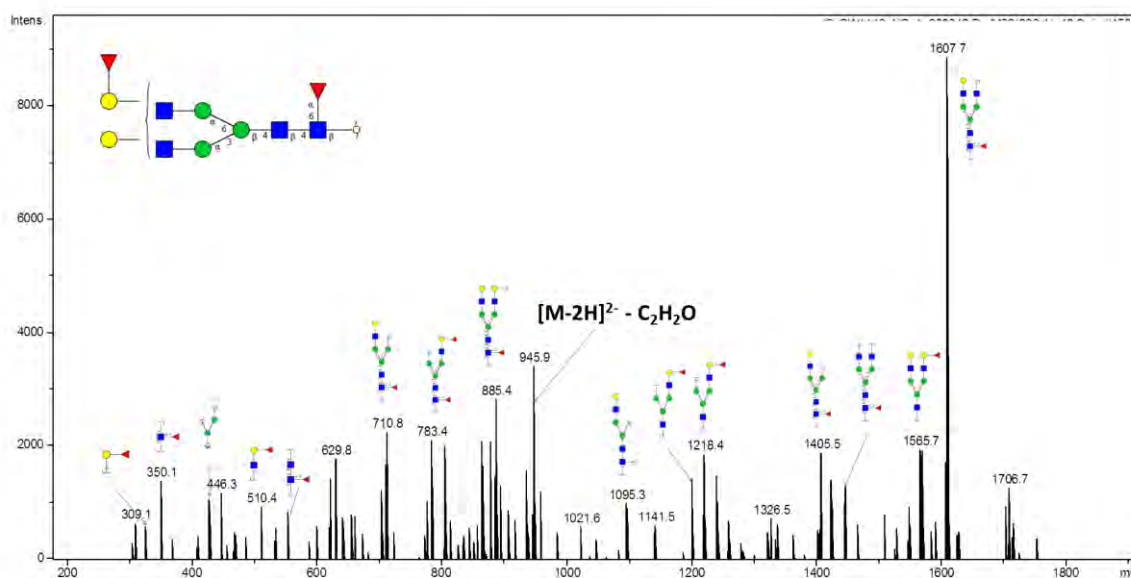
Glycan 84

Parent ion: *m/z* 966.5²⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction intended in between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme

Glycan 85

Parent ion: *m/z* 966.5²⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

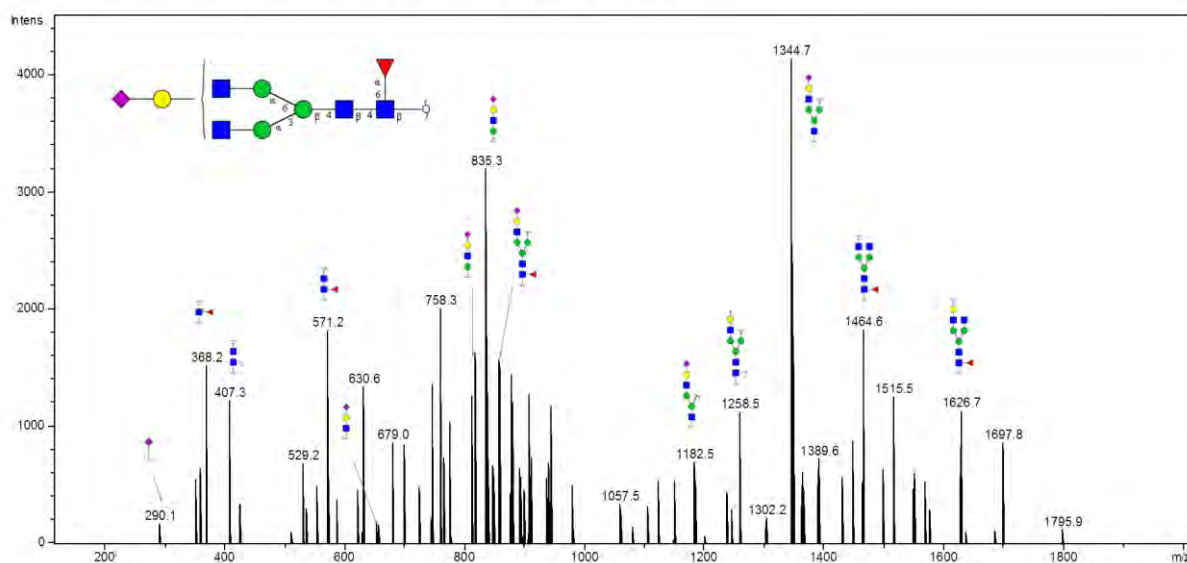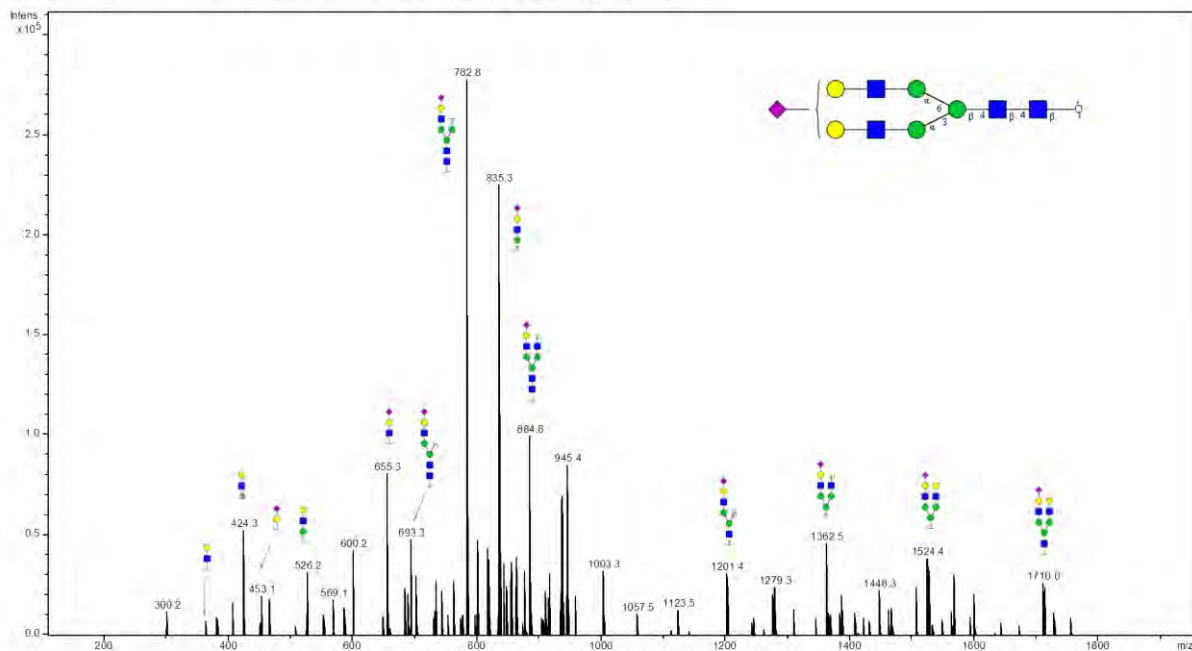Notes: No distinction intended in between 3-arm and 6-arm in glycan fragment scheme. Depiction of specific linkage of outer-arm fucose not intended.
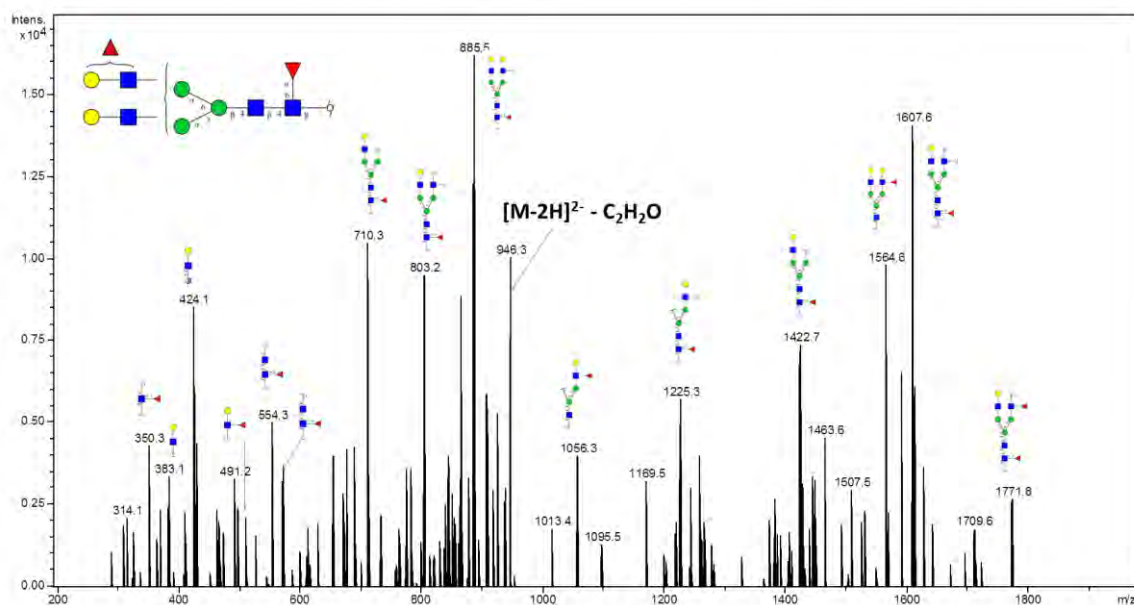
Glycan 86

Parent ion: *m/z* 974.4[2-]
Composition: (Hex)₃ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Depiction of specific linkage of outer-arm fucose not intended in fragment scheme (i.e. may be linked to Gal or GlcNAc)
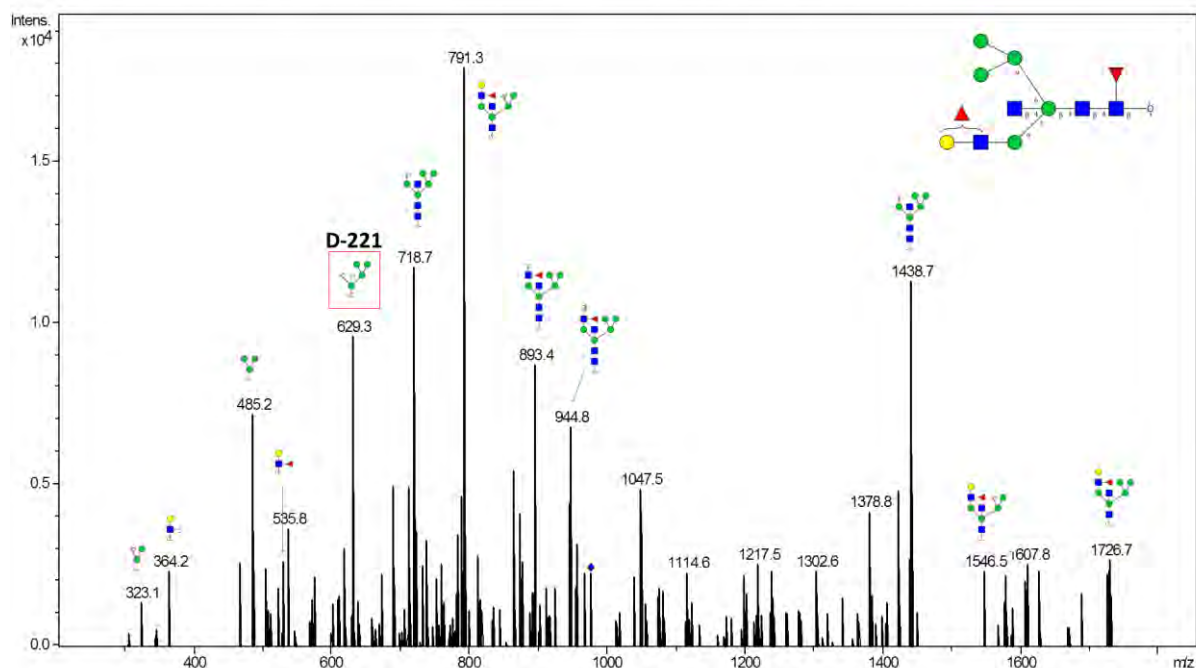
Glycan 87

Parent ion: *m/z* 974.4[2-]
Composition: (Hex)₃ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Depiction of specific linkage of outer-arm fucose not intended in fragment scheme.
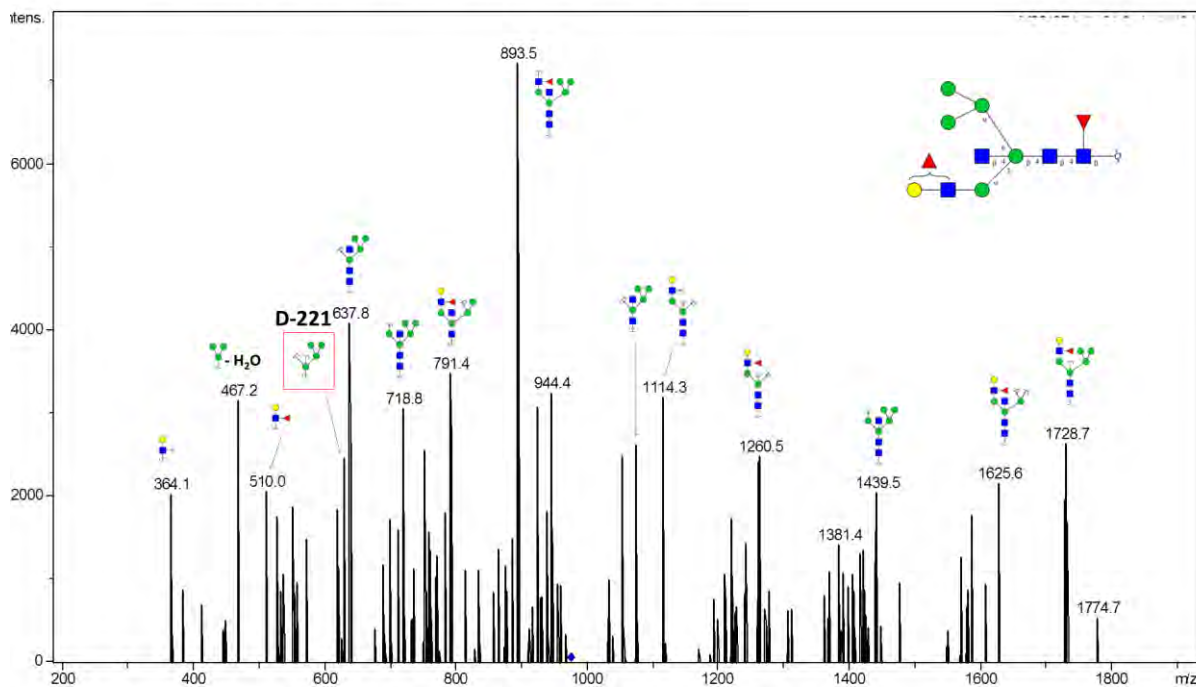
Glycan 88

Parent ion: *m/z* 974.4²⁻
Composition: (Hex)₃ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
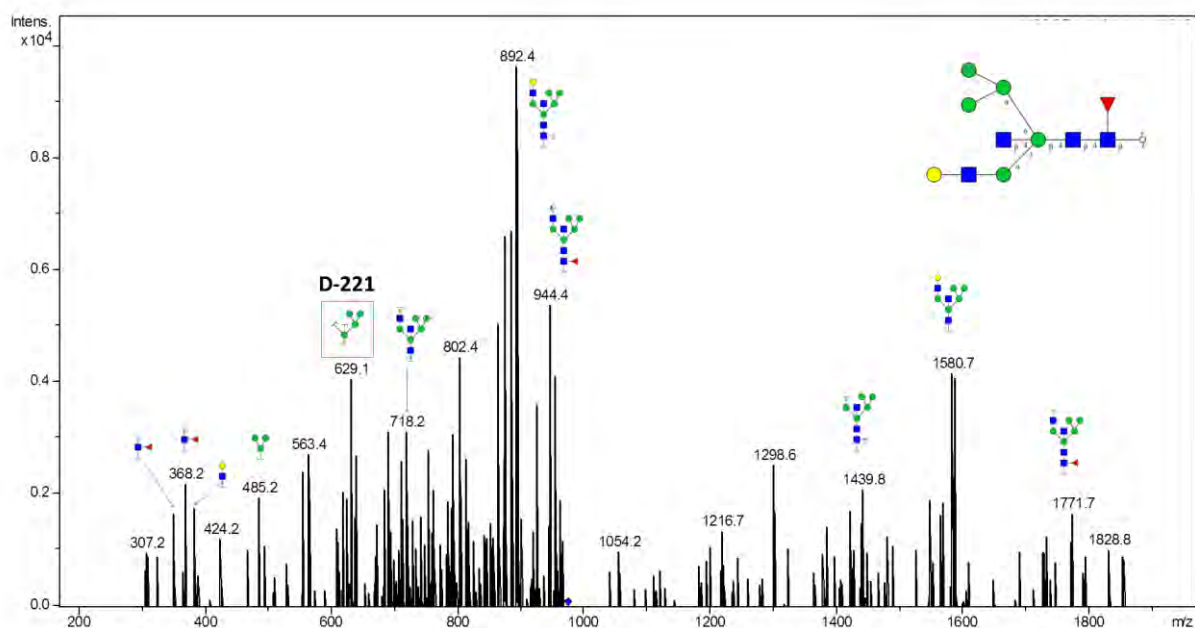
Glycan 89

Parent ion: *m/z* 974.4²⁻
Composition: (Hex)₃ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

162
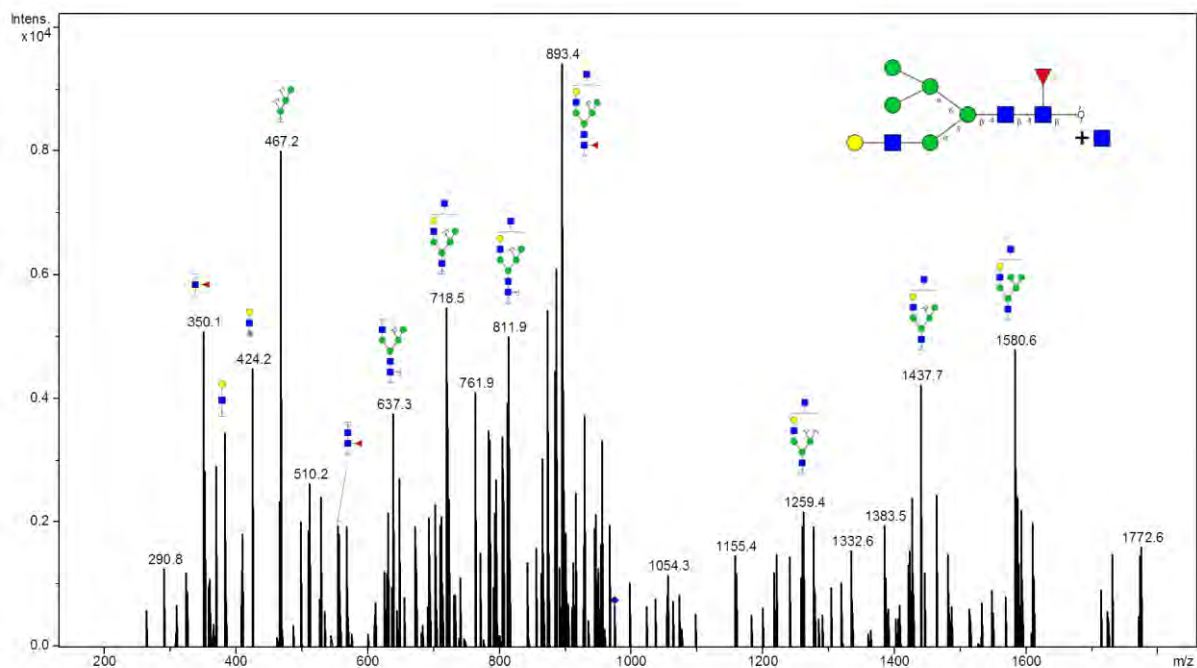
Glycan 90

Parent ion: *m/z* 974.4²⁻
Composition: (Hex)₃ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
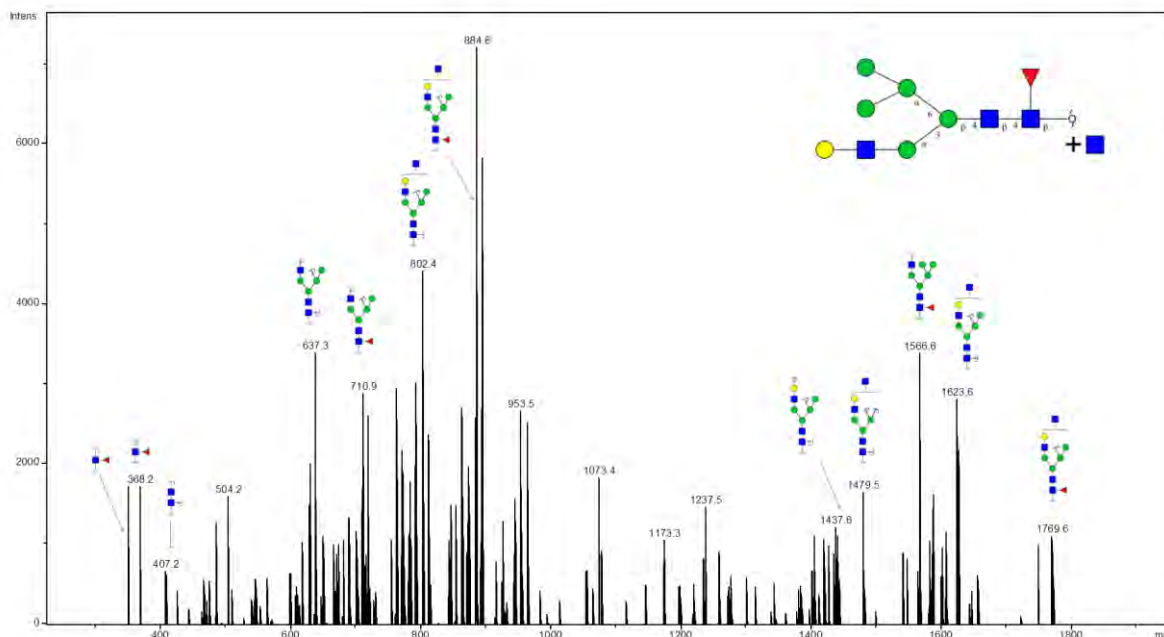
Glycan 91

Parent ion: *m/z* 986.9²⁻
Composition: (Hex)₁ (HexNAc)₃ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan representations. Depiction of specific linkage of outer-arm fucose not intended in fragment scheme.
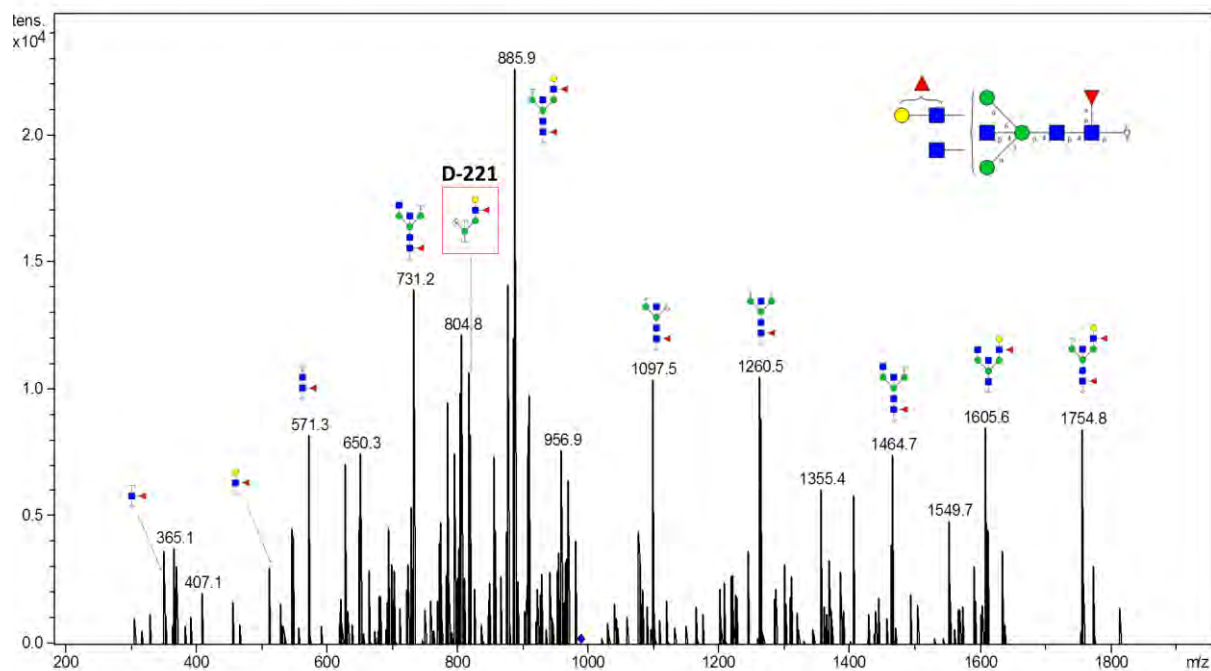
Glycan 92

Parent ion: *m/z* 986.9^{2-}
Composition: (Hex)$_1$ (HexNAc)$_3$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$

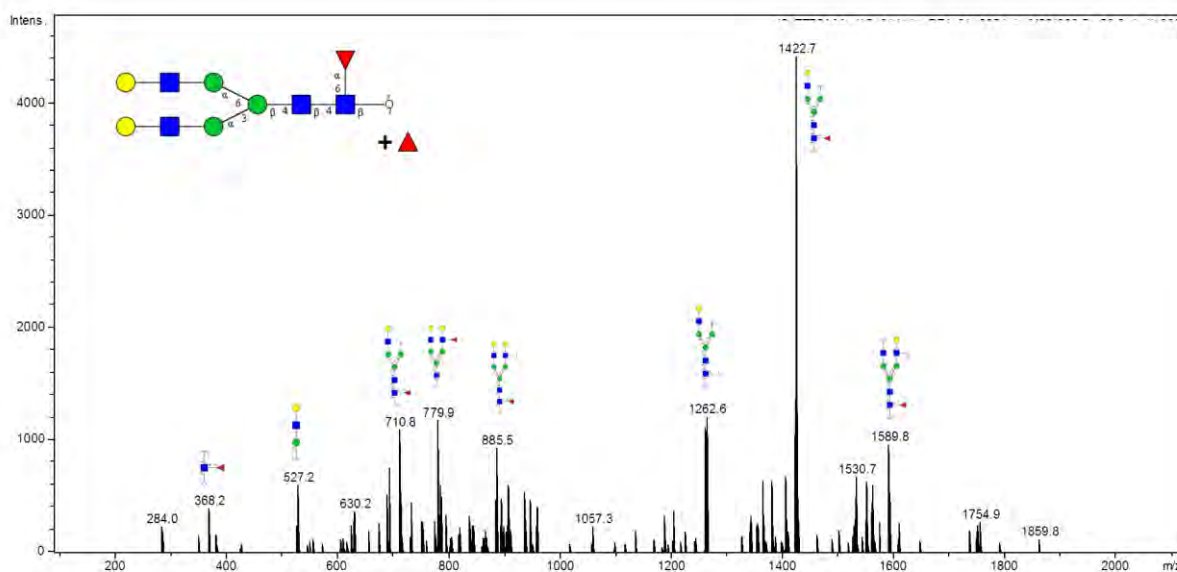Notes: No distinction intended between 3-arm and 6-arm in glycan representations.
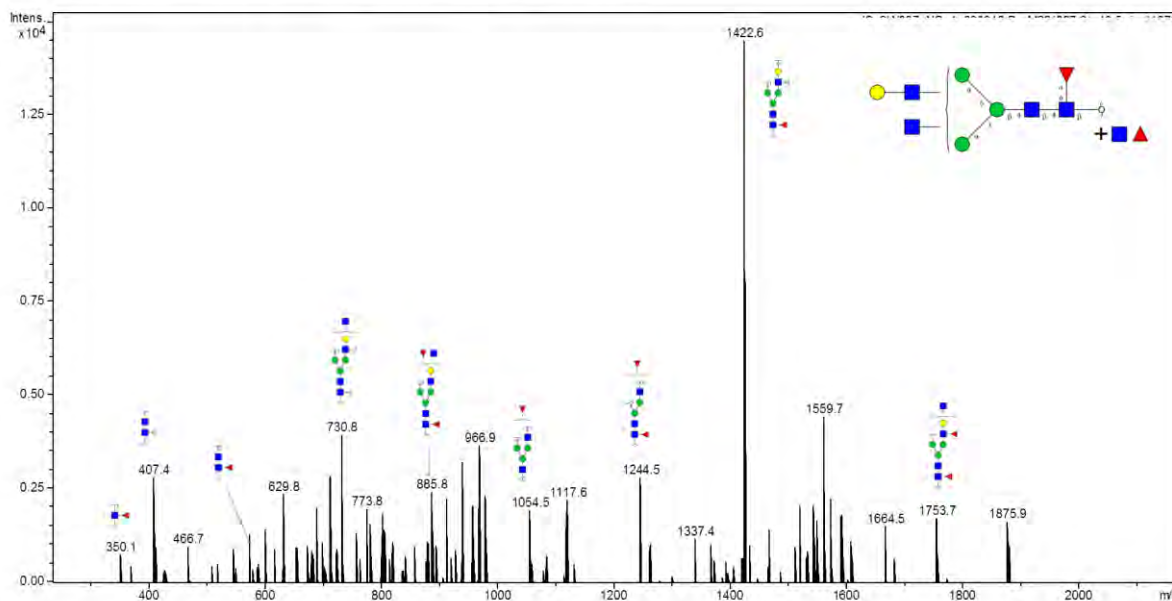
Glycan 93

Parent ion: *m/z* 994.9^{2-}
Composition: (Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction of 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan scheme (i.e. fucose can be linked to Gal or GlcNAc).
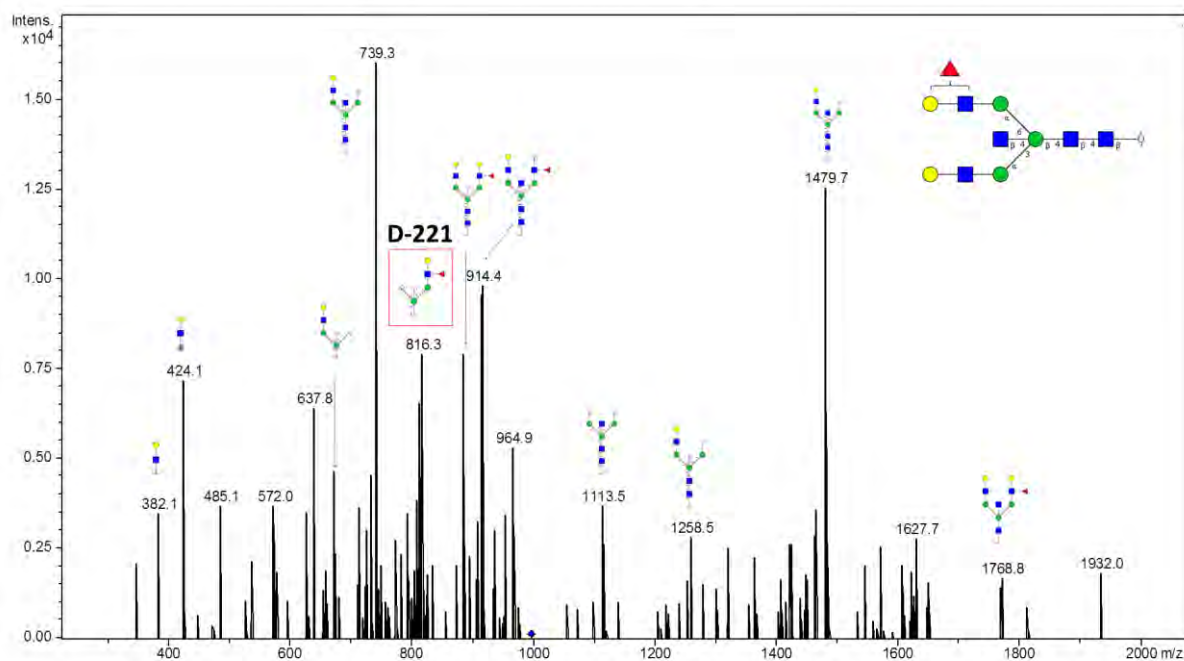
Glycan 94

Parent ion: *m/z* 994.9[2-]

Composition: (Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction of 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan scheme.
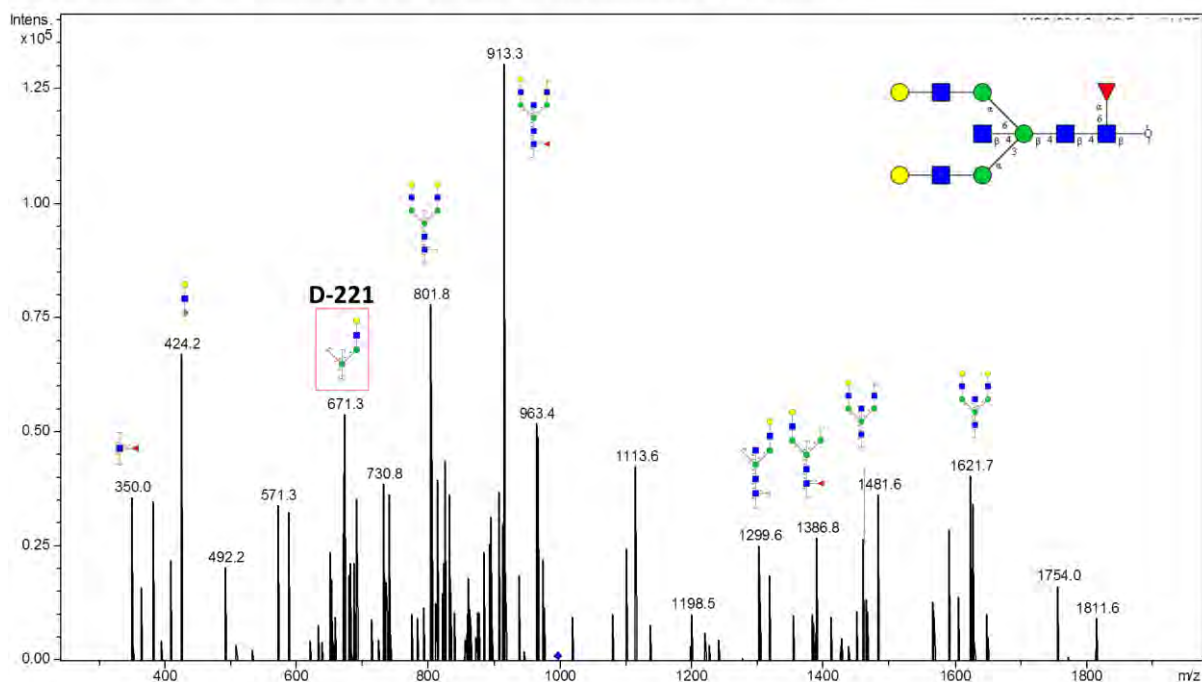
Glycan 95

Parent ion: *m/z* 994.9[2-]

Composition: (Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction of 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan scheme.
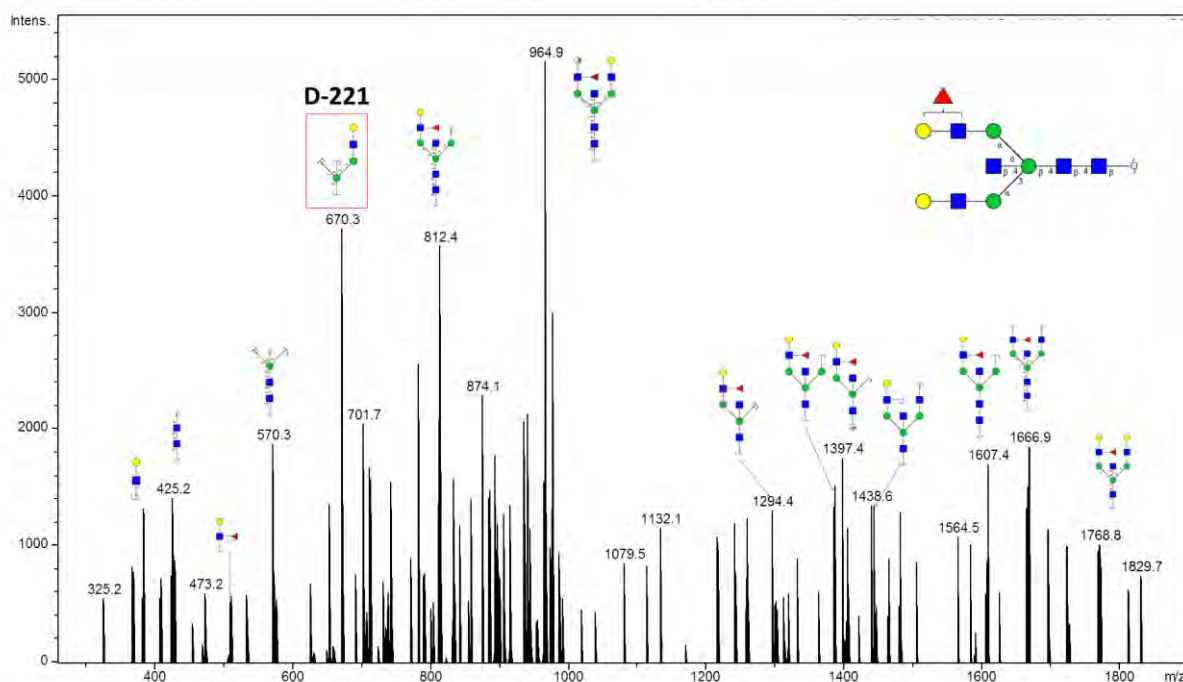
Glycan 96

Parent ion: *m/z* 994.9²⁻
Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction of 3-arm/6-arm not intended in glycan scheme.
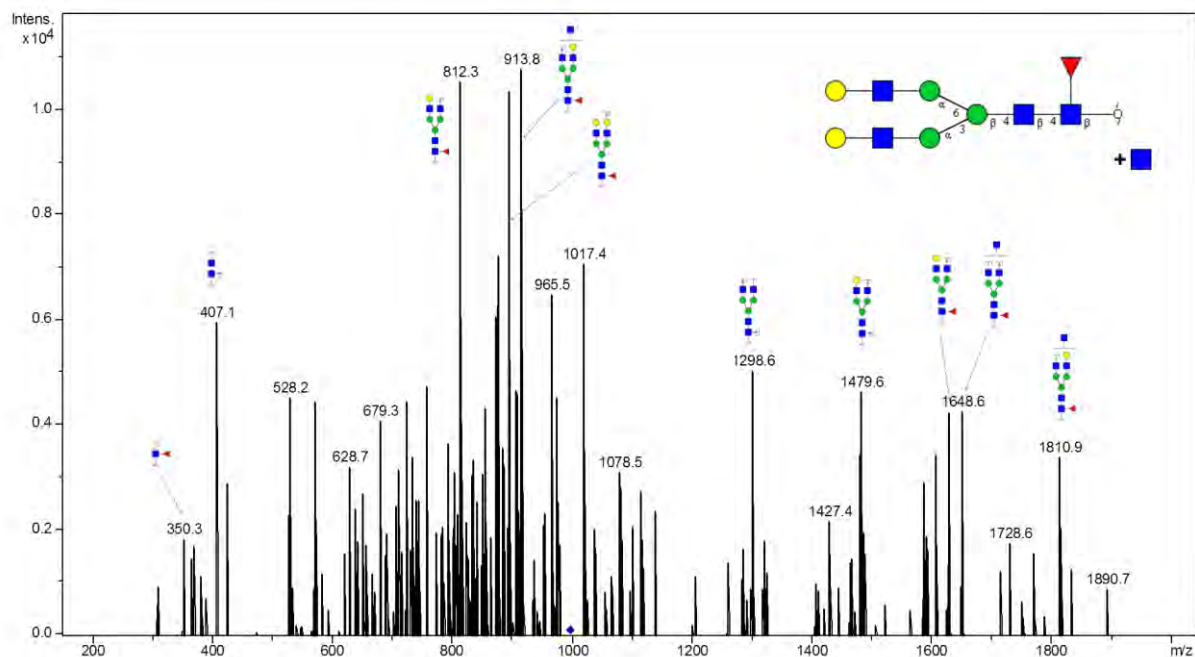
Glycan 97

Parent ion: *m/z* 1002.7²⁻
Composition: (Hex)₃ (HexNAc)₃ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm and 6-arm in glycan fragmentation scheme
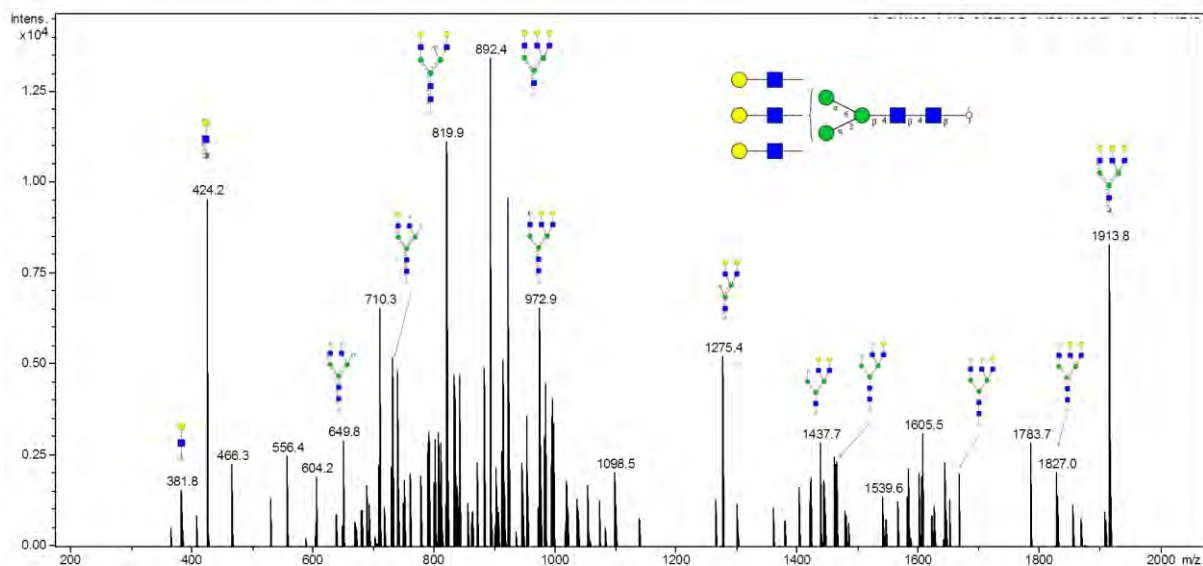
Glycan 98

Parent ion: *m/z* 1006.4²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₂ (Sulph)₁ + (Man)₃(GlcNAc)₂
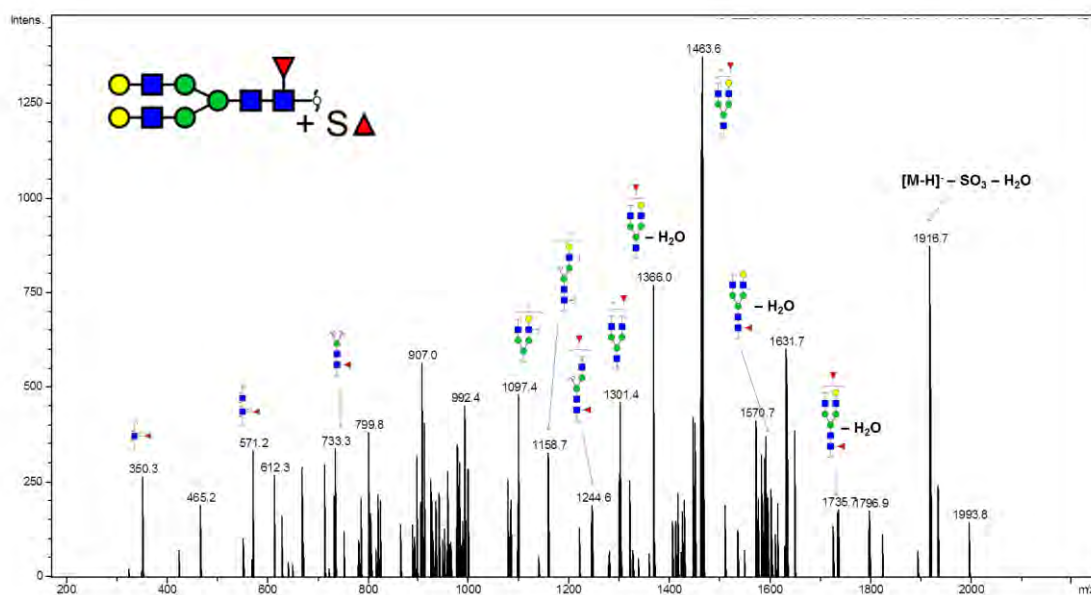
Glycan 99

Parent ion: *m/z* 1018.4²⁻

Composition: (Hex)₃ (HexNAc)₁ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Specific linkage of outer-arm fucose not intended in fragment scheme i.e. can be linked to Gal or GlcNAc
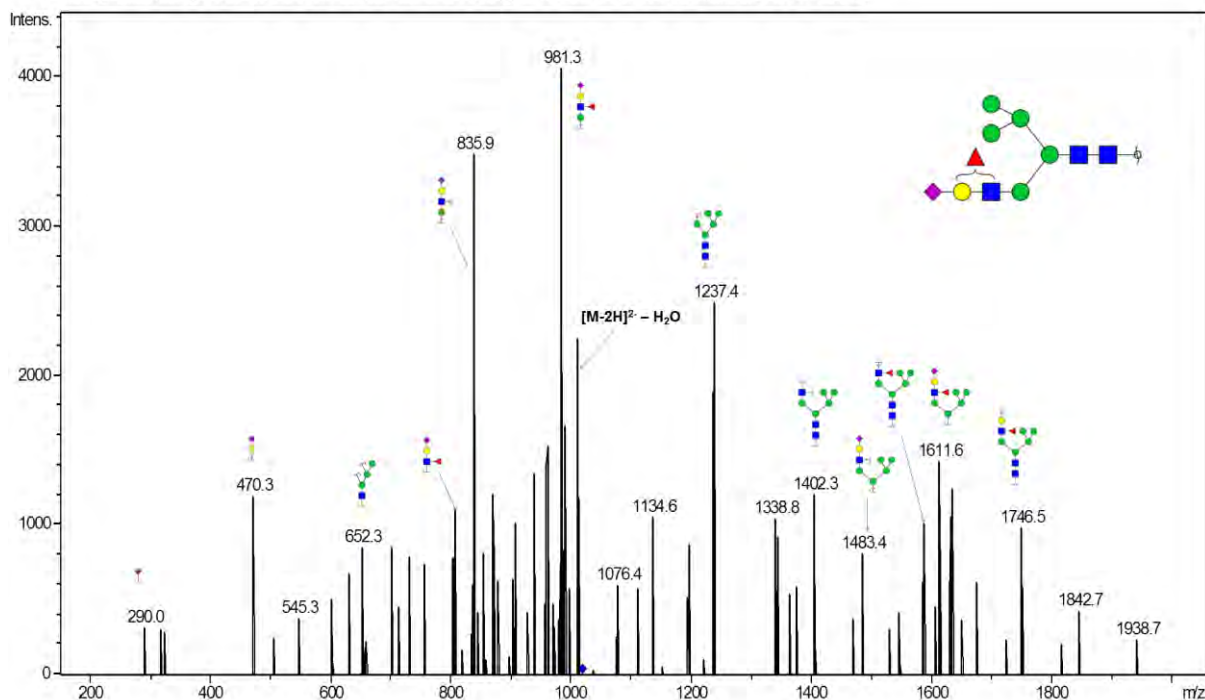
Glycan 100

**Parent ion:** *m/z* 1018.4^{2-}
Composition: (Hex)$_3$ (HexNAc)$_1$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
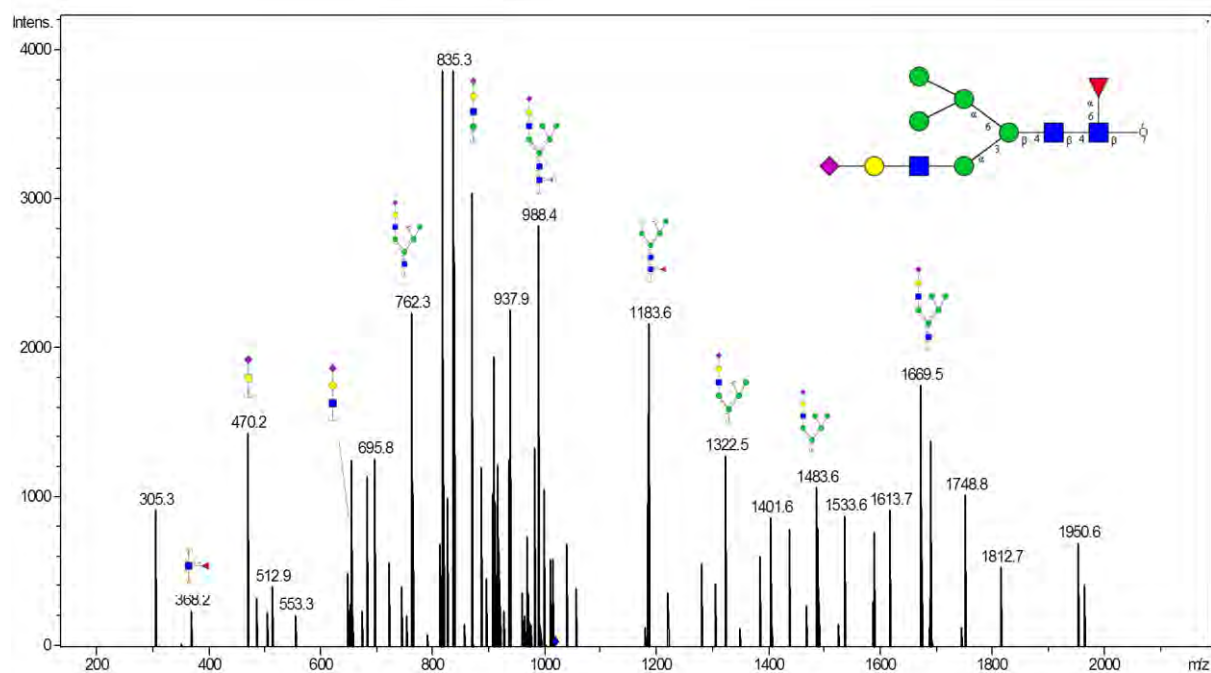
Glycan 101

Parent ion: *m/z* 1022.4^{2-}
Composition: (Hex)$_7$ + (Man)$_3$(GlcNAc)$_2$
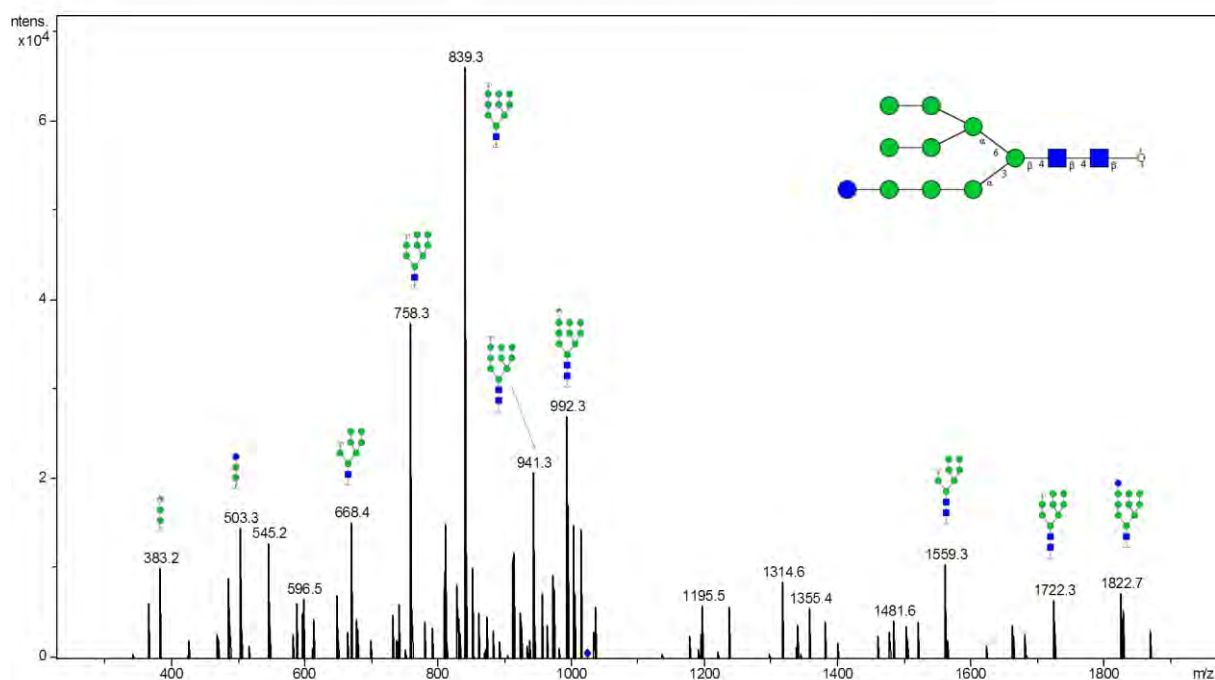
Glycan 102

Parent ion: *m/z* 1026.4^{2-}

Composition: (Hex)$_4$ (HexNAc)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
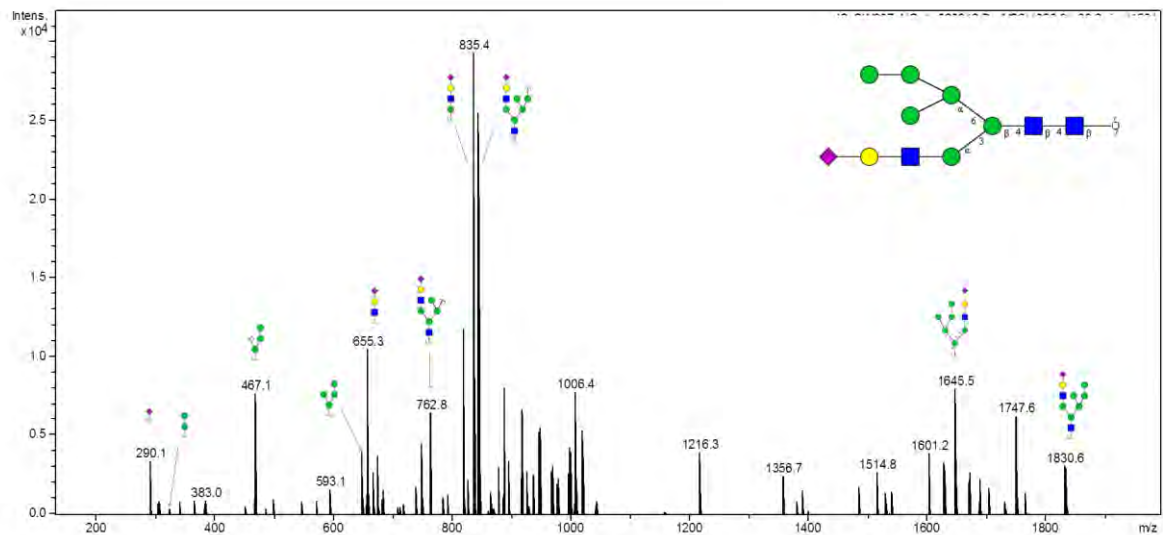
Glycan 103

Parent ion: *m/z* 1038.9^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Glycan 103 co-elutes with glycan 108. Distinction between 3-arm/6-arm and Specific linkage of outer-arm fucose not intended in glycan fragment scheme.
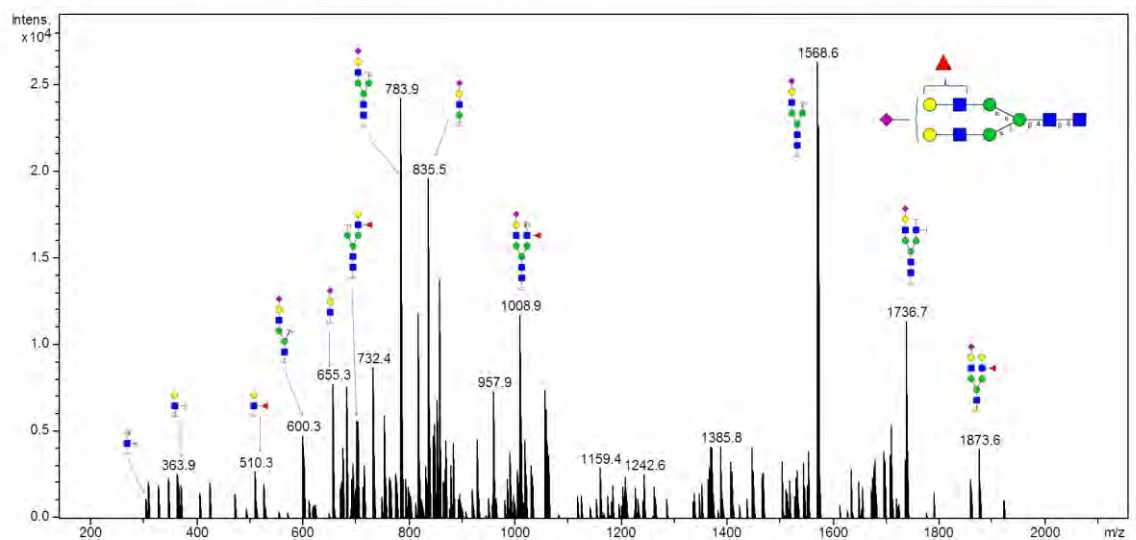
Glycan 104

Parent ion: *m/z* 1038.9[2-]

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme.
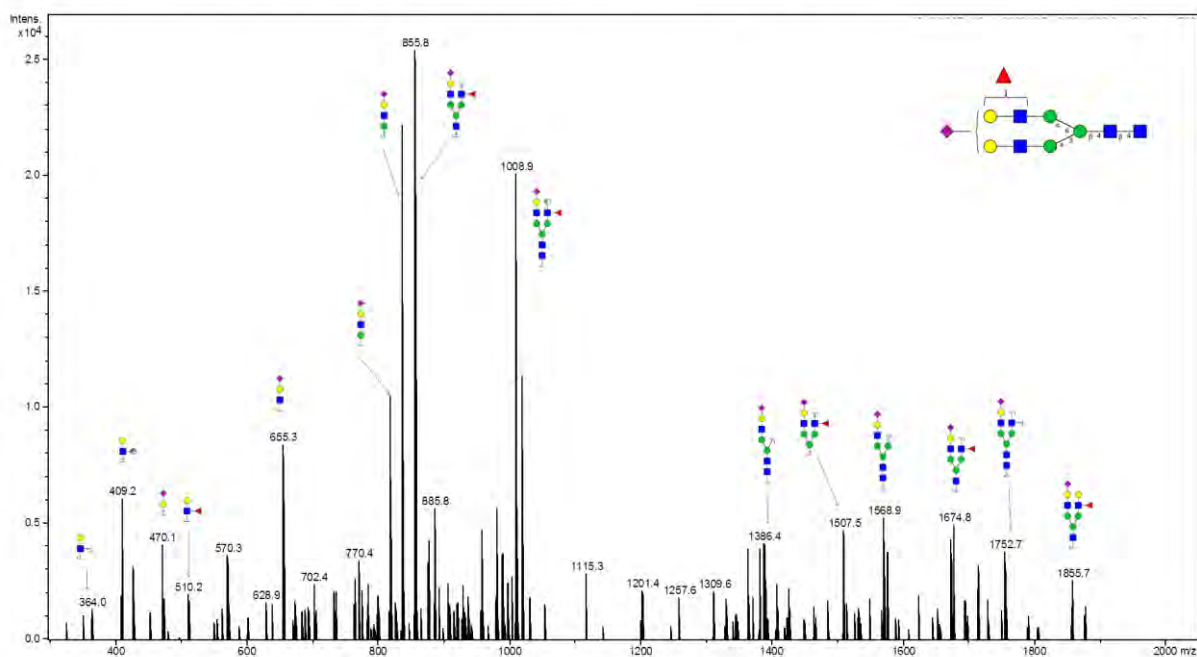
Glycan 105

Parent ion: *m/z* 1038.9[2-]

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme.

170

Glycan 106

Parent ion: *m/z* 1038.9²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme.
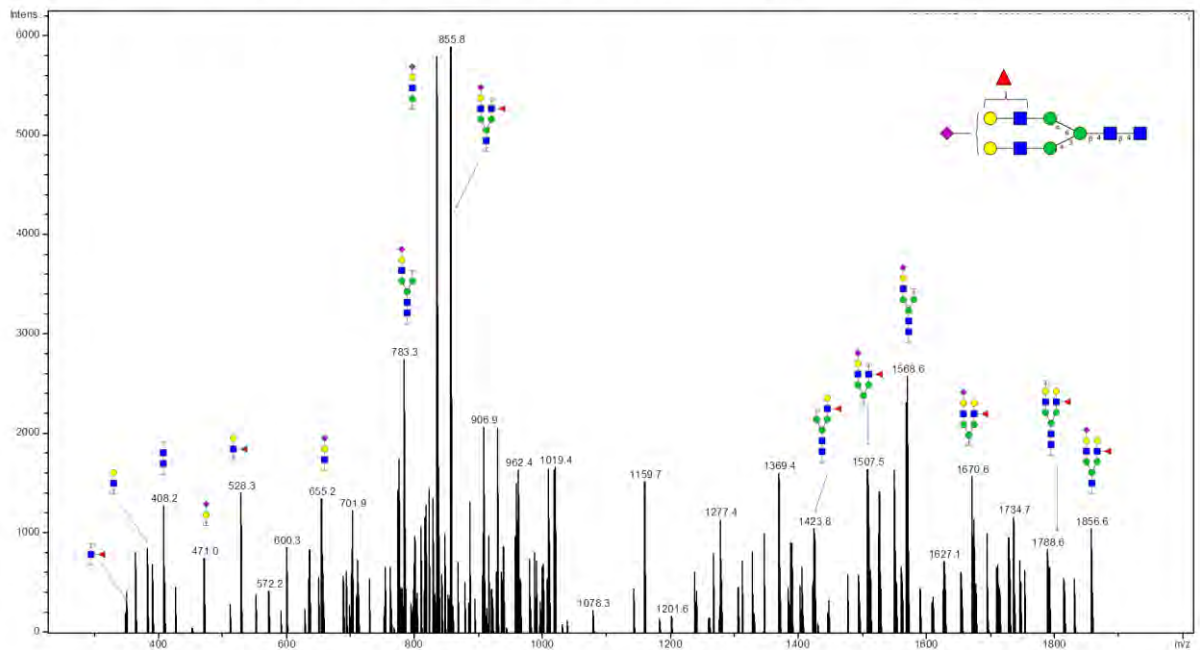
Glycan 107

Parent ion: *m/z* 1038.9²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme.

Glycan 108

Parent ion: *m/z* 1039.4²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Specific linkage of outer-arm fucose not intended in glycan fragment scheme.

Glycan 109

Parent ion: *m/z* 1039.4²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme.
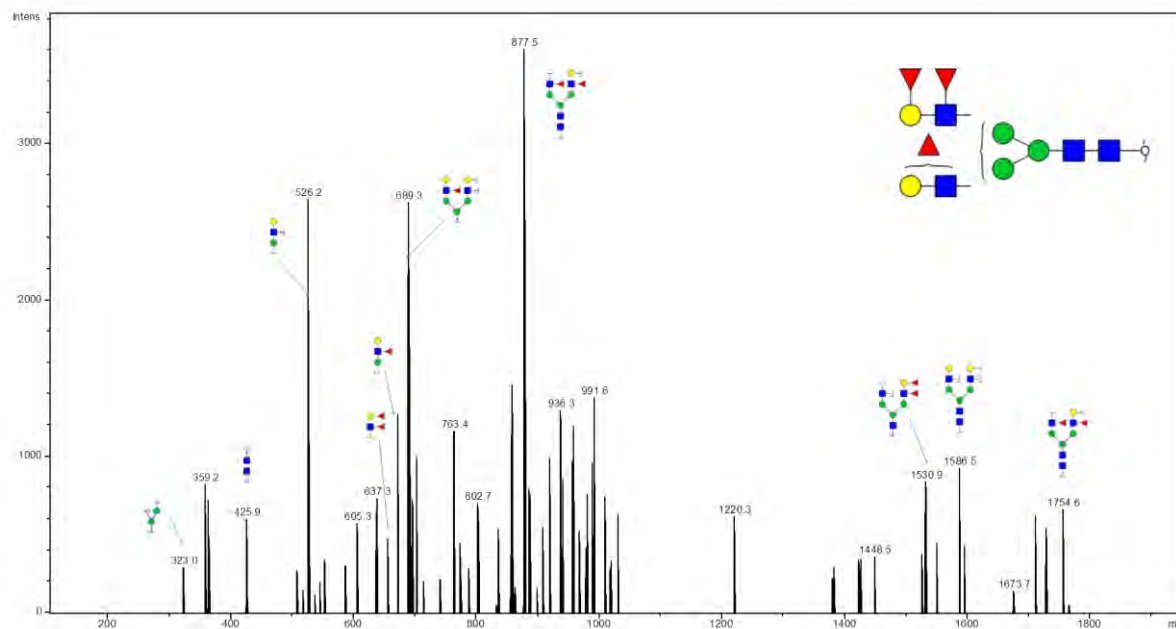
Glycan 110

Parent ion: *m/z* 1039.4^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Specific linkage of outer-arm fucose not intended in glycan fragment scheme.
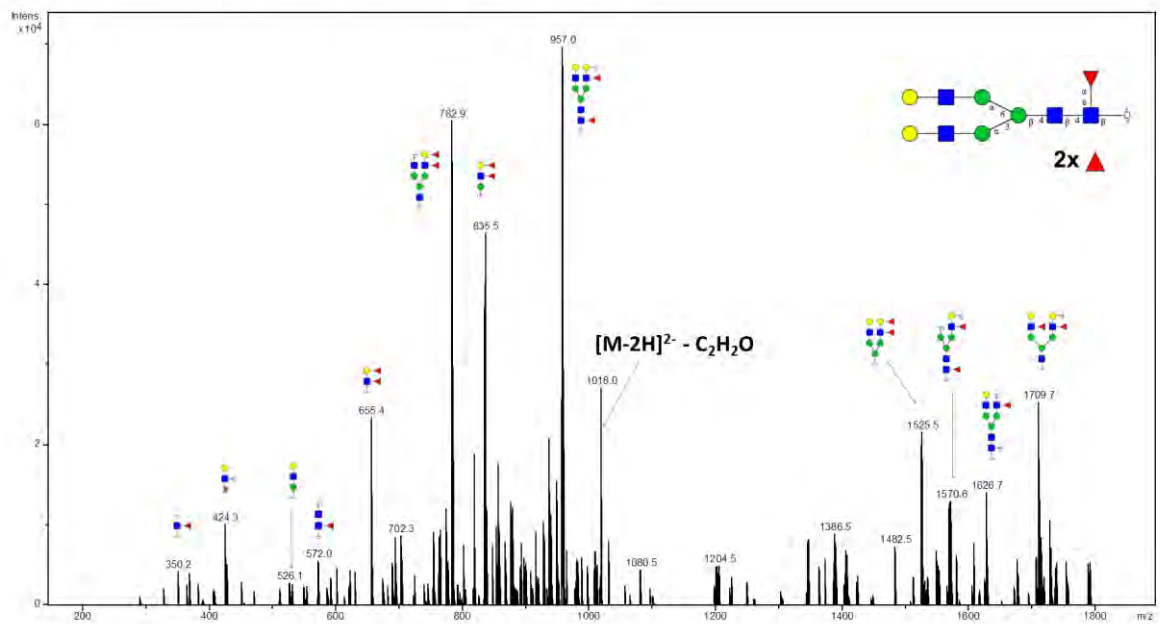
Glycan 111

Parent ion: *m/z* 1039.4^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Specific linkage of outer-arm fucose not intended in glycan fragment scheme.
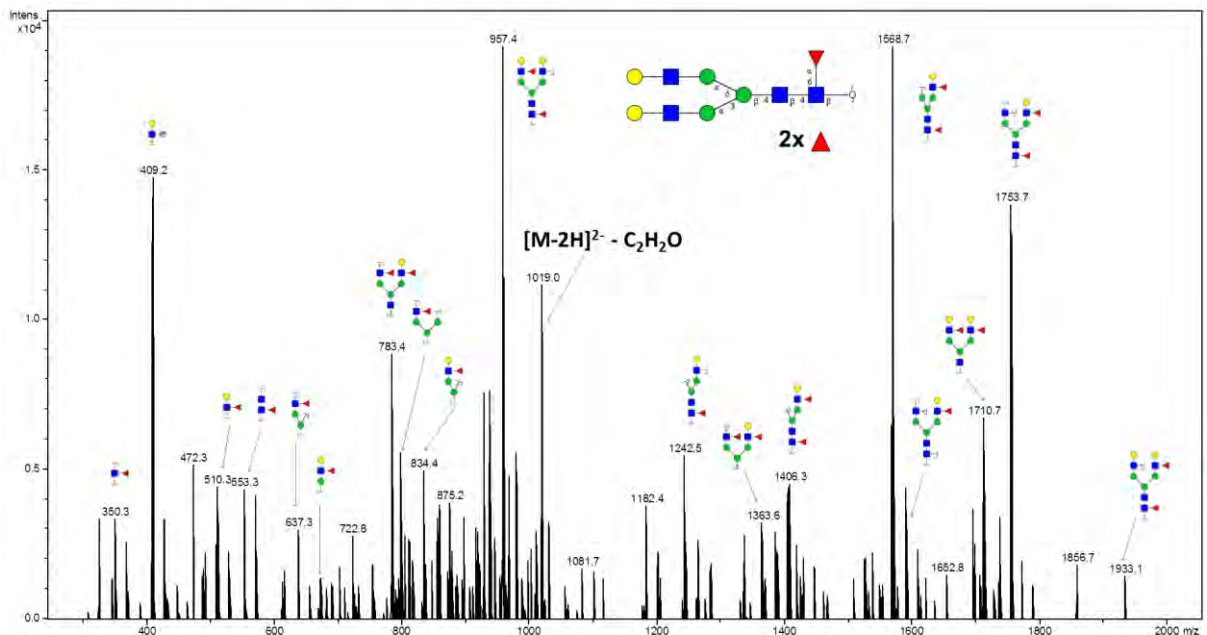
Glycan 112

Parent ion: *m/z* 1047.7^{2-}
Composition: (Hex)$_3$ (HexNAc)$_2$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$
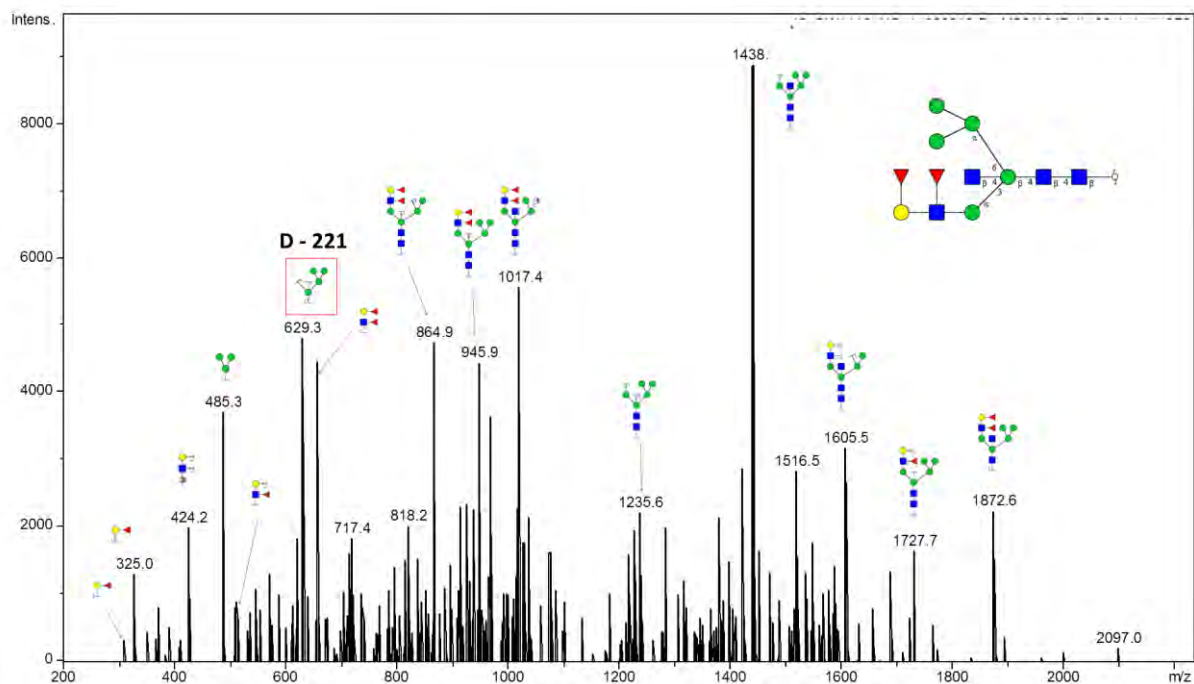
Glycan 113

Parent ion: *m/z* 1055.4^{2-}
Composition: (Hex)$_4$ (HexNAc)$_2$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
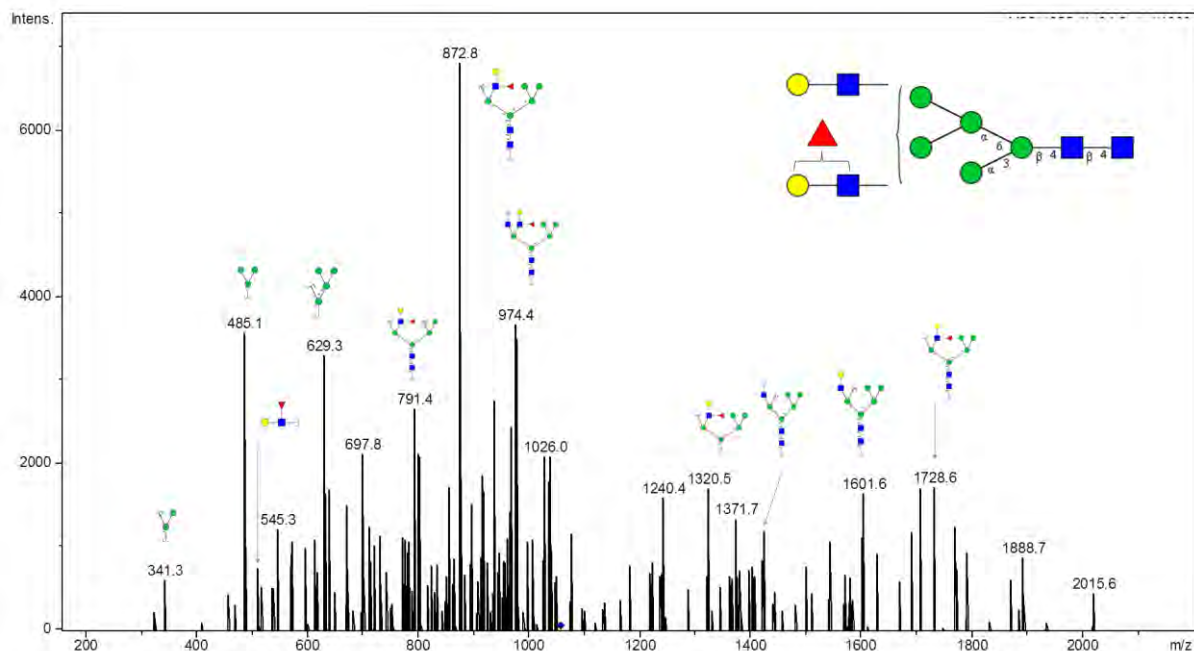
Glycan 114

Parent ion: *m/z* 1055.4²⁻
Composition: (Hex)₄ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
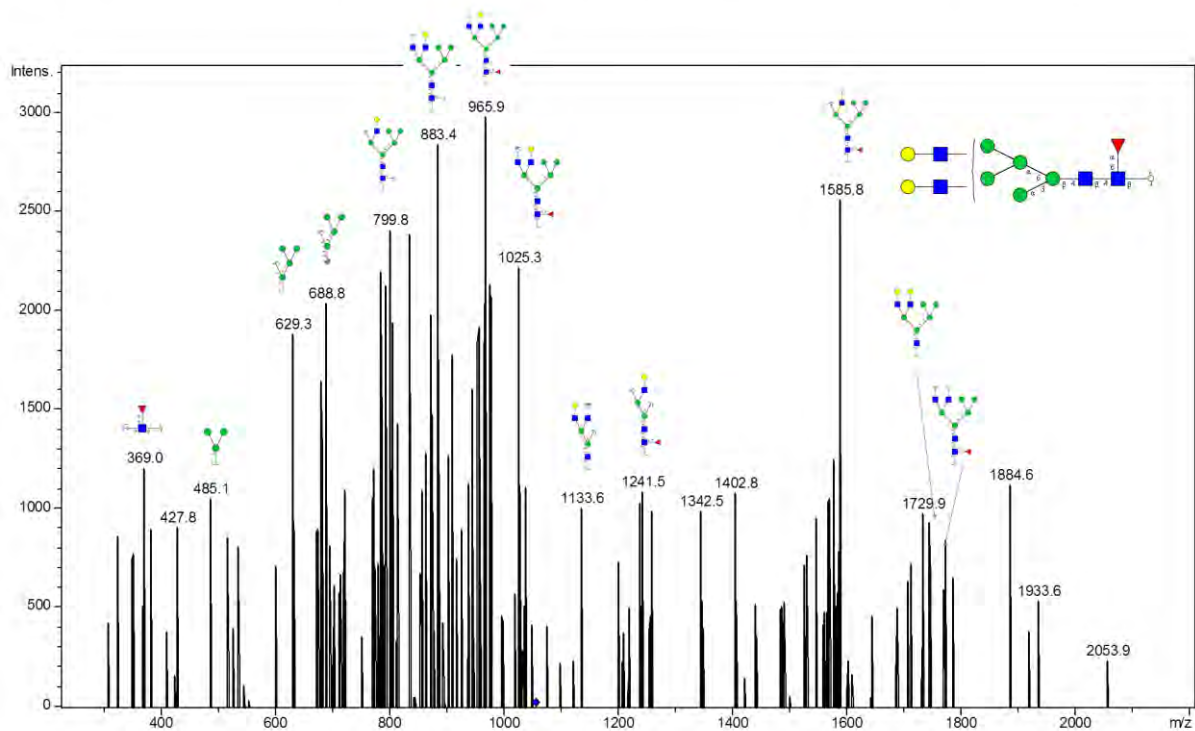
Glycan 115

Parent ion: *m/z* 1055.4²⁻
Composition: (Hex)₄ (HexNAc)₂ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Specific linkage of fucose not intended in glycan fragment scheme (i.e. may be outer-arm or core-fucose).
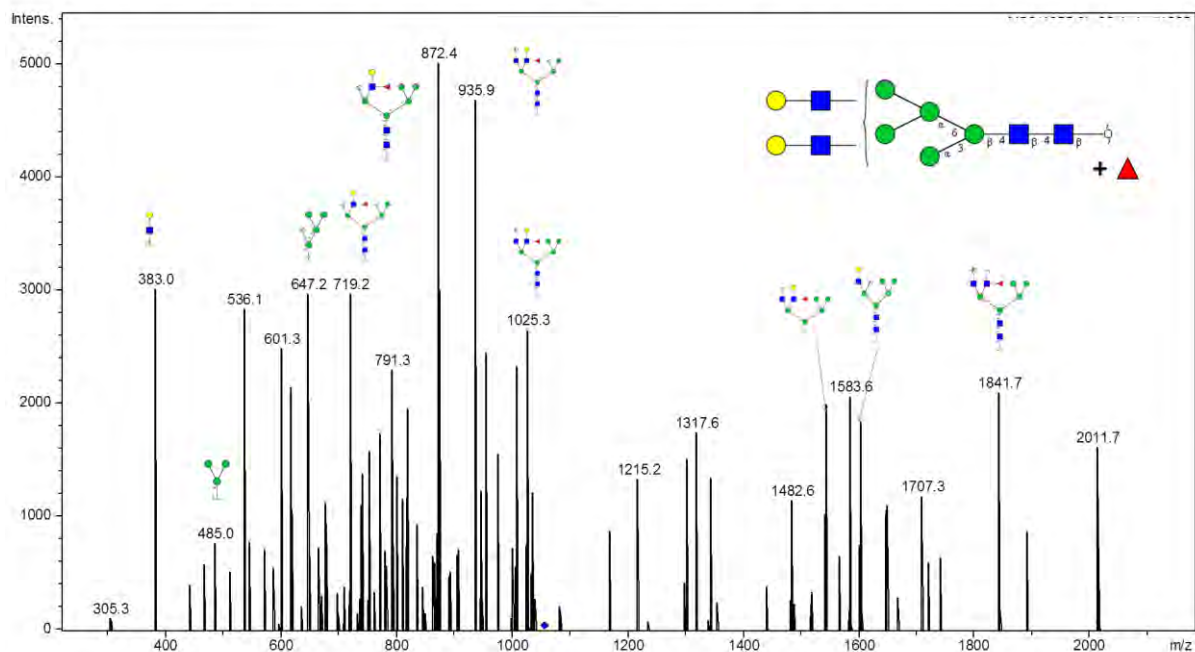
Glycan 116

Parent ion: *m/z* 1059.4[2-]

Composition: (Hex)$_1$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
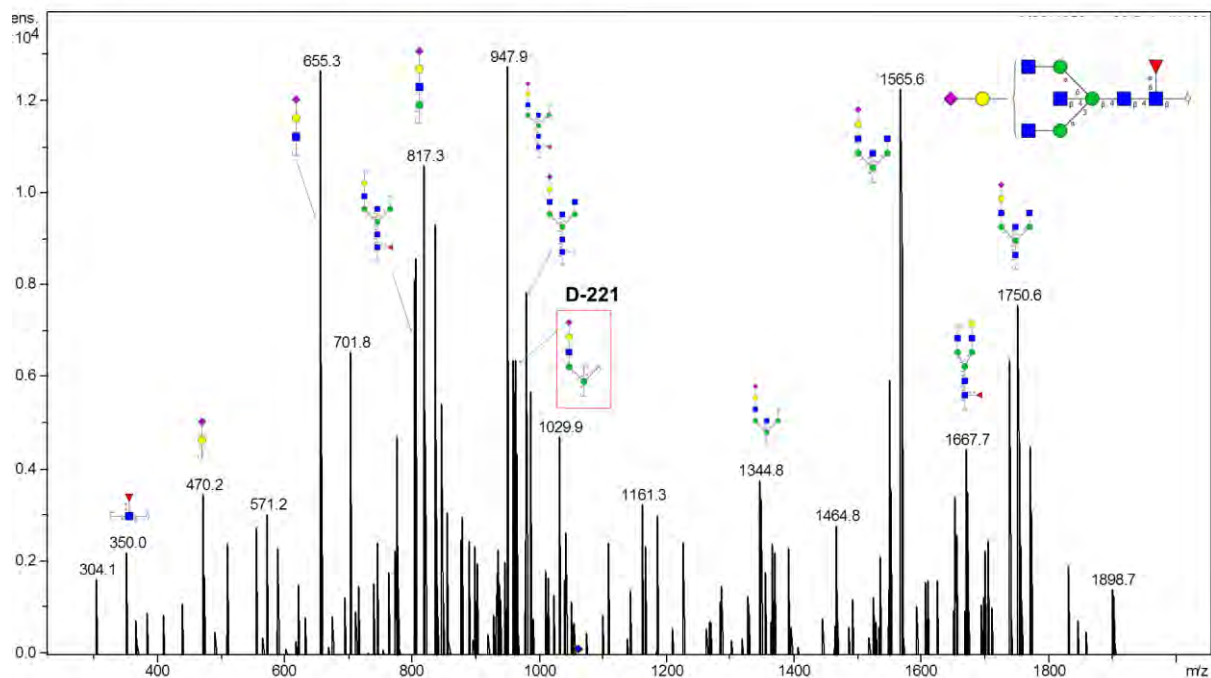
Glycan 117

Parent ion: *m/z* 1059.4[2-]

Composition: (Hex)$_1$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
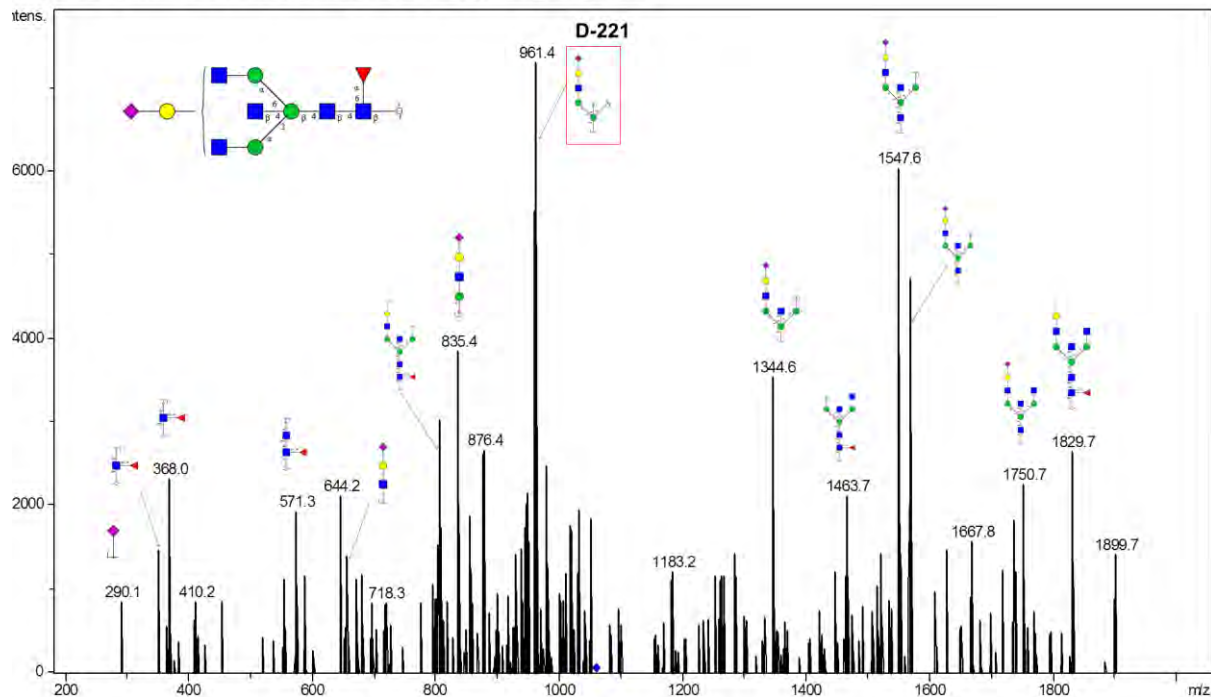
Glycan 118

Parent ion: *m/z* 1059.4²⁻

Composition: (Hex)₁ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
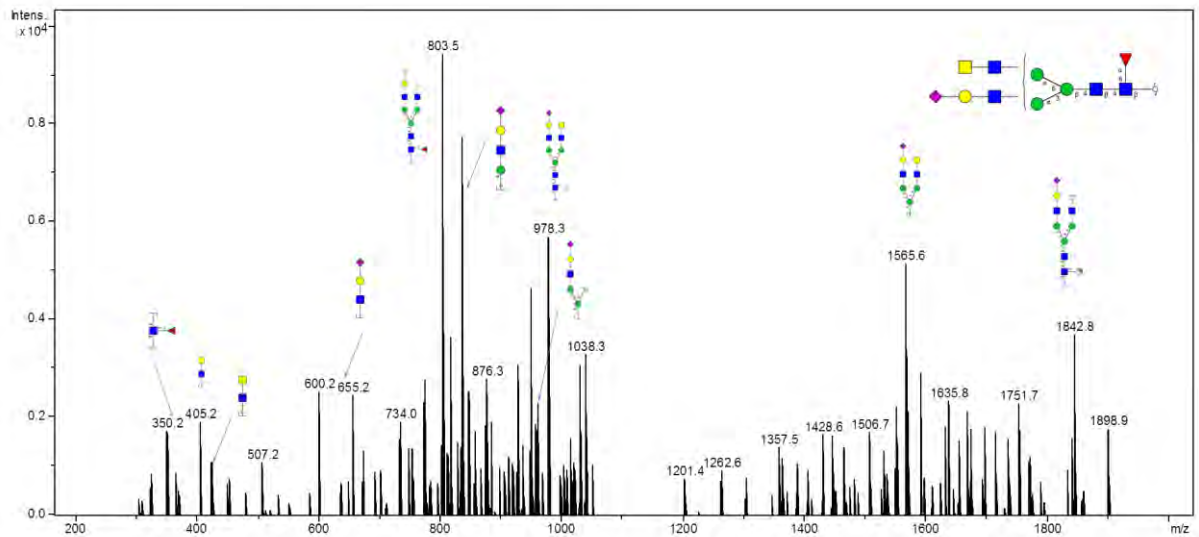
Glycan 120

Parent ion: *m/z* 1068.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
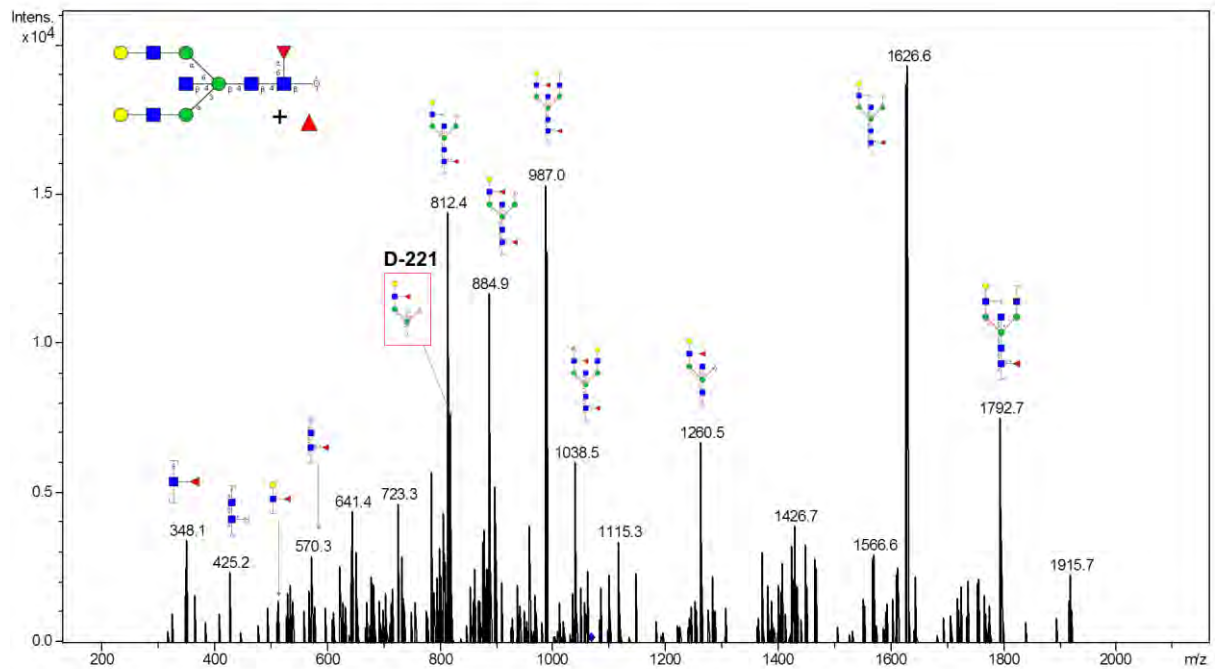
Glycan 121

Parent ion: *m/z* 1068.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
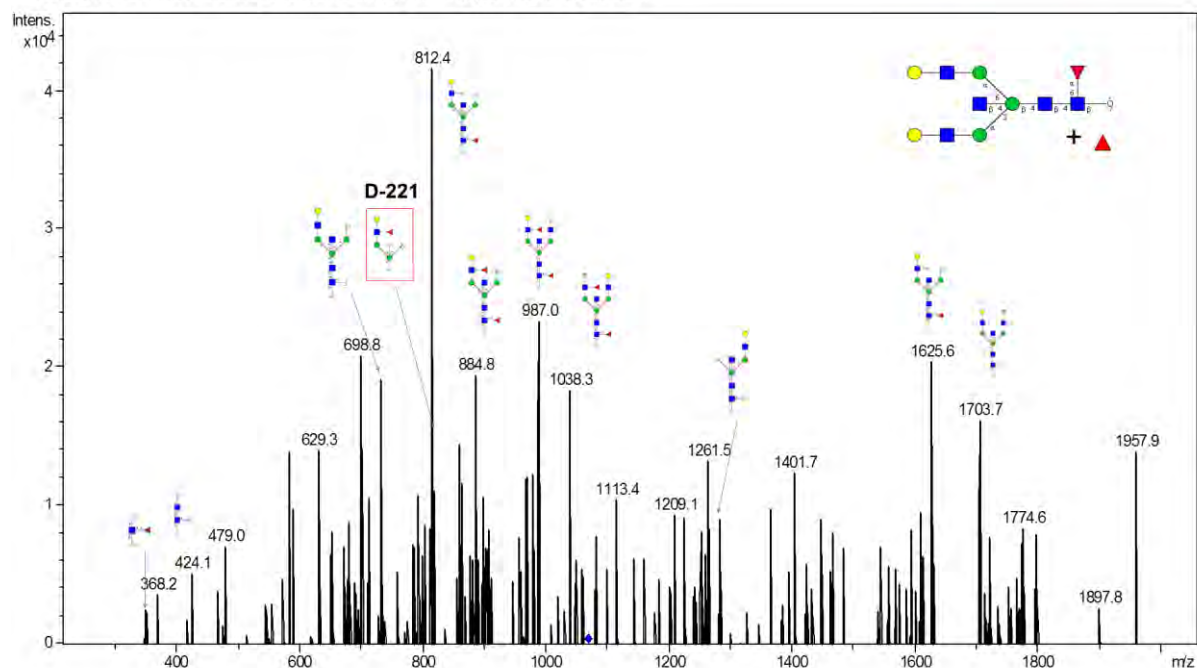
Glycan 122

Parent ion: *m/z* 1068.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
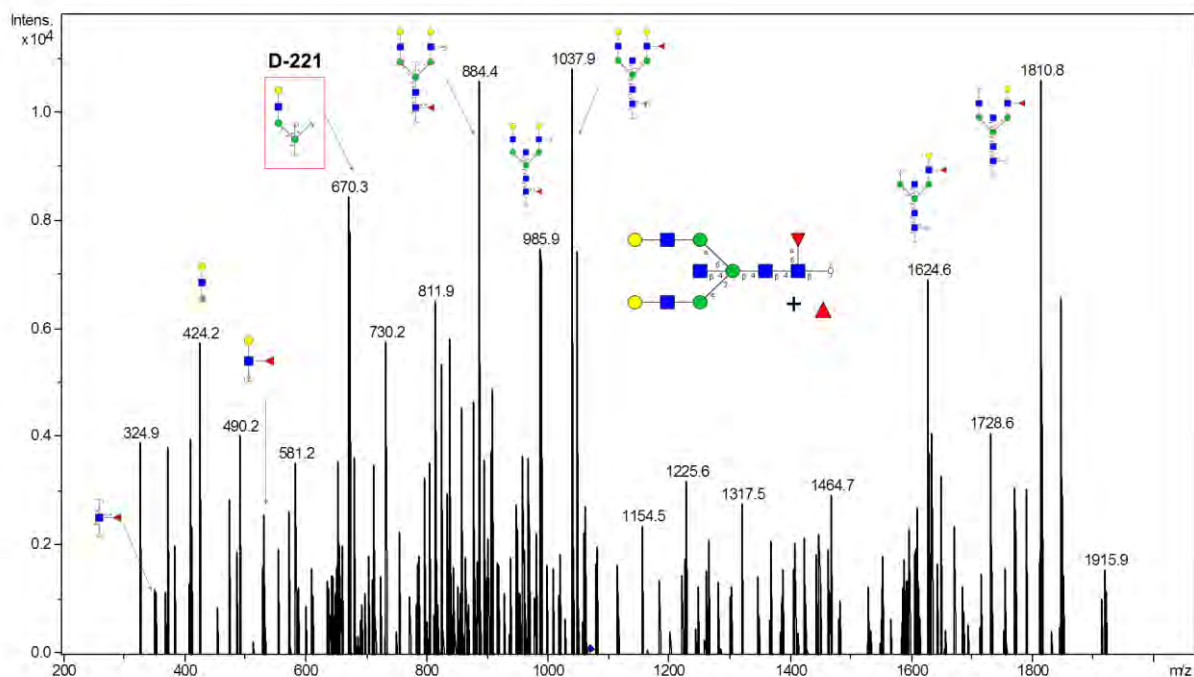
Glycan 123

Parent ion: *m/z* 1075.9[2-]

Composition: (Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
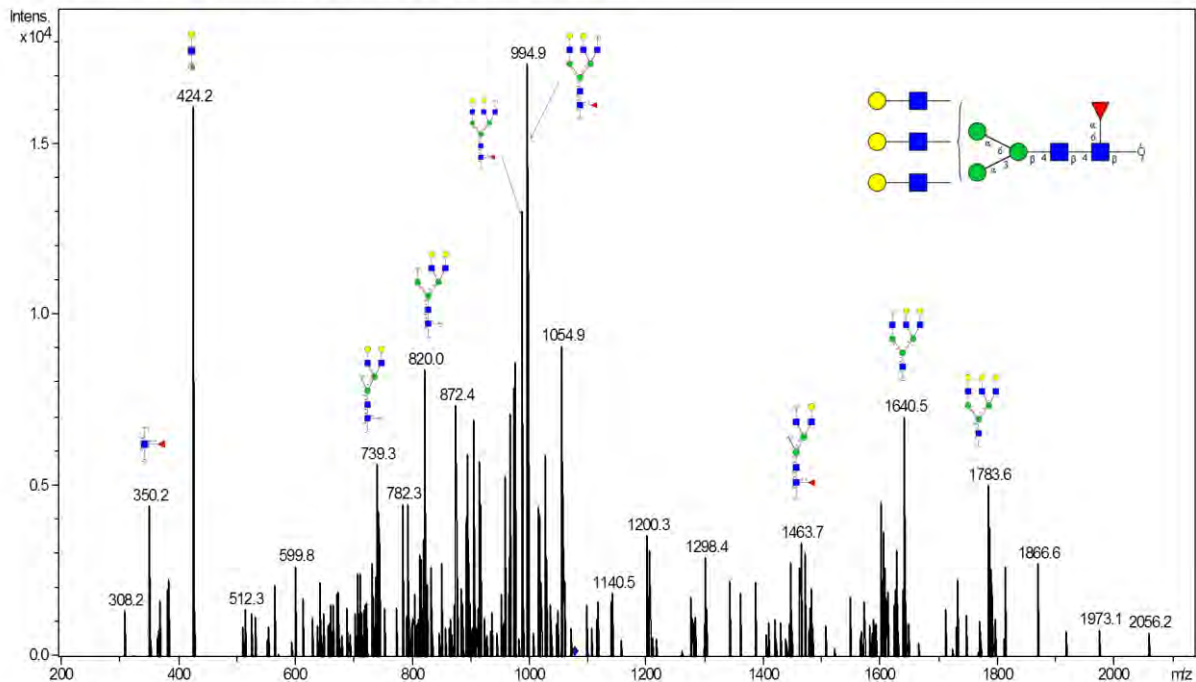
Glycan 124

Parent ion: *m/z* 1075.9[2-]

Composition: (Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
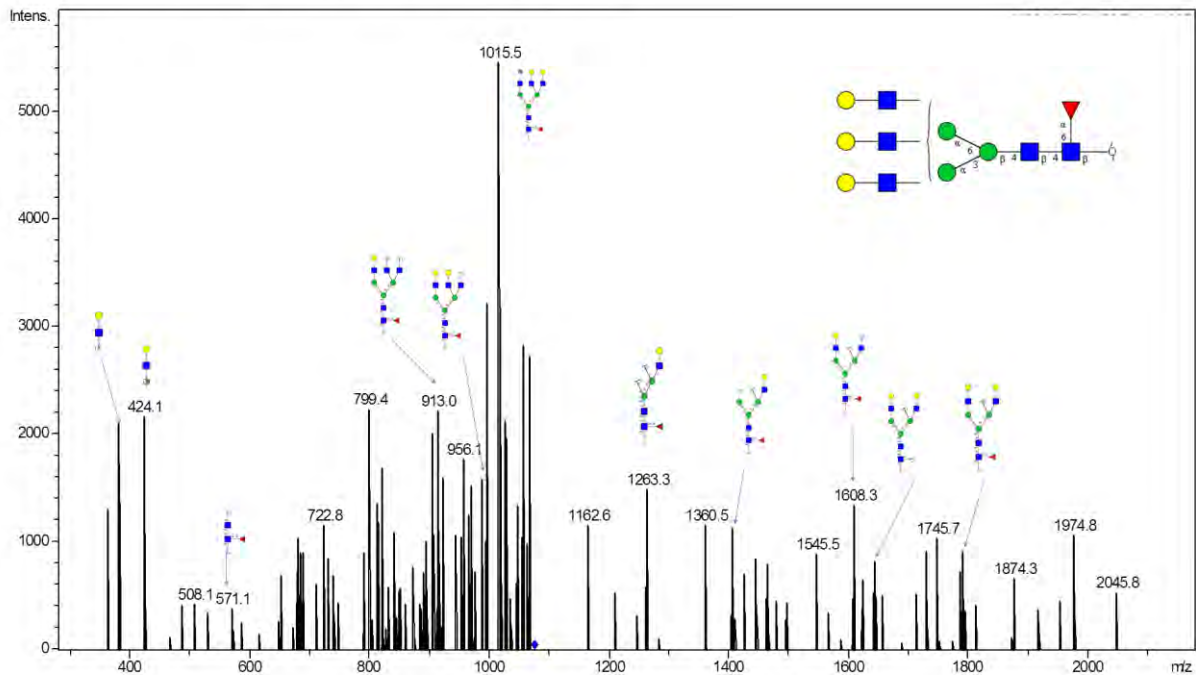
Glycan 125

Parent ion: *m/z* 1075.9^{2-}

Composition: (Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
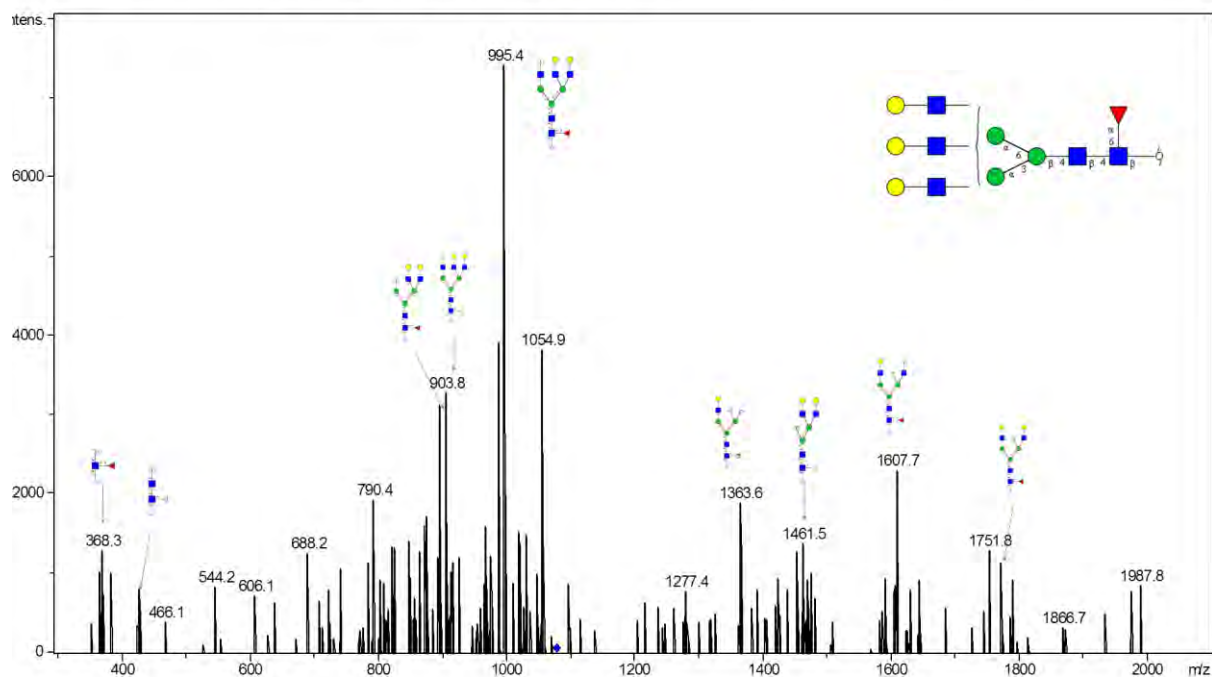
Glycan 126

Parent ion: *m/z* 1080.6^{2-}

Composition: (HexNAc)$_4$ (Deoxyhexose)$_3$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
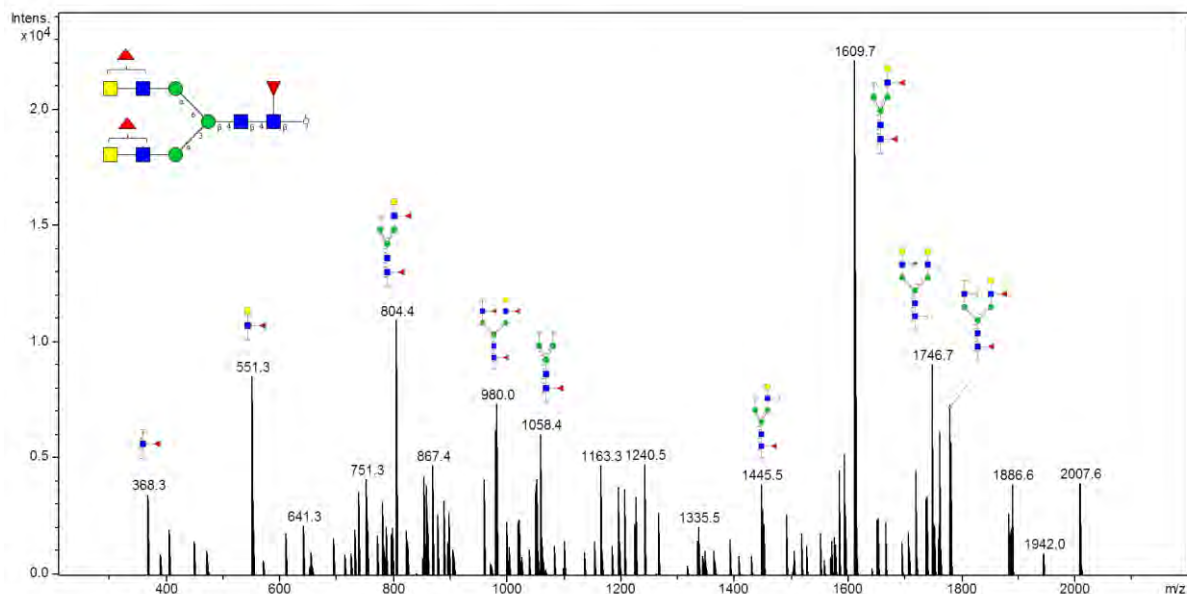
Glycan 127

Parent ion: *m/z* 1096.5²⁻

Composition: (Hex)₂ (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
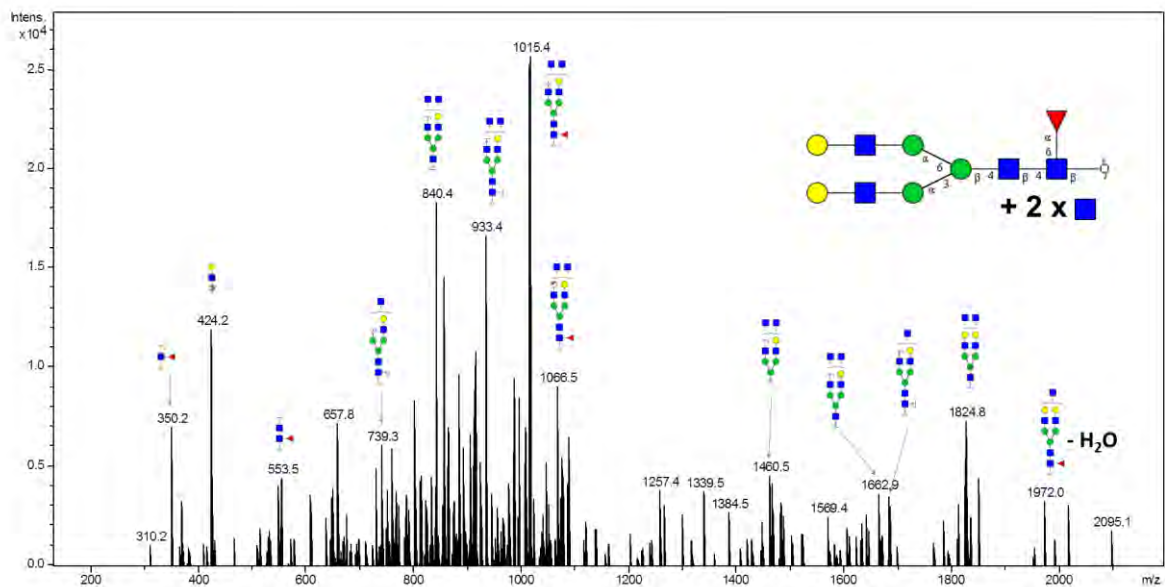
Glycan 128

Parent ion: *m/z* 1096.5²⁻

Composition: (Hex)₂ (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
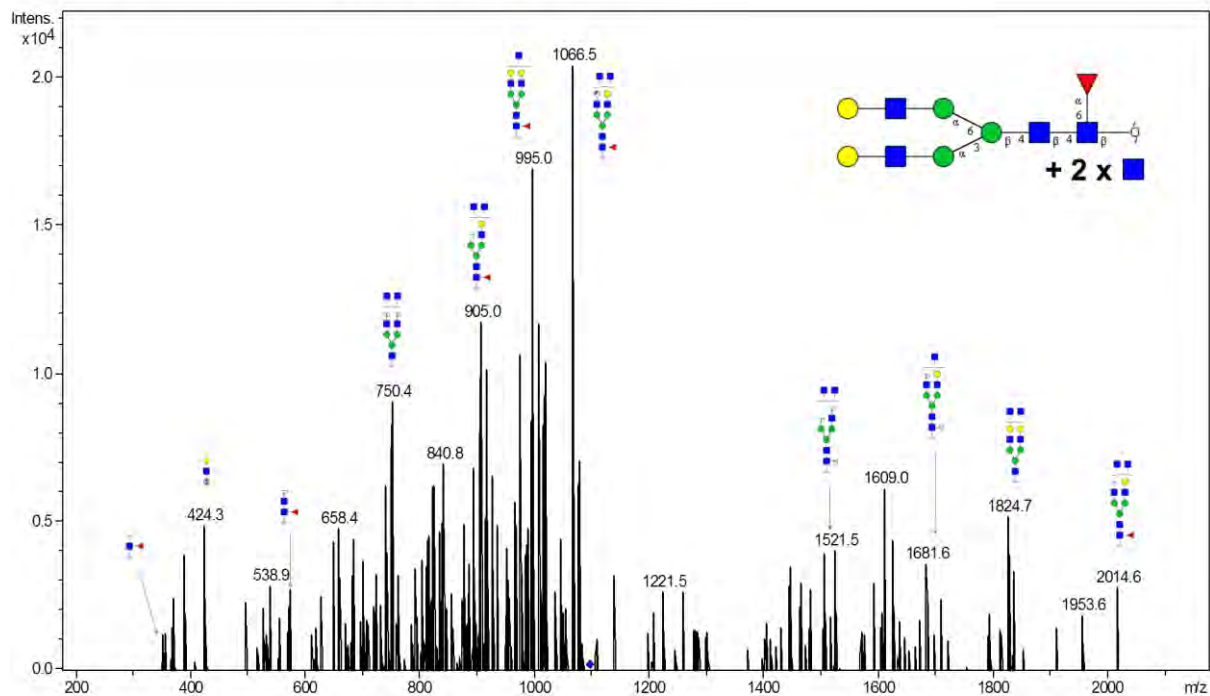
Glycan 129

Parent ion: *m/z* 1103.4^{2-}

Composition: (Hex)$_8$ + (Man)$_3$(GlcNAc)$_2$
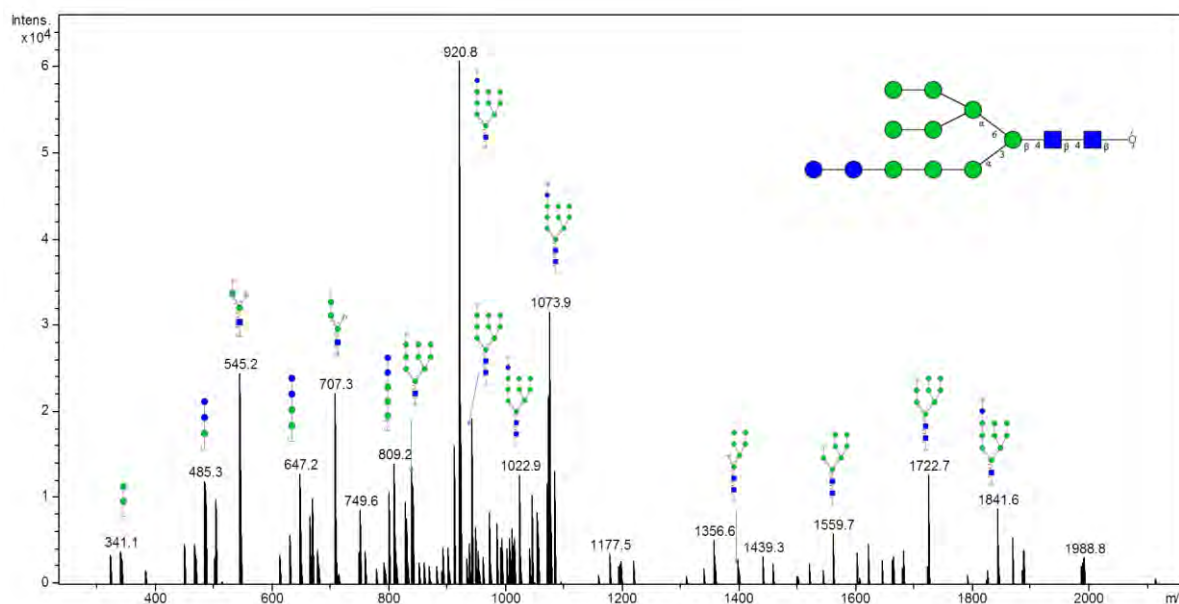
Glycan 130

Parent ion: *m/z* 1111.4^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
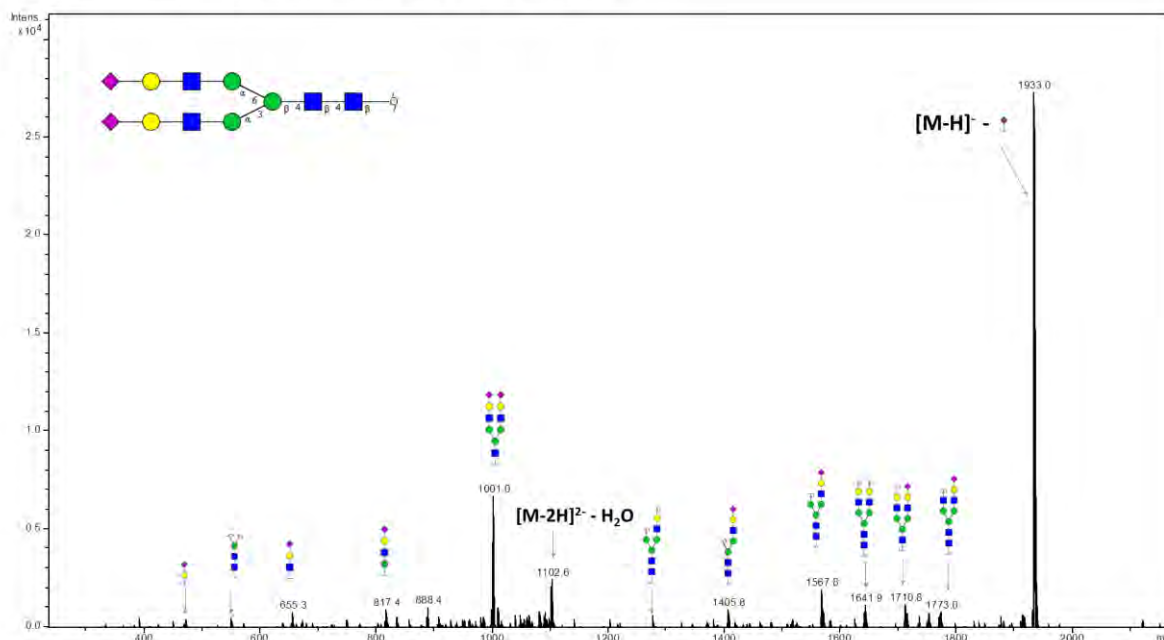
Glycan 131

Parent ion: *m/z* 1111.4²⁻

Composition: (Hex)₂ (HexNAc)₂ (NeuAc)₂ + (Man)₃(GlcNAc)₂

Notes: No distinction intended between 3-arm/6-arm in glycan fragment scheme.
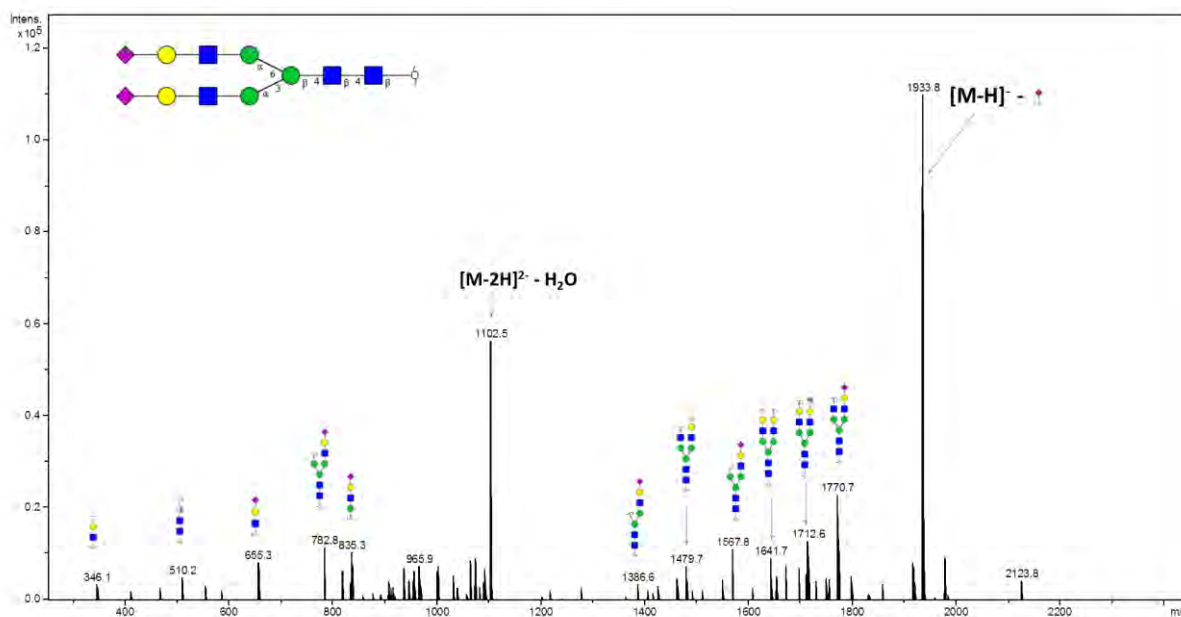
Glycan 132

Parent ion: *m/z* 1111.9²⁻

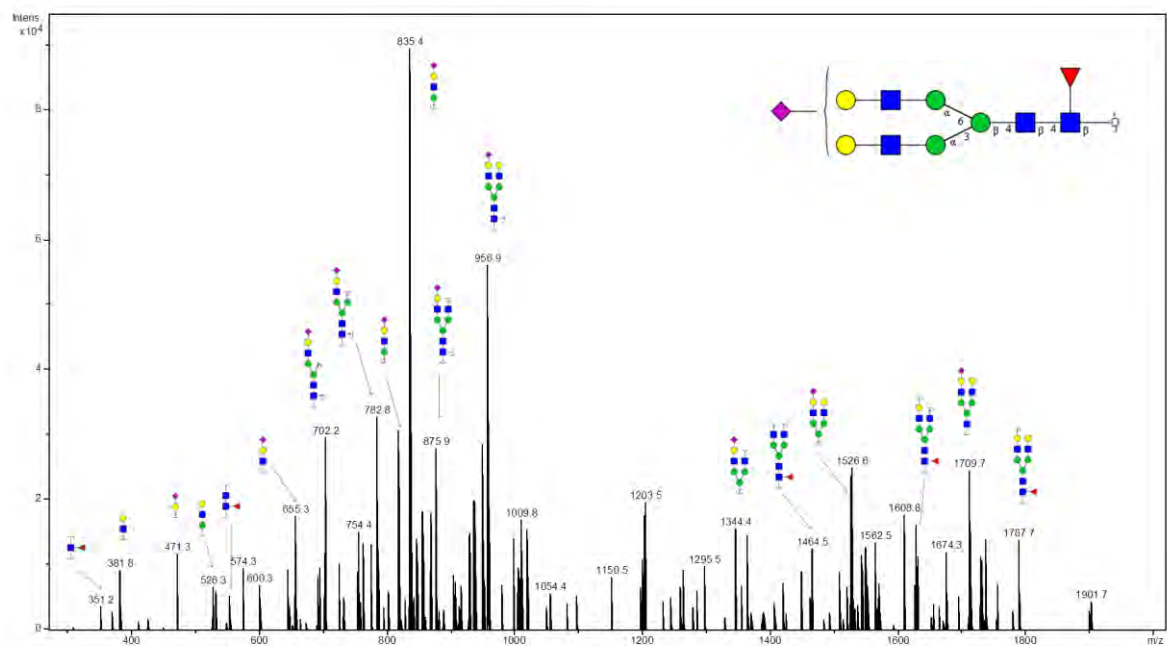**Composition:** (Hex)₂ (HexNAc)₂ (Deoxyhexose)₂ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended glycan fragment scheme..

183
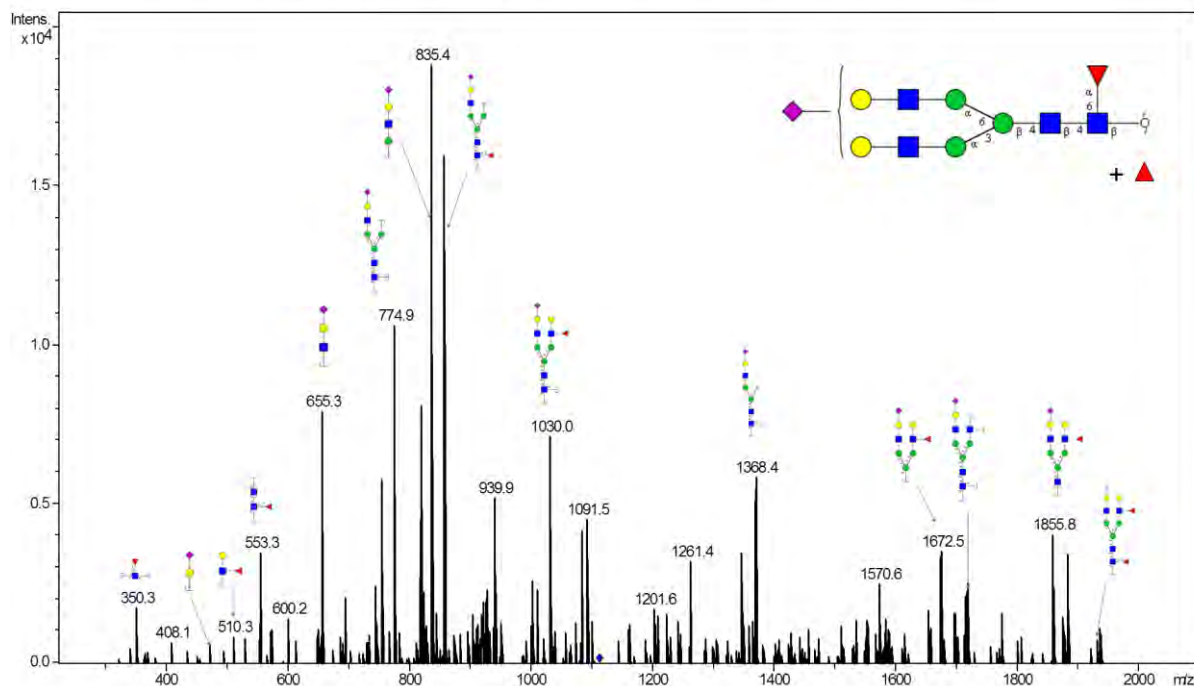
Glycan 133

Parent ion: *m/z* 1111.9^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended glycan fragment scheme..
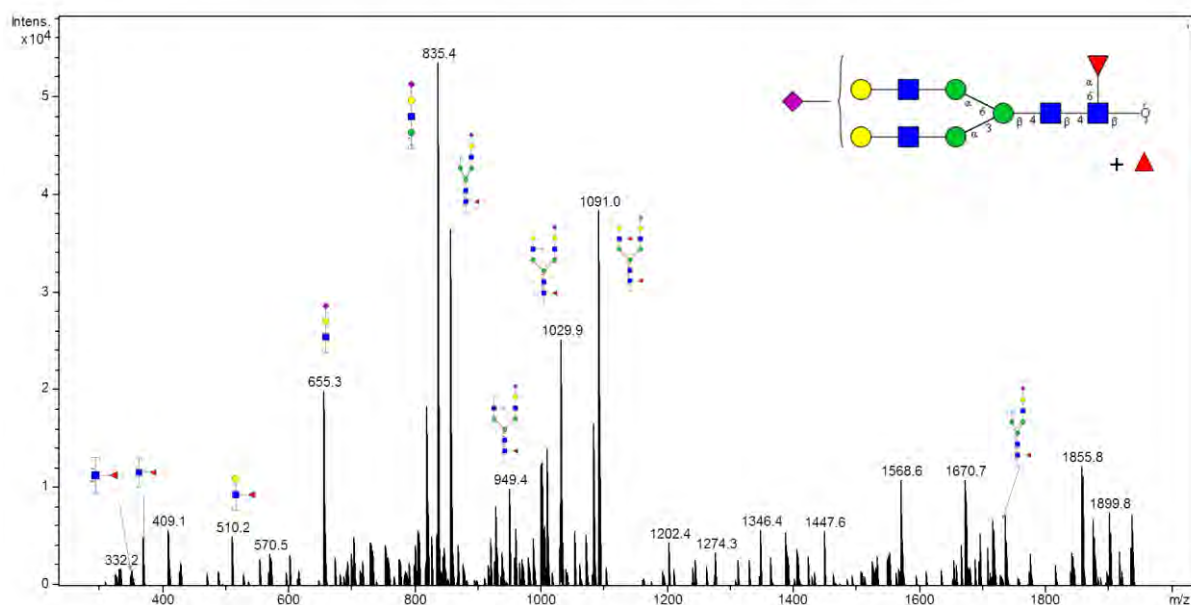
Glycan 134

Parent ion: *m/z* 1111.9^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended glycan fragment scheme..
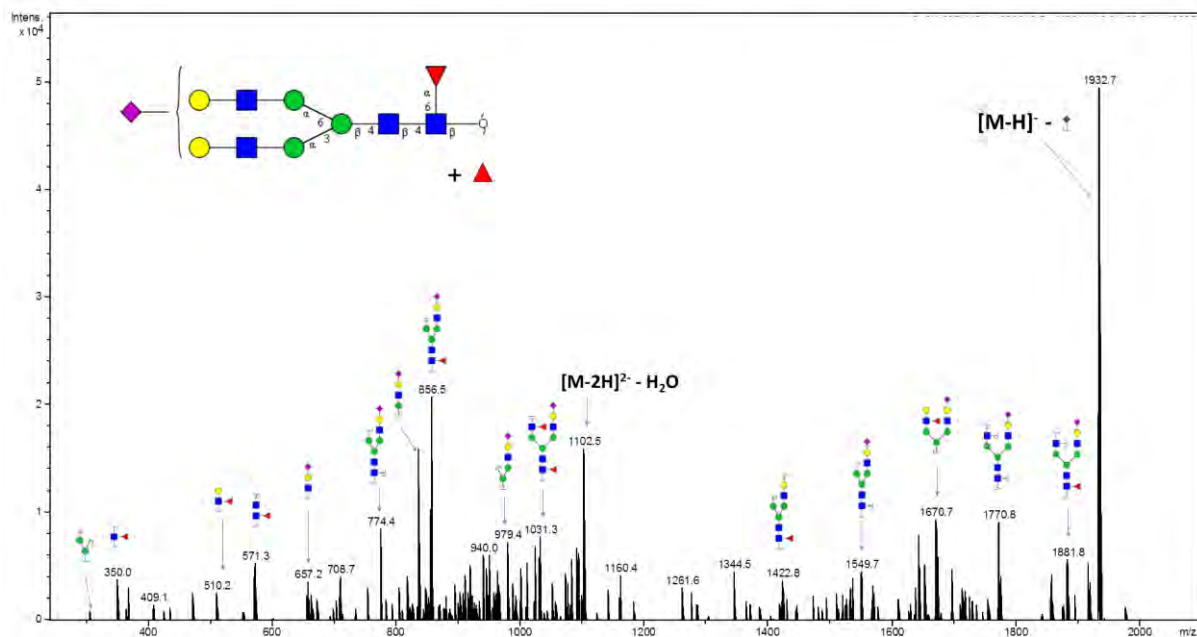
Glycan 135

Parent ion: *m/z* 1111.9^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
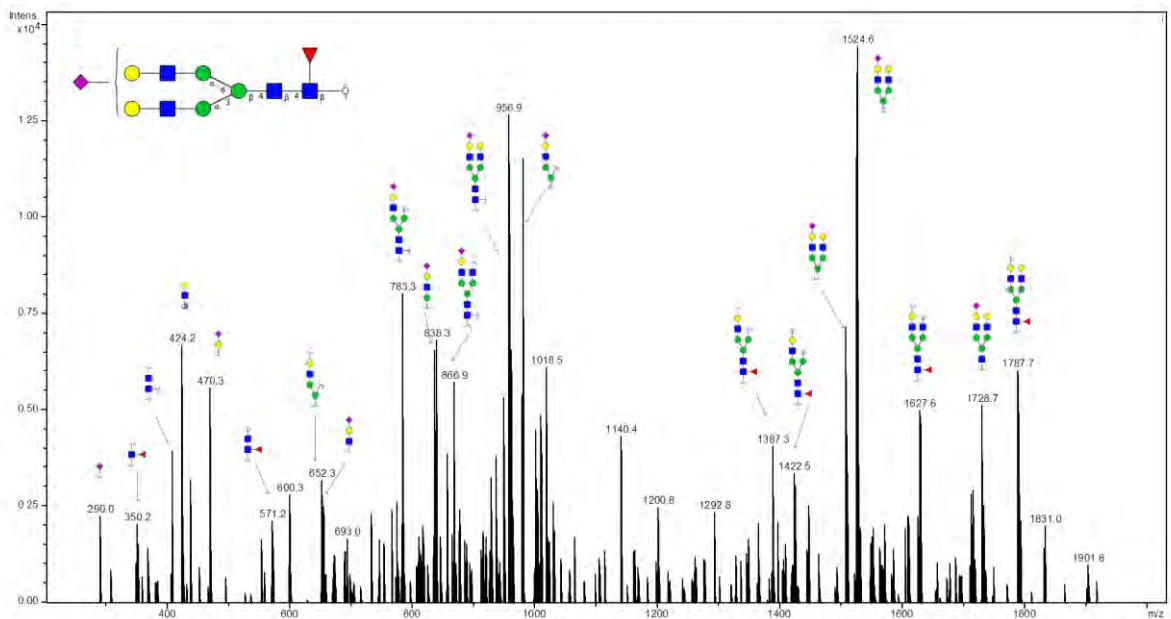
Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended glycan fragment scheme..
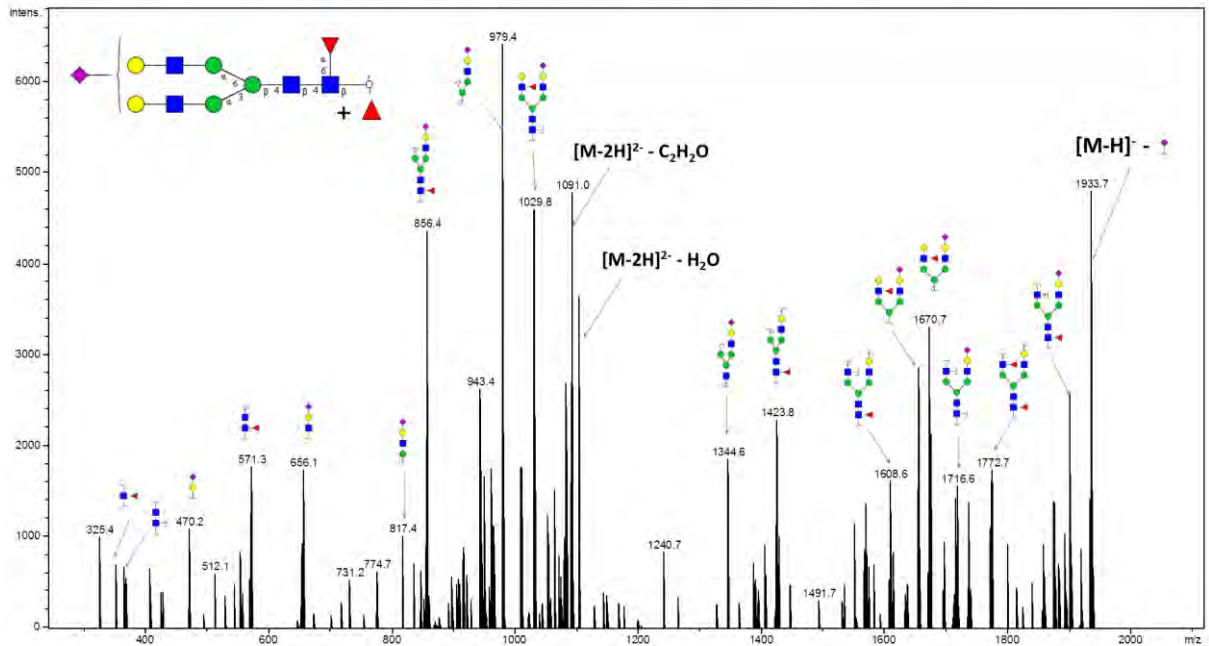
Glycan 136

Parent ion: *m/z* 1111.9^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
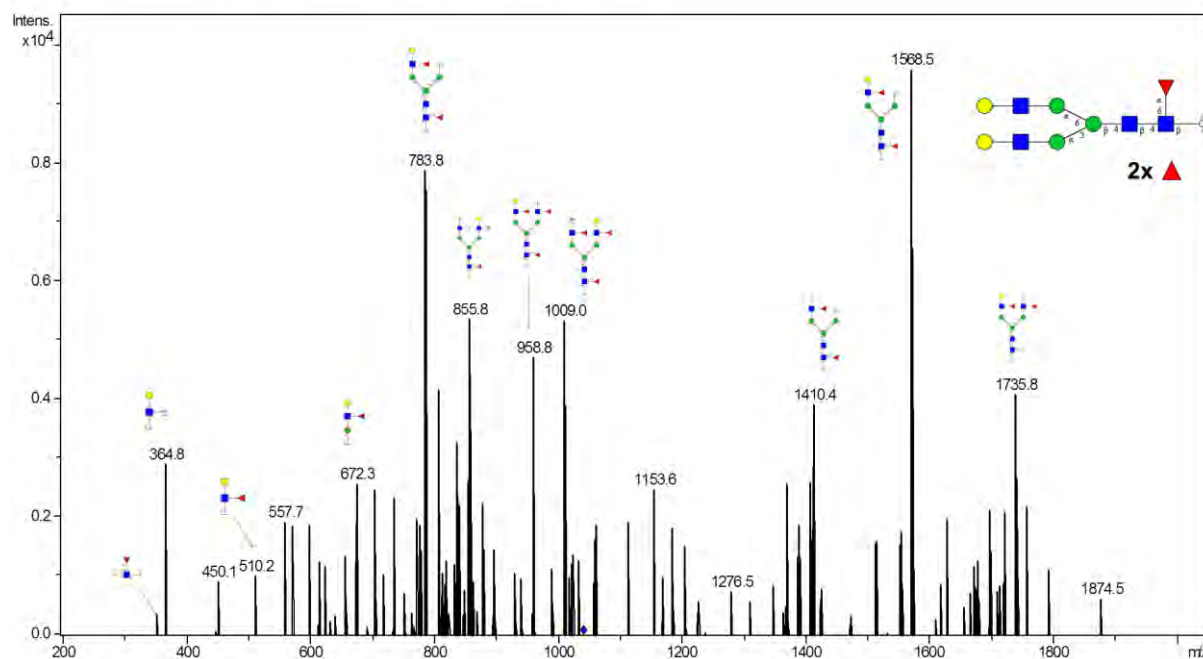
Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended glycan fragment scheme..
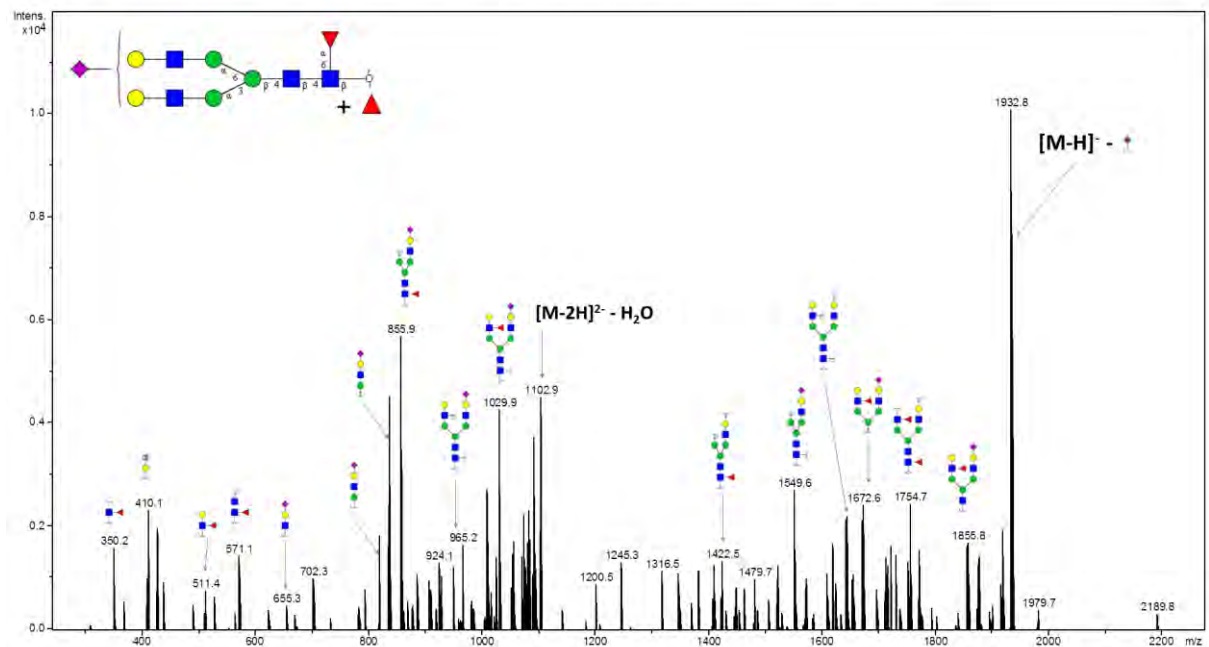
Glycan 137

Parent ion: *m/z* 1132.4²⁻

Composition: (Hex)₁ (HexNAc)₃ (Deoxyhexose)₂ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended glycan fragment scheme. Fuc, Fucose.

[M-2H]²⁻ - Fuc

[M-H]⁻ -
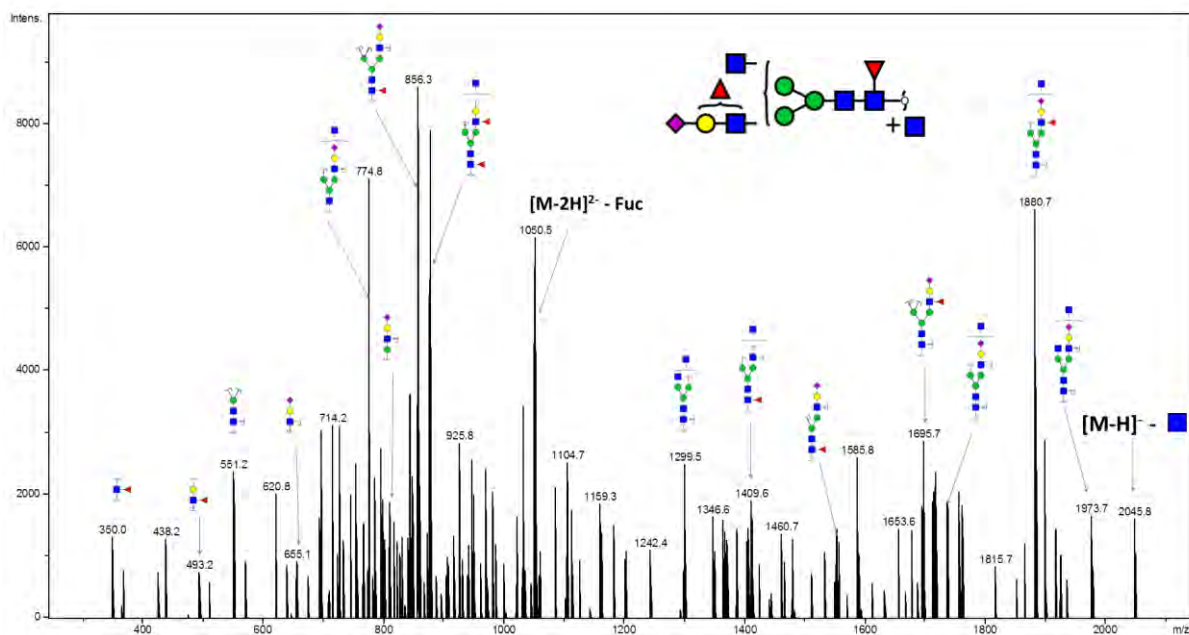
Glycan 138

Parent ion: *m/z* 1140.4²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended glycan fragment scheme. Hex, hexose.

[M-2H]²⁻ - Hex

[M-2H]²⁻ - C₂H₂O

[M-H]⁻ -
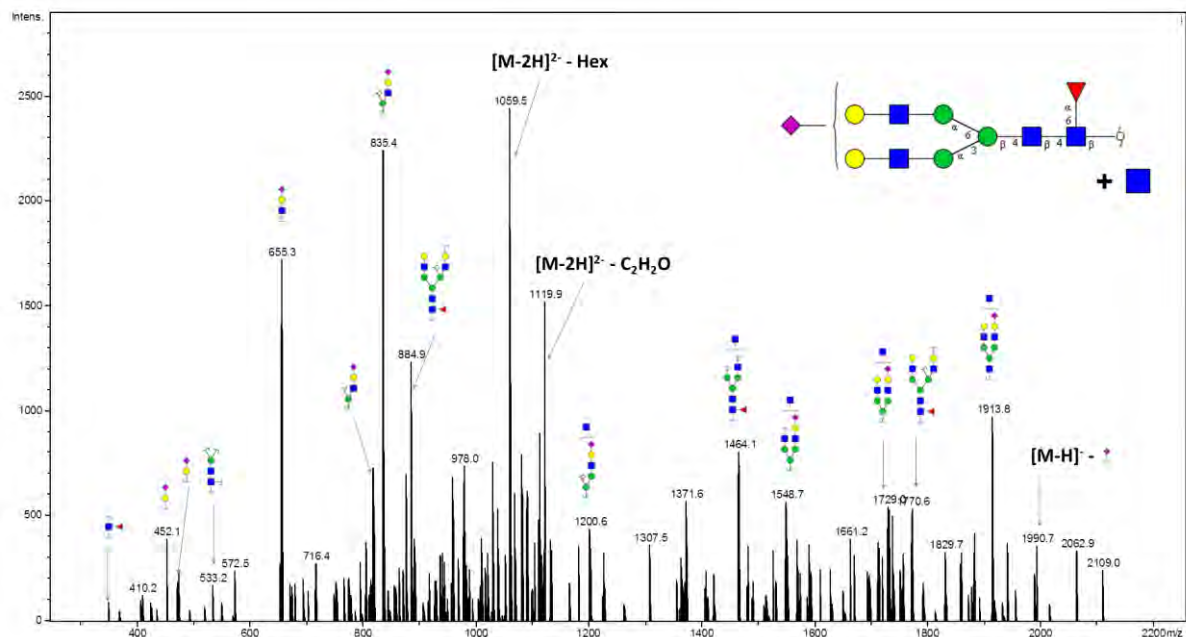
Glycan 139

Parent ion: *m/z* 1140.4[2-]

Composition: (Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended glycan fragment scheme. Fuc, Fucose.
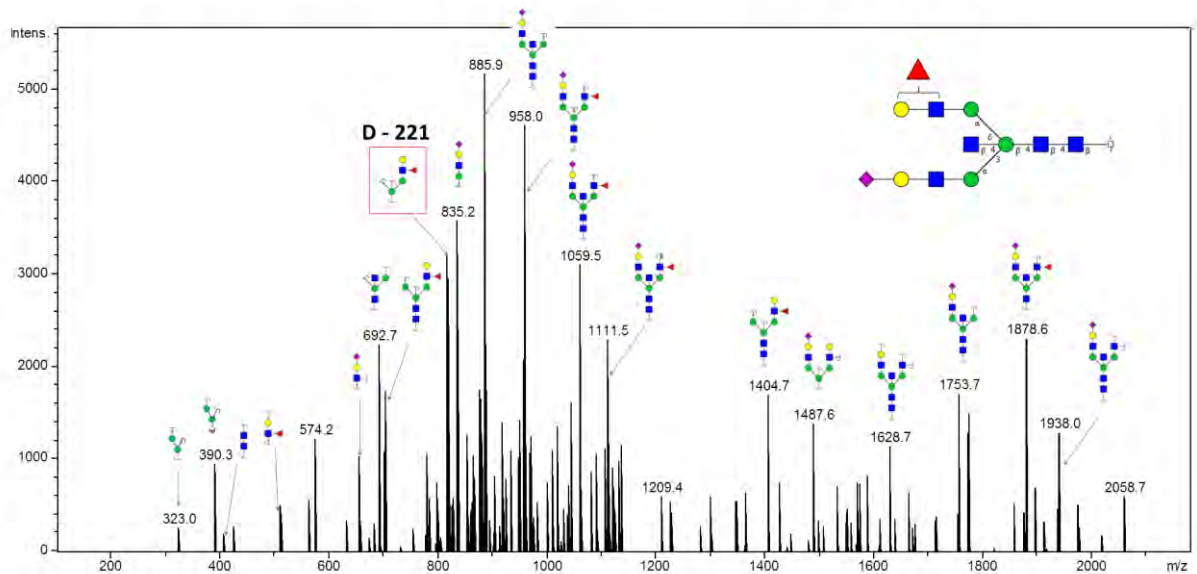
Glycan 140

Parent ion: *m/z* 1140.4[2-]

Composition: (Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended glycan fragment scheme.
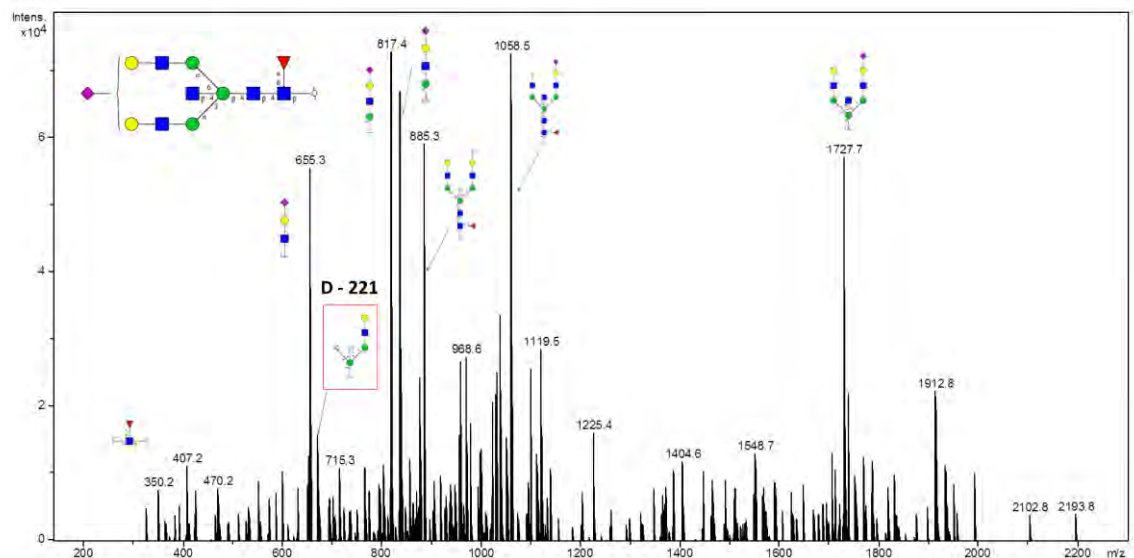
Glycan 141

Parent ion: *m/z* 1140.4²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended glycan fragment scheme.
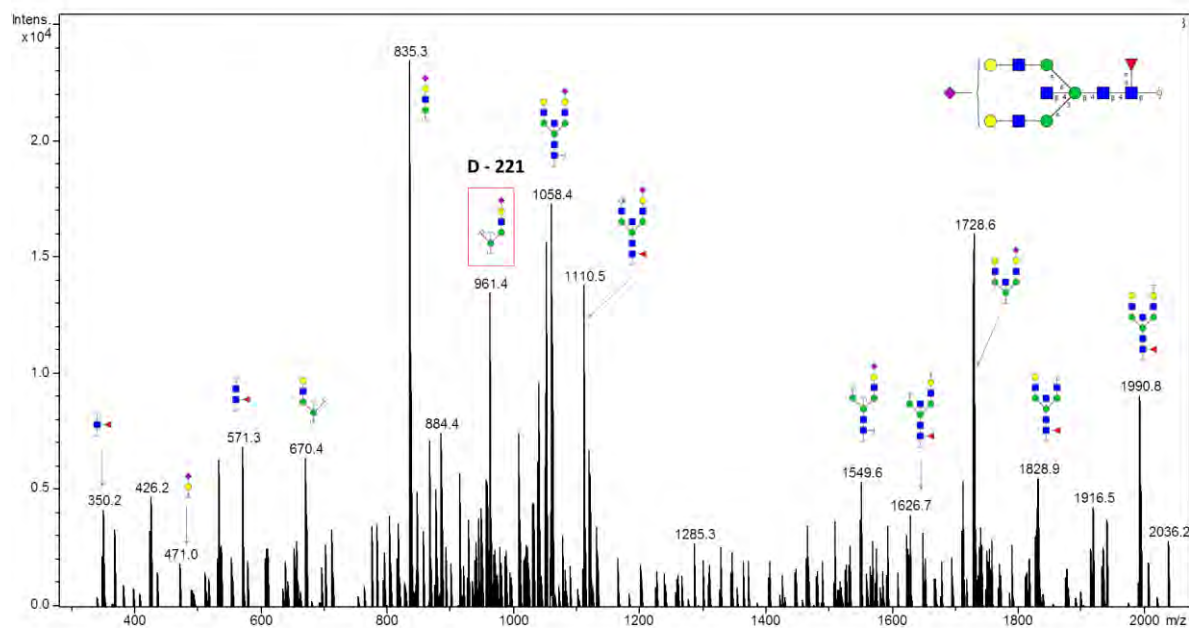
Glycan 142

Parent ion: *m/z* 1140.4²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended glycan fragment scheme.
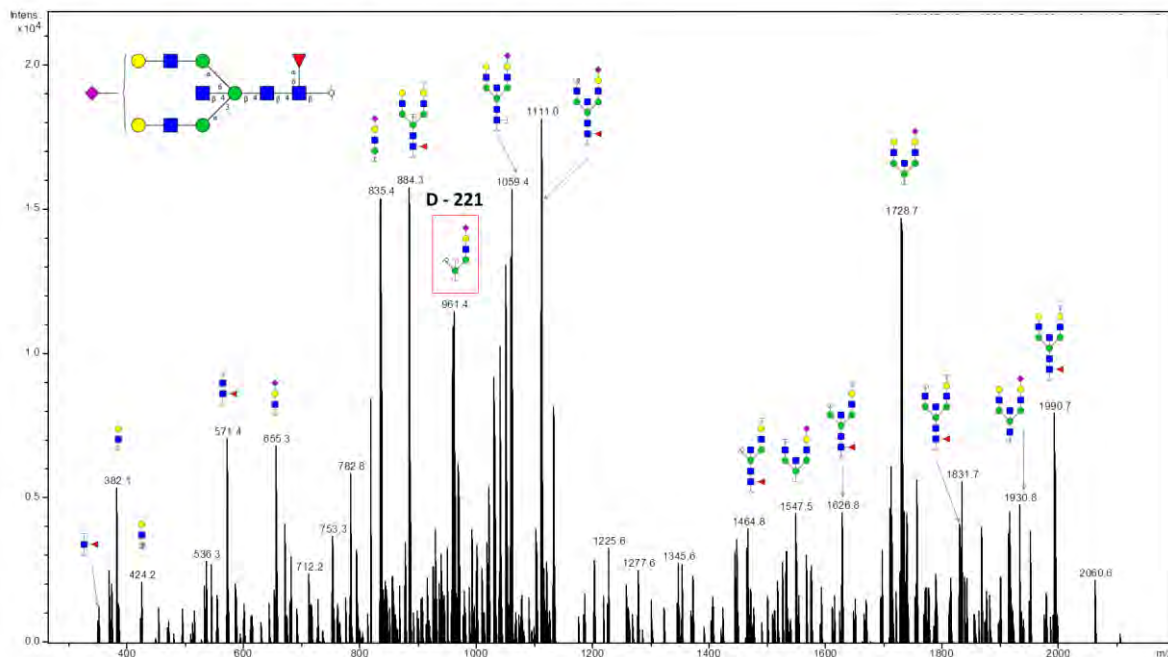
Glycan 143

Parent ion: *m/z* 1141.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₃ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended glycan fragment scheme.
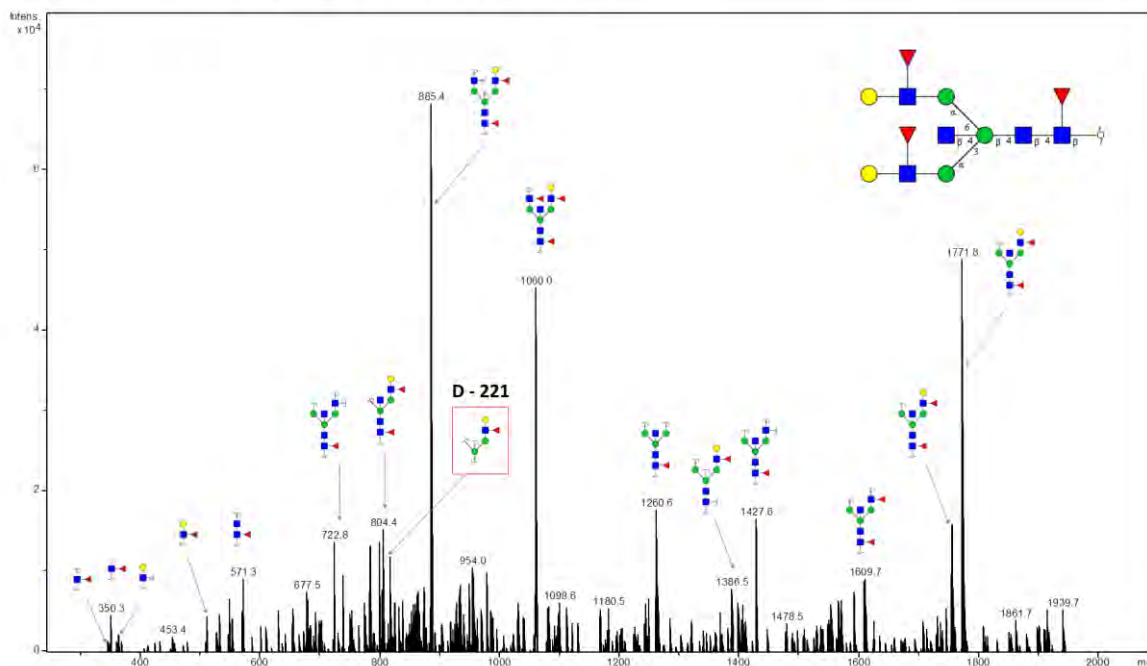
Glycan 144

Parent ion: *m/z* 1141.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₃ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme.
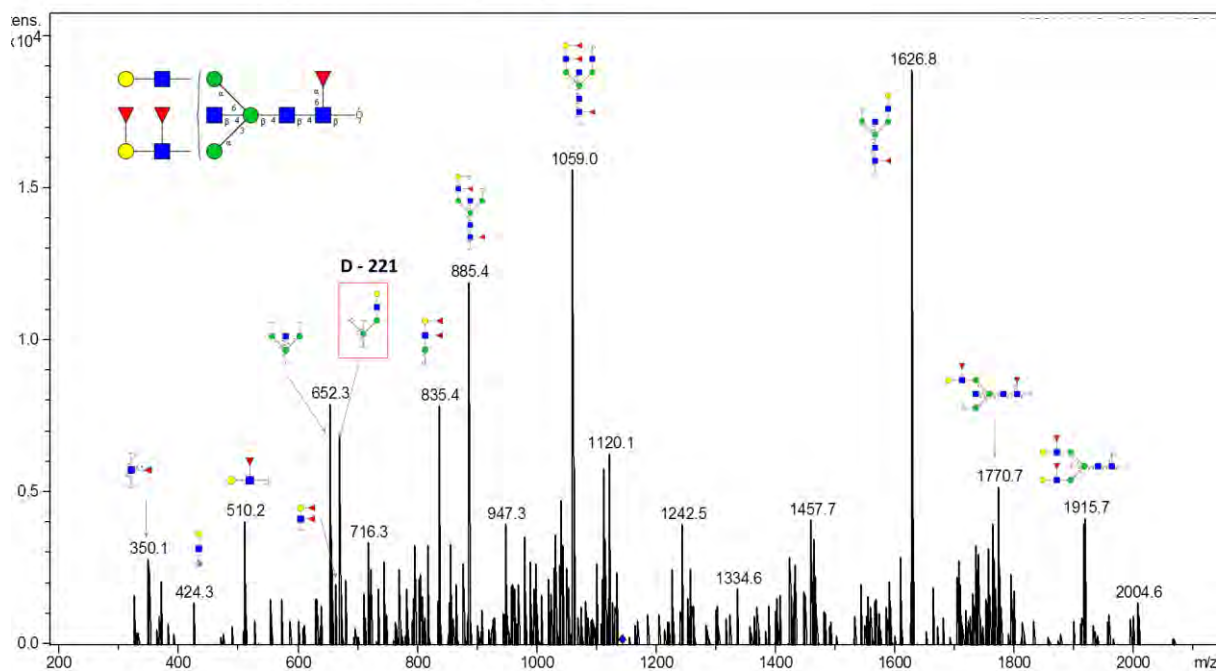
Glycan 145

Parent ion: *m/z* 1141.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₃ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme.
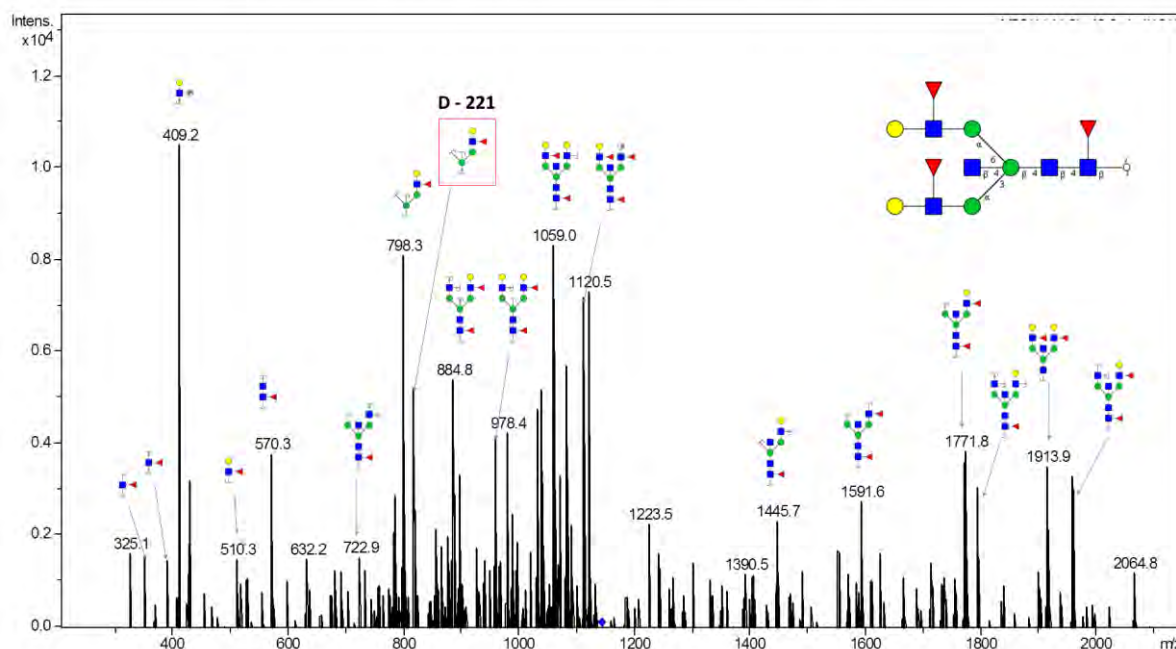
Glycan 146

Parent ion: *m/z* 1141.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₃ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm branching not intended in glycan fragment scheme. Depiction of specific branching also not intended.

190
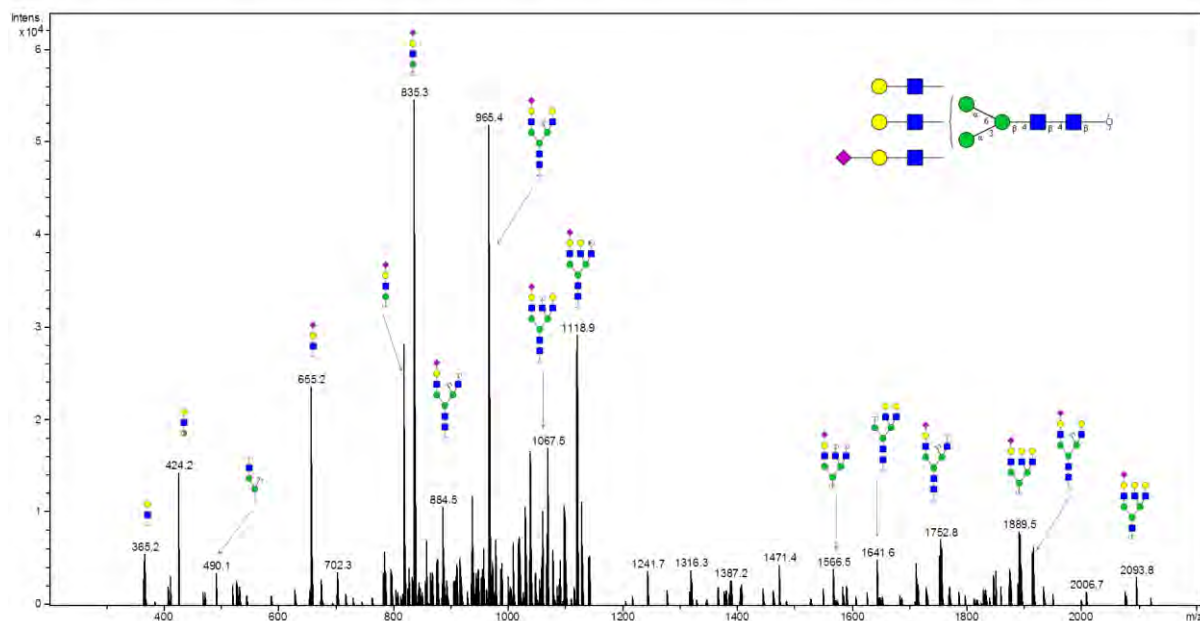
Glycan 147

Parent ion: *m/z* 1148.4²⁻

Composition: $(Hex)_3 (HexNAc)_3 (NeuAc)_1 + (Man)_3(GlcNAc)_2$

Notes: Distinction between 3-arm/6-arm branching not intended in glycan fragment scheme. Depiction of specific branching also not intended.
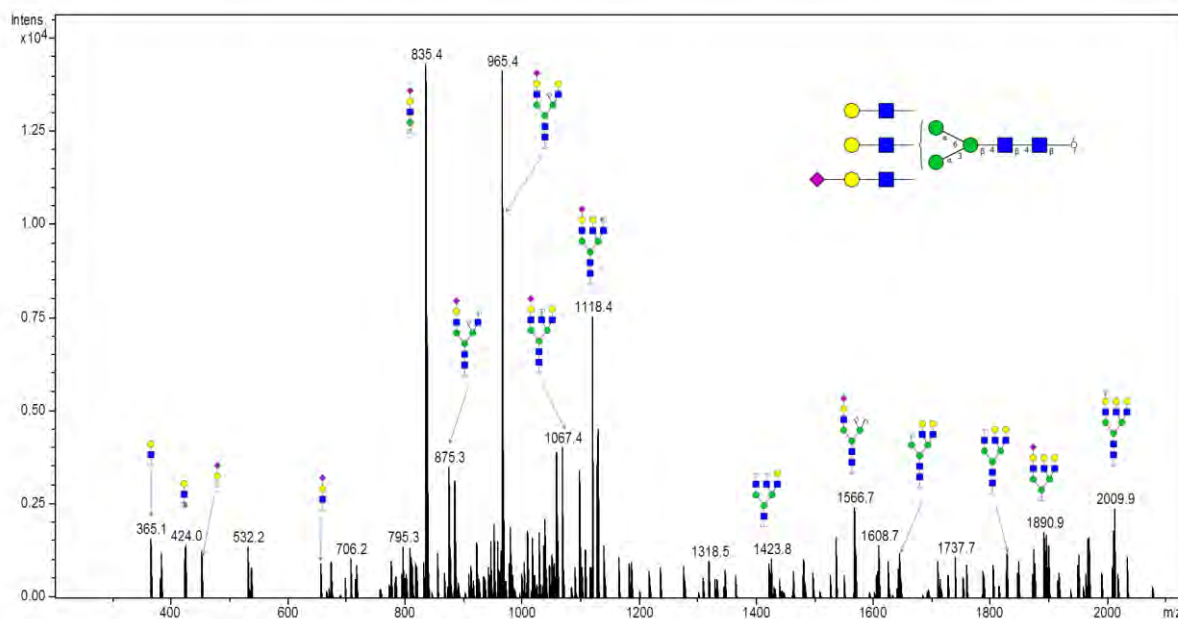
Glycan 148

Parent ion: *m/z* 1149.1²⁻

Composition: $(Hex)_3 (HexNAc)_3 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$

Notes: Distinction between 3-arm/6-arm branching not intended in glycan fragment scheme. Depiction of specific branching also not intended.
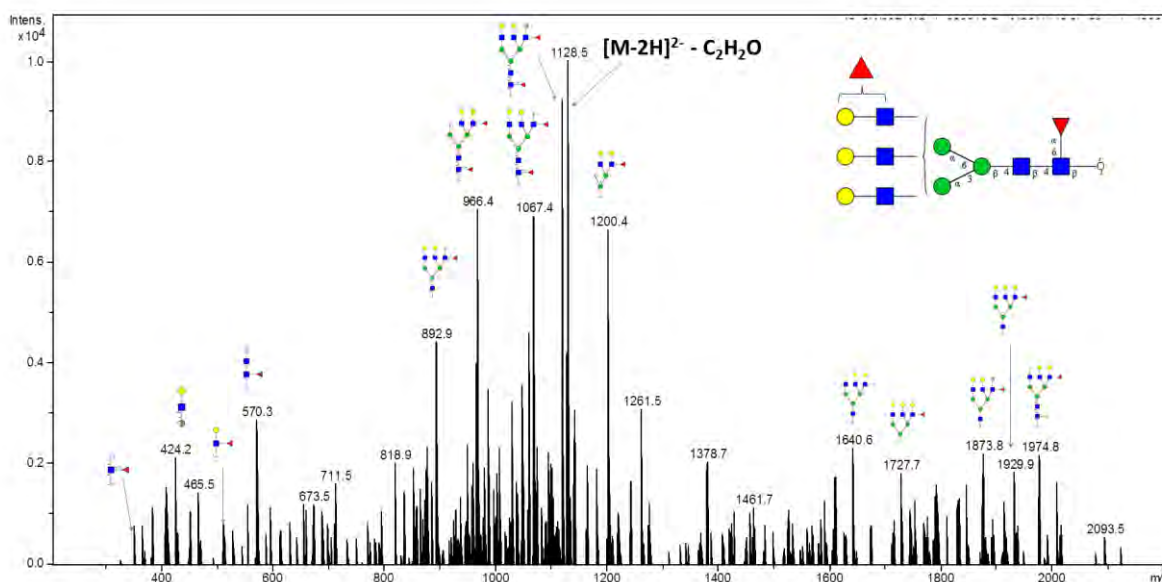
Glycan 149

Parent ion: *m/z* 1149.1²⁻

Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm branching and specific linkage for fucose not intended in glycan fragment scheme. Depiction of specific branching also not intended.
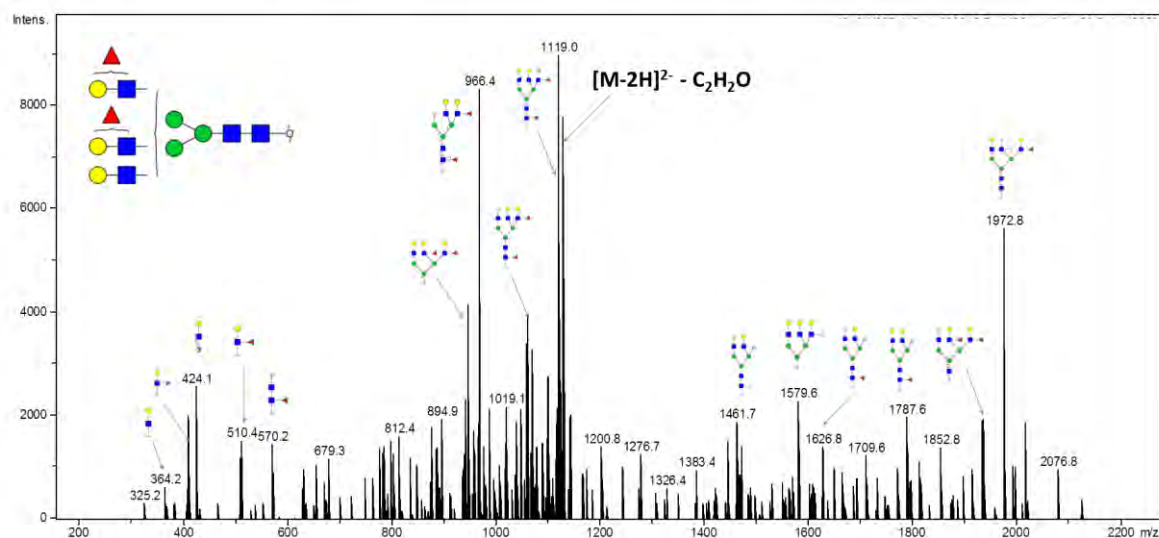
Glycan 150

Parent ion: *m/z* 1149.1²⁻

Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage for outer-arm fucose not intended in glycan fragment scheme. Depiction of specific branching also not intended.
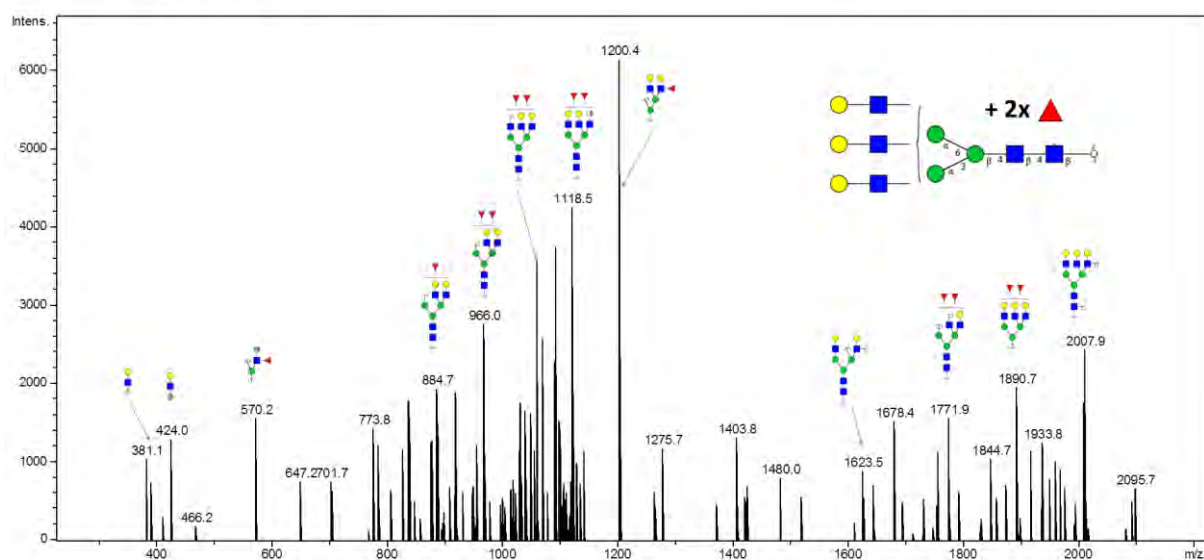
Glycan 151

Parent ion: *m/z* 1152.9²⁻

Composition: (HexNAc)₄ (Deoxyhexose)₂ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage for outer-arm fucose not intended in glycan fragment scheme.
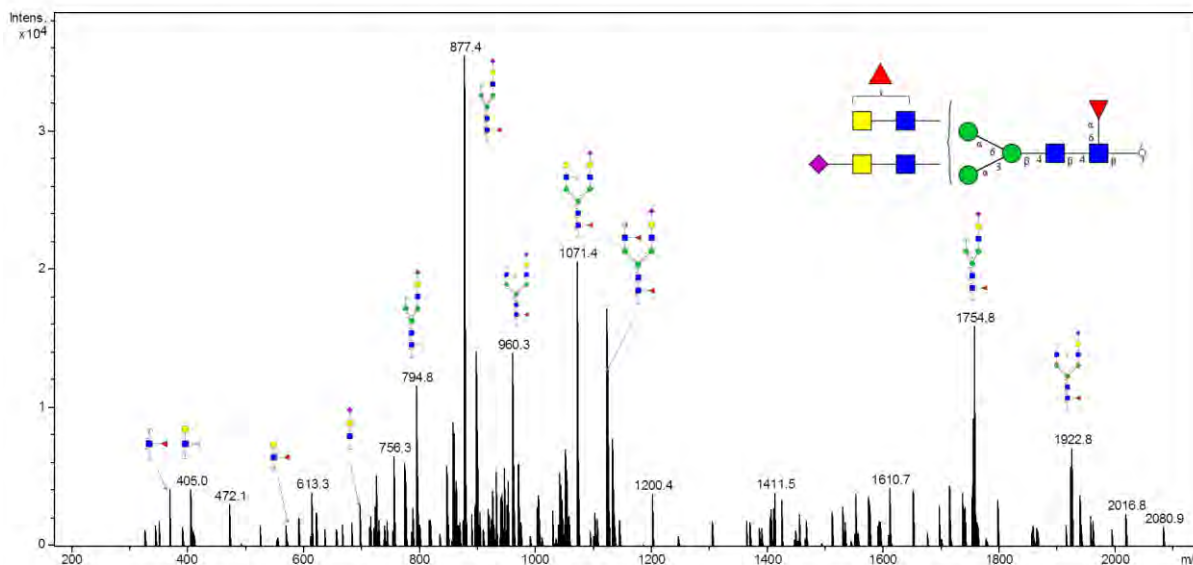
Glycan 152

Parent ion: *m/z* 1177.5²⁻

Composition: (Hex)₃ (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and depiction of specific branching not intended in glycan fragment scheme.
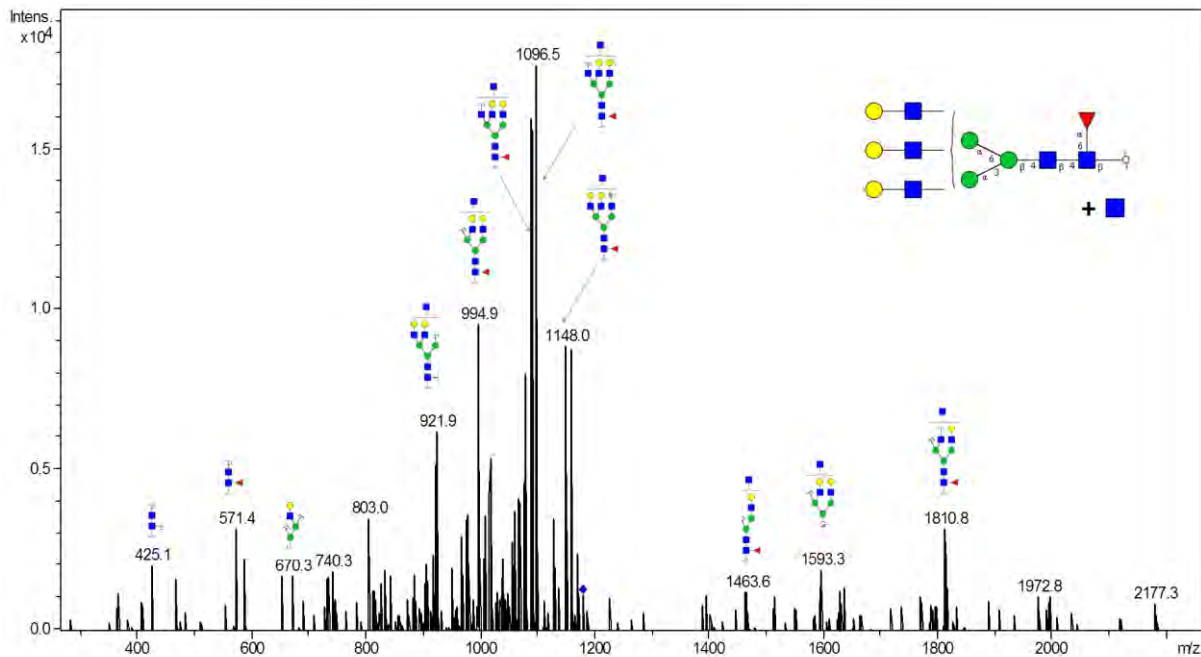
Glycan 153

Parent ion: *m/z* 1184.5^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme.
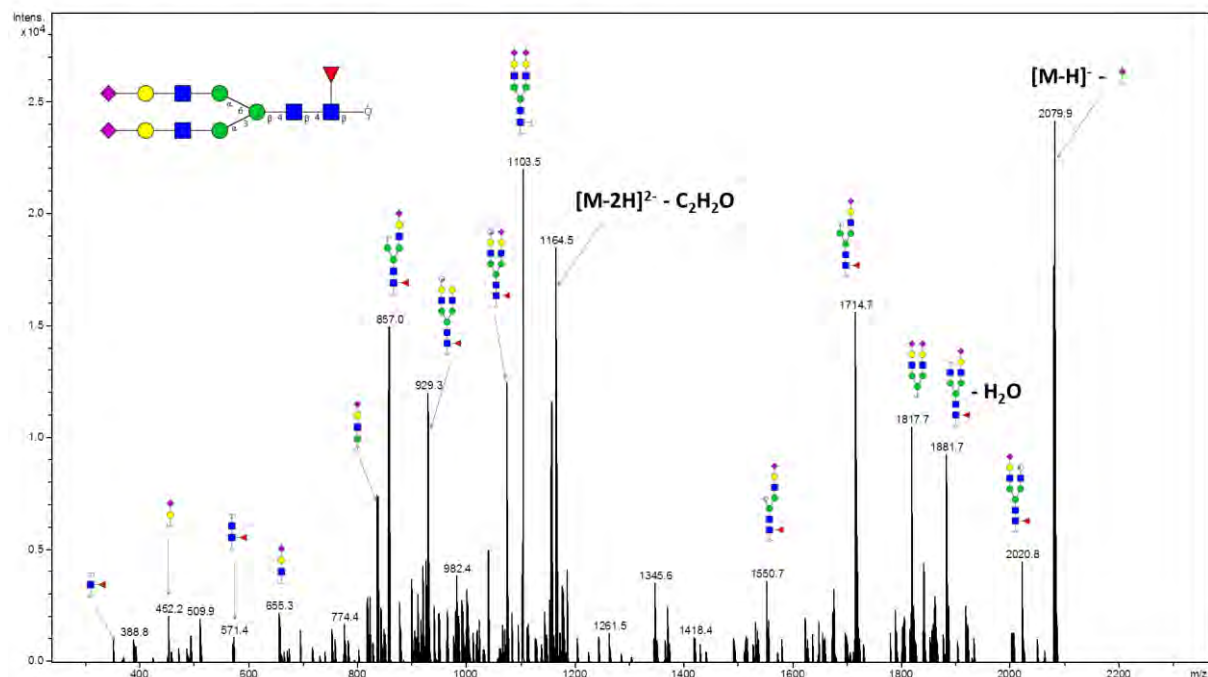
Glycan 154

Parent ion: *m/z* 1184.5^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme.
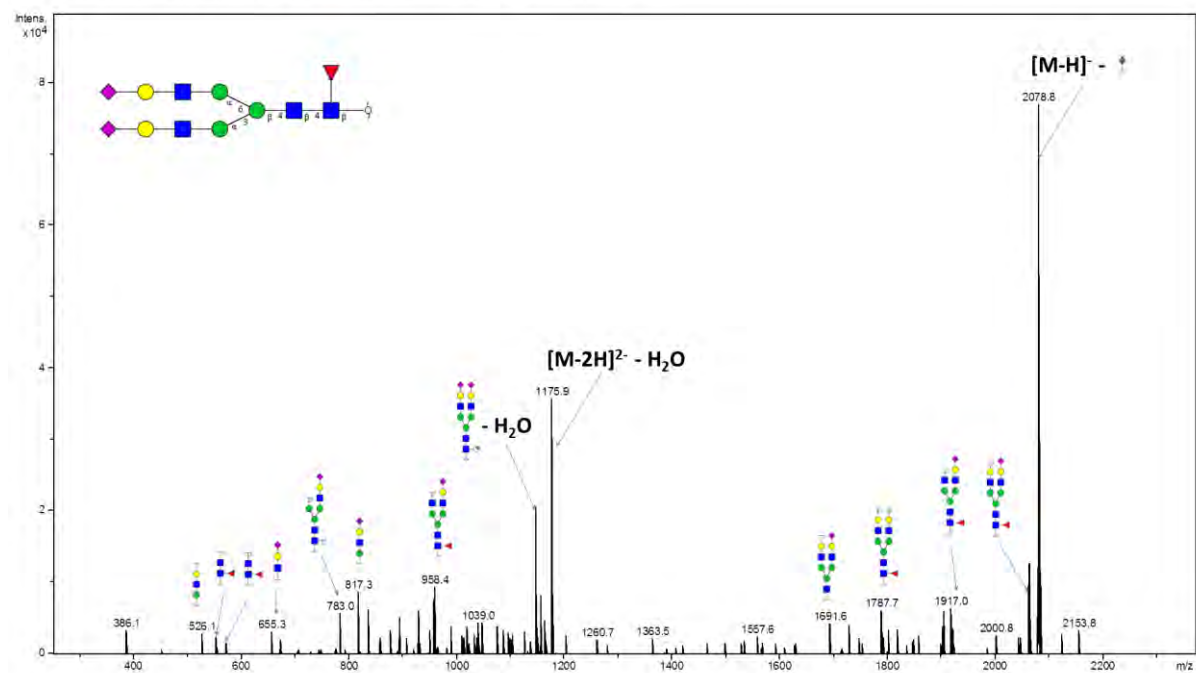
Glycan 155

Parent ion: *m/z* 1184.5^{2-}

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme.
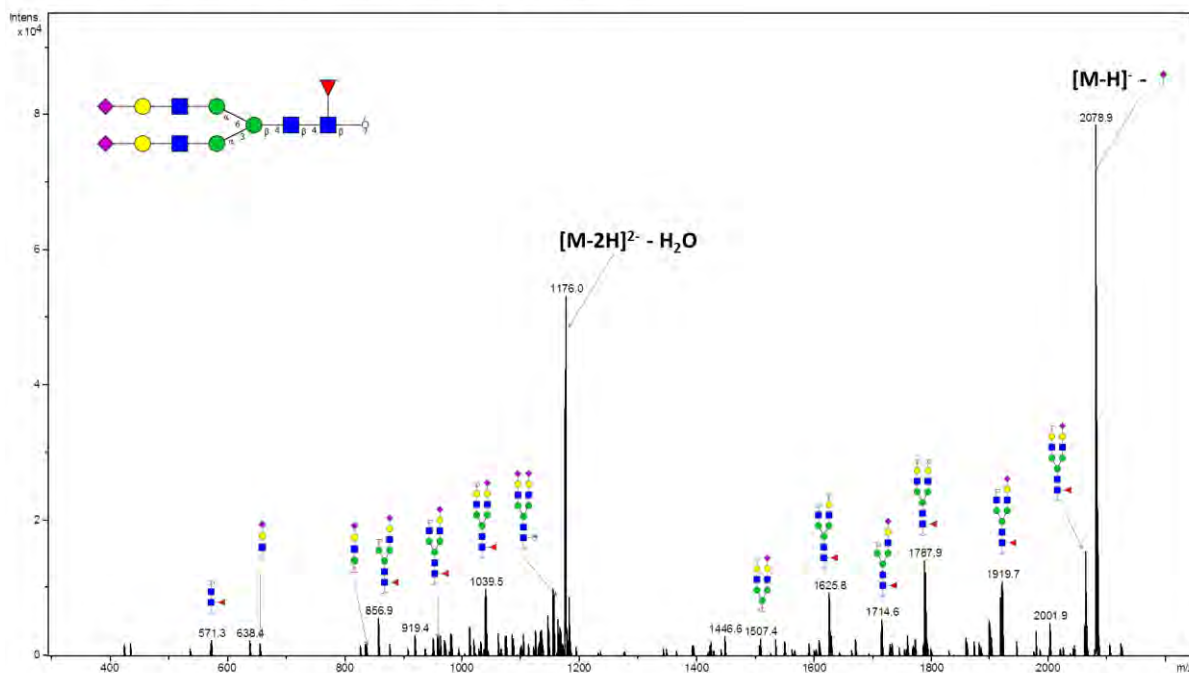
Glycan 158

Parent ion: *m/z* 1213.6^{2-}

Composition: (Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_2$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
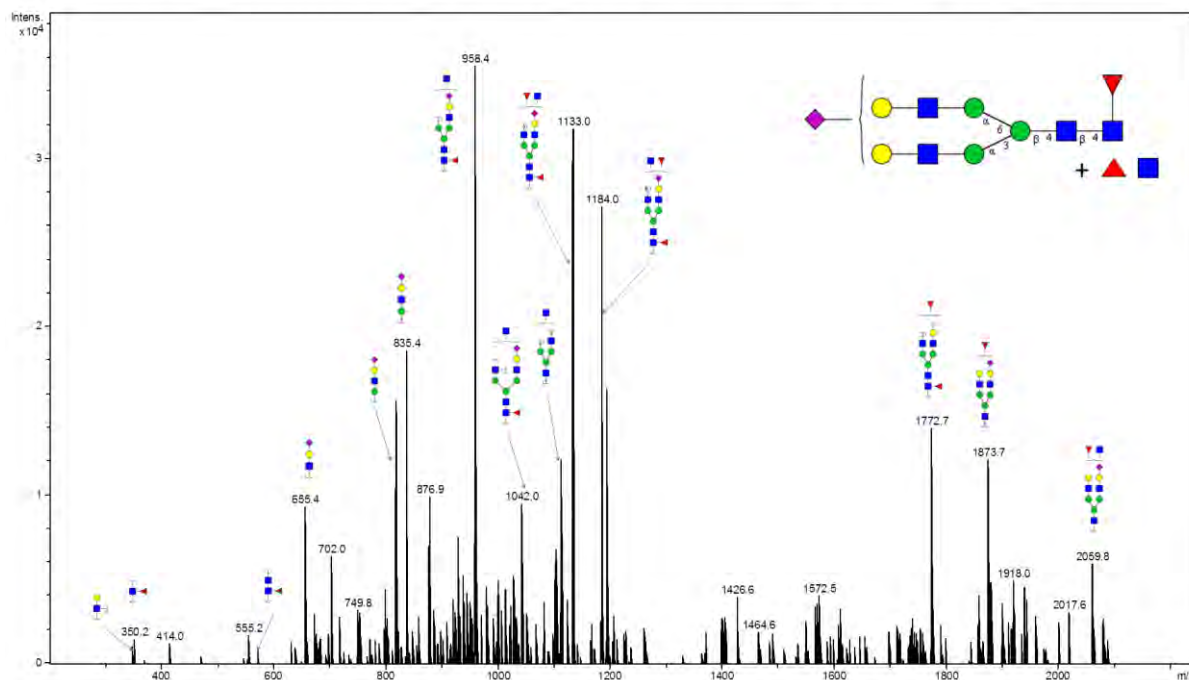
Glycan 159

Parent ion: *m/z* 1221.5^{2-}

Composition: (Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
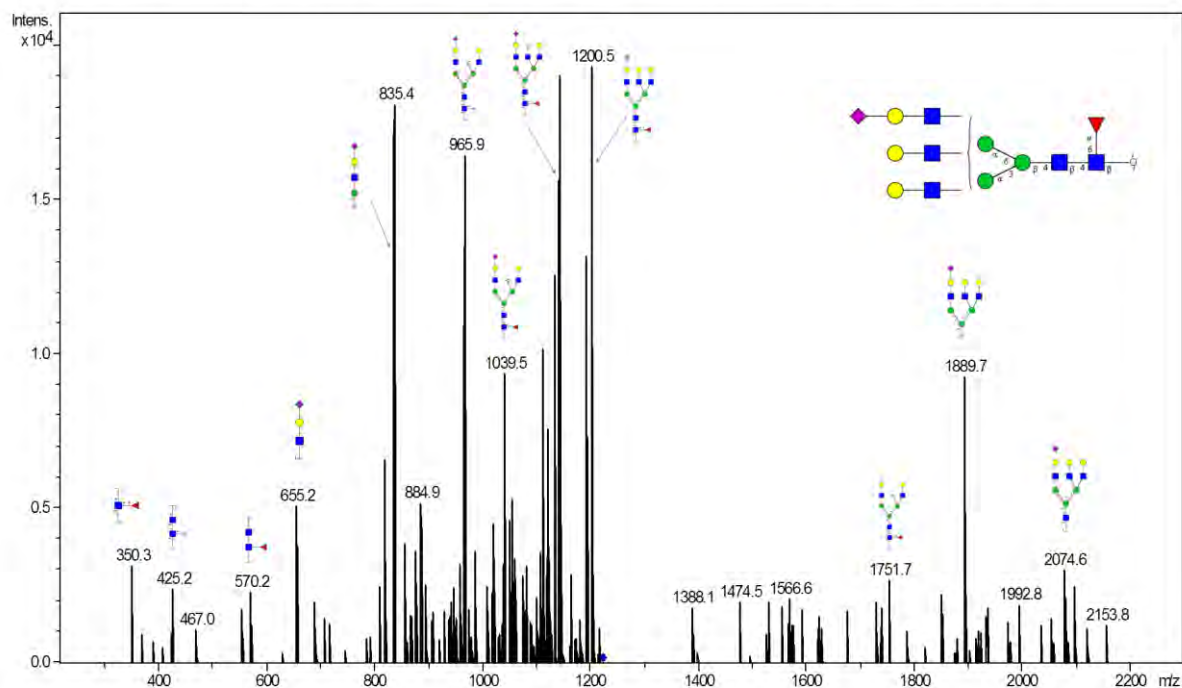
Glycan 160

Parent ion: *m/z* 1221.5^{2-}

Composition: (Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
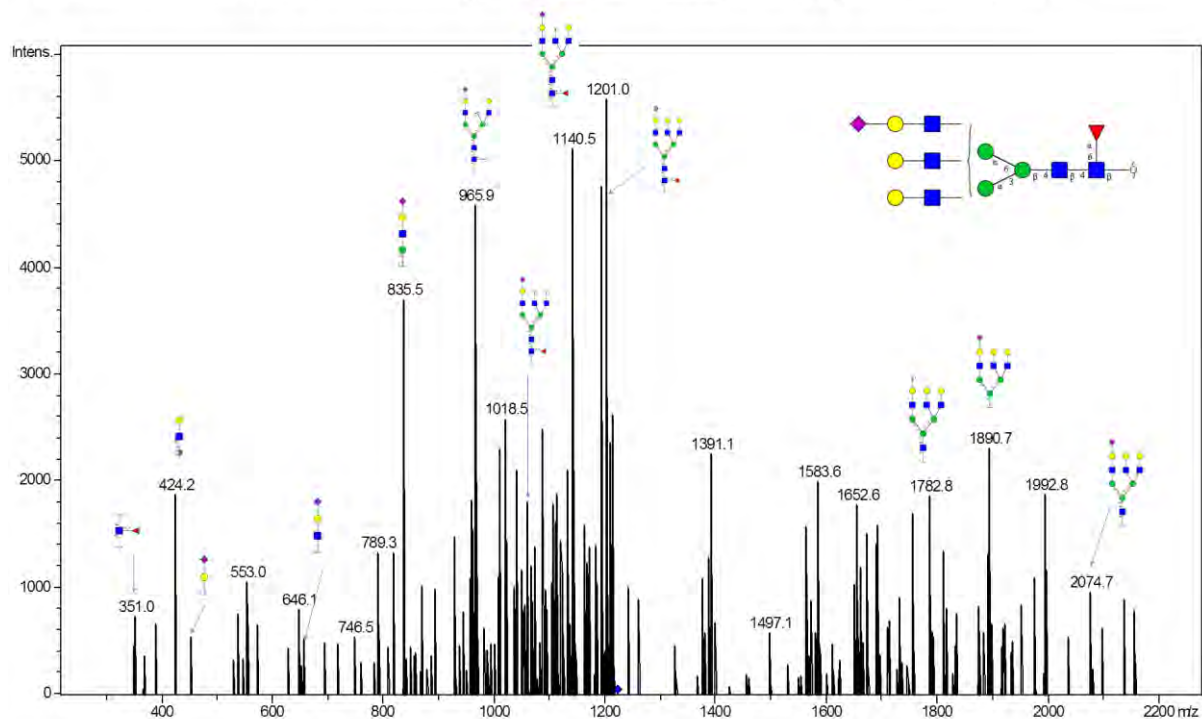
Glycan 161

Parent ion: *m/z* 1221.5²⁻

Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
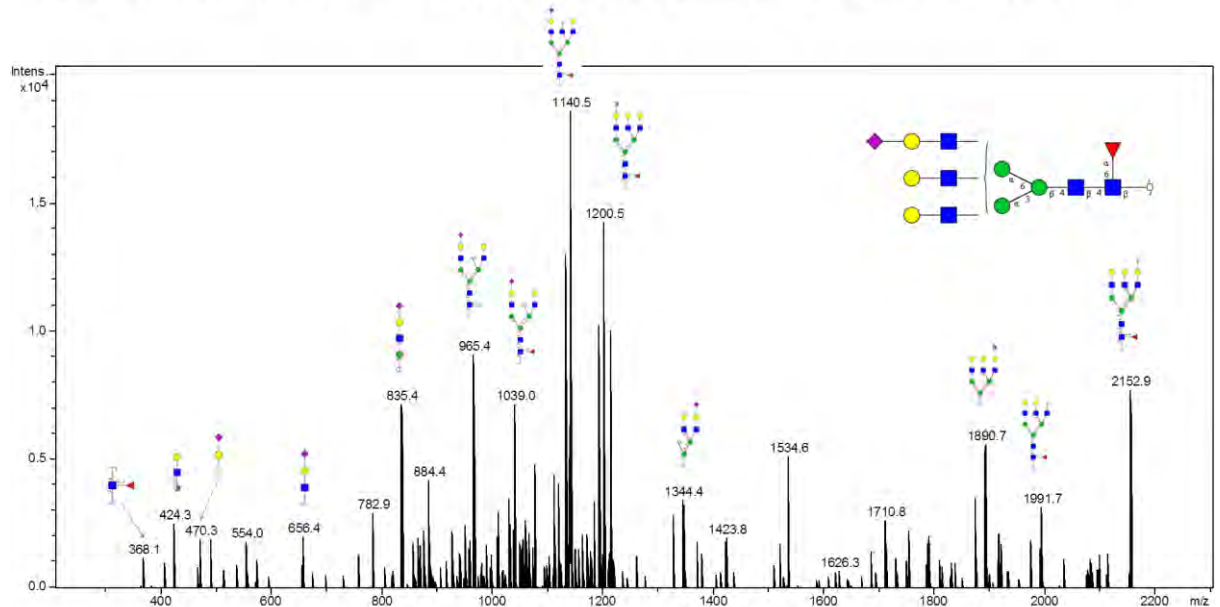
Glycan 162

Parent ion: *m/z* 1222.0²⁻

Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₃ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
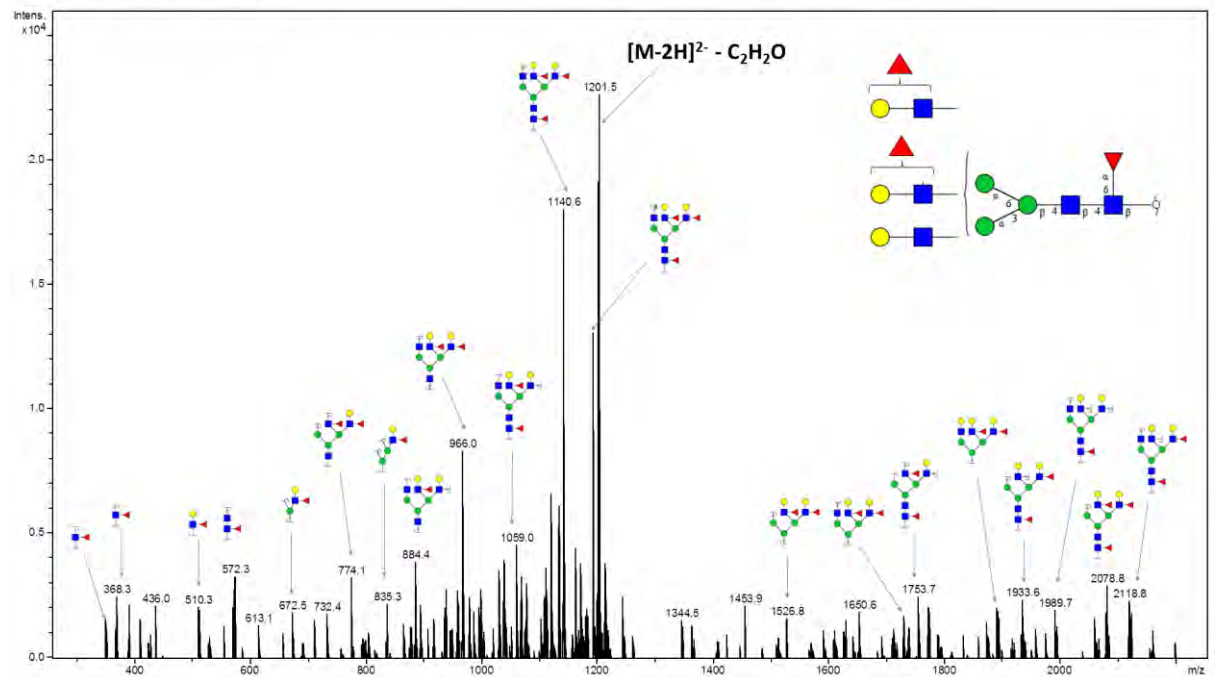
Glycan 163

Parent ion: *m/z* 1222.0²⁻

Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₃ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm and specific linkage of outer-arm fucose not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
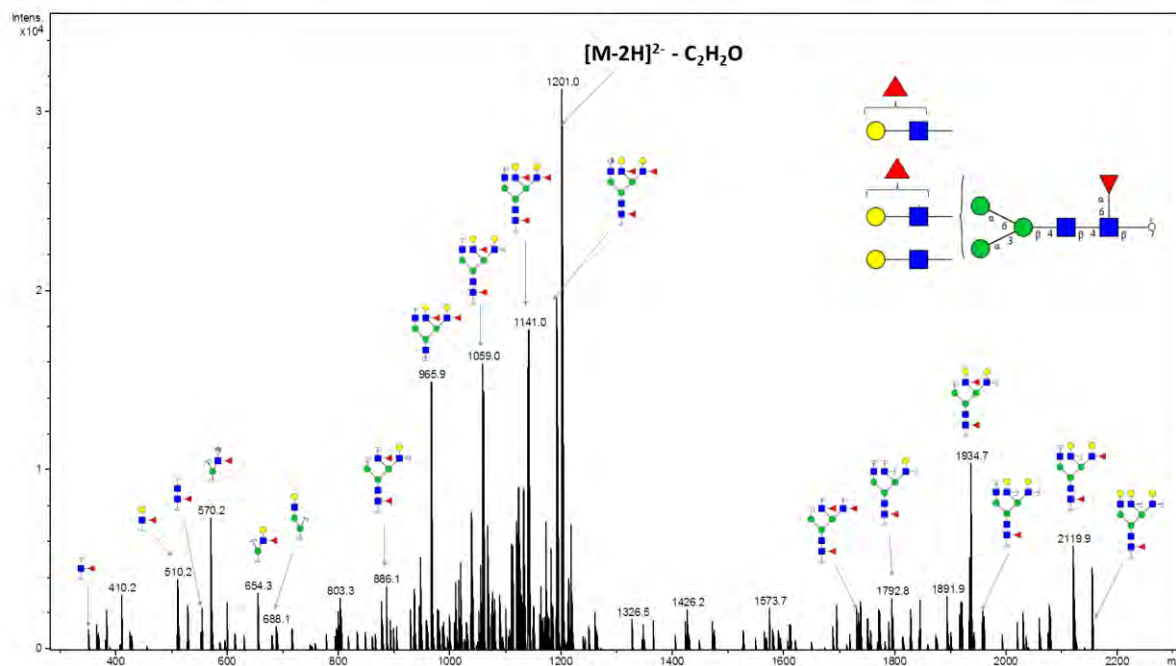
Glycan 166

Parent ion: *m/z* 1258.5²⁻

Composition: (Hex)₄ (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
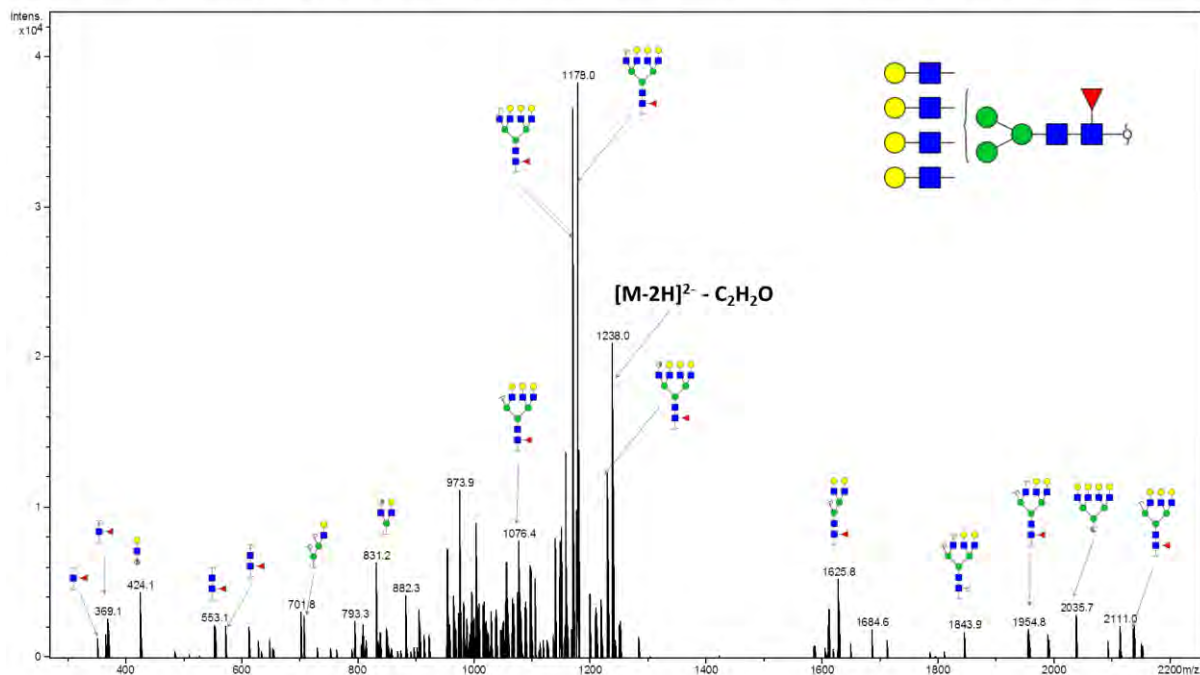
Glycan 167

Parent ion: *m/z* 1258.5²⁻

Composition: (Hex)₄ (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
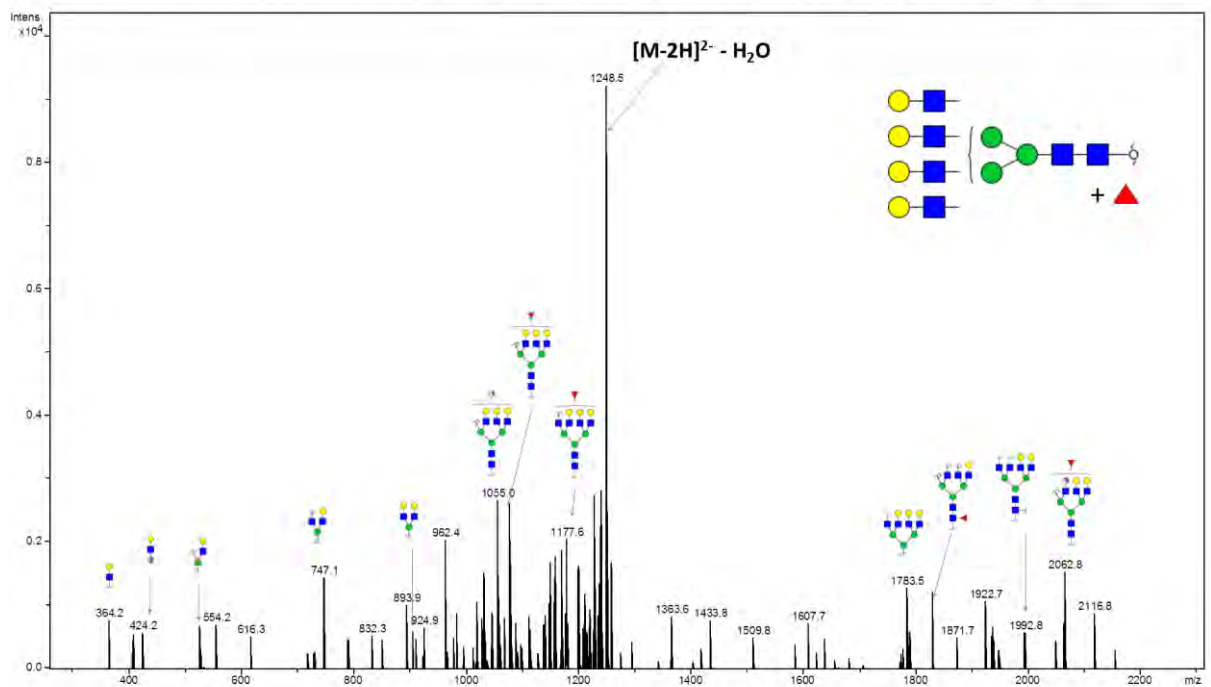
Glycan 168

Parent ion: *m/z* 1286.0²⁻

Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
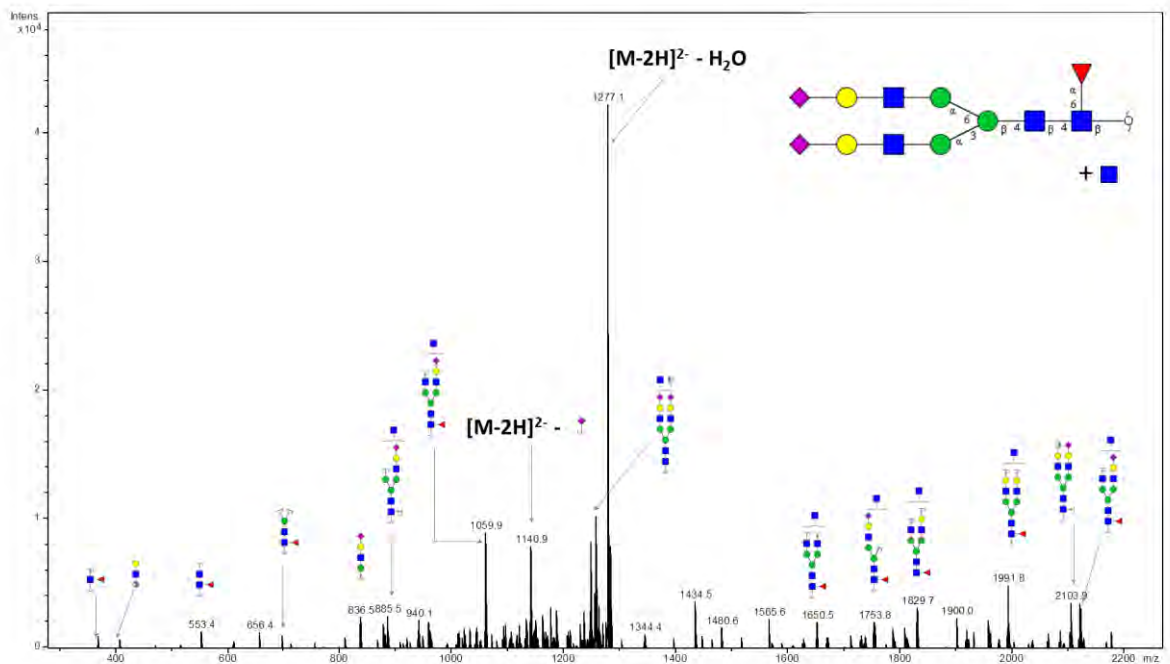
Glycan 169

Parent ion: *m/z* 1294.0²⁻

Composition: (Hex)₃ (HexNAc)₃ (NeuAc)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
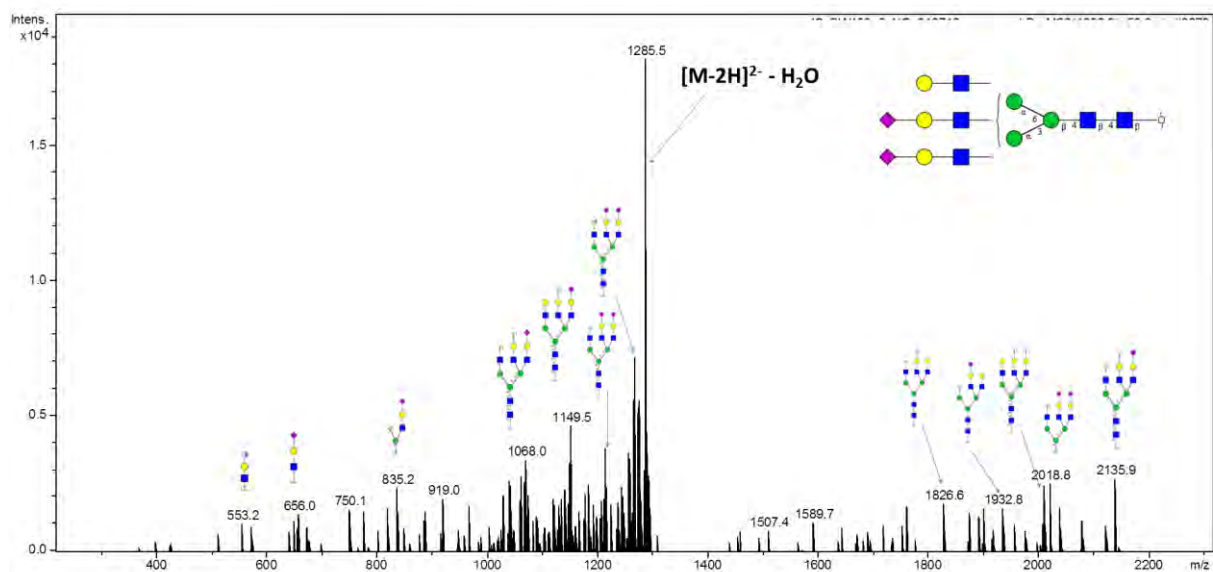
Glycan 170

Parent ion: *m/z* 1330.6²⁻

Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₃ (NeuAc)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
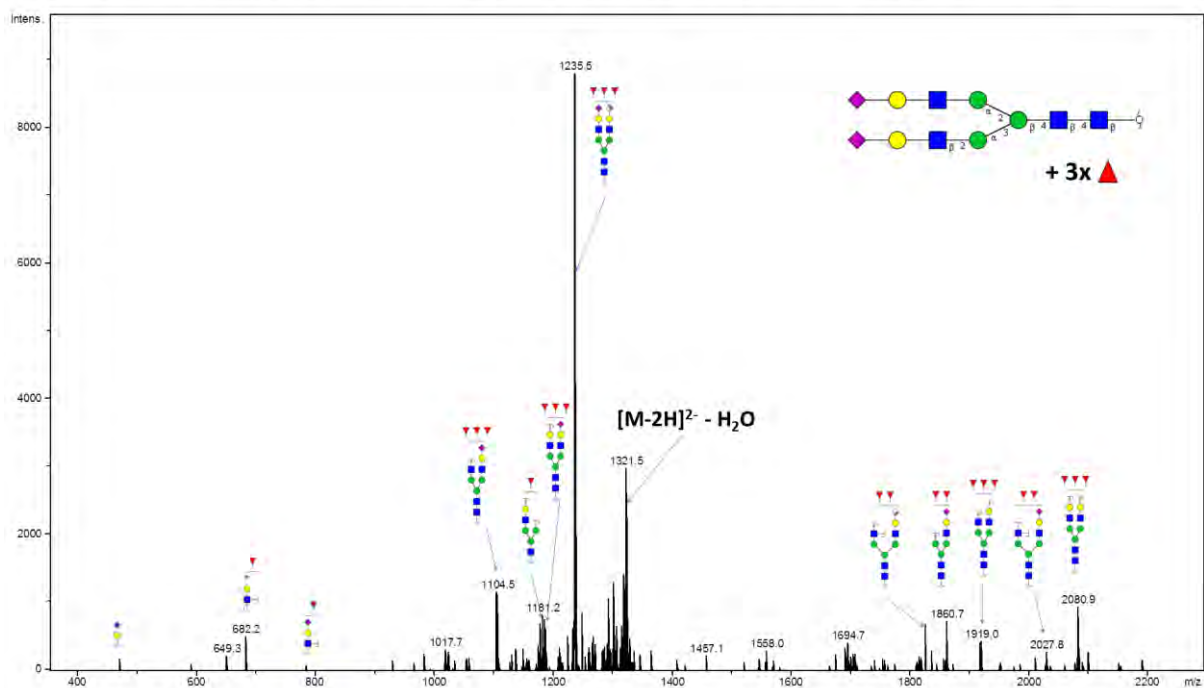
Glycan 171

Parent ion: *m/z* 1331.0²⁻

Composition: (Hex)₄ (HexNAc)₄ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
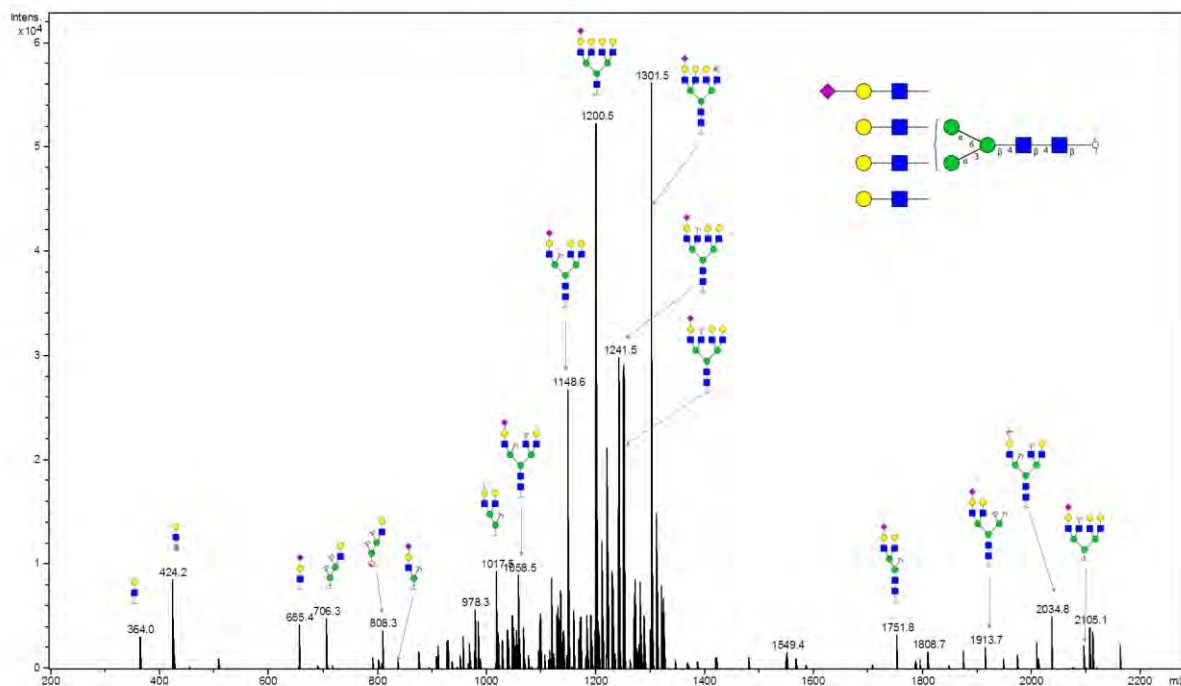
Glycan 172

Parent ion: *m/z* 1331.0²⁻

Composition: (Hex)₄ (HexNAc)₄ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
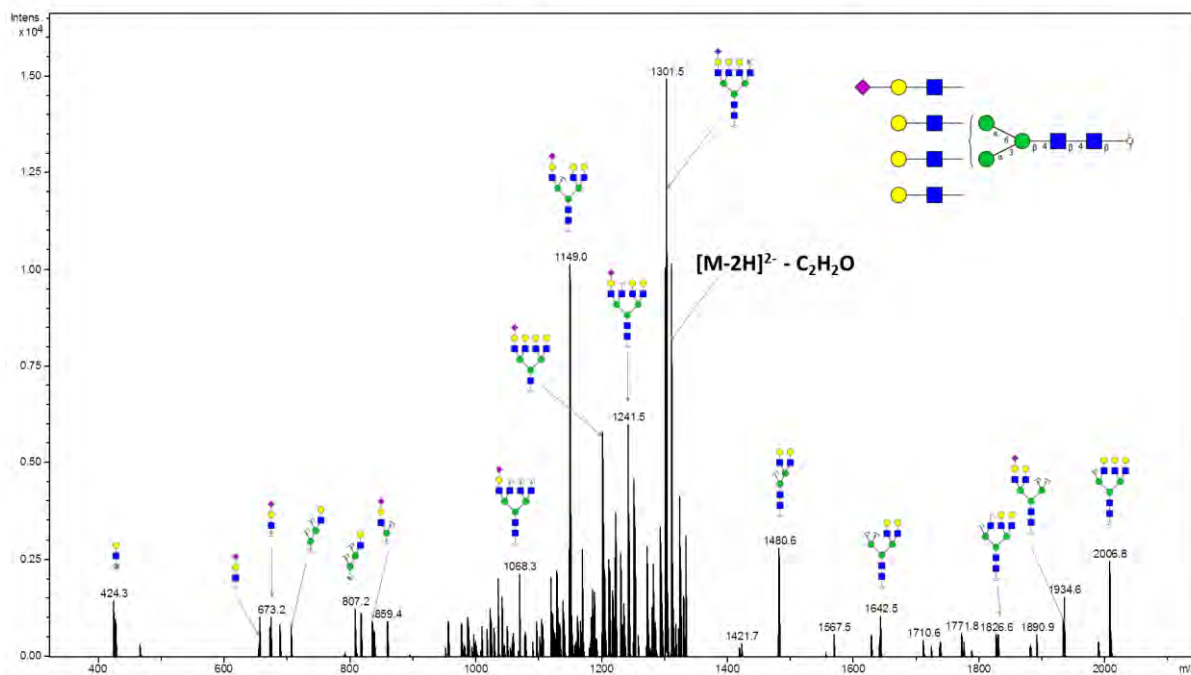
Glycan 173

Parent ion: *m/z* 1367.0²⁻

Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
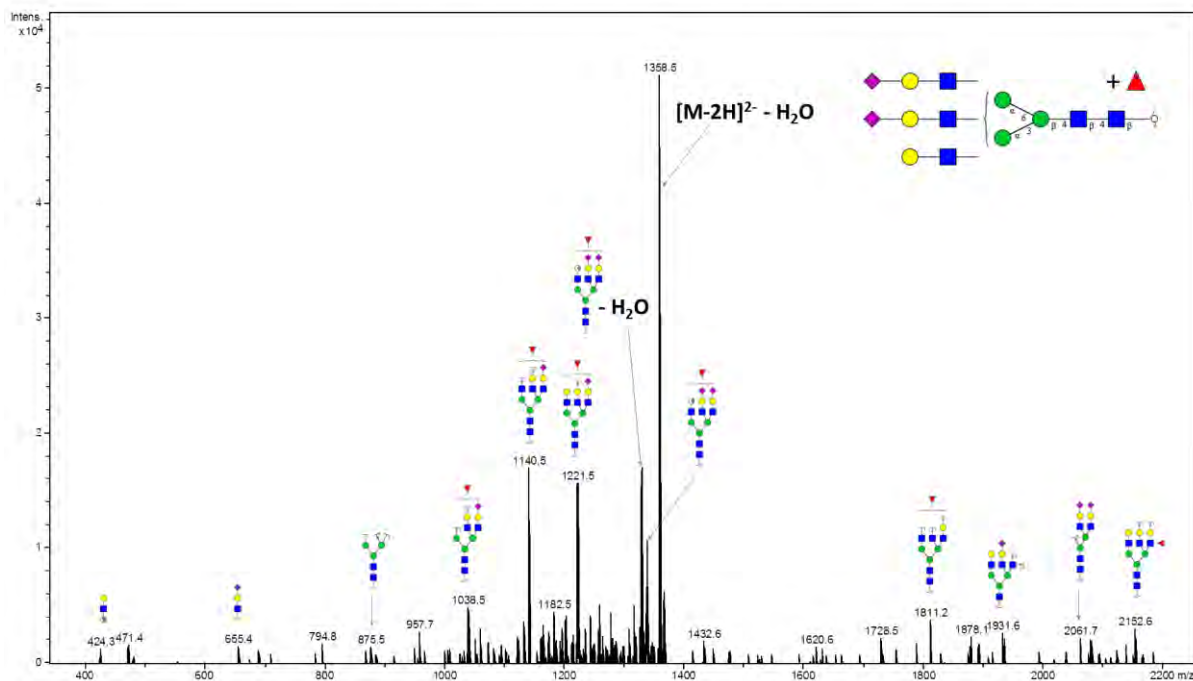
Glycan 174

Parent ion: *m/z* 1367.0²⁻

Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₂ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
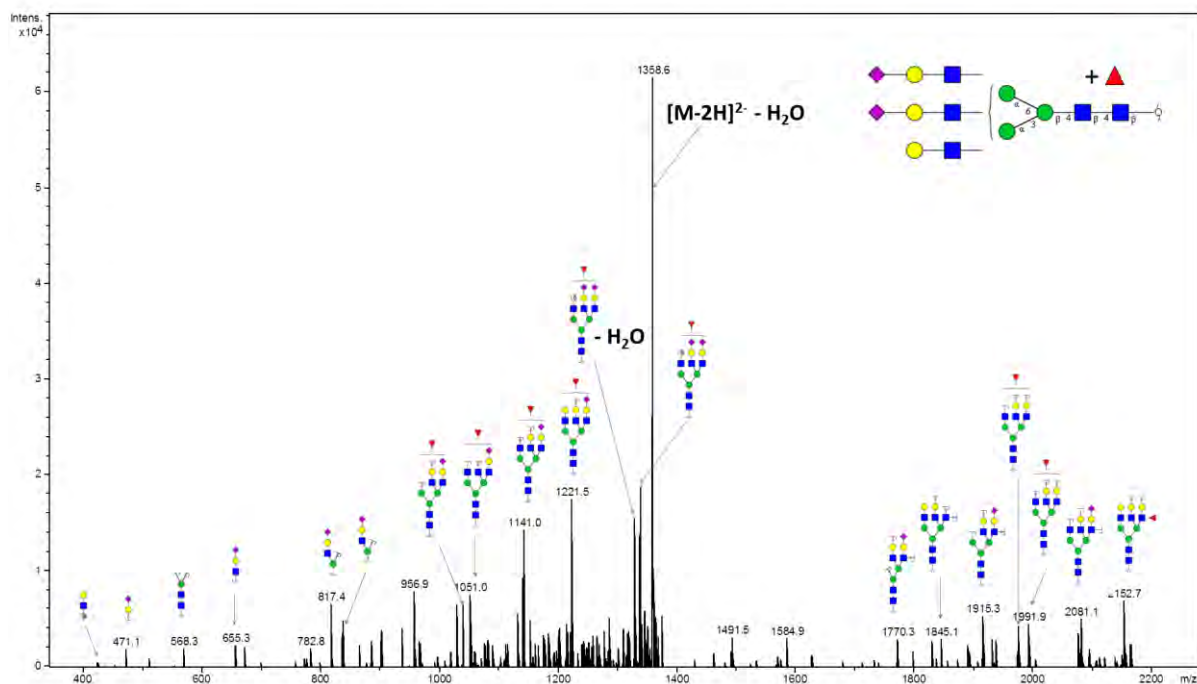
Glycan 175

Parent ion: *m/z* 1404.0²⁻

Composition: (Hex)₄ (HexNAc)₄ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.
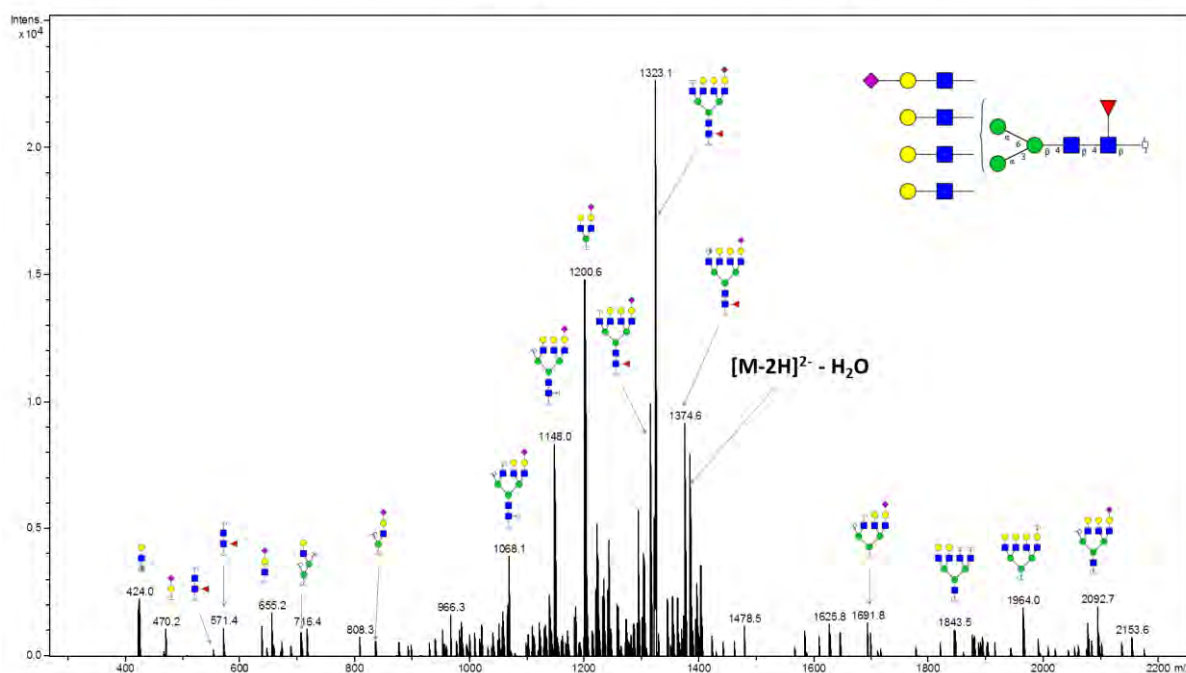
Glycan 176

Parent ion: *m/z* 1404.0²⁻

Composition: (Hex)₄ (HexNAc)₄ (Deoxyhexose)₁ (NeuAc)₁ + (Man)₃(GlcNAc)₂

Notes: Distinction between 3-arm/6-arm not intended in glycan fragment scheme. Additional isomers of fragments cannot be excluded.

Glycan 178

Parent ion: *m/z* 1512.6^{2-}

Composition: (Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (NeuAc)$_3$ + (Man)$_3$(GlcNAc)$_2$

Notes: Structural assignment based on related (precusor) glycan 173.

Glycan 179

Parent ion: *m/z* 1549.7^{2-}

Composition: (Hex)$_4$ (HexNAc)$_4$ (Deoxyhexose)$_1$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$

Notes: Structural assignment based on related (precusor) glycan 173.
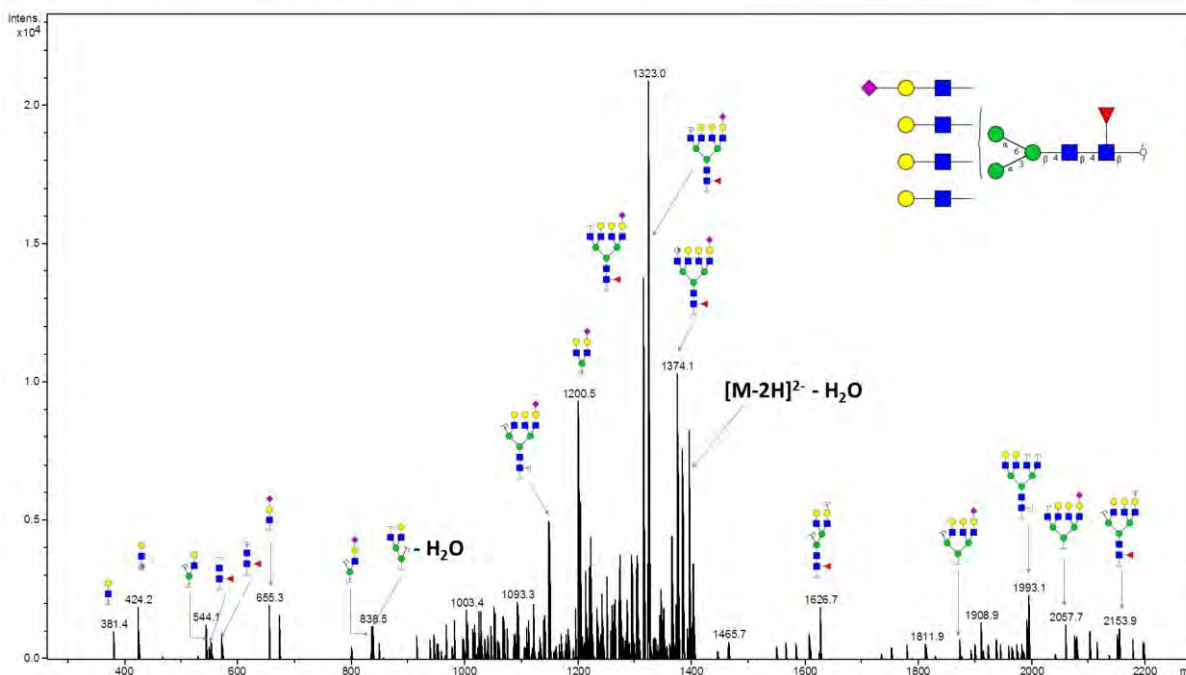
Supplementary Figure 2. Annotated MS/MS for *O*-glycans. *O*-glycans have been numbered according to Supplementary Table 4. Glycan schemes were derived from GlycoWorkbench. Annotation was based on the presence of structural feature ions and common knowledge of known glycan synthetic pathways. Additional structural isomers cannot be excluded. Structural schemes not intended to show specific linkages (linkages not shown by linkage angles). Where ambiguity is indicated in parent structure, for example in the assignment of fucose, depiction of specific linkages are not intended for corresponding glycan fragments.

Glycan 1
Parent ion: *m/z* 513.3⁻
Composition: (HexNAc)₁ (NeuAc)₁
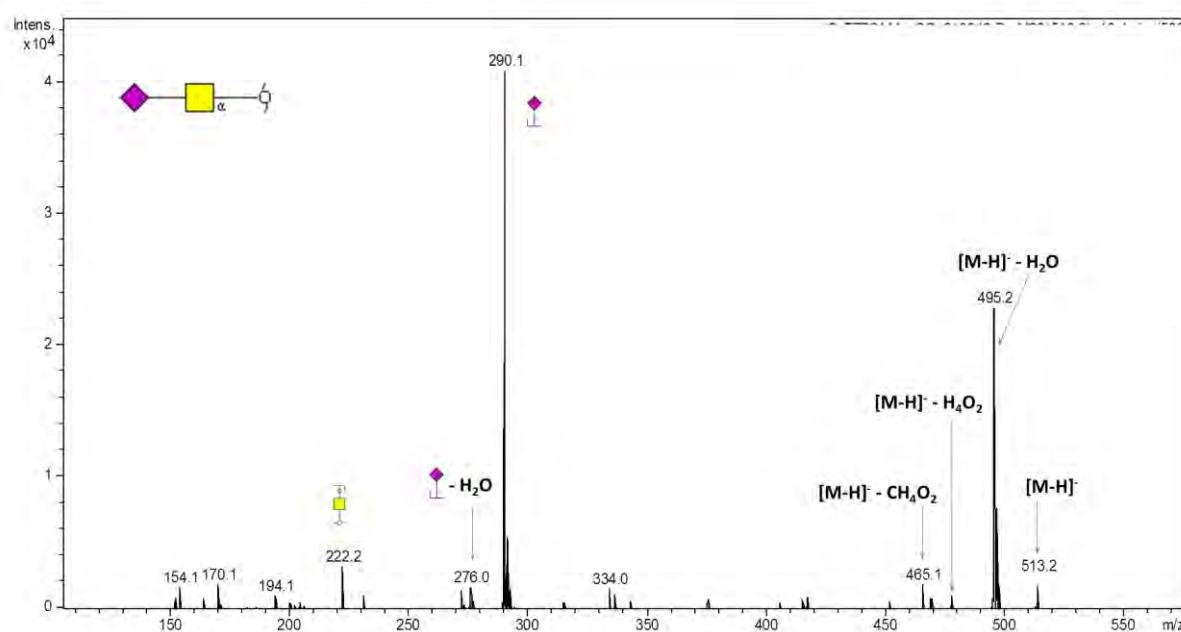
Glycan 2

Parent ion: *m/z* 587.3⁻
Composition: (Hex)₁ (HexNAc)₂
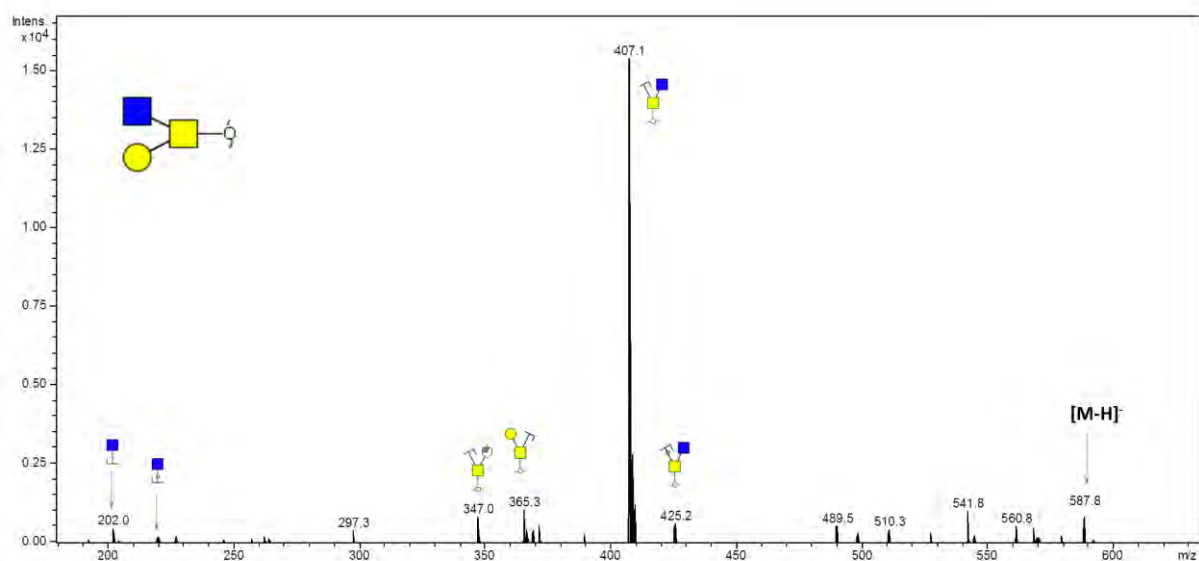
Glycan 3

Parent ion: *m/z* 587.3⁻
Composition: (Hex)₁ (HexNAc)₂
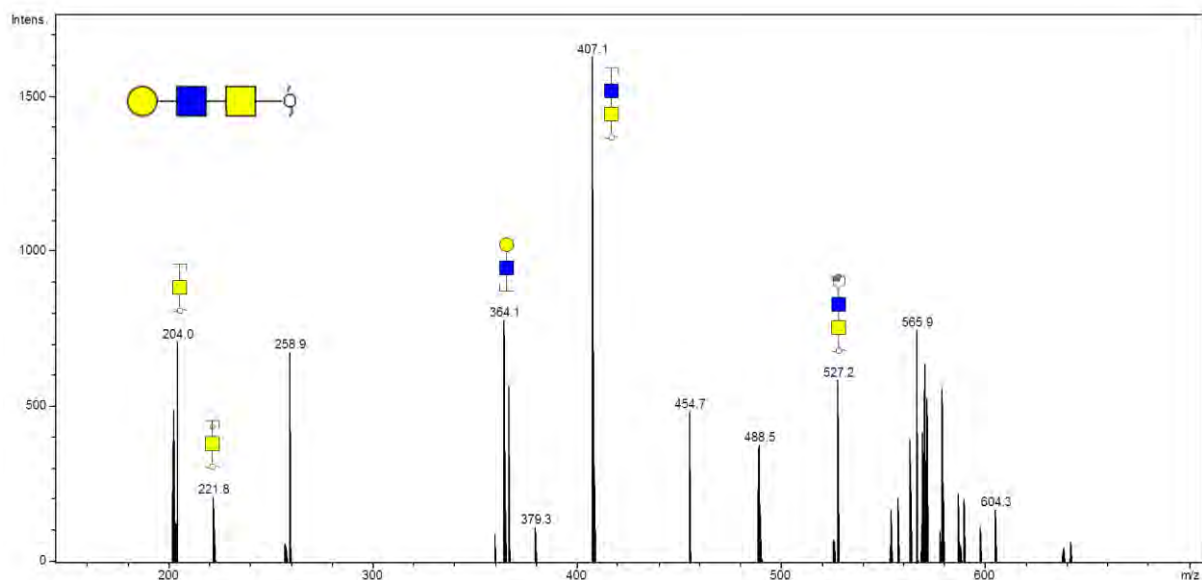
Glycan 4

Parent ion: *m/z* 675.3⁻
Composition: (Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_1$
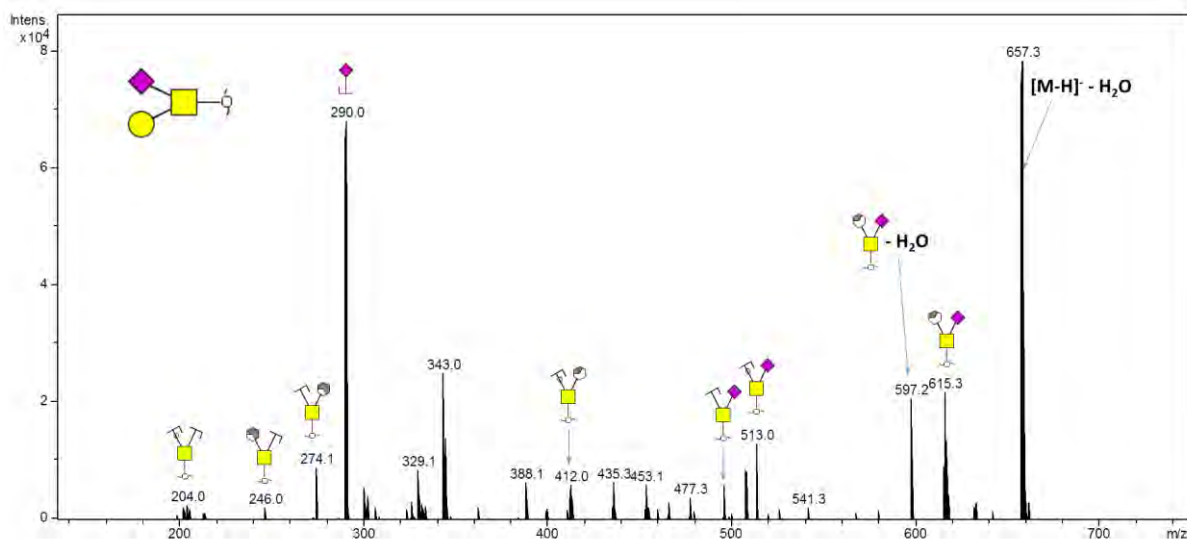
Glycan 5

Parent ion: *m/z* 675.3⁻
Composition: (Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_1$
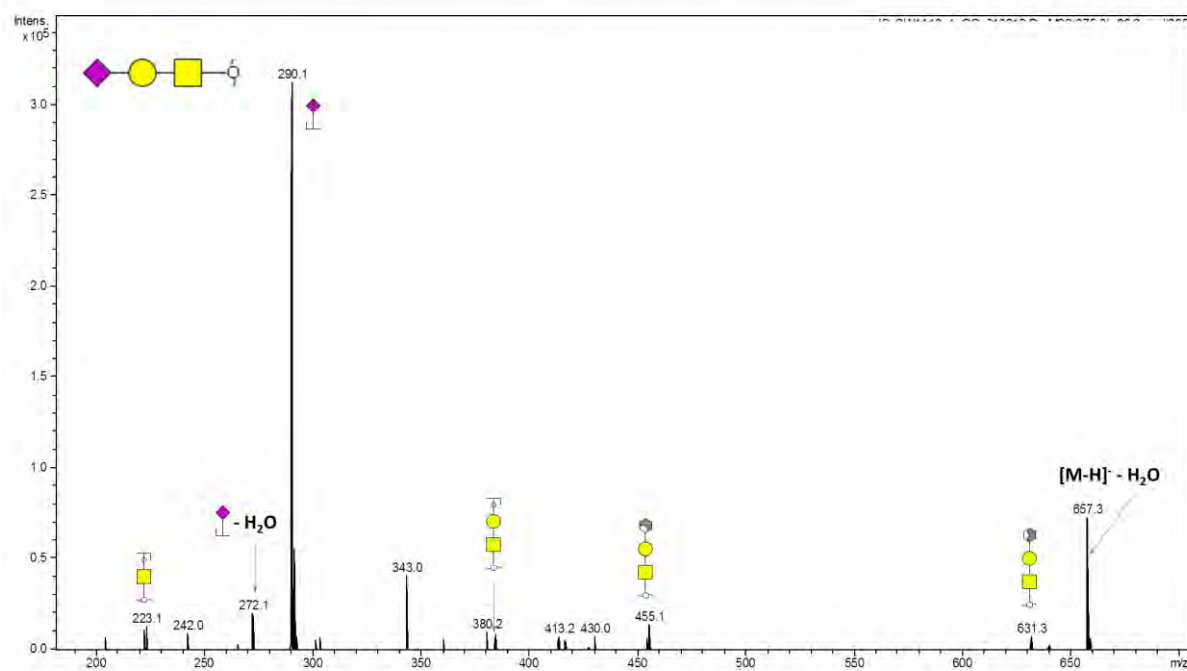
Glycan 6

Parent ion: *m/z* 716.2⁻
Composition: (HexNAc)₂ (NeuAc)₁
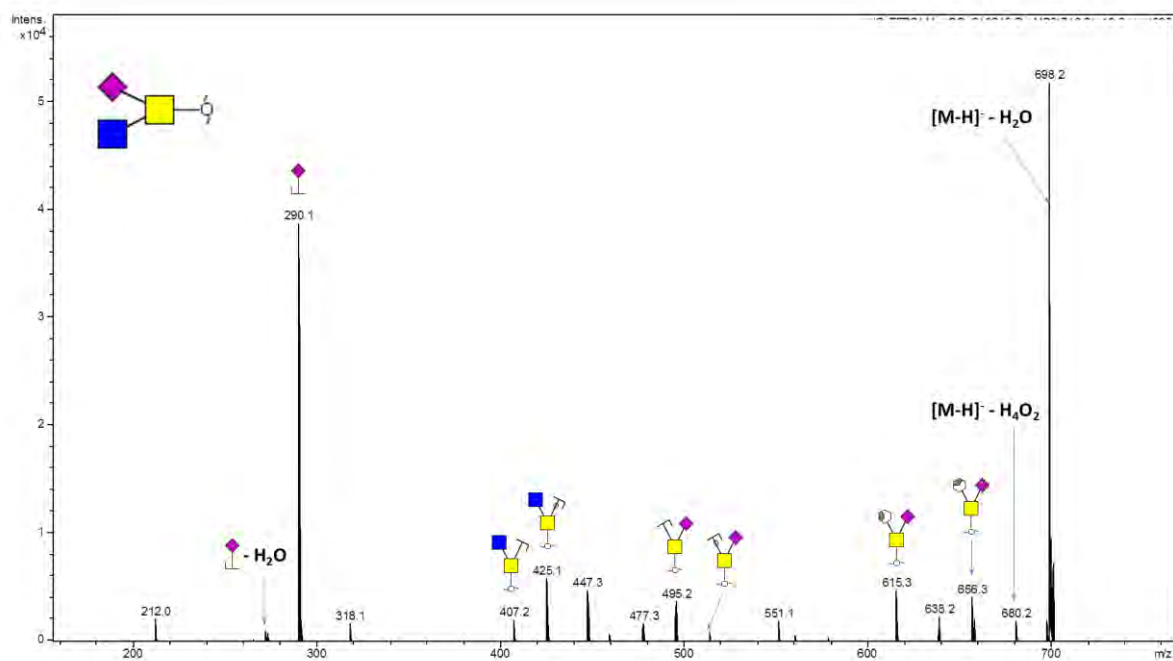
Glycan 7

Parent ion: *m/z* 716.2⁻
Composition: (HexNAc)₂ (NeuAc)₁
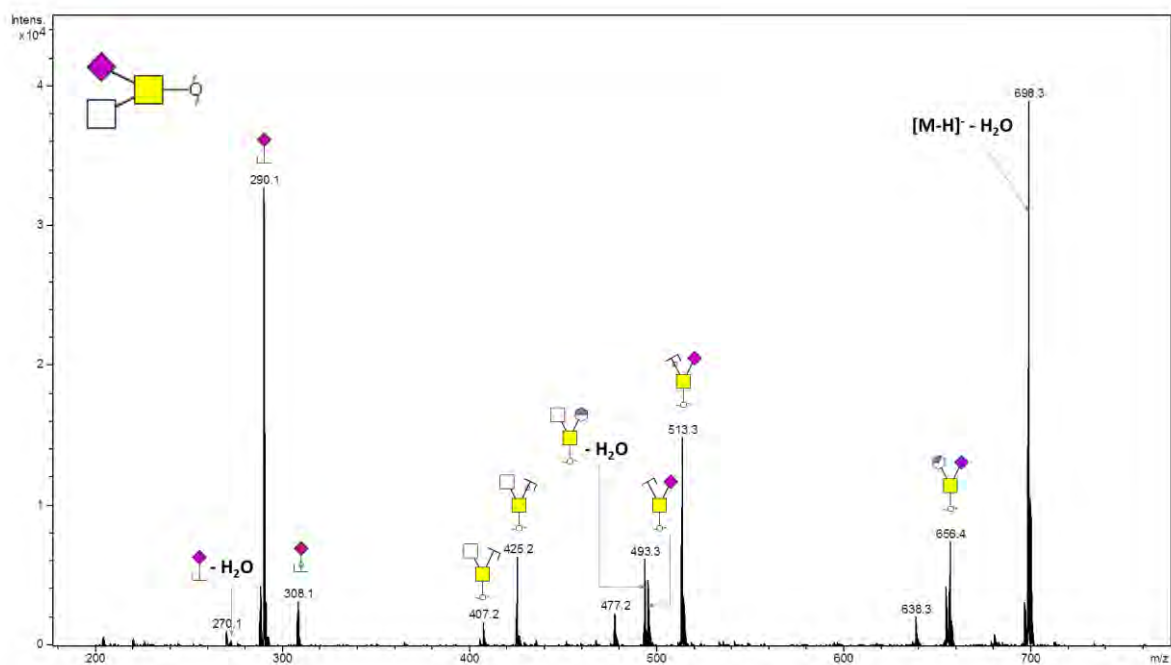
Glycan 8

Parent ion: *m/z* 716.2⁻
Composition: (HexNAc)₂ (NeuAc)₁
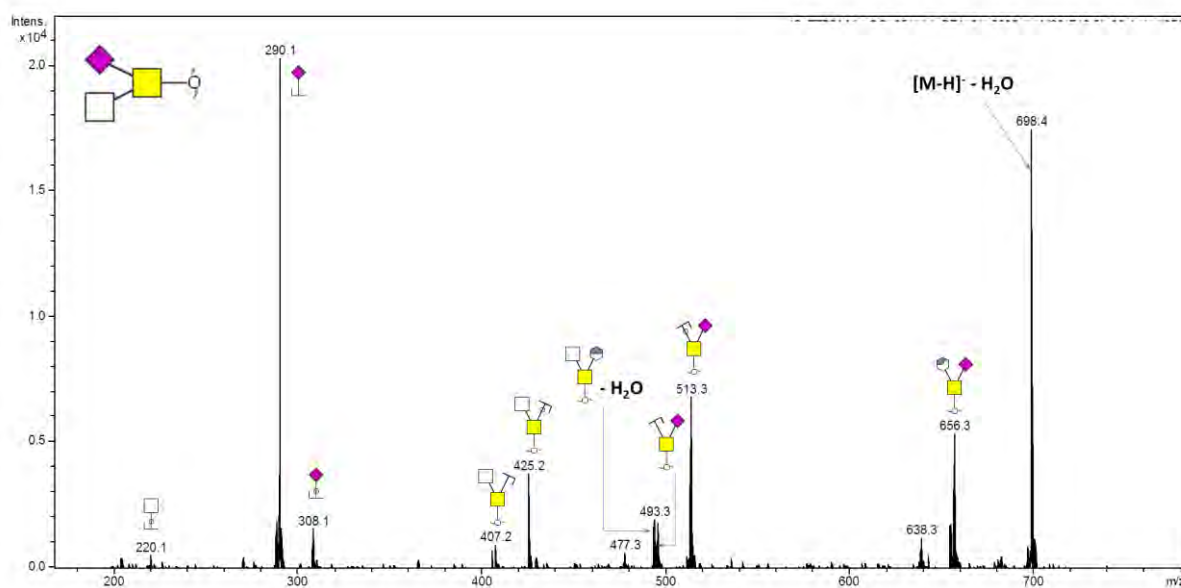
Glycan 9

Parent ion: *m/z* 749.2⁻
Composition: (Hex)₂ (HexNAc)₂
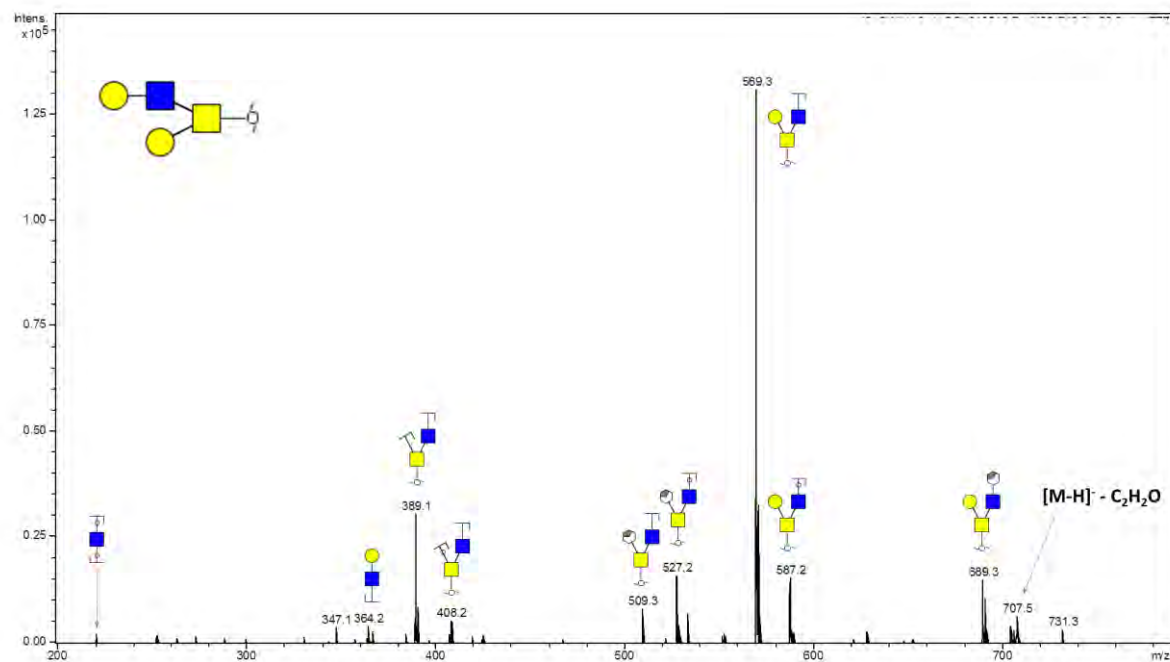
Glycan 10

Parent ion: *m/z* 829.3⁻
Composition: (Hex)₂ (HexNAc)₂ (Sulph)₁
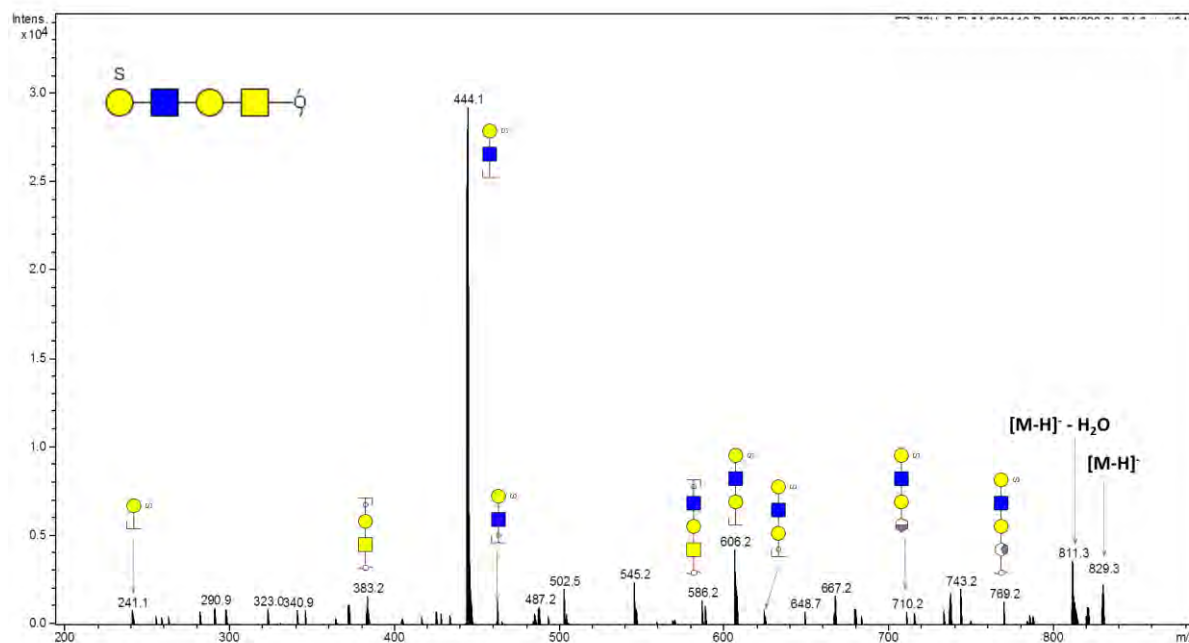
Glycan 11

Parent ion: *m/z* 878.4⁻
Composition: (Hex)₁ (HexNAc)₂ (NeuAc)₁
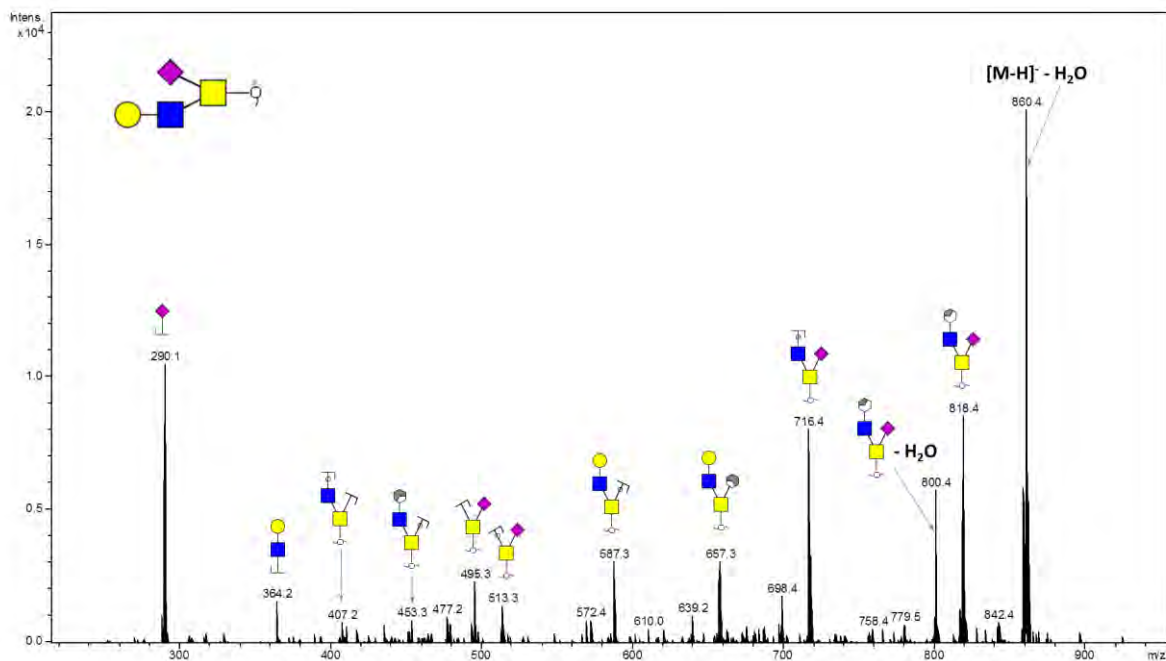
Glycan 12

Parent ion: *m/z* 878.4⁻
Composition: (Hex)₁ (HexNAc)₂ (NeuAc)₁
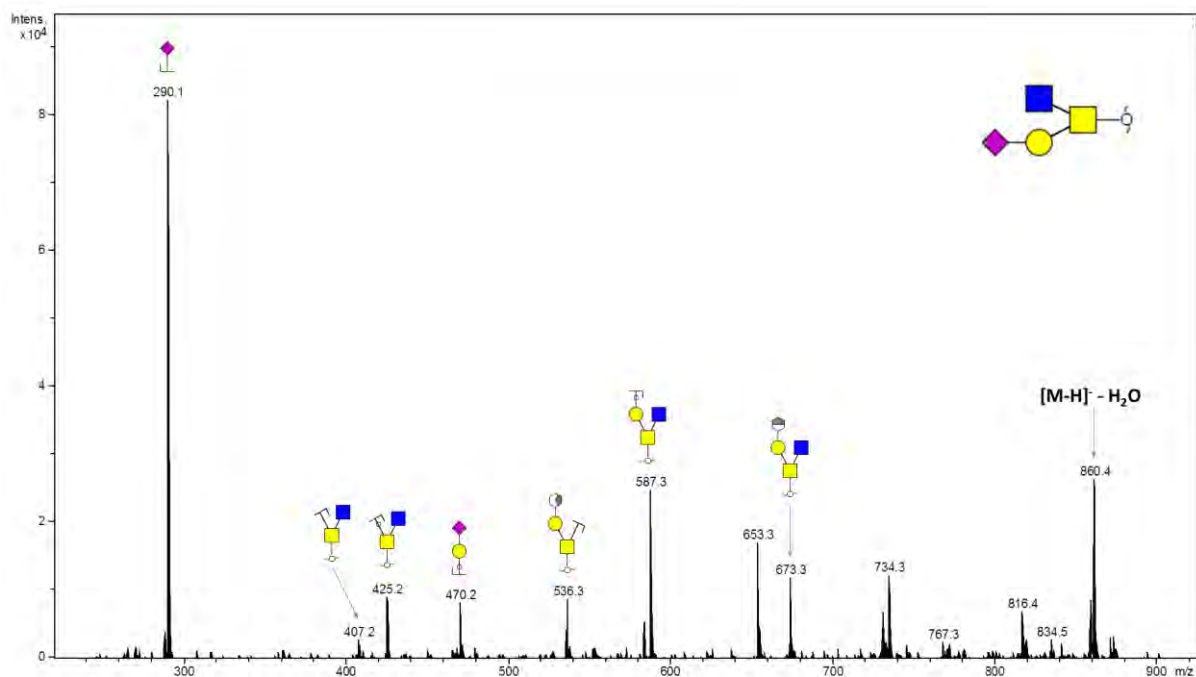
Glycan 13

Parent ion: *m/z* 895.3⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁
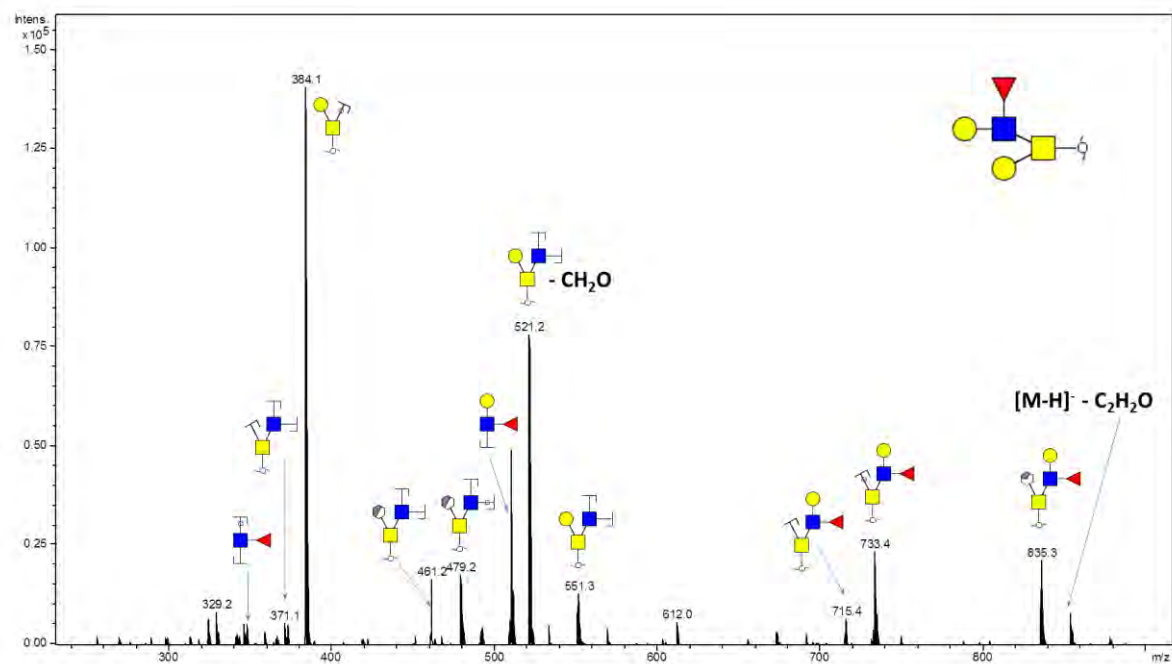
Glycan 14

Parent ion: *m/z* 913.0^{2-}
Composition: (Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_2$ (NeuAc)$_2$
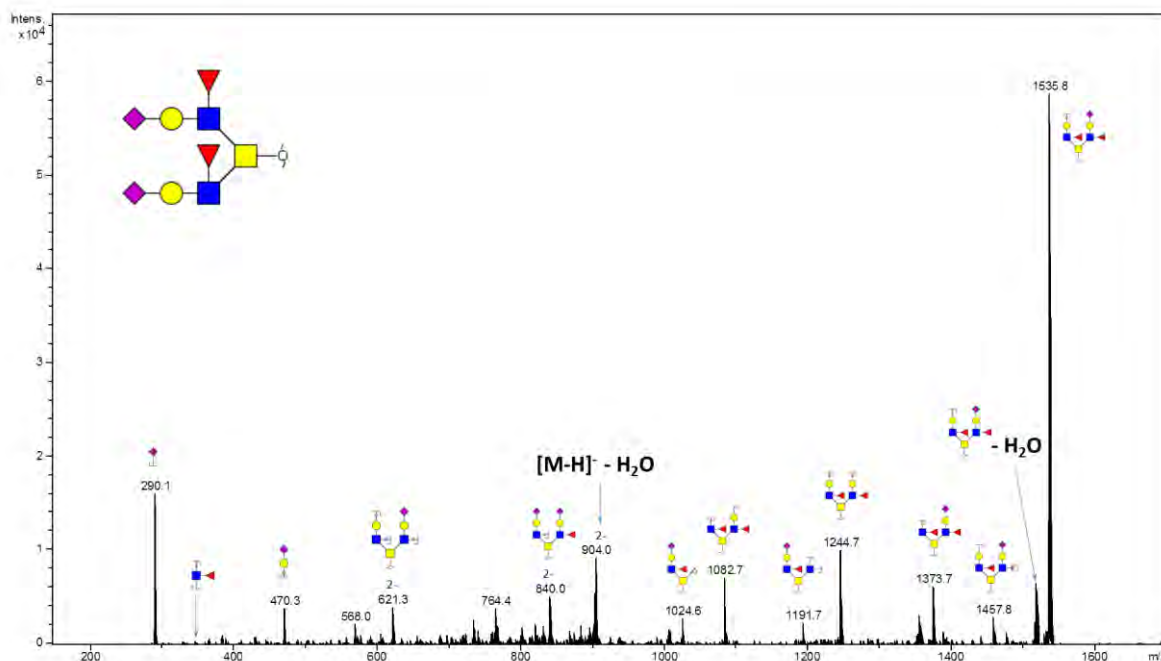
Glycan 15

Parent ion: *m/z* 966.3$^-$
Composition: (Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_2$

Notes: Sialylated form of glycan 4 and 5.
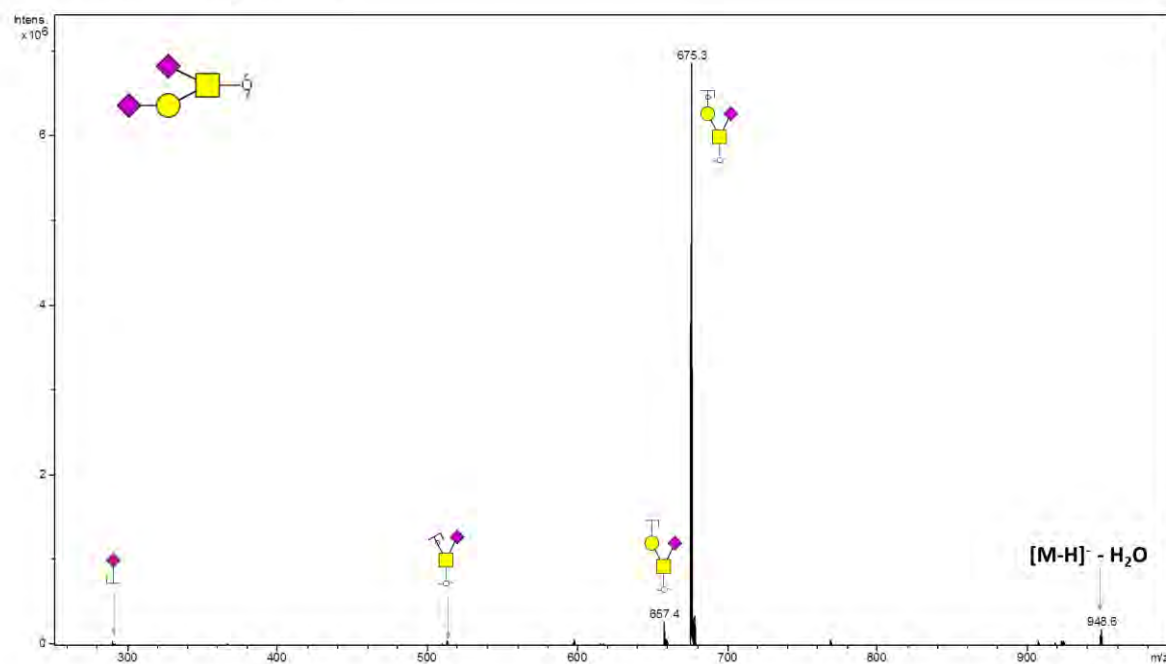
Glycan 16

Parent ion: *m/z* 966.3⁻
Composition: (Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_2$

Notes: Sialylated form of glycan 4 and 5.
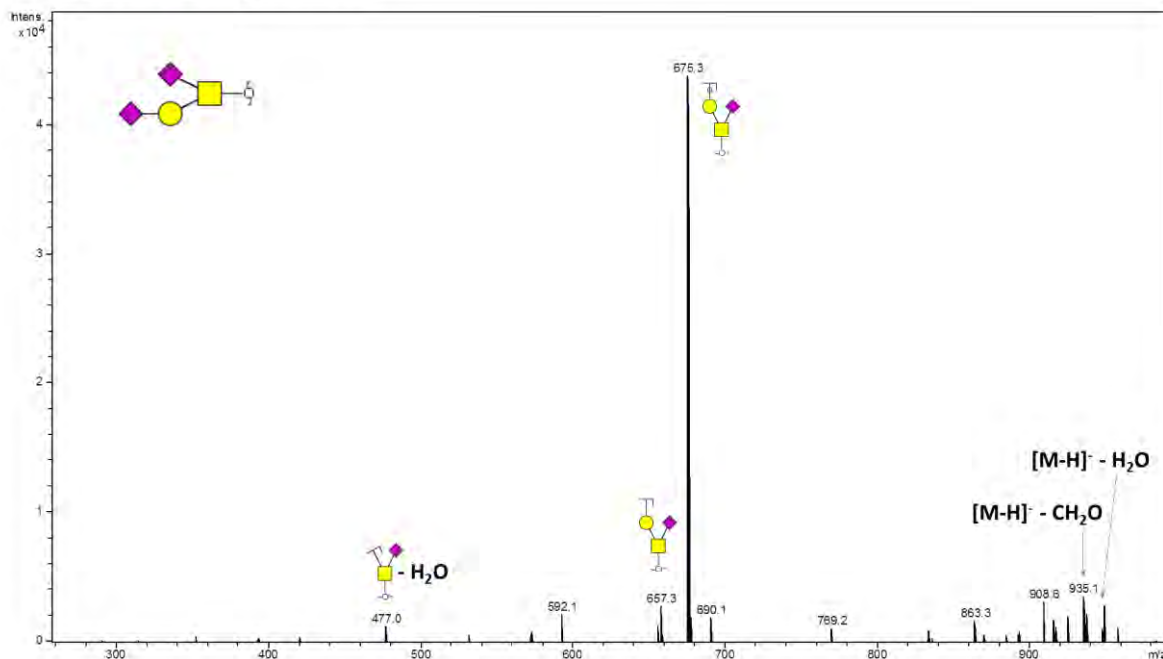
Glycan 17

Parent ion: *m/z* 975.4⁻
Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (Sulph)$_1$
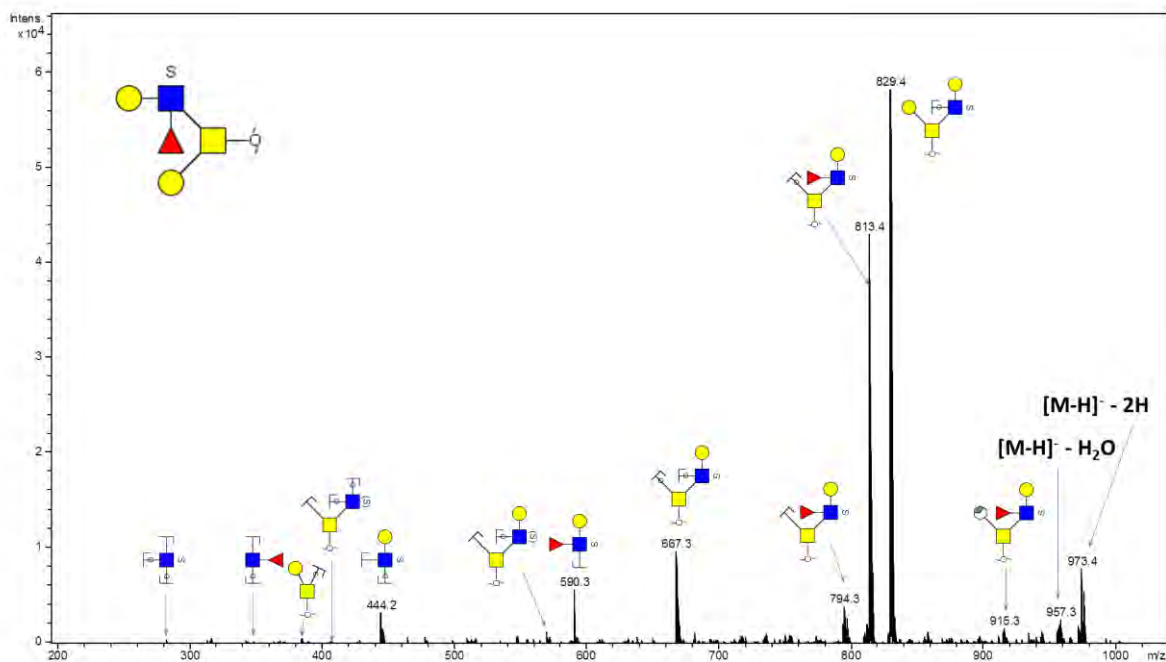
Glycan 18

Parent ion: *m/z* 975.4⁻
Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (Sulph)$_1$

Notes: Specific position of sulphate not intended in glycan fragment scheme
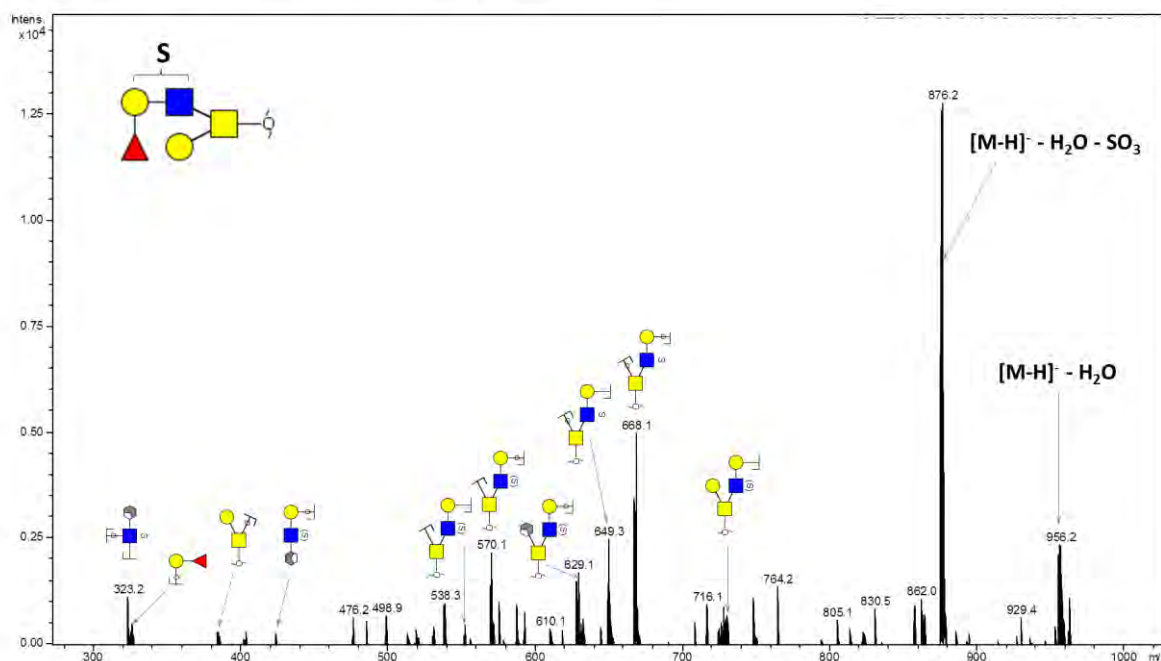
Glycan 19

Parent ion: *m/z* 975.4²⁻
Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (Sulph)$_1$
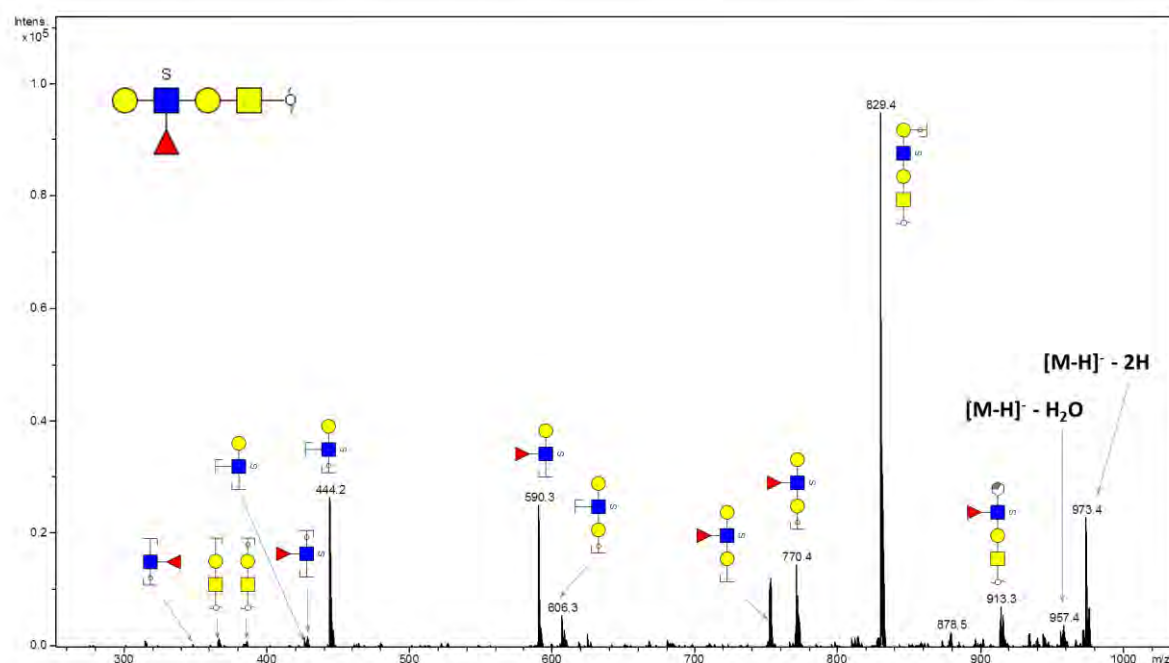
Glycan 20

Parent ion: *m/z* 1016.4²⁻
Composition: (Hex)₁ (HexNAc)₃ (Deoxyhexose)₁ (Sulph)₁
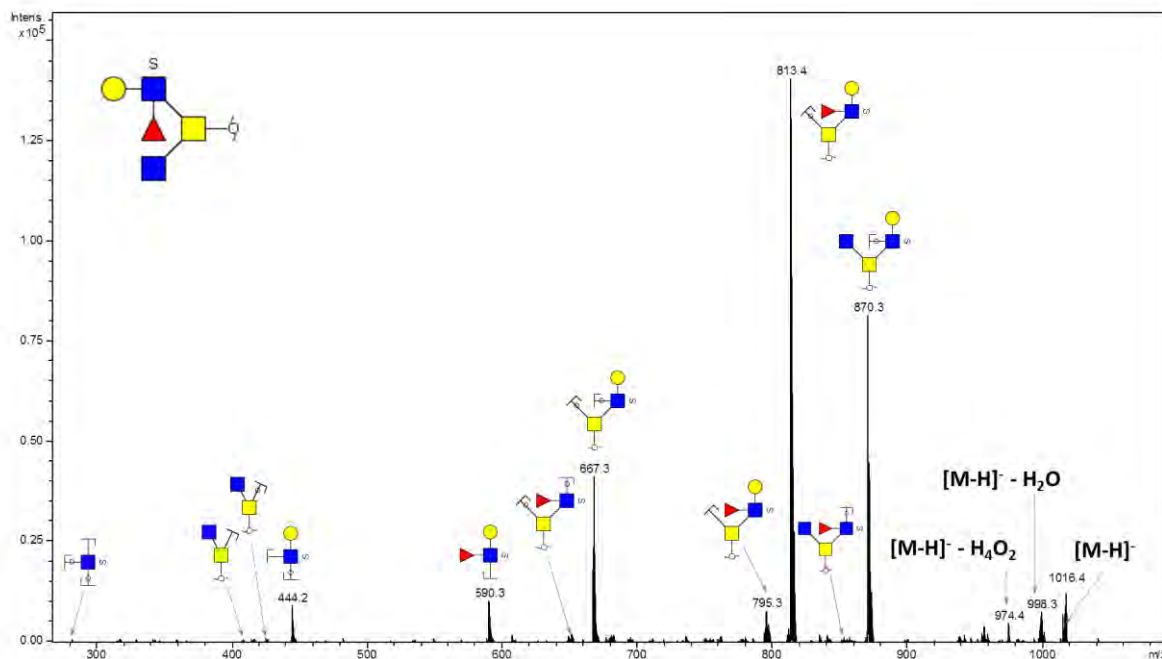
Glycan 21

Parent ion: *m/z* 1024.5²⁻
Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁
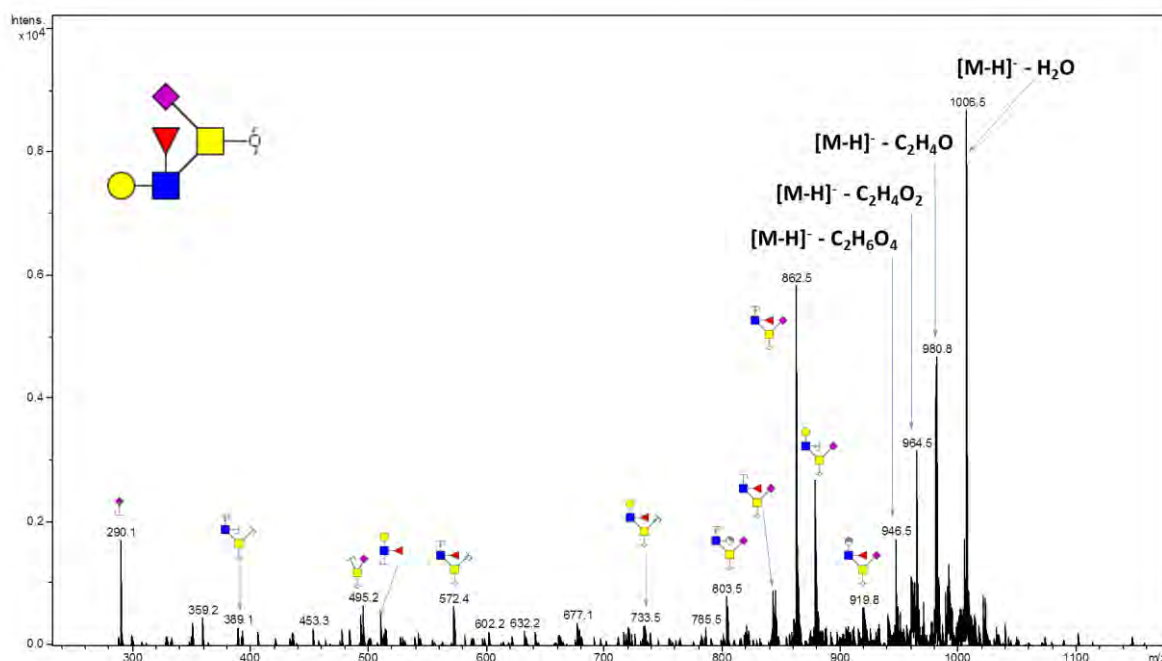
Glycan 22

Parent ion: *m/z* 1024.5⁻

Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁
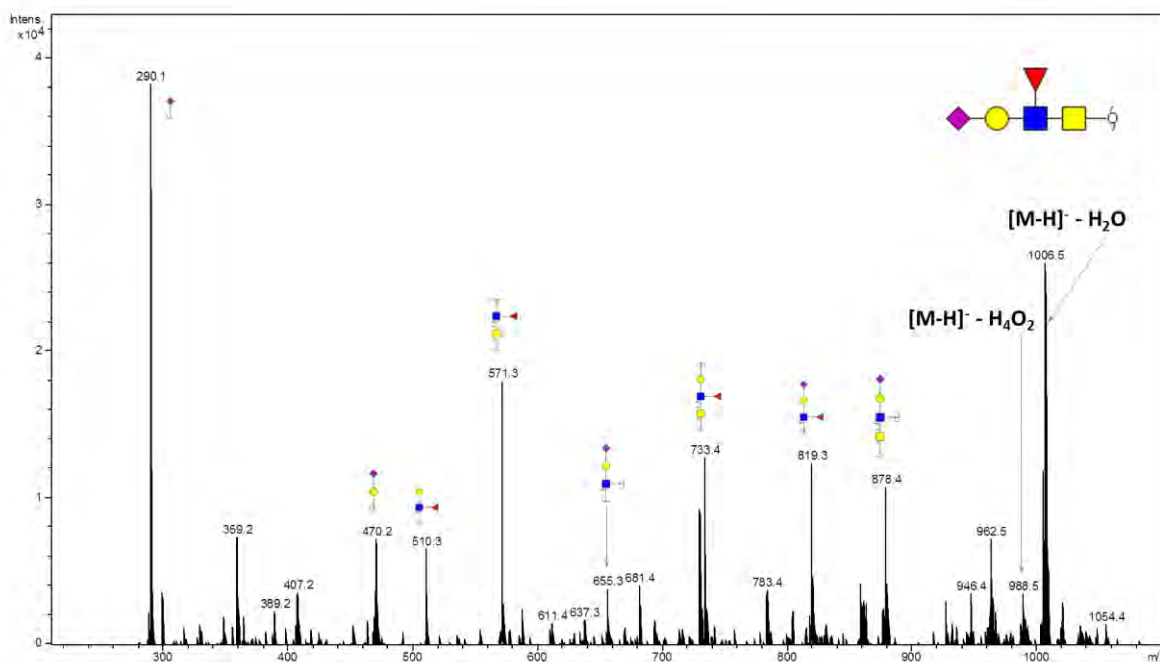
Glycan 23

Parent ion: *m/z* 1040.5⁻

Composition: (Hex)₂ (HexNAc)₂ (NeuAc)₁

Notes: Specific position of sialic acid not intended in glycan fragment scheme (i.e. sialic acid can be on either Gal).
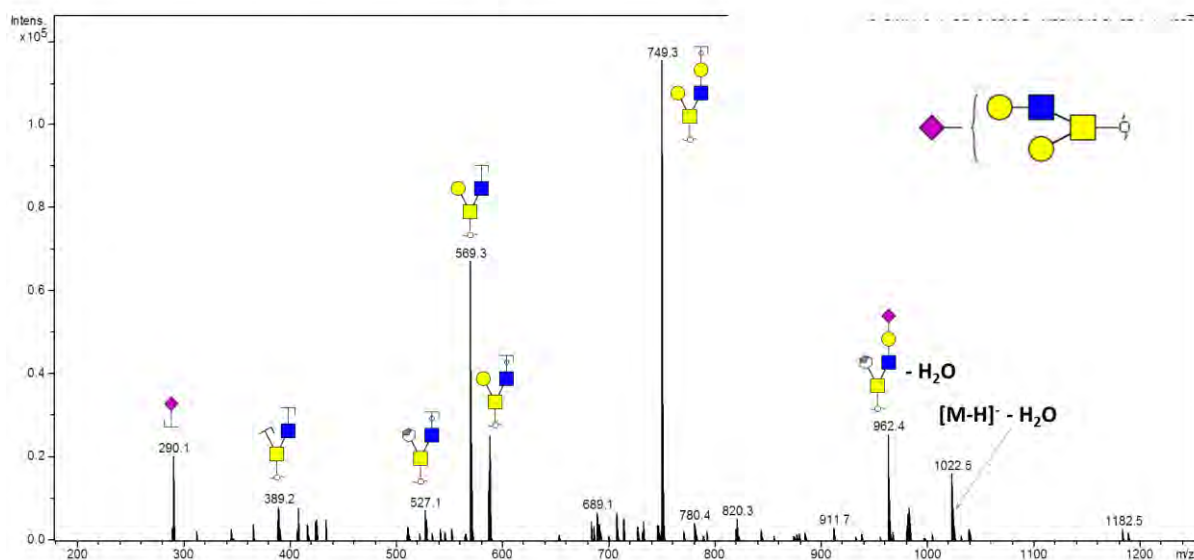
Glycan 24

Parent ion: *m/z* 1040.5⁻
Composition: (Hex)₂ (HexNAc)₂ (NeuAc)₁

Notes: Specific position of sialic acid not intended in glycan fragment scheme (i.e. sialic acid can be on either Gal).
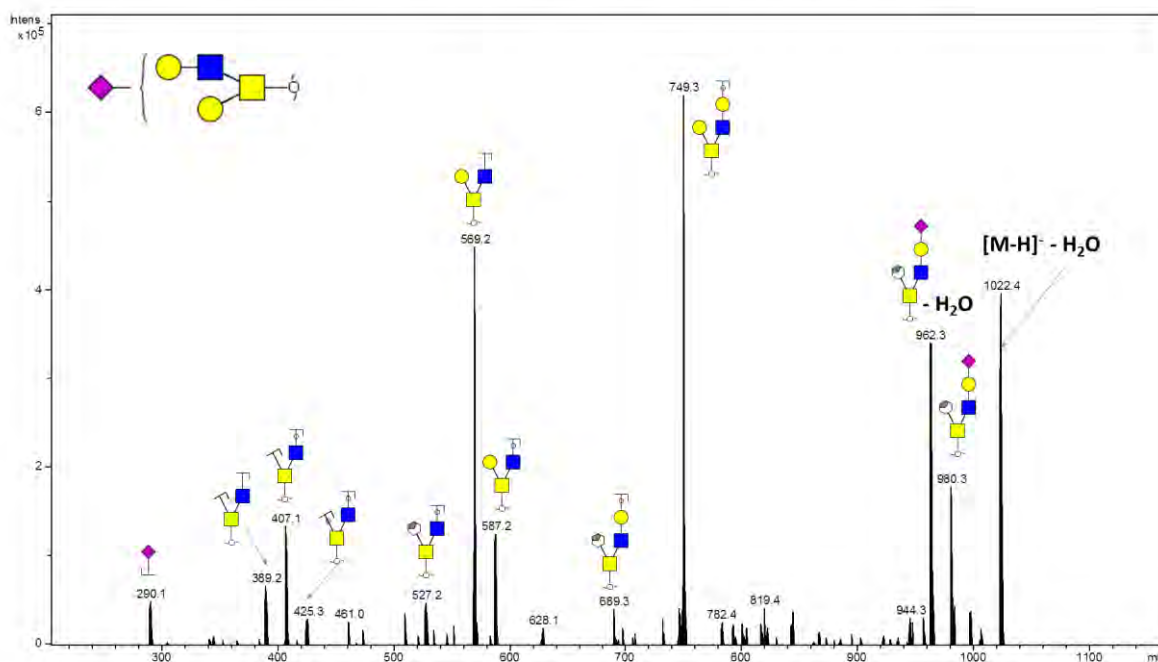
Glycan 25

Parent ion: *m/z* 1040.5⁻
Composition: (Hex)₂ (HexNAc)₂ (NeuAc)₁

Notes: Specific position of sialic acid not intended in glycan fragment scheme (i.e. sialic acid can be on either Gal).
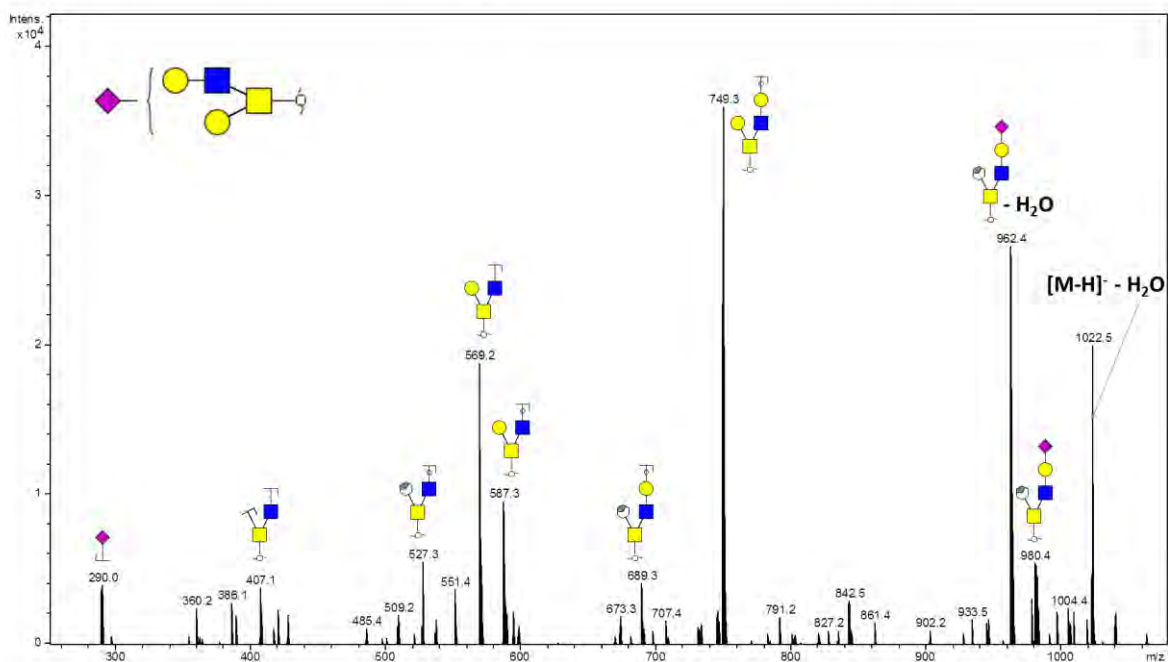
Glycan 26

Parent ion: *m/z* 1081.5⁻
Composition: (Hex)₁ (HexNAc)₃ (NeuAc)₁

Notes: Specific position of Gal not intended in glycan fragment scheme (i.e. Gal can be on either GlcNAc).
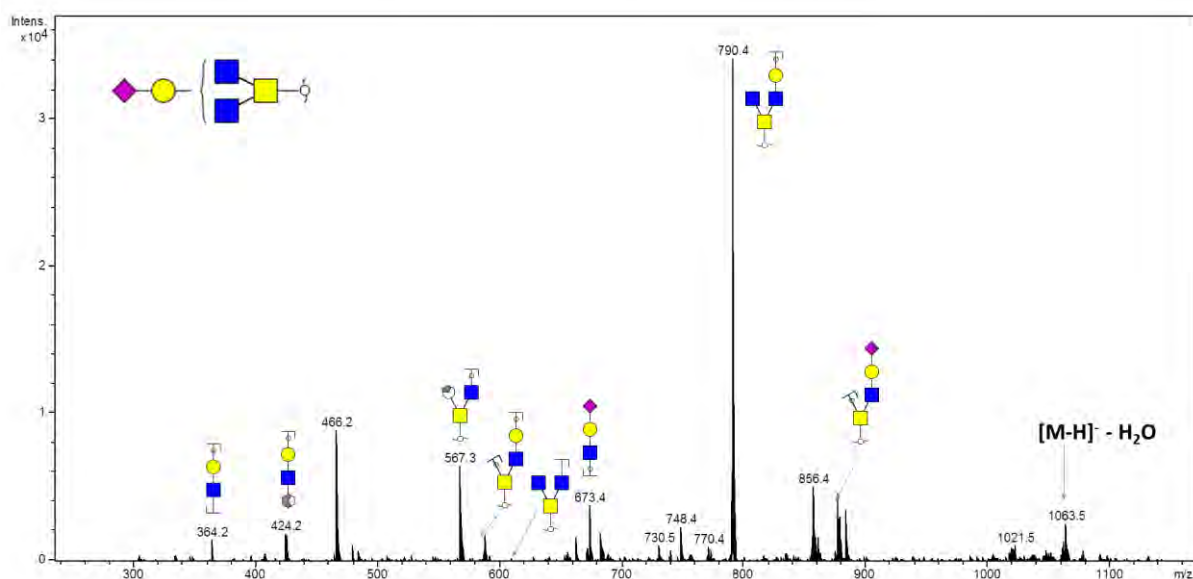
Glycan 27

Parent ion: *m/z* 1081.5⁻
Composition: (Hex)₁ (HexNAc)₃ (NeuAc)₁

Notes: Specific position of Gal not intended in glycan fragment scheme (i.e. Gal can be on either GlcNAc).
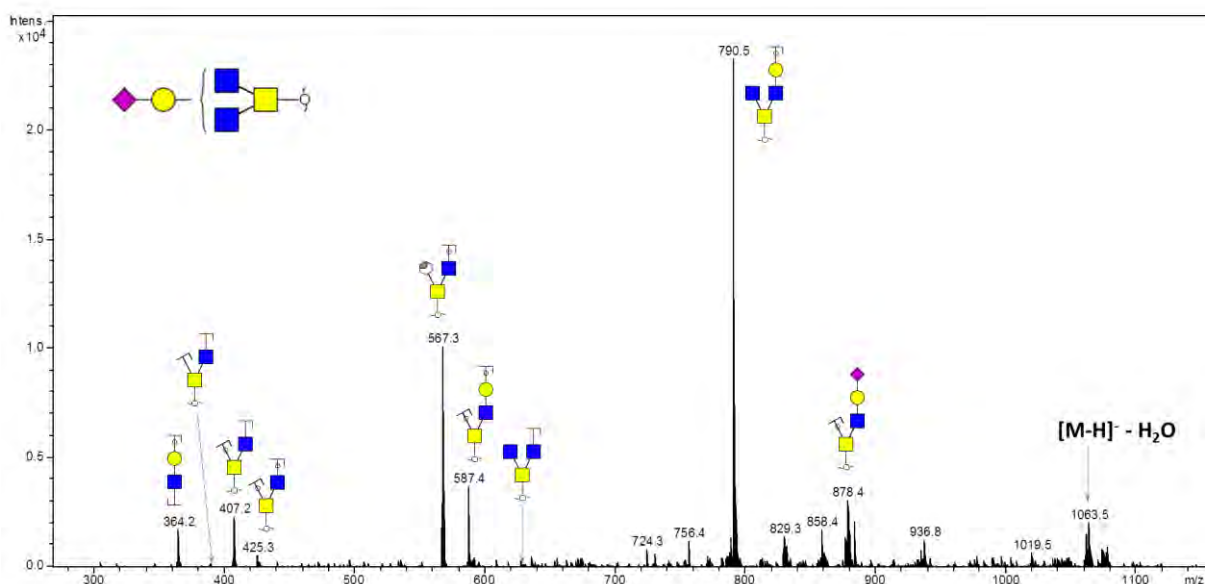
Glycan 28

Parent ion: *m/z* 1081.5⁻
Composition: (Hex)₁ (HexNAc)₃ (NeuAc)₁

Notes: Specific position of Gal not intended in glycan fragment scheme (i.e. Gal can be on either GlcNAc).
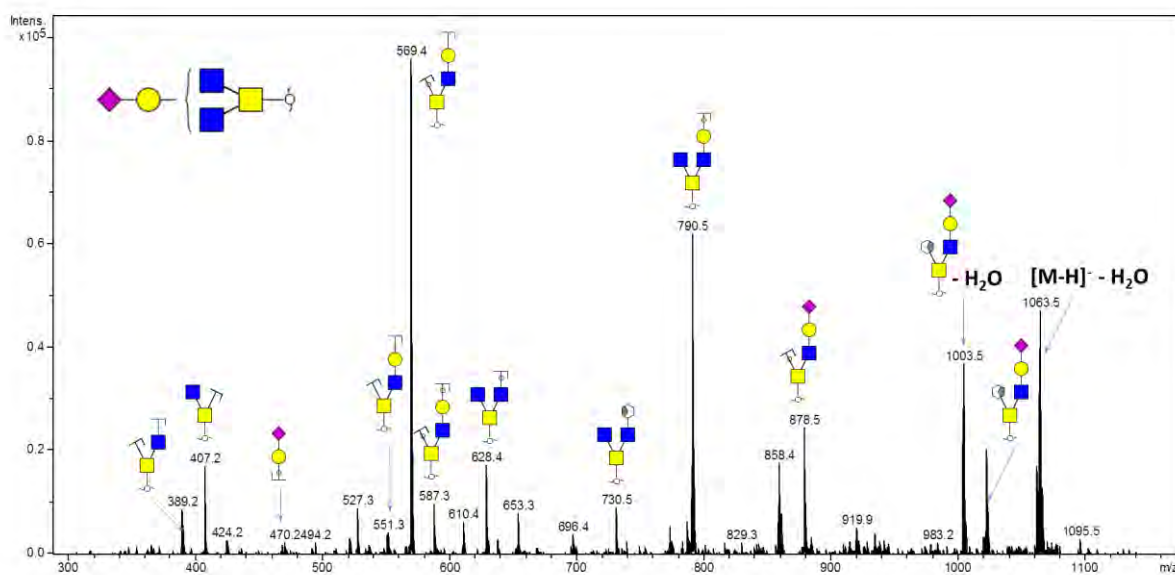
Glycan 29

Parent ion: *m/z* 551.8²⁻
Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ (Sulph)₁
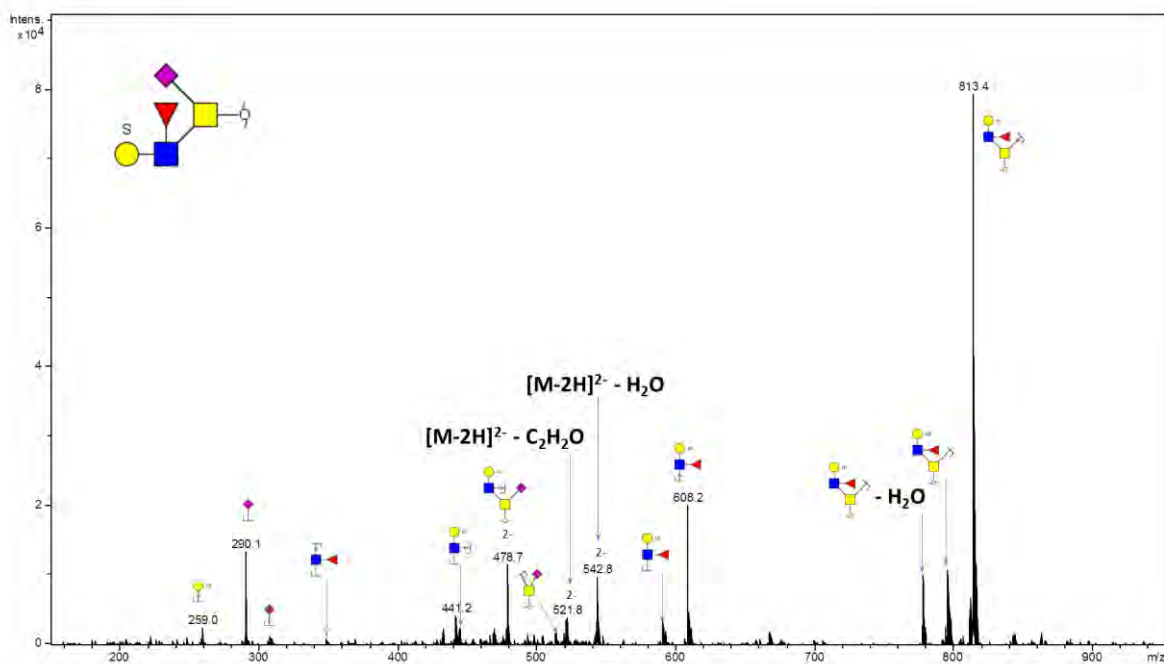
Glycan 30

Parent ion: *m/z* 1114.5⁻
Composition: (Hex)3 (HexNAc)₃
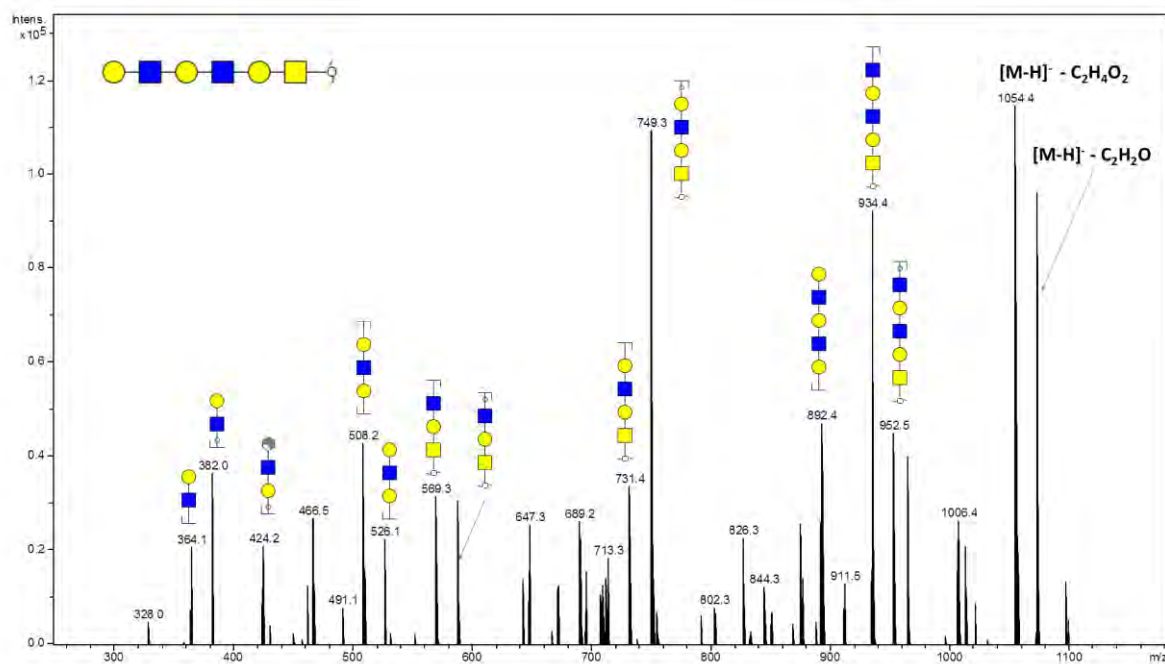
Glycan 31

Parent ion: *m/z* 1120.5⁻
Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ (Sulph)₁

Notes: Related to glycan 10.
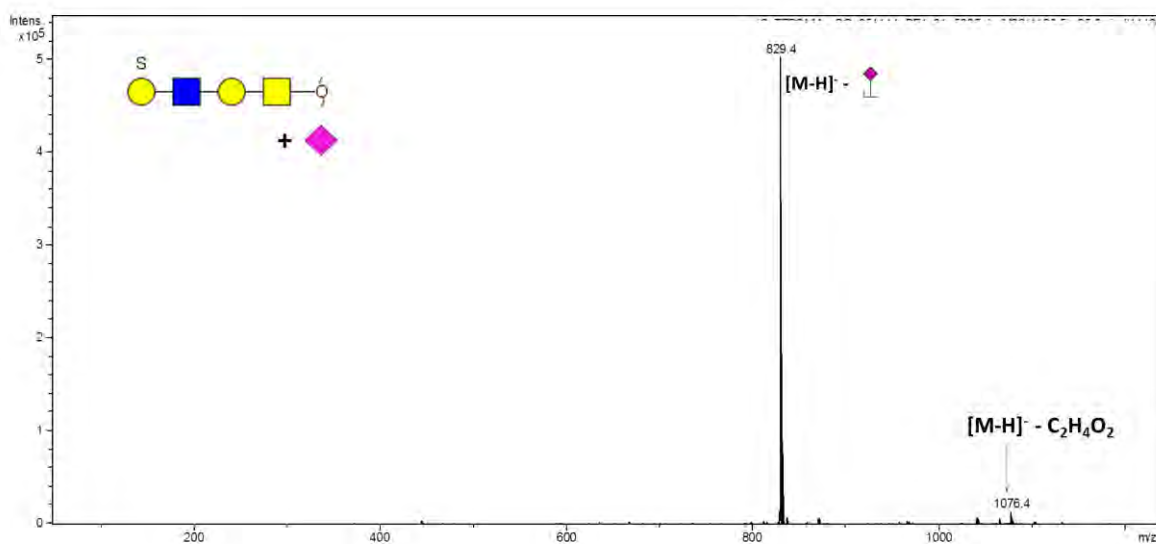
Glycan 32

Parent ion: *m/z* 584.3²⁻
Composition (Hex)₁ (HexNAc)₂ (NeuAc)₂
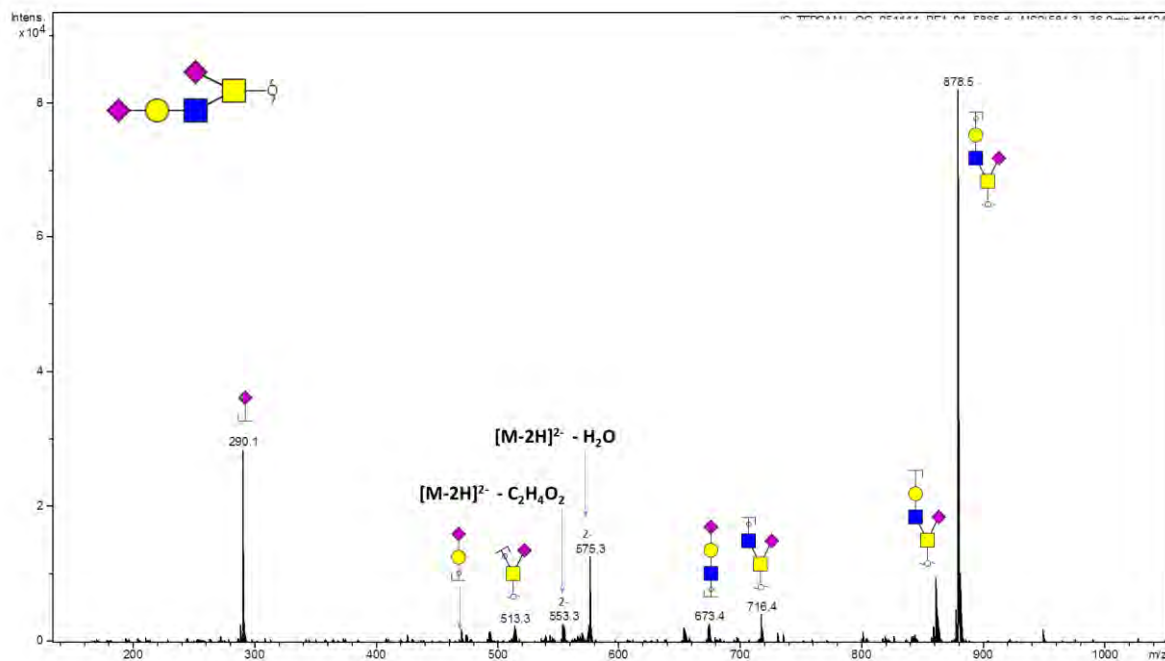
Glycan 33

Parent ion: *m/z* 1186.4⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁
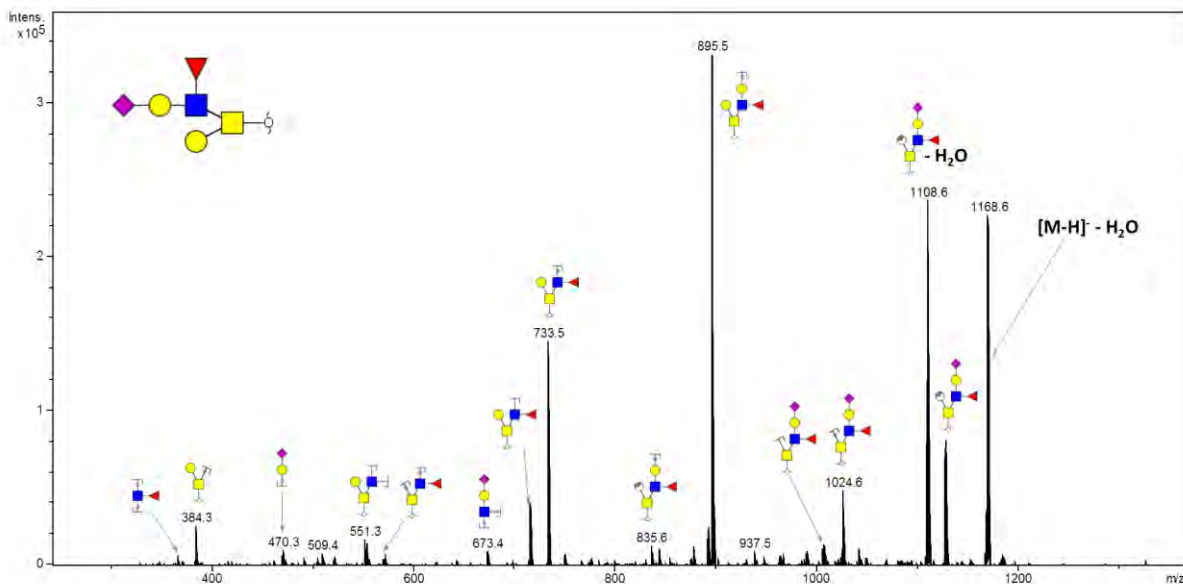
Glycan 34

Parent ion: *m/z* 1186.4⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁

Notes: Specific position of sialic acid not intended in glycan fragment scheme (i.e. sialic acid can be on either Gal). Asterisked peak likely to be from interfering compound.
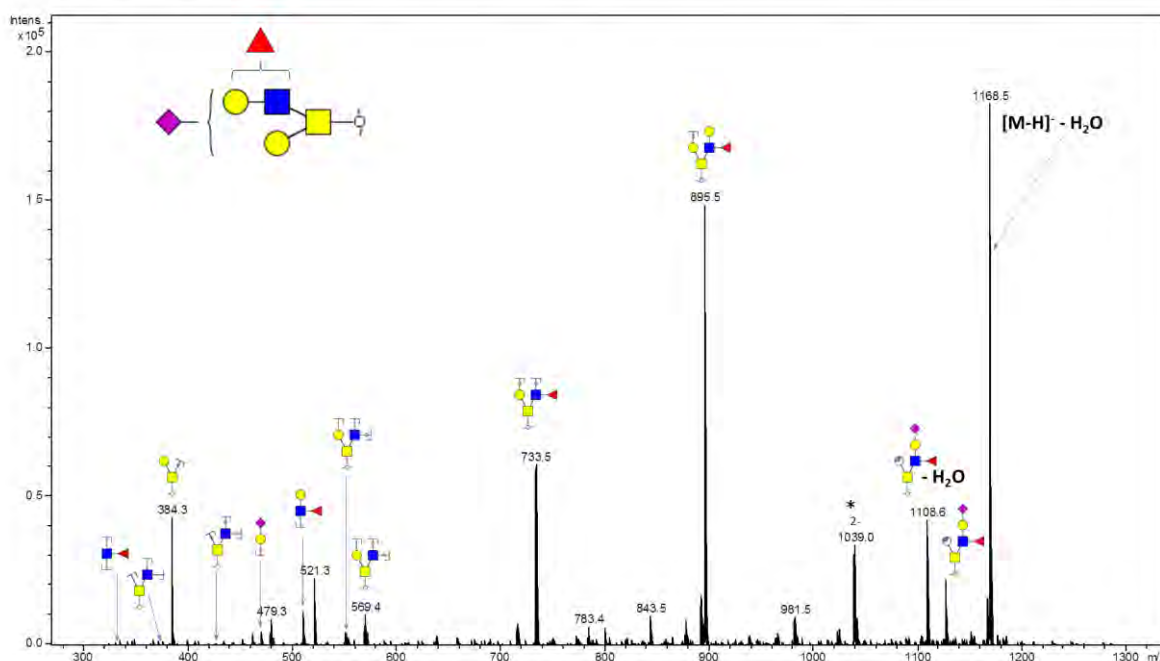
Glycan 35

Parent ion: *m/z* 1227.7⁻
Composition: (Hex)₁ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₁

Notes: Specific position of Gal and Fuc not intended in glycan fragment scheme (i.e. Gal can be on either GlcNAc, Fuc can be on either GlcNac).
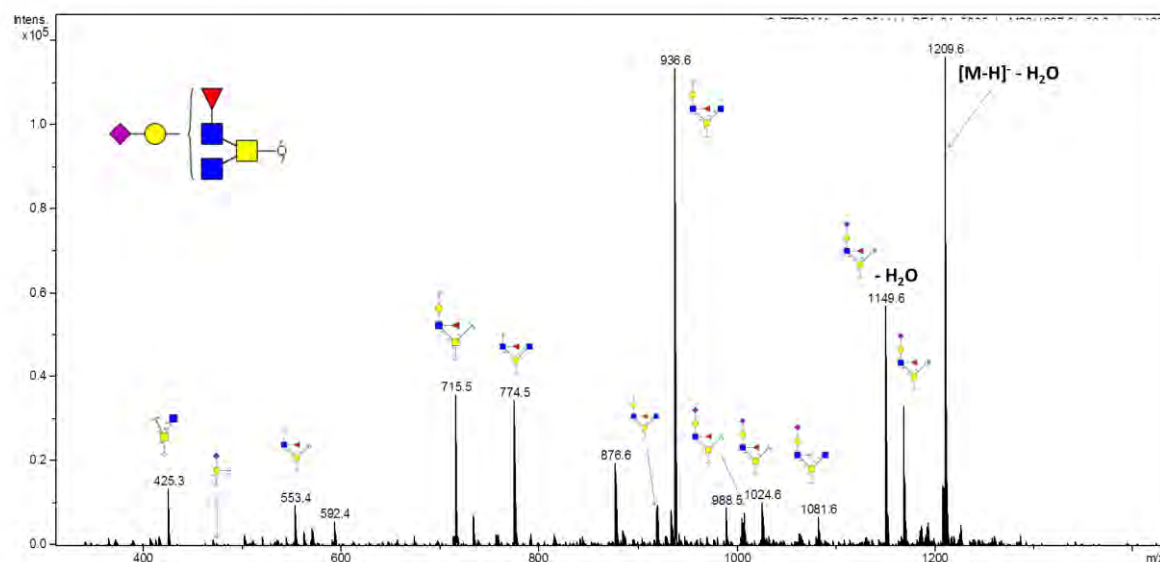
Glycan 36

Parent ion: *m/z* 1260.5⁻
Composition: (Hex)₃ (HexNAc)₃ (Deoxyhexose)₁
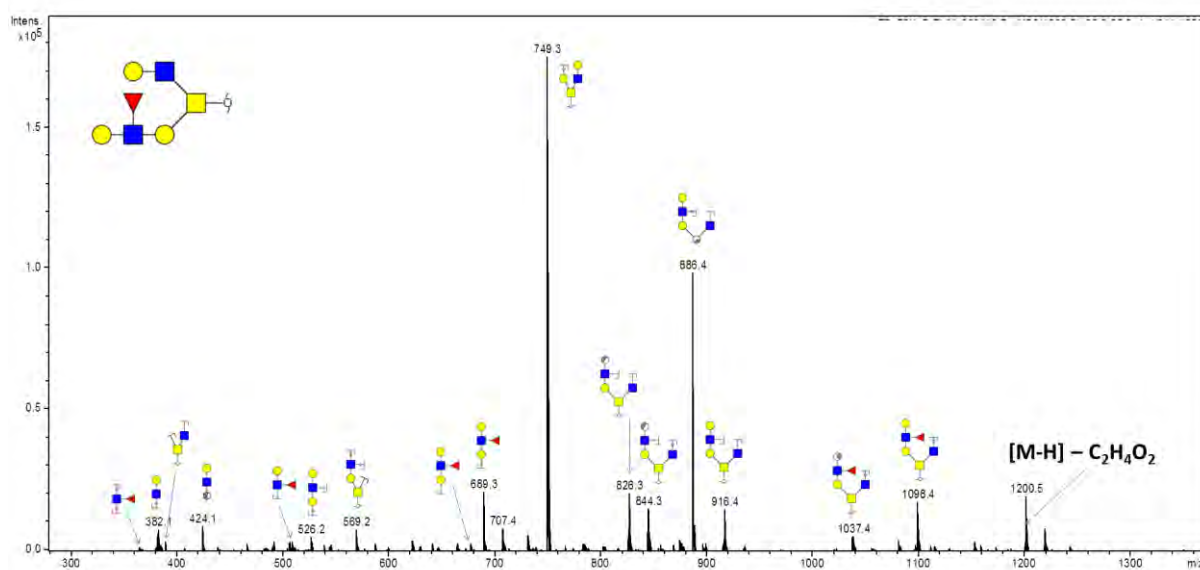
Glycan 37

Parent ion: *m/z* 1266.4⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ (Sulph)₁

Notes: Related to glycan 17.
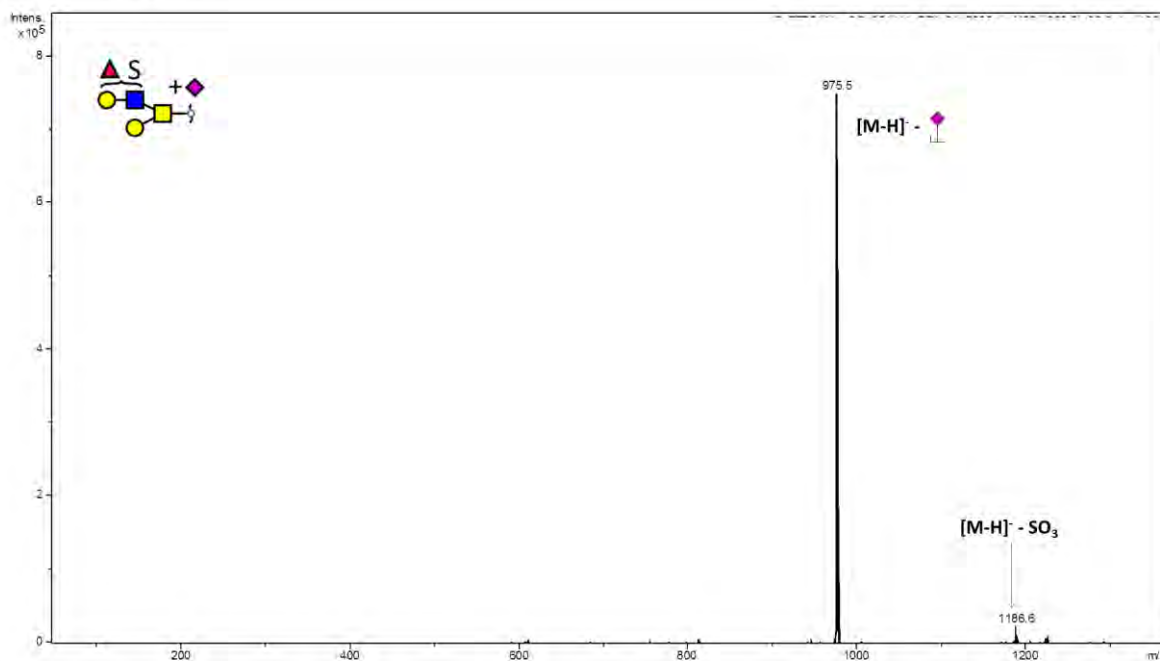
Glycan 38

Parent ion: *m/z* 1266.4⁻

Composition: (Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$ (Sulph)$_1$

Notes: Related to glycan 17.
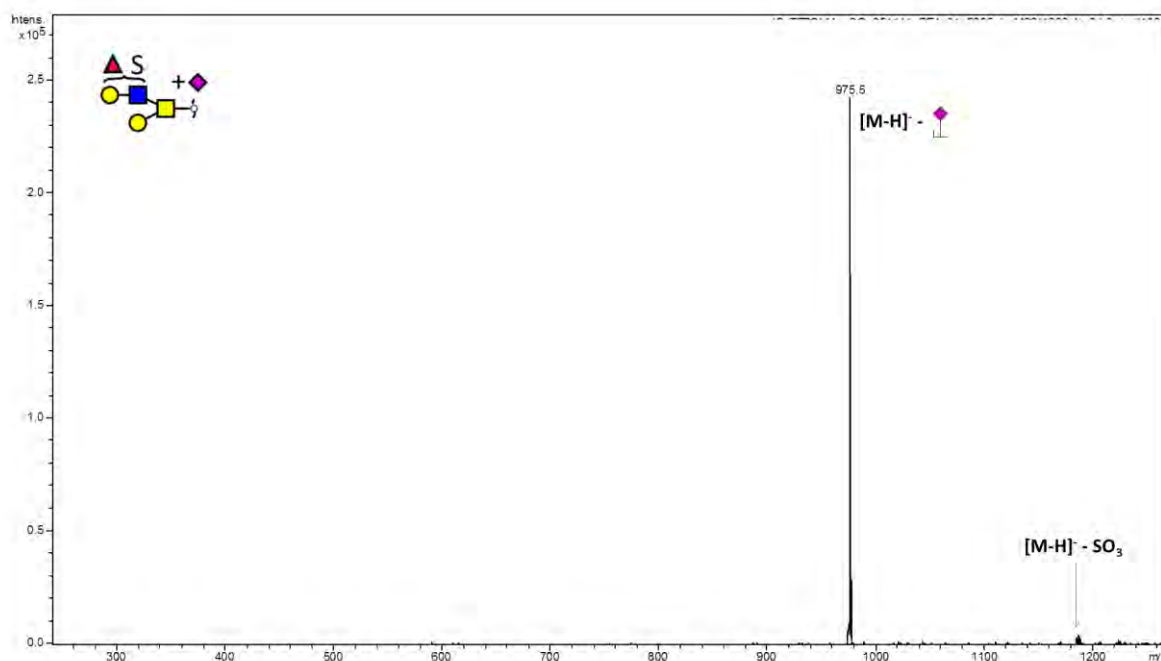
Glycan 39

Parent ion: *m/z* 657.4²⁻

Composition: (Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_2$
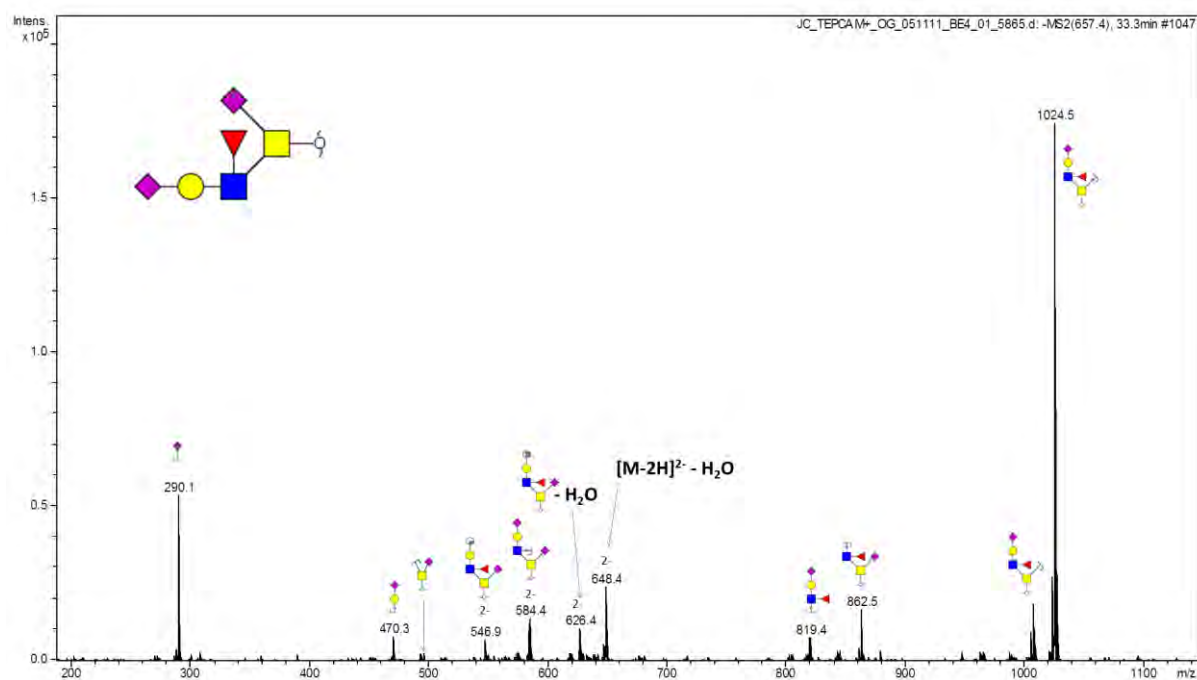
Glycan 40

Parent ion: *m/z* 1331.6⁻
Composition: (Hex)₂ (HexNAc)₂ (NeuAc)₂
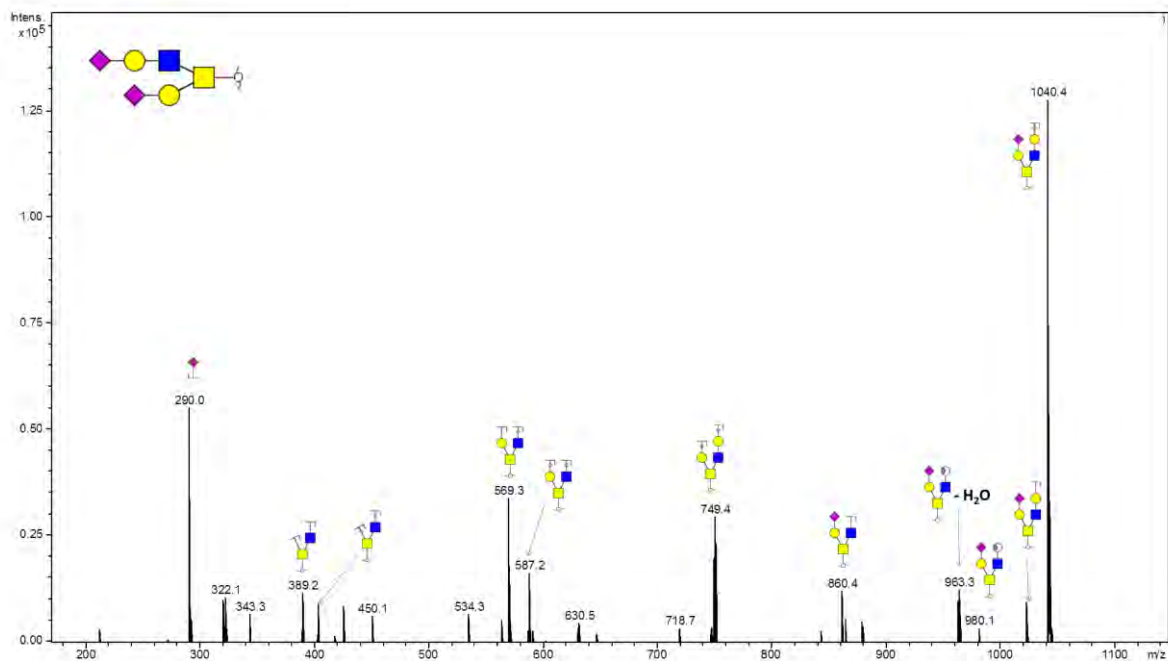
Glycan 41

Parent ion: *m/z* 685.9²⁻
Composition: (Hex)₁ (HexNAc)₃ (NeuAc)₂
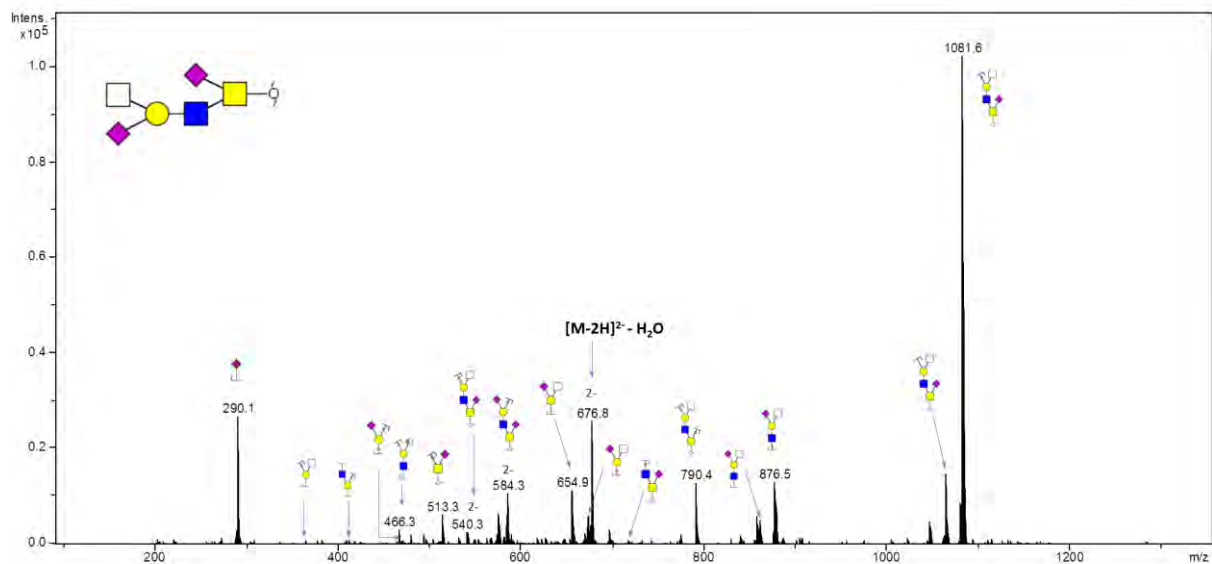
Glycan 42

Parent ion: *m/z* 738.3²⁻
Composition: (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₂
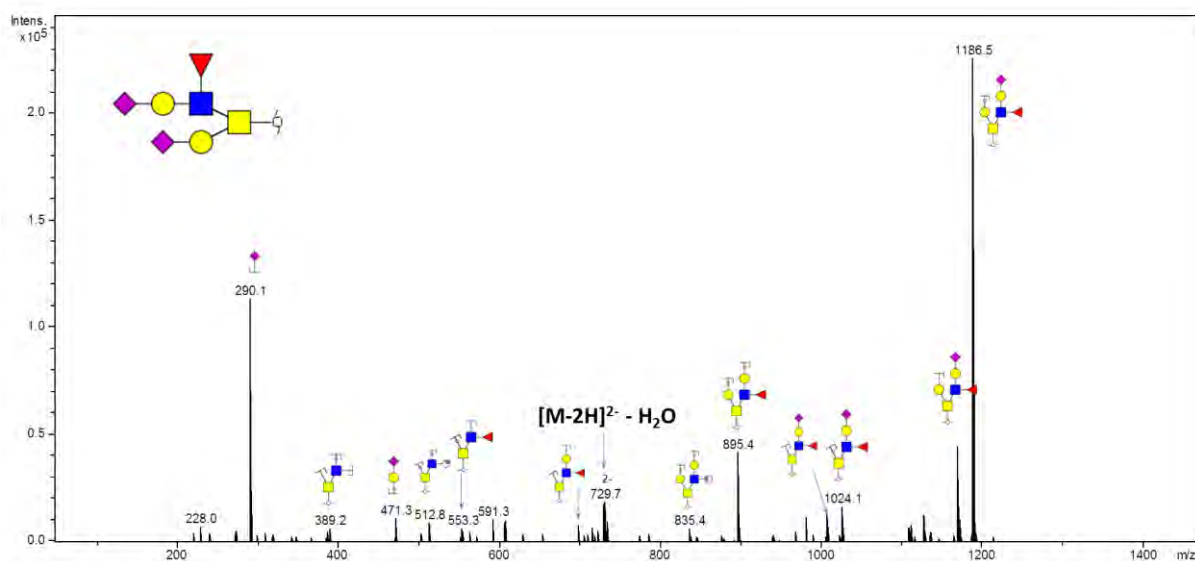
Glycan 43

Parent ion: *m/z* 840.0²⁻
Composition: (Hex)₂ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₂

Notes: Specific linkage of fucose not intended in glycan scheme (i.e. can be on either GlcNAc).
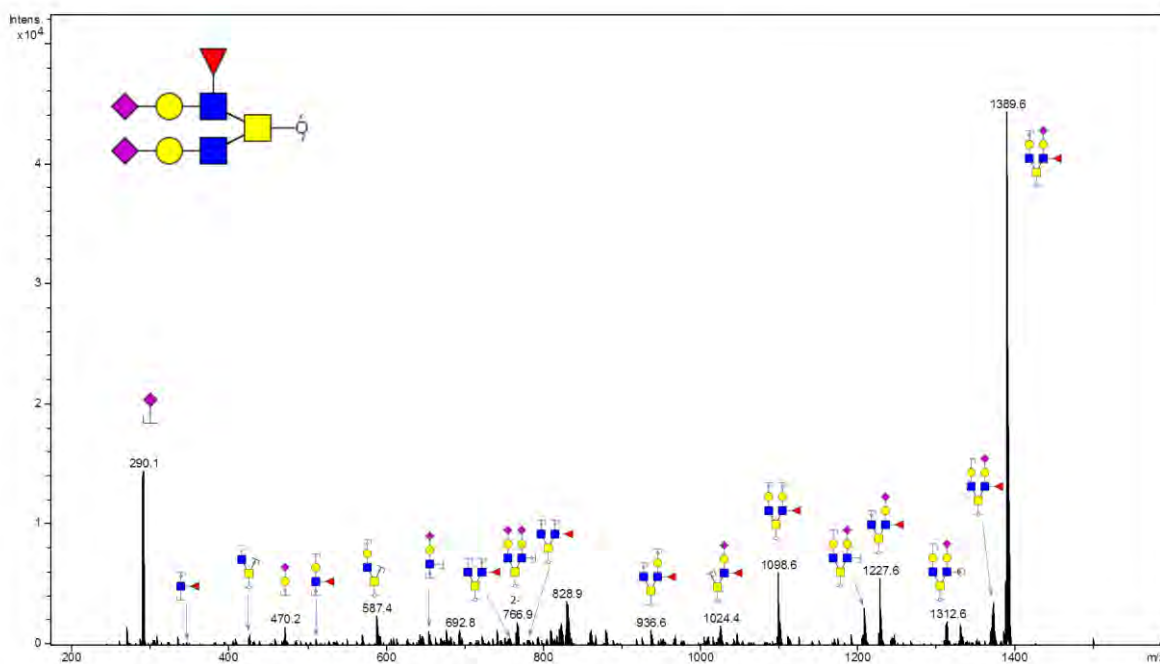
Chapter 3

Membrane glycosylation profiling of epithelial cells from patient-matched normal and tumour colonic tissues

3 Membrane glycosylation profiling of epithelial cells from patient-matched normal and tumour colonic tissues

3.1 Introduction

The findings presented from Chapter 2 suggest there are glycosylation differences between cellular model systems and the tumour tissues which they represent. In particular, potential glycan markers may be underexpressed or not detected in the cell model systems. Therefore, it is optimal to include clinical tissue samples during biomarker screening which more accurately capture the *in vivo* glycosylation status of the patient. To this end, there have been many large-scale antibody-based studies performing targeted glycan detection on colonic tissues from colorectal (CRC) patients (detailed in the introduction; Chapter 1). However, there have been few studies utilising mass spectrometry glycome profiling to investigate protein glycosylation in CRC tissues.

For *N*-glycosylation, Balog *et. al.* [38] performed a notable recent study on the *N*-glycosylation profiles of proteins from 13 patient-matched normal and tumour tissues using LC-MS/MS. In that work, paucimannosidic, sulfated and sialyl-Lewis type *N*-glycan structures were significantly increased while structures with bisecting GlcNAc were decreased in the tumours. On the other hand, *O*-glycosylation profiling has been focused on mucins. In a study using tissues from the tumours, transitional mucosa and normal mucosa from three CRC patients, sialyl-Tn was confirmed to be increased as well as a core 3 sialyl-Lewis X hexasaccharide [102]. Although these studies add to the growing evidence that glycosylation is altered during cancer, there is limited knowledge about the glycosylation profiles of membrane proteins, which is the focus of this thesis. Moreover, there have been no studies on protein glycosylation where epithelial cells have been immunoenriched from tissue samples prior to glycan analysis. Therefore, the aim of this chapter was to use the methods established in Chapter 2 to obtain *N*-glycosylation and *O*-glycosylation profiles from the epithelial cells of normal and pair-matched tumour tissues of up to six patients using immunoaffinity enrichment and LC-MS/MS.

3.2 Materials and methods

Tissue collection

Written and informed consent were obtained from patients under the same Ethics approval given by Sydney South Area Health Services (as Chapter Two) with the protocol number X08-0614. Matched normal and tumour tissues from six patients have been used for this study with normal mucosae sampled at least 10 cm away from tumours. The tissues were sampled during surgical resection and the clinicopathological characteristics are given in Table 1. The *N*- and *O*-glycan profiles of patient one were published earlier as a glycomic comparison between cell lines and tumour tissues [219]

(Chapter 2 of this thesis). *N*-glycosylation profiles were acquired from all six patients and described in this Chapter. However, in order to maximise glycomic and proteomic information, *O*-glycans were profiled from three patients only, while membrane proteomic profiles were analysed from the remaining three patients following *N*-glycan profiling (described in Chapter 4). A diagram showing sample preparation and experimental workflow for this Chapter is shown in Fig 1.

Table 1. Clinicopathological characteristics of patient cohort analysed in this thesis. The *N*- and *O*-glycan profiles of patients 1-3 will be analysed in this Chapter. [a]*N*- and *O*-glycan profiles of this sample has been previously published [219] (Chapter 2). [b]Profiles of membrane proteins were obtained from these patients after *N*-glycan profiles were obtained. Membrane protein profiles of these patients have been presented in Chapter 4.

Patient no.	Age	Gender	ACPS Stage	TNM stage	Blood Group
1[a]	76	F	B1	T3 N0 M0	o-
2	84	F	B1	T3 N0 M0	o+
3	67	F	B2	T4 N0 M0	o+
4[b]	74	F	B1	T3 N0 M0	o+
5[b]	96	F	B1	T3 N0 M0	o+
6[b]	79	M	B1	T3 N0 M0	a+
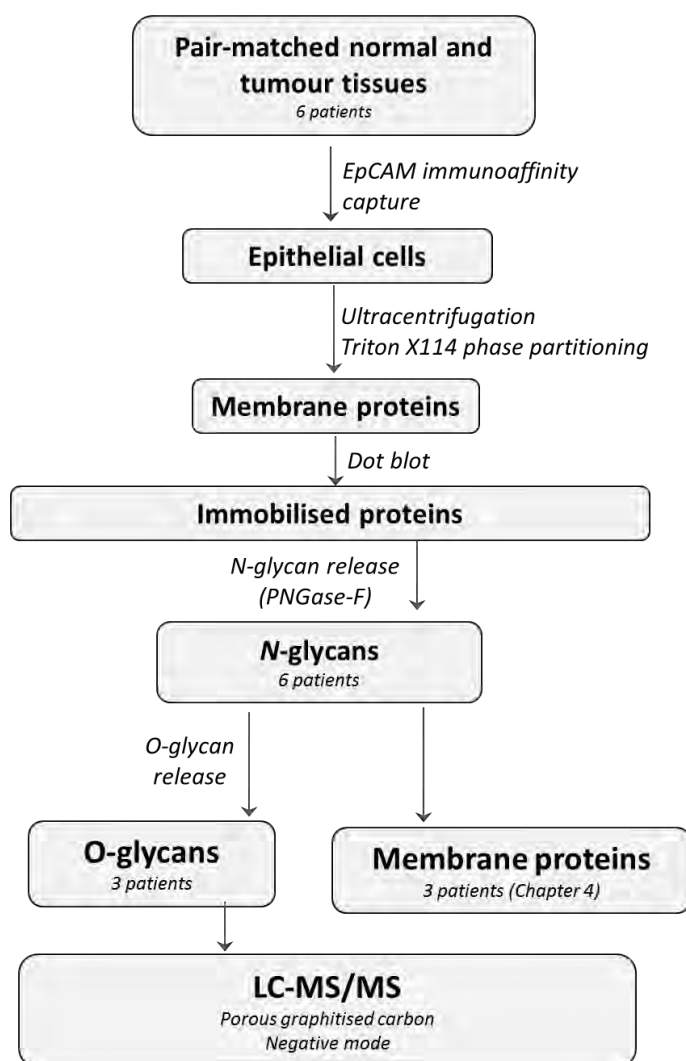
Figure 1. Sample preparation and experimental workflow for pair-matched *N*- and *O*-glycan profiling of epithelial cells from normal and tumour colonic tissues presented in this chapter.

Epithelial cell and membrane protein enrichment

The materials and methods for epithelial cell and membrane protein enrichment were basically the same as detailed in Chapter 2. Immunoaffinity capture by magnetic beads conjugated to epithelial marker EpCAM was used to enrich for epithelial cells from tissue, similar to Zhou *et. al.* [220]. However, it should be noted that Zhou *et. al.* utilised a negative selection method to enrich for epithelial/stromal cells (by removing of leukocytes using CD45-conjugated magnetic beads). In contrast, a positive selection method for enrichment of only epithelial cells was employed in this thesis. As mentioned previously, membrane protein enrichment was carried out by sodium carbonate stripping, ultracentrifugation and Triton X114 phase partitioning [221](Fig 1) in order to remove cytosolic proteins. Due to limited sample recovery, protein quantitation was not performed and glycan normalisation was carried out post-data acquisition using total glycan abundance (more details below).

N- and O-glycosylation profiling by LC-MS/MS

With the exception that normal (as well as tumour samples) were used, there was little difference in sample preparation as described in Chapter 2 [121]. LC-MS/MS and data analysis approaches also remained the same. Please refer to Chapter 2 for other details.

Relative quantification of glycans and statistical analysis

N- and *O*-glycans were normalised against total glycan abundance and relatively quantified by calculation of percentage area-under-the curve (AUC, calculated as (glycan area/sum of all glycan areas) x 100%) in Compass DataAnalysis (v4, Bruker Daltonics, USA) as previously detailed in Chapter 2. Two-tailed paired student's t-test was applied to determine statistical significance between normal and tumour samples. A p-value of less than 0.05 is considered statistically significant. Differences were considered differential if fold changes were greater than 1.5 with a p-value of < 0.05. For statistical differences between individual glycans, a value of 0.01 was added to all percentage areas to account for null values.

3.3   Results and discussion

In this chapter, the *N*- and *O*-glycosylation profiles of epithelial cells from patient-matched normal and tumour colon tissues were compared, as there is limited knowledge about changes in membrane protein glycosylation profiles during CRC carcinogenesis. Given that the most overt changes in the cell line and tumour comparison (Chapter 2) were *O*-glycan changes, such as the expression of sialyl-Tn, the differences in *O*-glycosylation will be first discussed below, followed by *N*-glycosylation changes.

O-glycan profiling of membrane proteins from normal and tumour colonic tissues

In total, 48 *O*-glycans were quantified on the epithelial membrane proteins of pair-matched normal and tumour colonic tissues from three patients. Of these, 43 O-glycans were structurally characterised by manual annotation of MS2. Proposed glycan compositions have been given for the remaining 5 glycans. The majority of the glycan structures observed were previously characterised in the one tissue sample using MS2 (Chapter 2). However, there were five structures not characterised earlier in this sample and their annotated MS2 spectra are shown in Supplementary Fig 1. The percentage abundance of all quantified *O*-glycans in normal and tumour tissue epithelial proteins is shown in Supplementary Table 1. A representative example of the *O*-glycan profiles from normal and cancer tissue membrane proteins observed in this study is shown in Fig 2a.

The glycan structures characterised in this study have all been previously reported to be in the *O*-glycan profiles of mucin 2 isolated from 25 normal colonic biopsies by Larsson *et. al.* [122]. When

compared to glycan profiles reported in that same study, the overall normal glycan profile was similar although the number of *O*-glycans described was higher with over 100 *O*-glycans reported in the previous study. Interestingly, the number of glycan structures in another study by Robbe-Masselot *et. al.* comparing mucin 2 glycans in normal, transitional mucosa and tumours of three patients found a similar number of glycan structures (approximately 40)[102] compared to this study. Since both Larsson *et. al.* [122] and Robbe-Masselot *et. al.* [102] analysed mucin *O*-glycosylation profiles, differences in sample preparation was also compared to determine whether it was a contributing factor in the number of *O*-glycans reported. However, both studies utilised similar approaches consisting of density gradient centrifugation with guanidinium chloride for the mucin enrichment, followed by negative mode LC-MS/MS. Therefore, the differences in number of *O*-glycans reported between these studies are less likely to be simply due to changes in sample preparation. Having mentioned this, differences in glycan loading cannot be excluded, although neither study quantified total carbohydrate content analysed and only one study mentioned wet weight of biopsy analysed (~5-10 mg by Larsson *et. al.[122]*). In addition, the patient cohort was notably larger in the study performed by Larsson *et. al,* compared to Robbe-Masselot *et. al.* and this study, which may have contributed to the increased glycan heterogeneity observed. Thus, discrepancies in the number of *O*-glycans reported by between various studies (including this Chapter) are most likely explained by a combination of differences in patient group size and/or glycan loading.

Despite these differences, overt relative *O*-glycan changes can be observed between normal and tumour tissues (Fig 2a) in this study. However, although these changes were not statistically significant (p-value < 0.05), sialyl-Tn, $(HexNAc)_1 (NeuAc)_1$, with *m/z* 513.3 (Fig 2d), and the sialyl Lewis X hexasaccharide $(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_2$ with *m/z* 1315 (Fig 2e), showed a trend towards increased expression in the tumour compared to normal tissue (increased by 17-fold and 1.3-fold respectively), which is consistent with previous observations of their upregulation in tumours [78, 102]. As discussed above, it is likely that the small patient group (n=3) of this study contributed to the lack of statistical significance. However, the large patient variance in glycan expression is also likely to be an additional contributing factor. This is especially likely in the case of sialyl-Tn, which has been shown upregulated in tumours but not usually detected in normal tissues in numerous other studies with larger patient cohorts [37, 78, 80, 169, 170].
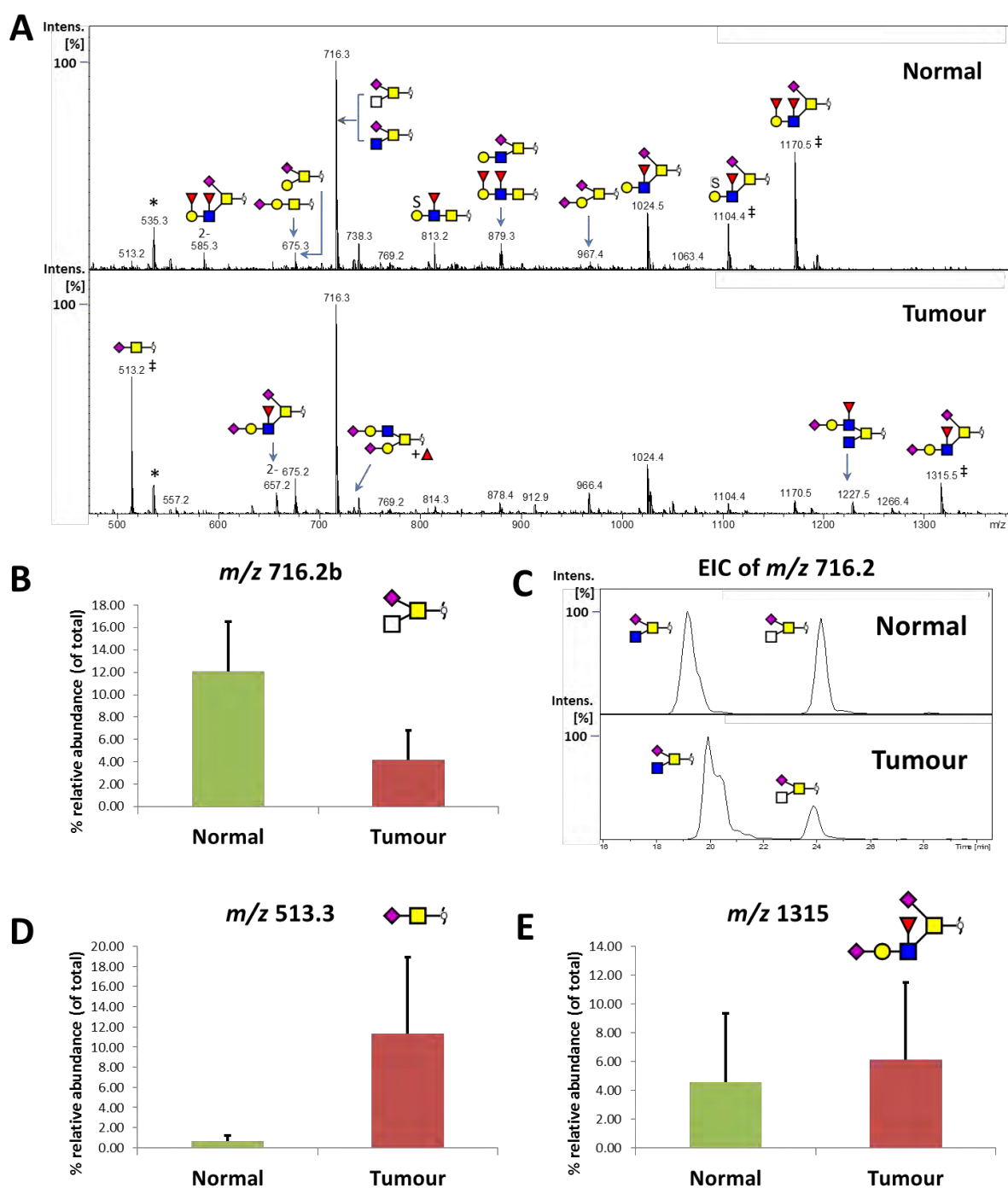
Figure 2. *O*-glycan profiling of membrane proteins and glycan changes from patient-matched normal and tumour colonic tissues. Glycans have been depicted using CFG symbol nomenclature (Chapter 1). Relative abundances were determined by % area-under-the-curve of total *O*-glycans. Bar charts depict data shown as mean ± SD of three patients. **A)** Example *O*-glycan profiles using average MS representing glycans across total elution time. Major glycan features have been labelled. Peaks are singly charged unless otherwise indicated. Glycan changes can be observed although many were not considered statistically significant (examples labelled with ‡) by two-tailed paired student's t-test. *Asterisked peaks do not correspond to glycans (as determined by MS2). **B)** The second isomer of *m/z* 716.2 was the only glycan found to be differentially expressed in colon cancer tissue

(downregulated by 2.9-fold), using criteria of 1.5-fold change and p-value < 0.05. **C)** Representative example of extracted ion chromatogram (EIC) of *m/z* 716.2 showing relative decrease in second eluted isomer in tumours. **D)** Well-known cancer marker sialyl-Tn was found to be increased 17-fold in the tissue, although difference was not significant (p = 0.120) due to tumour expression variability. **E)** Potential glycan marker with *m/z* 1315 also showed an upward trend in tumours (1.3-fold, p = 0.749).
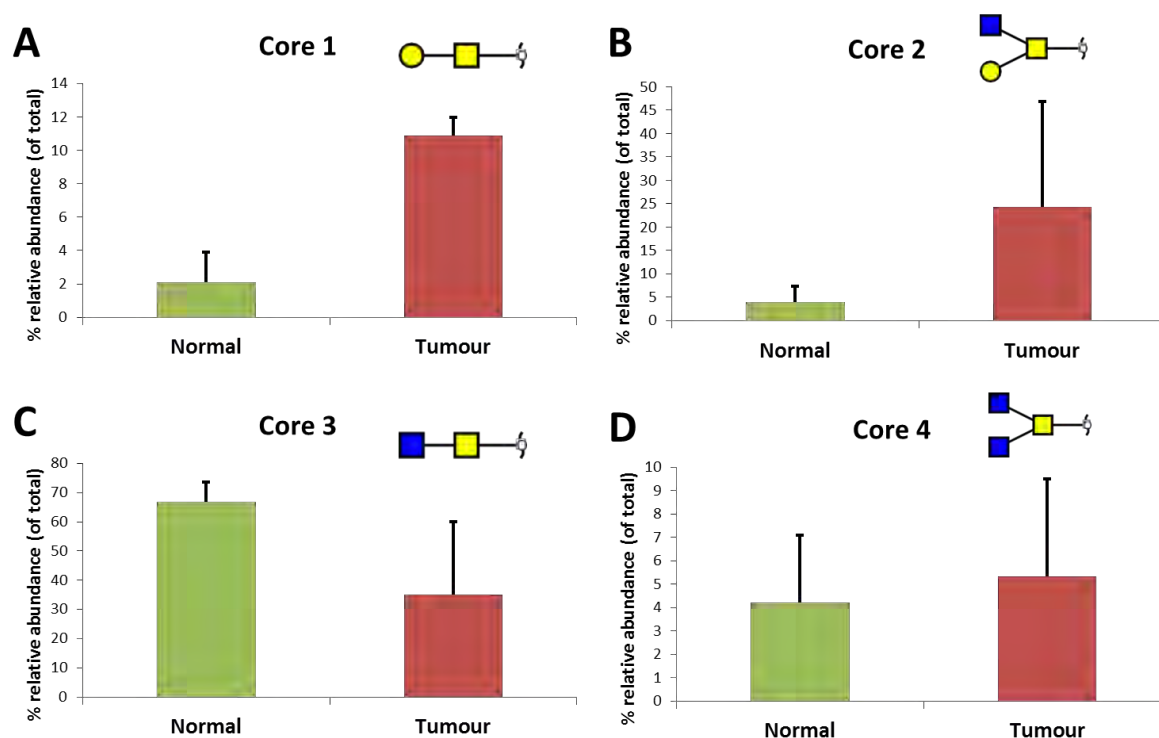
Figure 3. Relative abundances of *O*-glycans based on cores 1-4. Relative abundances were determined by % area-under-the-curve of total *O*-glycans. Data shown as mean ± SD of three patients. Statistical significance was calculated using two-tailed paired student's t-test. **A)** Core 1 glycans were increased in tumours (5.1-fold, p = 0.03). **B)** Core 2 glycans showed an increased trend in the tumours (6.2-fold, p = 0.28). **C)** Core 3 glycans showed a decreased trend in the tumours (1.9-fold, p = 0.20). **D)** Little change was observed for core 4 glycans between normal and tumour colonic tissues (1.25-fold, p = 0.77).

In Chapter 2, sialyl-Tn was shown to be upregulated and correlated with mucin expression [102, 122] in the cell lines, especially mucin-2, mucin-5B and mucin-6, and thus these mucins were proposed as potential carriers of these glycans. Moreover, the sialyl Lewis X hexasaccharide is also likely to be carried by mucins since it was first discovered as a potential glycan marker on mucins isolated from colonic biopsies [102]. Therefore, the increased expression of these glycans strongly suggests that mucins were present in the enriched membrane proteins from tissue, which was confirmed in the proteomic study on three other patients (Chapter 4). Several mucins were detected across all three

patients analysed (Supp Table 2, Chapter 4) but only mucin-2 and mucin-13 were consistently found in all patients (with a min. of 1 peptide). When these two mucins were quantified, neither was found to be differentially expressed (Fig 4, criteria of 1.5-fold change and $p < 0.05$) although there was a decreasing trend in mucin-2 expression in the tumours (2.1-fold, Fig 4a). Taken together, this indicates that the increased expression of sialyl-Tn and the sialyl Lewis X hexasaccharide markers are not a simple direct result of increased expression of mucin-2 or mucin-13. Rather, alterations in glycan-to-mucin ratios resulting from changes in the glycan biosynthetic pathway may be more crucial for overall membrane expression of these cancer antigens.

Strikingly, the isomer of *m/z* 716, equating to a composition of $(HexNAc)_2 (NeuAc)_1$ and eluting second from the PGC column (Fig 2c), was generally found to be decreased in abundance by 2.9-fold in tumours compared to normal colon tissues (Fig 2b) for the first time. The elution order of this glycan has been previously established under similar conditions [222]. As determined in Chapter 2, the first isomer is a core 3 *O*-linked structure with the second isomer based on core 5/6/7 which has been reported to be found in the colon [22, 223]. Though little is known about synthesis of these cores, it is reasonable to propose that enzymes involved in their synthesis have been downregulated or their activity inhibited in the tumours. Other factors such as limited substrates are less likely given that only GalNAc and GlcNAc are required for core 5/6/7 formation and the other glycans expressed also use these substrates. Nevertheless, given the small number of patients analysed (n=3), further verification is critical to establish whether this novel finding is reproducible in larger sample sizes. It would also be interesting to determine whether the expression of this glycan is correlated with clinicopathological variables such as tumour stage, lymphatic invasion and cell proliferation, which may also indicate a functional role in malignancy. Thus, the decreased expression of this glycan is intriguing and warrants further investigation.

In addition to analysis of individual glycan structures, the abundance of different core *O*-glycan structures based on core 1-4 was also determined (Fig 3). Core 1 glycans (Fig 3a) were found to be differentially upregulated (5.1-fold) in the tumours ($p = 0.03$). For core 2 (Fig 3b), there was a trend towards increased expression (6.2-fold) in the tumours though increased expression was not significant ($p = 0.28$). The expressions of core 3 and 4 *O*-glycans were also not found to be significantly altered although core 3 glycans showed a decreased trend (1.9-fold, $p = 0.20$) while there was little change in core 4 glycan structures in the tumours. While it was noted that the above changes in glycan cores were generally not significant, the trends exhibited between normal and tumour colonic tissues still provide some information on the complex dynamics of *O*-glycan synthesis, although as mentioned before, glycan synthesis results from an intricate interplay of factors such as substrate availability [224], nucleotide sugar transporter activity [225] and

glycosyltransferase expression [226, 227] (see also Chapter 2). For example, multiple glycosyltransferases may compete for the same substrate, such as Tn antigen, for the formation of core 1, core 3 and sialyl-Tn (Fig 6a, Chapter 2). In particular, in colon cancer cell cultures, C1GalT1 and ST6GalNAc1 were identified as competing glycosyltranferases with important roles in sialyl-Tn expression (Fig 6f, Chapter 2).Overall, an increase in both C1GalT1 and ST6GalNAc1 activity may explain the increased abundance of core 1 glycans and upward trend in sialyl-Tn expression in the tumours.
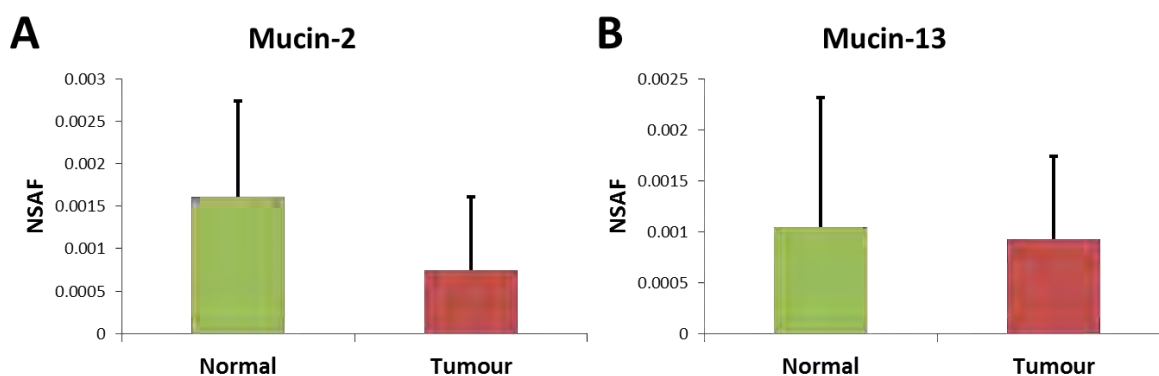
Figure 4. Label-free quantification of mucins from membrane proteomic study of patient-matched normal and tumour colonic tissues (details in Chapter 4) using normalised spectral abundance factors (NSAF). Data shown as mean ± SD of three patients. **A)** Mucin-2 expression. **B)** Mucin-13 expression. Neither mucin was found to be differentially expressed. Although several other mucins were detected (e.g. mucin-1, mucin-4 and mucin-5B), these were not quantified as they were not consistently detected across all patients (min. 1 pep).

On the other hand, core 3 glycans exhibited a relative decrease in the tumours compared to normal tissue (1.9-fold, $p = 0.20$, Fig 3c), potentially suggesting a lower core 3 synthase activity in the tumours, which has been reported earlier [152]. Notably, this decrease in core 3 glycans appears to mirror the increase in core 1 glycans suggesting a reciprocal relationship between the corresponding enzymes similar to previous reports [228]. Taking all glycan changes together, there appears to be a shift towards increased C1GalT1 and/or ST6GalNAc1 activity with a corresponding decrease in activity of core 3/5/6/7-related enzymes in the tumours. Overall, *O*-glycan alterations were found associated with CRC which also implies alterations of the corresponding enzymes involved in their synthesis. While observations were not necessarily statistically significant, glycan changes generally followed expected trends including an increase in sialyl-Tn expression, which is consistent with previous studies and confirms its status as a glycan marker of CRC. In the following section, the *N*-glycan profiles of membrane proteins from pair-matched normal and tumour colonic tissues will be compared.

N-glycan membrane profiling from normal and tumour colonic tissues

In this study, N-glycan membrane protein profiles from epithelial-enriched normal and pair-matched tumours in the colon were obtained and analysed from 6 patients (Table 1). A total of 90 N-glycans were found and quantified by integration of their areas-under-the-curve across all 6 patients. The relative abundances (% of total area-under-the-curve) of these glycans are shown in Supp. Table 2. Out of these 90 structures, 82 were characterised by manual annotation of MS2. For the other 8 N-glycans (not structurally characterised), proposed glycan compositions have been given (Supp. Table 2). Aside from these, most of the glycan structures were previously characterised using MS2 in Chapter 2. However, 8 new structures were also found in addition. Their annotated MS2 have been provided in Supp. Fig 2. Representative N-glycan profiles from normal and tumour tissues are shown in Fig 5a.

In comparison with previous literature, only one other study has analysed protein N-glycans from colonic tissues in-depth using mass spectrometry [38]. In this report by Balog *et. al.* [38], N-glycan profiles were obtained from 13 normal controls and their corresponding colon tumour tissues. The compositions of 245 N-glycans were reported in total to be present in the tissues which is a higher number compared to this study. However, this is unsurprising given the size of patient group analysed was considerably smaller (13 vs 6 analysed in this study) which may reduce glycan heterogeneity observed. Nevertheless, differences in sample preparation between the study by Balog *et. al.* and this investigation may also contribute to differences in N-glycan number. In the former study, a general protein extraction was performed using chloroform:methanol precipitation. In addition, a mixed cell population was likely analysed since cell-type enrichment was not conducted prior to glycan analysis. In contrast, in this study, epithelial cells were first enriched using EpCAM-conjugated beads. Subsequently, membrane proteins were enriched using sodium carbonate stripping, ultracentrifugation and Triton X114 phase partitioning (Fig 1). Thus, the discrepancy in the number of N-glycans may be reflective of different approaches in sample preparation used. Indeed, sample loss during the additional processing steps required for epithelial cell and membrane protein enrichment performed in this study cannot be excluded, which may result in lower overall glycan abundance and coverage. Alternatively, there may be lower N-glycan heterogeneity in epithelial membrane proteins compared to cell N-glycome of proteins as a whole and this may also explain the lower number of N-glycans observed in this study.
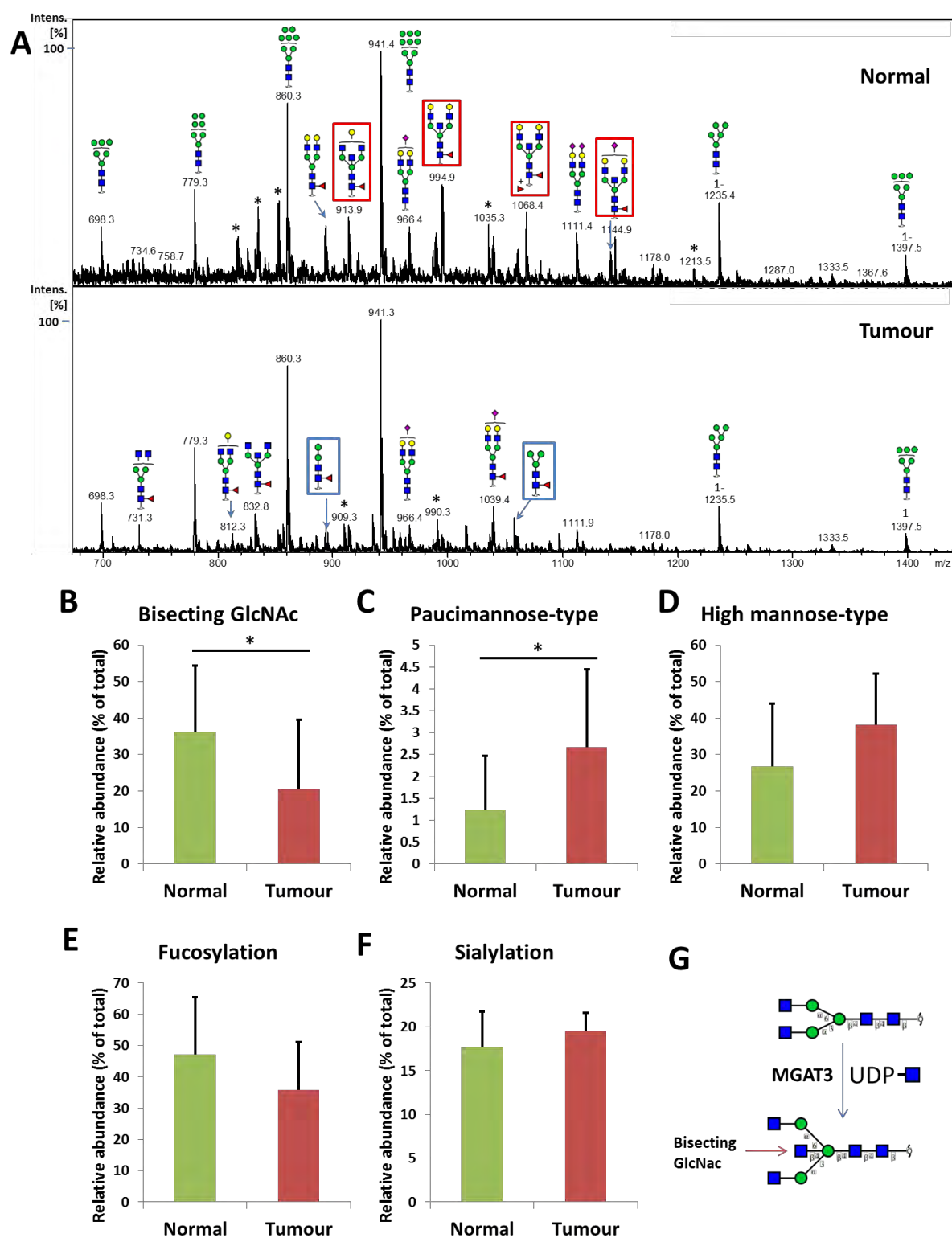
Figure 5. *N*-glycan changes between epithelial cells from normal and pair-matched colonic tissues.
A) Example *N*-glycan profile (average MS) from normal and tumour tissue. Major glycans have been labelled. Peaks are doubly charged unless otherwise indicated. Bisected *N*-glycans (examples boxed in red) were generally lower in intensity in the tumour samples (see also Fig 5B). In contrast, paucimannose *N*-glycans (examples boxed in blue) were more highly expressed in the tumours (see also Fig 5C). *Asterisked peaks do not correspond to glycans (as determined by MS2). **B-F)** Relative abundance of different *N*-glycan types for normal and tumour colonic tissues (by % area-under-the-

curve of total *N*-glycans). Glycans were considered differentially changed if there was a fold-change > 1.5 and p-value < 0.05 by two-tailed paired student's t-test. **B)** Bisecting-GlcNAc *N*-glycans were found to be significantly decreased in tumours (1.75-fold, p = 0.017). **C)** Paucimannose-type *N*-glycans (defined as glycans with compositions (Hex)$_{1-3}$ (HexNAc)$_2$ (Deoxyhexose)$_{0-1}$) were observed to be significantly increased in tumours (2.15-fold, p = 0.022). **D)** High mannose-type *N*-glycans, **E)** fucosylated *N*-glycans (fucose may be attached at the core or outer-arm), and **F)** sialylated *N*-glycans were not found to be significantly altered between normal and tumours colon tissues. **G)** Synthesis of bisecting GlcNAc structures occurs by the action of GlcNAcT-III enzyme (encoded by MGAT3 gene), which adds a GlcNAc residue to central mannose of chitobiose core.

In spite of this, the findings presented here have been generally consistent with the previous study [38]. High mannose-type, fucosylated and sialylated *N*-glycans (Fig 5d-f) were not reported as differentially expressed between normal and tumour colon tissues in both studies. However, alterations in bisecting and paucimannosidic *N*-glycans (Fig 5b & c) between normal and tumour tissues were reported from both investigations. Bisecting *N*-glycans have been defined previously (in Chapter 2) as *N*-glycans with a GlcNAc attached to the central mannose of the *N*-glycan chitobiose core. Bisecting *N*-glycans are assigned when a diagnostic fragment ion (D-221) is highly abundant in the MS2 [139, 142](Fig 7). Paucimannosidic structures can be defined as truncated *N*-glycans containing one to three mannose residues [(Man)$_{1-3}$ (HexNAc)$_2$] which may or may not be core-fucosylated. A similar definition has also been employed in the previous study by Balog *et. al.* [38]. Here, it was observed that bisecting *N*-glycans were significantly decreased in the tumours (by 1.76-fold; p = 0.017, Fig 5b). In contrast, paucimannose structures were at low abundance but were increased in the tumours (by 2.15 fold; p = 0.022, Fig 5c) compared to the matched normal tissue. Both these findings are in agreement with the earlier report by Balog *et. al.* [38] and provide further evidence for the association of bisecting and paucimannosidic structures with CRC.

The changes in paucimannosidic and bisecting *N*-glycans also imply alterations in the glycan synthetic pathways underlying their expression. In particular, glycan synthetic enzymes (glycosyltransferases and glycosidases) are implicated. In regards to bisecting *N*-glycans, the addition of bisecting GlcNAc is catalysed by GlcNAcT-III enzyme (encoded by the MGAT3 gene, Fig 5g). Thus, the reduced expression of bisecting *N*-glycans in the tumours suggests a corresponding decrease in the activity of GlcNAcT-III. Synthesis of paucimannosidic glycans begins similarly to other N-glycans with N-acetylglucosaminyltransferase I (GlcNAcT I) transferring a GlcNAc residue onto (Man)$_5$(GlcNAc)$_2$[229] which is followed by the removal of two mannose residues by α-mannosidase II [25, 230, 231] (Fig 6a). The terminal GlcNAc residue (transferred by GlcNAcT I) is then removed by another hexosaminidase to produce (Man)$_3$(GlcNAc)$_2$ [232]. The increased expression of paucimannosidic

glycans therefore implies higher activity of hexosaminidase and α-mannosidase II. Alternatively, these trimmed structures may result from action of lysosomal exoglycosidases, such as α-mannosidases [233], N-acetylglucosaminidases [233], which have been shown to be upregulated in colorectal tumours compared to normal controls. Indeed, it has been proposed that the internal lysosome containing these upregulated glycosidases may undergo exocytosis in malignant cells, releasing their intracellular contents, thus enabling cell surface glycan modification of tumours[234]. Although the mechanism behind the increased expression of paucimannose structures in these CRC tissues remain to be elucidated, the upregulation of paucimannosidic glycans in CRC clearly indicates dysregulation of glycan synthetic pathways leading to their formation. This finding also gives evidence to support the expression of paucimannosidic glycans in humans which were previously considered to be expressed only in plants and invertebrates [25, 235].

At the level of individual *N*-glycans, 12 differentially altered *N*-glycans between normal and tumour colonic tissues were identified in this study in total (Table 2). With the exception of one glycan (glycan 65), these structures have not been identified as differentially expressed between normal and tumours colon tissues in the earlier study by Balog *et. al.* [38]. Of these 12 glycans, 5 were bisected *N*-glycans (glycans 65, 94, 120, 121 and 143; Table 2) and their MS2, with the diagnostic ion for bisecting GlcNAc, are shown in Fig 7. All were differentially decreased in the tumour (2.77 to 12.97-fold, Table 2) which agrees with the overall trend of decreased bisecting *N*-glycans in colon cancer observed in this and the previous study [38]. In addition, a paucimannose structure with composition $(Man)_3 (HexNAc)_2 (Deoxyhexose)_1$ was observed as differentially increased in the tumours (2.44-fold, glycan 3, Table 2). The structure of this glycan is similar to the paucimannosidic glycan previously reported as increased in tumours [38], which is essentially the same except it contained one less mannose residue $((Man)_2 (HexNAc)_2 (Deoxyhexose)_1)$. Both of these paucimannosidic structures were found to be differentially increased in the tumours compared to normal adjacent tissue (> 2-fold) and therefore are potential biomarker candidates for CRC.

Table 2. Statistically significant differentially expressed *N*-glycans on membrane proteins from epithelial cells of patient-matched normal and tumour colonic tissues (p < 0.05, > 1.5-fold change). Statistical significance was determined by two-tailed paired student's t-test. Additional isomers cannot be excluded. Glycans have been numbered and previously characterised according to Supp Fig 5 in Chapter 2 except glycan 189 which is located in Supp Fig 2 of this Chapter. n.d., not detected.

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure	Fold change (T/N)	P-value (T-test)	Normal (% area)	Tumour (% area)
37	n.d.	856.4	1713.8	0.18	(Hex)1 (HexNAc)1 (Deoxyhexose)1 (NeuAc)1 + (Man)3(GlcNAc)2		4.03	0.0207	0.41 ± 0.371	1.67 ± 1.136
36	n.d.	856.3	1713.6	-0.02	(Hex)1 (HexNAc)1 (Deoxyhexose)1 (NeuAc)1 + (Man)3(GlcNAc)2		3.18	0.0445	0.07 ± 0.051	0.21 ± 0.15
189	1260.5	n.d.	1260.5	0.03	(HexNAc)1 (Deoxyhexose)1 + (Man)3(GlcNAc)2		2.56	0.0295	0.23 ± 0.408	0.59 ± 0.627
3	1057.4	n.d.	1057.4	0.01	(Hex)3 (HexNAc)2 (Deoxyhexose)1		2.44	0.0236	0.77 ± 0.856	1.88 ± 1.507
17	n.d.	783.5	1568	0.44	(Hex)1 (HexNAc)1 (NeuAc)1 + (Man)3(GlcNAc)2		2.27	0.0239	0.44 ± 0.245	0.99 ± 0.393
138	n.d.	1140.4	2281.8	-0.03	(Hex)2 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1 + (Man)3(GlcNAc)2		-2.51	0.0077	0.47 ± 0.396	0.19 ± 0.252
120	n.d.	1068	2137	0.20	(Hex)2 (HexNAc)3 (Deoxyhexose)2 + (Man)3(GlcNAc)2		-2.77	0.0483	0.23 ± 0.216	0.08 ± 0.104
65	n.d.	913.9	1828.8	0.12	(Hex)1 (HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2		-3.31	0.0161	4.33 ± 2.493	1.31 ± 1.45
143	n.d.	1141	2283	0.15	(Hex)2 (HexNAc)3 (Deoxyhexose)3 + (Man)3(GlcNAc)2		-3.34	0.0139	1.57 ± 1.199	0.47 ± 0.694
152	n.d.	1177.5	2356	0.13	(Hex)3 (HexNAc)4 (Deoxyhexose)1 + (Man)3(GlcNAc)2		-7.00	0.0455	0.49 ± 0.451	0.07 ± 0.071
94	n.d.	994.9	1990.8	0.06	(Hex)2 (HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2		-7.44	0.0093	6.84 ± 4.173	0.92 ± 0.817
121	n.d.	1068	2137	0.20	(Hex)2 (HexNAc)3 (Deoxyhexose)2 + (Man)3(GlcNAc)2		-12.97	0.0252	5.46 ± 4.446	0.42 ± 0.637
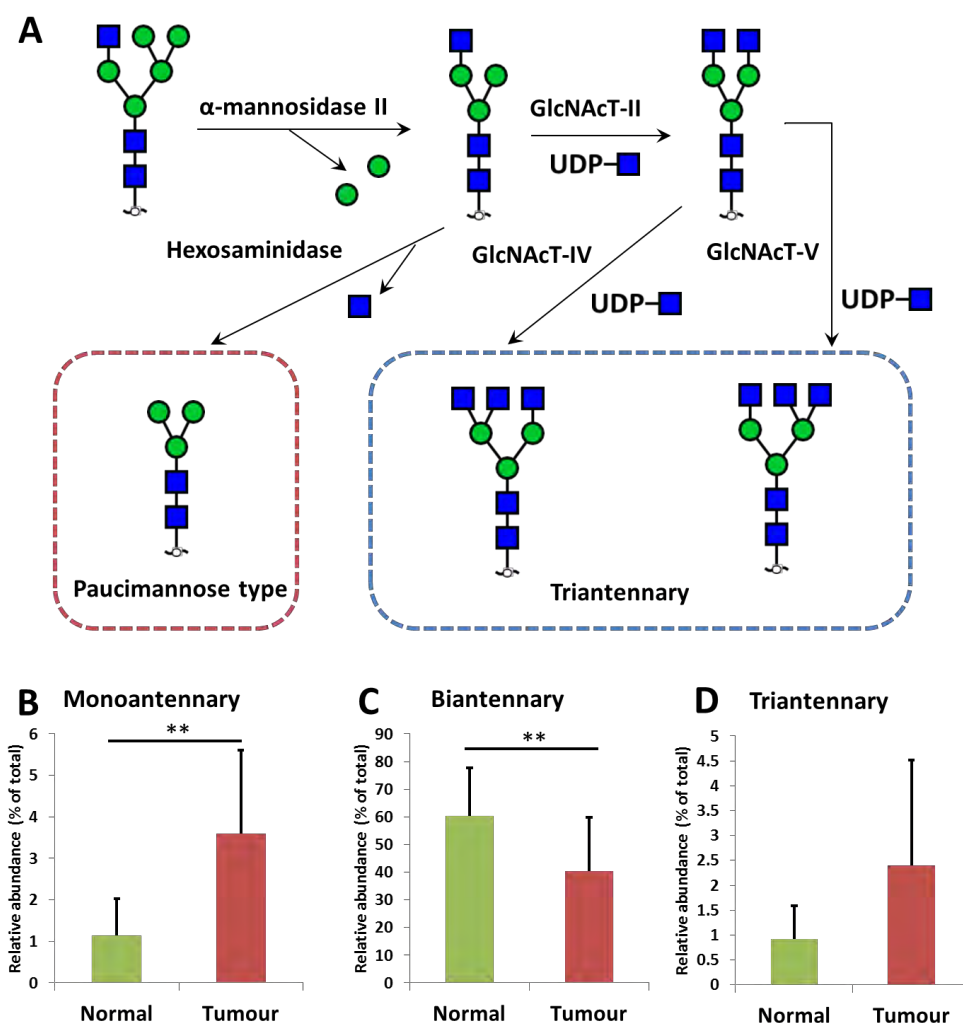
Figure 6. Branching of complex *N*-glycans in normal and tumour colonic tissues. Relative abundances were calculated by % area-under-the-curve of total *N*-glycans. Data shown as mean ± SD of three patients. **A)** Schematic showing biosynthesis of branched and paucimannosidic *N*-glycans. Triantennary *N*-glycans can be further elongated to tetraantennary *N*-glycans. Adapted from [25]. **B)** Monoantennary glycans were significantly increased in the tumours (3.2-fold, p = 0.0083), **C)** Biantennary glycans were significantly decreased in the tumours (1.5-fold, p = 0.0041). **D)** Triantennary glycans showed a trend of increased expression in the tumours (2.6-fold, p = 0.13).
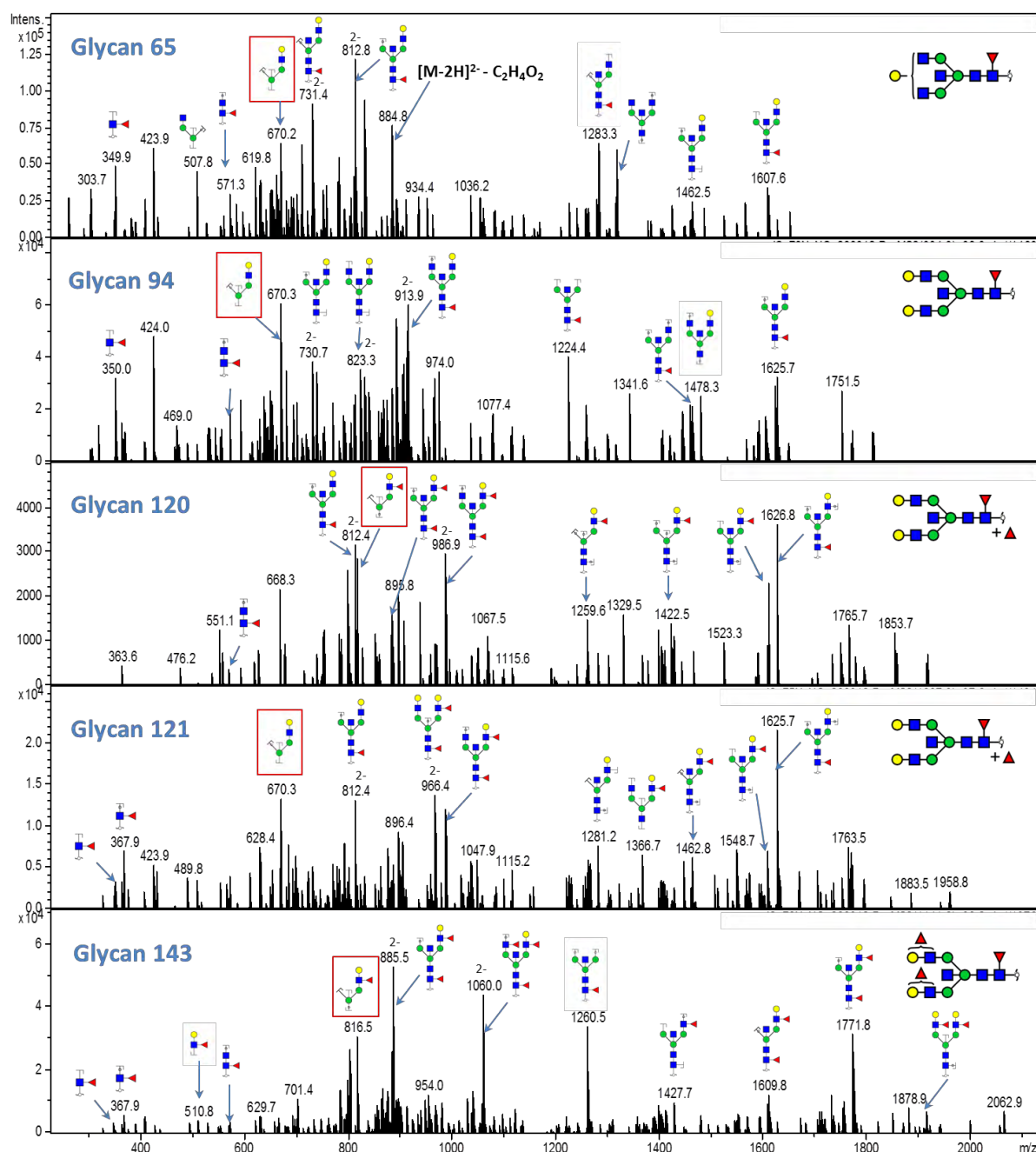
Figure 7. Annotated MS2 of the 5 differentially expressed bisecting *N*-glycans from CRC tissues (from Table 2). Glycans have been numbered according to Table 2 and depicted using CFG notation. Glycan schemes have been derived from GlycoWorkbench. Peaks are singly charged unless otherwise indicated. Diagnostic fragment ions for the verification of bisecting GlcNAc (D-221) have been boxed. Location of outer-arm fucose not defined by glycan fragment (fucose may be linked to galactose or GlcNAc).

Intriguingly, four sialylated *N*-glycans were also found to be differentially expressed in the tumours, although they were at low abundance. Three were upregulated (glycan 17, 36 and 37; 2.27-, 3.18- and 4.03-fold respectively; Table 2) while the fourth was downregulated (glycan 138, 2.51-fold; Table 2). This is in contrast to the previous study by Balog *et. al.* [38] in which no sialylated *N*-glycans were

found to be increased in tumours, while only one sialylated *N*-glycan, with composition $(Hex)_2$ $(HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$ was decreased in the tumours.

The three sialylated *N*-glycans identified as upregulated in CRC tumours for the first time were monoantennary, while the underexpressed sialylated *N*-glycan was biantennary (Table 2). The increase in monoantennary structures was concordant with the overall increase in monoantennary structures in the tumours (3.2-fold, Fig 6a) which was also found to be highly significant (p = 0.008). This was mirrored by decreased expression of biantennary *N*-glycans in the tumours (1.5-fold, p = 0.004, Fig 6b). Interestingly, these changes did not preclude the synthesis of more highly branched structures such as triantennary and an upward trend in triantennary *N*-glycans was observed in the tumours (2.6-fold, Fig 6c). As triantennary structures were synthesised, this also suggests that UDP-GlcNAc (nucleotide sugar donor) was not limiting in the tumours. Taken together, these observations indicate a highly complex dynamic in the regulation of glycan branching. The regulation of *N*-glycan branching has not been previously suggested in other studies on CRC, although branching has been reported as upregulated in other cancers, such as breast [107] and liver [236]. Although further experimentation is required to validate the changes in glycan synthetic pathways, the differential expression of these glycans nonetheless indicates they may also be glycan markers of tumorigenesis in CRC.

In addition to the above changes in the glycan biosynthetic pathways, differential protein expression was also considered as a factor which may impact on global membrane *N*-glycosylation. This is detailed further in Chapter 4 where the membrane proteome of a subset of patients was analysed using mass spectrometry (Fig 1). Interestingly, only 9 glycoproteins were differentially expressed out of more than 1600 proteins quantified (Table 2, Chapter 4), suggesting there is little change in glycoprotein expression. Though further work is needed to determine whether *N*-glycosylation changes occur proteome-wide or on specific proteins, the small alterations to glycoprotein expression overall indicates that changes in the glycan biosynthetic pathways are likely more critical for changes in *N*-glycosylation than differential glycoprotein expression.

Nuanced glycan changes as a hallmark of tissue systems

One salient feature of the glycan profiles in this study was the relatively nuanced changes between normal and tumour tissue which largely consist of changes in glycan abundance, rather than presence/absence as seen in the initial study comparing cell model systems and cancer tissues. For example, cancer markers sialyl-Tn and sialyl Lewis X hexasaccharide with *m/z* 1315 were absent for the majority of CRC cell lines but were detected in high amounts in the tumours in Chapter 2. In contrast, in this study, very low expression of sialyl-Tn can also be detected even in the healthy control tissues (Fig 2d), which illustrates that changes are more quantitative in nature (changes in

abundance), rather than qualitative (presence/absence). This is further supported by changes observed in differentially expressed *N*-glycans, where some glycans were detected in low abundance (<1%) but were nonetheless present. In addition, the variability in the expression of sialyl-Tn (Fig 2d) and the hexasaccharide with *m/z* 1315 (Fig 2e) demonstrates that patient variance of glycan expression may be large. The degree of variation observed also underscores the need for larger patient groups for higher statistical power. In light of this, the smaller patient group (n=3) analysed for *O*-glycans, which is half the size of cohort analysed for *N*-glycans (n=6) may contribute to the low number of *O*-glycans found to be statistically differentially expressed. Despite this, the findings presented suggests that quantitative glycan changes are relatively common in tissue systems, and that these alterations are more highly nuanced than the overt changes shown in the comparison between cell lines and tumour tissues (Chapter 2).

Additionally, only a minority (3/12) of the differentially expressed *N*-glycans had abundances above 3% on average (glycan 65, 94 and 121) and some had average abundances of approximately 0.2% (in either normal or tumour tissue type). Compared to high-mannose glycans, such as $(Man)_9 (GlcNAc)_2$ which are expressed at >10% of the total glycans on average (Supp Table 2), differentially expressed glycans appear to be of lower abundance. This appears to follow the trends observed in proteomics, where candidate markers are often found in proteins expressed in low abundance [237]. Given that up to three parent precursor ions are selected for MS2 in this workflow, additional biomarkers may be present which remain uncharacterised. Nonetheless, this study indicates there may be subtle changes in glycosylation and its regulation, especially of the protein *N*-glycans. These glycan changes alone are potential biomarker candidates in CRC tissue, and may serve as molecular features which help confirm tissue malignancy. With additional study, it may also be determined whether these glycan structures are predictive for patient response to chemotherapy. It may also be possible to use these potential markers in conjunction with other known cancer associated molecular changes, such as an increase in carcinoembryonic antigen [238] in order to create a biomarker panel with increased sensitivity and specificity. Given the intricate regulation required for the synthesis of these glycans, there may also be functional roles for these glycans such as in the fine-tuning of cell signalling and cell adhesion, which remain to be further investigated.

Perspectives and limitations of current LC-MS/MS approach in glycan analysis

Although the overall glycan profiles obtained in this thesis are very comparable to previous studies using a similar approach [102, 122], it was noted that there were a small number of sulfated glycans which were detected previously that were not observed in this study. Examples include glycans with composition $(SO3^-)_1(Hex)_1(HexNAc)_2$ and $(SO3^-)_1 (Hex)_2(HexNAc)_3(Deoxyhexose)_2$, which correspond with [M-H]- ions of *m/z* 667 and 1324, respectively. Moreover, other studies analysing *O*-glycans

released from porcine stomach and intestinal mucin show a higher number of sulfated glycans than have been observed here [126, 239]. It is likely that differences in both sample preparation and mass spectrometry analysis methods underlie these differences, although cross-species differences cannot be excluded (in other words, porcine colonic mucins may be inherently more sulfated than human mucins). In terms of sample preparation, separation of glycans into neutral, acidic and sulfated fractions [239, 240] by anion exchange followed by analysis in a QTOF instrument [126], may facilitate more favourable detection of sulfated glycans. This is because lower mass diagnostic ions indicative of sulfate are more likely to be observed in a QTOF instrument than ion trap due to the so-called "one-third rule" of ion traps [241]. Examples of these diagnostic fragments include m/z 97 (HSO_4^-), 139 and 199 [126]. Other ions that indicate position of sulfate such as m/z 241 [Hex-SO3]- and m/z 282 [HexNAc-SO3]⁻, however, are compatible with observation under experimental condition used in this study (for example, see O-glycan 44). Overall, utilising a different experimental and mass spectrometric approach (such as the above) may increase the number of sulfated glycans identified, although it is unclear whether the fractionation method is feasible with biopsy-sized colonic tissue that have undergone epithelial enrichment as described here.

It is also acknowledged that the LC-MS/MS approach employed for glycan analysis (for both chapter 2 and 3) results in glycan structures with a degree of ambiguity. For example, the position of the fucose (in the non-reducing outer-arm antennae vs core) can be differentiated by presence of diagnostic fragments such as m/z 330/348/510 for outer-arm fucose and fragments with m/z 350/368/571 for core-fucose [139]. However, it is frequently unclear where the fucose is specifically positioned on the non-reducing antennae due to limited availability of diagnostic ions with lower m/z. Although the ion with m/z 510 corresponding to $(Hex)_1(HexNAc)_1(Fuc)_1$ is relatively common, other ions with lower m/z such as 348 and 325, that indicates fucose on GlcNAc and on the Gal respectively, are not as frequently observed. Therefore, this results in structural ambiguity where Lewis A/X (with fucose attached to GlcNAc) and H-epitopes (with fucose attached to Gal) cannot be clearly distinguished.

One approach that may overcome this limitation is the selection and subsequent fragmentation of the ion m/z 510 by MS3. Depending on the fragments generated (as mentioned above), the Lewis and H-epitopes may be differentiated. A complementary approach is the use of specific fucosidases such as AMF (almond meal α-fucosidase) that cleaves α1-3/4 fucose or BKF (bovine kidney α-fucosidase) that releases α(1-2/6) fucose from the non-reducing end [129]. Depending on the glycan, the digested products are expected to have reproducible retention time/mass shifts which give further information regarding linkage of the fucose. While it is clear that the above suggestions can

be carried out in the future direction of this project, given the time-consuming nature of glycan analysis and limited availability of tissues, these have not been performed in this preliminary study.

3.4   Summary and final remarks

In conclusion, *N*- and *O*-glycans from the membrane proteins of epithelial cells from patient-derived normal and colorectal cancer tissue have been profiled. The pair-matched normal and tumour *O*-glycan profiles of 3 patients and *N*-glycan profiles of 6 patients were obtained respectively. For *O*-glycans, sialyl-Tn and a core 3 sialyl-Lewis X hexasaccharide increased in the tumour tissue as previously reported. However, these changes were not statistically significant - most likely due to the small patient cohort and large variance amongst patients. Nevertheless, we did find a sialylated trisaccharide with composition $(HexNAc)_2 (NeuAc)_1$ as significantly decreased in the tumours, which has not been described before. In regards to *O*-glycan cores, an inverse relationship was observed between core 1 and core 3 with core 1 being increased in the tumours and vice versa for core 3. Overall, the *O*-glycan changes observed imply the regulation of C1GalT1/ST6GalNAc1 activity and a potential decrease in the enzymes involved in core 3/5/6/7 synthesis in the tumours.

On the other hand, paucimannosidic *N*-glycans were found to be increased, while bisecting *N*-glycans were found decreased in the tumours, which is in agreement with the previous protein *N*-glycan profile reported [38]. Monoantennary *N*-glycans were also significantly increased and biantennary *N*-glycans were significantly decreased in the tumours. These changes were reflected in the 12 differentially expressed *N*-glycans, which included 5 downregulated bisected and 4 monoantennary *N*-glycans. A paucimannose structure, $(Man)_3 (HexNAc)_2 (Deoxyhexose)_1$, was also upregulated in the tumours. All but one of these *N*-linked glycan structures are novel identified glycan changes. Their differential expression potentially implicates altered activity and regulation of relevant enzymes, such as Mgat3, α-mannosidase II and *N*-acetylglucosaminyltransferases (GlcNAcT IV and V). Overall in this study, glycan changes have been identified in CRC tissues which may be novel candidate diagnostic markers. These and the implicated changes to their underlying synthetic pathways also warrant further investigation as they may be targeted for therapeutic intervention. Furthermore, this study has also demonstrated a workflow whereby the membrane glycosylation of epithelial cells can be analysed. It is envisioned that this may be employed as a tool for glycan profiling of epithelial cells in other cancer types.

3.5 Supplemental information

Supplementary Tables (over page)

Note: Not all glycans from Chapter 2 were detected. T.b.c, to be confirmed.

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure(s)	P-value (T-test)	Fold change (T/N)	Normal (% area)	Tumour (% area)
1	513.3		513.3	0.106	$(HexNAc)_1 (NeuAc)_1$		0.1199	17.13	0.66 ± 0.52	11.34 ± 7.57
2	587.3		587.3	0.069	$(Hex)_1 (HexNAc)_2$		0.9019	1.22	1.08 ± 1.74	1.32 ± 1.74
3	587.3		587.3	0.069	$(Hex)_1 (HexNAc)_2$		0.2673	-11.40	1.04 ± 1.03	0.09 ± 0.08
4	675.3		675.3	0.053	$(Hex)_1 (HexNAc)_1 (NeuAc)_1$		0.0776	4.51	1.18 ± 0.74	5.32 ± 1.96
5	675.3		675.3	0.053	$(Hex)_1 (HexNAc)_1 (NeuAc)_1$		0.4946	3.55	0.51 ± 0.87	1.83 ± 2.34
6	716.2		716.2	-0.072	$(HexNAc)_2 (NeuAc)_1$		0.8693	-1.10	18.18 ± 5.18	16.60 ± 14.12
7	716.2		716.2	-0.072	$(HexNAc)_2 (NeuAc)_1$		0.0218	-2.90	12.06 ± 4.44	4.16 ± 2.62
8	716.2		716.2	-0.072	$(HexNAc)_2 (NeuAc)_1$		0.4226	43.29	0.01 ± 0.00	0.43 ± 0.73
9	749.2		749.2	-0.082	$(Hex)_2 (HexNAc)_2$		0.6579	2.03	0.11 ± 0.18	0.23 ± 0.29
10	829.3		829.3	0.06	$(Hex)_2 (HexNAc)_2 (Sulph)_1$		0.4226	11.54	0.01 ± 0.00	0.12 ± 0.18
11	878.4		878.4	0.074	$(Hex)_1 (HexNAc)_2 (NeuAc)_1$		0.5050	-2.88	1.30 ± 2.24	0.45 ± 0.47
12	878.4		878.4	0.074	$(Hex)_1 (HexNAc)_2 (NeuAc)_1$		0.4172	5.79	1.20 ± 2.06	6.96 ± 8.88
13	895.4		895.4	0.059	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1$		0.1964	-1.52	0.23 ± 0.21	0.15 ± 0.14

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure(s)	P-value (T-test)	Fold change (T/N)	Normal (% area)	Tumour (% area)
14		913	1827	0.331	(Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_2$ (NeuAc)$_2$		0.1856	3.99	0.28 ± 0.26	1.13 ± 0.78
15	966.4		966.4	0.058	(Hex)$_1$ (HexNAc)$_1$ (NeuAc)$_2$		0.1331	8.18	0.31 ± 0.34	2.57 ± 1.79
17	975.4		975.4	0.102	(Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (Sulph)$_1$		0.4702	6.57	0.40 ± 0.35	2.63 ± 4.03
19	975.4		975.4	0.102	(Hex)$_2$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (Sulph)$_1$		0.4226	67.45	0.01 ± 0.00	0.67 ± 1.15
20	1016.4		1016.4	0.075	(Hex)$_1$ (HexNAc)$_3$ (Deoxyhexose)$_1$ (Sulph)$_1$		0.6685	1.71	0.36 ± 0.32	0.62 ± 0.59
21	1024.5		1024.5	0.116	(Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$		0.2449	-3.82	11.50 ± 8.08	3.01 ± 4.21
22	1024.5		1024.5	0.116	(Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (NeuAc)$_1$		0.4226	9.72	0.12 ± 0.19	1.17 ± 2.00
23	1040.5	519.7	1040.5	0.121	(Hex)$_2$ (HexNAc)$_2$ (NeuAc)$_1$		0.2722	5.62	0.27 ± 0.30	1.54 ± 1.16
24	1040.5	519.7	1040.5	0.121	(Hex)$_2$ (HexNAc)$_2$ (NeuAc)$_1$		0.3041	42.74	0.01 ± 0.00	0.43 ± 0.53
26	1081.5		1081.5	0.095	(Hex)$_1$ (HexNAc)$_3$ (NeuAc)$_1$		0.4226	-2.76	1.32 ± 2.28	0.48 ± 0.82
27	1081.5		1081.5	0.095	(Hex)$_1$ (HexNAc)$_3$ (NeuAc)$_1$		0.4850	-6.37	1.81 ± 2.87	0.28 ± 0.30
28	1081.5		1081.5	0.095	(Hex)$_1$ (HexNAc)$_3$ (NeuAc)$_1$		0.4226	5.86	0.12 ± 0.19	0.69 ± 1.17
29	1104.4	551.8	1104.6	0.259	(Hex)1 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 (Sulph)1		0.1982	-5.60	10.81 ± 7.05	1.93 ± 1.27
31	1120.5		1120.5	0.164	(Hex)$_2$ (HexNAc)$_2$ (NeuAc)$_1$ (Sulph)$_1$		0.3381	2.98	0.15 ± 0.24	0.45 ± 0.66
32	1169.6	584.3	1169.6	0.179	(Hex)$_1$ (HexNAc)$_2$ (NeuAc)$_2$		0.8850	-1.07	2.21 ± 2.97	2.06 ± 2.25

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure(s)	P-value (T-test)	Fold change (T/N)	Normal (% area)	Tumour (% area)
33	1186.4		1186.4	-0.036	(Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁		0.4226	54.07	0.01 ± 0.00	0.54 ± 0.92
34	1186.4		1186.4	-0.036	(Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁		0.1328	185.50	0.01 ± 0.00	1.85 ± 1.30
35	1227.7		1227.7	0.237	(Hex)₁ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₁		0.2130	100.72	0.01 ± 0.00	1.01 ± 0.96
37	1266.4	632.7	1266.4	0.006	(Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₁ (Sulph)₁		0.2185	7.52	0.16 ± 0.18	1.21 ± 0.92
38	1266.4	632.7	1266.4	0.006	(Hex)₂ (HexNAc)₂ (Deoxyhexose)₂ (NeuAc)₁ (Sulph)₁		0.2355	55.42	0.01 ± 0.00	0.55 ± 0.56
39	1315.7	657.4	1315.7	0.221	(Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₂		0.7494	1.34	4.55 ± 4.76	6.12 ± 5.35
40	1331.6	665.3	1331.6	0.126	(Hex)₂ (HexNAc)₂ (NeuAc)₂		0.3577	84.63	0.01 ± 0.00	0.85 ± 1.22
41	1372.6	685.9	1372.6	0.099	(Hex)₁ (HexNAc)₃ (NeuAc)₂		0.3207	-3.12	9.43 ± 10.25	3.02 ± 2.18
42	1477.6	738.3	1477.6	0.068	(Hex)₂ (HexNAc)₂ (Deoxyhexose)₂ (NeuAc)₂		0.4518	12.47	0.49 ± 0.59	6.11 ± 10.33
43		840	1681	0.388	(Hex)₂ (HexNAc)₃ (Deoxyhexose)₁ (NeuAc)₂		0.0810	3.01	0.39 ± 0.43	1.17 ± 0.60
44	813.3		813.3	0.055	Hex)₁ (HexNAc)₂ (Deoxyhexose)₁ (Sulph)₁	S⁴, or S⁶	0.2054	-9.03	4.59 ± 3.16	0.51 ± 0.68
45	879.5		879.5	0.154	(Hex)₁ (HexNAc)₂ (Deoxyhexose)₂		0.4226	-7.15	0.39 ± 0.65	0.05 ± 0.08
46	879.5		879.5	0.154	(Hex)₁ (HexNAc)₂ (Deoxyhexose)₂		0.4226	-100.81	1.01 ± 1.73	0.01 ± 0.00
47		868.4	1737.8	0.167	(Hex)₂ (HexNAc)₄ (NeuAc)₂	T.b.c.	0.9927	1.02	2.12 ± 3.09	2.16 ± 3.68
48		778.3	1557.6	0.111	(Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₂ (Sulph)₁	T.b.c.	0.4139	36.00	0.11 ± 0.17	3.94 ± 6.67

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure(s)	P-value (T-test)	Fold change (T/N)	Normal (% area)	Tumour (% area)
49	667		667	0.013	$(Hex)_1 (HexNAc)_2 (Sulph)_1$		0.4226	9.86	0.01 ± 0.00	0.10 ± 0.15
50	958.3		958.3	0.017	$(Hex)_1 (HexNAc)_2 (NeuAc)_1 (Sulph)_1$	T.b.c.	0.4226	3.89	0.39 ± 0.65	1.50 ± 2.58
51	1014		1014	0.272	$(Hex)_2 (HexNAc)_4 (NeuAc)_3$	T.b.c.	0.3397	-18.50	3.02 ± 4.24	0.16 ± 0.27
52	1170.5	585.3	1170.5	0.058	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_2 (NeuAc)_1$		0.4226	-13.25	6.47 ± 11.19	0.49 ± 0.83
53		1095	2191	0.219	$(Hex)_3 (HexNAc)_4 (NeuAc)_3$	T.b.c.	0.4226	45.01	0.01 ± 0.00	0.45 ± 0.76

Supp Table 2. Relative quantification of *N*-glycans of membrane proteins from normal and patient-matched tumour colon tissues (n=6). Glycans have been numbered according to Chapter 2 (except glycans not previous characterised, see Supp Fig 2). For statistical analysis, a value of 0.01 has been added to percentage areas to account for glycans below detection. Values are expressed as mean percentage area-under-the-curve ± SD. P-value calculated using two-tailed paired student's t-test. A p-value < 0.05 was required for statistical significance.

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure	P-value (T-test)	Fold change (T/N)	Normal (% Area)	Tumour (% Area)
1	895.5		895.5	0.159	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1$		0.1454	1.66	0.49 ± 0.39	0.81 ± 0.40
3	1057.4		1057.4	0.006	$(Hex)_3 (HexNAc)_2 (Deoxyhexose)_1$		0.0236	2.44	0.77 ± 0.86	1.88 ± 1.51
4	1235.6		1235.6	0.158	$(Hex)_2 + (Man)_3 (GlcNAc)_2$		0.1466	1.29	3.15 ± 2.34	4.05 ± 2.18
5	1397.6	698.3	1397.6	0.105	$(Hex)_3 + (Man)_3 (GlcNAc)_2$		0.8226	1.05	2.90 ± 1.59	3.04 ± 1.27
6		698.3	1397.6	0.105	$(Hex)_3 + (Man)_3 (GlcNAc)_2$		0.3632	12.84	0.01 ± 0.00	0.13 ± 0.29
11		731.3	1463.6	0.047	$(HexNAc)_2 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.0956	1.70	0.75 ± 0.67	1.27 ± 0.96
14		759.9	1520.8	0.226	$(HexNAc)_3 + (Man)_3 (GlcNAc)_2$		0.0963	-1.27	1.34 ± 0.96	1.05 ± 0.93
15		779.4	1559.8	0.253	$(Hex)_4 + (Man)_3 (GlcNAc)_2$		0.1265	2.18	0.94 ± 0.53	2.04 ± 1.40
16		779.4	1559.8	0.253	$(Hex)_4 + (Man)_3 (GlcNAc)_2$		0.0802	1.42	3.12 ± 2.37	4.41 ± 2.30
17		783.5	1568	0.436	$(Hex)_1 (HexNAc)_1 (NeuAc)_1 + (Man)_3 (GlcNAc)_2$		0.0239	2.27	0.44 ± 0.25	0.99 ± 0.39
20		791.8	1584.6	0.021	$(Hex)_2 (HexNAc)_1 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.1907	1.58	0.10 ± 0.14	0.16 ± 0.14
22		799.8	1600.6	0.026	$(Hex)_3 (HexNAc)_1 + (Man)_3 (GlcNAc)_2$		0.4120	1.23	0.44 ± 0.16	0.54 ± 0.21
26		812.3	1625.6	-0.005	$(Hex)_1 (HexNAc)_2 (Deoxyhexose)_1 + (Man)_3 (GlcNAc)_2$		0.7290	1.10	0.73 ± 0.53	0.81 ± 0.41
27		820.4	1641.8	0.199	$(Hex)_2 (HexNAc)_2 + (Man)_3 (GlcNAc)_2$		0.8482	1.08	0.48 ± 0.40	0.52 ± 0.46

Supp Table 2 (cont).

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure	P-value (T-test)	Fold change (T/N)	Normal (% Area)	Tumour (% Area)
28		820.4	1641.8	0.199	$(Hex)_2(HexNAc)_2 + (Man)_3(GlcNAc)_2$		0.6856	1.16	0.55 ± 0.60	0.64 ± 0.84
29		820.4	1641.8	0.199	$(Hex)_2(HexNAc)_2 + (Man)_3(GlcNAc)_2$		0.8885	1.02	0.74 ± 0.48	0.76 ± 0.59
30		832.8	1666.6	-0.031	$(HexNAc)_3(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.5735	-1.29	3.17 ± 3.19	2.45 ± 2.99
31		832.8	1666.6	-0.031	$(HexNAc)_3(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.0691	18.54	0.05 ± 0.11	1.00 ± 0.98
34		840.9	1682.8	0.173	$(Hex)_1(HexNAc)_3 + (Man)_3(GlcNAc)_2$		0.2651	-1.12	2.33 ± 3.67	2.09 ± 4.05
36		856.3	1713.6	-0.021	$(Hex)_1(HexNAc)_1(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.0445	3.18	0.07 ± 0.05	0.21 ± 0.15
37		856.4	1713.8	0.178	$(Hex)_1(HexNAc)_1(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.0207	4.03	0.41 ± 0.37	1.67 ± 1.14
39		860.3	1721.6	0	$(Hex)_5 + (Man)_3(GlcNAc)_2$		0.2135	1.45	6.03 ± 4.28	8.77 ± 4.02
42		864.4	1729.8	0.183	$(Hex)_2(HexNAc)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.2815	1.32	0.92 ± 0.42	1.21 ± 0.29
45		872.9	1746.8	0.168	$(Hex)_3(HexNAc)_1(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.8503	-1.06	0.51 ± 0.30	0.48 ± 0.35
48		885.4	1771.8	0.136	$(Hex)_1(HexNAc)_2(Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.3056	2.28	0.45 ± 0.68	1.03 ± 1.07
49		885.4	1771.8	0.136	$(Hex)_1(HexNAc)_2(Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.0595	44.34	0.01 ± 0.00	0.44 ± 0.44
54		893.4	1787.8	0.141	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.1798	-2.11	1.84 ± 2.11	0.87 ± 1.13
55		893.4	1787.8	0.141	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.2013	2.70	0.12 ± 0.18	0.33 ± 0.45
56		893.4	1787.8	0.141	$(Hex)_2(HexNAc)_1(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.0732	-2.13	3.36 ± 3.26	1.58 ± 1.44
63		913.9	1828.8	0.115	$(Hex)_1(HexNAc)_3(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.7656	1.29	0.69 ± 0.98	0.89 ± 0.94

Supp Table 2 (cont).

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure	P-value (T-test)	Fold change (T/N)	Normal (% Area)	Tumour (% Area)
65		913.9	1828.8	0.115	(Hex)$_1$(HexNAc)$_3$(Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.0161	-3.31	4.33 ± 2.49	1.31 ± 1.45
66		921.9	1844.8	0.12	(Hex)$_2$(HexNAc)$_3$ + (Man)$_3$(GlcNAc)$_2$		0.7999	1.04	7.21 ± 15.10	7.50 ± 17.68
69		937.3	1875.6	-0.074	(Hex)$_2$(HexNAc)$_1$(Deoxyhexose)$_1$(NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.0993	2.40	0.19 ± 0.20	0.46 ± 0.21
72		941.4	1883.8	0.147	(Hex)$_6$ + (Man)$_3$(GlcNAc)$_2$		0.0970	1.45	10.28 ± 6.24	14.86 ± 6.58
73		941.4	1883.8	0.147	(Hex)$_6$ + (Man)$_3$(GlcNAc)$_2$		0.1611	2.02	0.47 ± 0.49	0.96 ± 0.98
74		945.4	1891.8	0.131	(Hex)$_3$(HexNAc)$_1$(NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.1032	1.38	1.19 ± 0.54	1.64 ± 0.65
79		957.8		-0.1	(Hex)$_1$(HexNAc)$_2$(Deoxyhexose)$_1$(NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.0758	2.76	0.20 ± 0.26	0.55 ± 0.56
81		965.9	1932.8	0.104	(Hex)$_2$(HexNAc)$_2$(NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.2778	-1.49	2.83 ± 1.53	1.90 ± 0.50
82		965.9	1932.8	0.104	(Hex)$_2$(HexNAc)$_2$(NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.6441	-1.07	0.80 ± 0.62	0.75 ± 0.69
84		966.5	1934	0.284	(Hex)$_2$(HexNAc)$_2$(Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$		0.2275	-2.33	1.38 ± 2.01	0.59 ± 0.70
91		986.9	1974.8	0.057	(Hex)$_1$(HexNAc)$_3$(Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$		0.6169	-1.59	1.57 ± 2.37	0.99 ± 1.12
92		986.9	1974.8	0.057	(Hex)$_1$(HexNAc)$_3$(Deoxyhexose)$_3$ + (Man)$_3$(GlcNAc)$_2$		0.7612	1.28	0.66 ± 0.56	0.85 ± 0.94
93		994.9	1990.8	0.062	(Hex)$_2$(HexNAc)$_3$(Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.0933	-7.99	0.35 ± 0.38	0.04 ± 0.06
94		994.9	1990.8	0.062	(Hex)$_2$(HexNAc)$_3$(Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.0093	-7.44	6.84 ± 4.17	0.92 ± 0.82
95		994.9	1990.8	0.062	(Hex)$_2$(HexNAc)$_3$(Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.3632	-1.78	0.19 ± 0.43	0.11 ± 0.23
96		994.9	1990.8	0.062	(Hex)$_2$(HexNAc)$_3$(Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$		0.4471	-1.72	0.32 ± 0.46	0.19 ± 0.10

Supp Table 2 (cont).

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure	P-value (T-test)	Fold change (T/N)	Normal (% Area)	Tumour (% Area)
98		1007.5	2016	0.23	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_2 (Sulph)_1 + (Man)_3(GlcNAc)_2$		0.8321	1.20	0.22 ± 0.33	0.26 ± 0.40
100		1018.4	2037.8	0.073	$(Hex)_3 (HexNAc)_1 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.4497	1.79	0.07 ± 0.09	0.12 ± 0.13
101		1022.4	2045.8	0.094	$(Hex)_7 + (Man)_3(GlcNAc)_2$		0.1285	1.41	0.39 ± 0.22	0.55 ± 0.19
103		1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.4212	-1.65	0.34 ± 0.34	0.21 ± 0.20
105		1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.3641	1.29	1.06 ± 0.57	1.36 ± 0.76
107		1038.9	2078.8	0.046	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.3194	-1.32	1.53 ± 0.62	1.16 ± 0.55
108		1039.5	2080	0.226	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.6288	1.23	0.63 ± 0.63	0.77 ± 1.13
119		1060	2121	0.199	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$	T.b.c.	0.4459	-1.57	1.10 ± 1.29	0.70 ± 0.55
120		1068	2137	0.204	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.0483	-2.77	0.23 ± 0.22	0.08 ± 0.10
121		1068	2137	0.204	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_2 + (Man)_3(GlcNAc)_2$		0.0252	-12.97	5.46 ± 4.45	0.42 ± 0.64
123		1075.9	2152.8	0.009	$(Hex)_3 (HexNAc)_3 (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.5272	-1.35	0.38 ± 0.32	0.28 ± 0.14
126		1080.6	2162.2	0.372	$(HexNAc)_4 (Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.6882	-1.14	1.15 ± 0.72	1.01 ± 1.02
130		1111.5	2224	0.209	$(Hex)_2 (HexNAc)_2 (NeuAc)_2 + (Man)_3(GlcNAc)_2$		0.3089	-1.24	4.34 ± 1.50	3.50 ± 1.12
132		1111.9	2224.8	-0.011	$(Hex)_2 (HexNAc)_2 (Deoxyhexose)_2 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.1782	2.39	0.42 ± 0.52	1.00 ± 1.16
137		1132.4	2265.8	-0.037	$(Hex)_1 (HexNAc)_3 (Deoxyhexose)_2 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.6224	1.64	0.18 ± 0.34	0.30 ± 0.46
138		1140.4	2281.8	-0.032	$(Hex)_2 (HexNAc)_3 (Deoxyhexose)_1 (NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.0077	-2.51	0.47 ± 0.40	0.19 ± 0.25

Glycan #	$[M-H]^-_{exp}$	$[M-2H]^{2-}_{exp}$	$[M-H]^-_{calc}$	$\Delta[M-H]^-$	Composition	Proposed structure	P-value (T-test)	Fold change (T/N)	Normal (% Area)	Tumour (% Area)
140		1140.4	2281.8	-0.032	$(Hex)_2(HexNAc)_3(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.5554	-1.58	0.23 ± 0.32	0.15 ± 0.17
143		1141	2283	0.146	$(Hex)_2(HexNAc)_3(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.0139	-3.34	1.57 ± 1.20	0.47 ± 0.69
144		1141	2283	0.146	$(Hex)_2(HexNAc)_3(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.6810	1.52	0.18 ± 0.42	0.28 ± 0.33
145		1141	2283	0.146	$(Hex)_2(HexNAc)_3(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.1884	18.19	0.01 ± 0.00	0.18 ± 0.28
152		1177.5	2356	0.13	$(Hex)_3(HexNAc)_4(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.0455	-7.00	0.49 ± 0.45	0.07 ± 0.07
153		1184.5	2370	0.151	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_2 + (Man)_3(GlcNAc)_2$		0.4520	1.34	0.47 ± 0.43	0.63 ± 0.30
154		1184.5	2370	0.151	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_2 + (Man)_3(GlcNAc)_2$		0.5416	-1.35	0.39 ± 0.43	0.29 ± 0.08
155		1184.5	2370	0.151	$(Hex)_2(HexNAc)_2(Deoxyhexose)_1(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.5657	-1.26	0.53 ± 0.38	0.42 ± 0.15
157		1205.6	2412.2	0.304	$(Hex)_1(HexNAc)_3(Deoxyhexose)_3(NeuAc)_1 + (Man)_3(GlcNAc)_2$	T.b.c.	0.9950	1.00	0.47 ± 0.81	0.48 ± 0.56
158		1213.6	2428.2	0.309	$(Hex)_2(HexNAc)_3(Deoxyhexose)_2(NeuAc)_1 + (Man)_3(GlcNAc)_2$		0.7888	-1.14	0.60 ± 0.49	0.53 ± 0.48
163		1222	2445	0.093	$(Hex)_3(HexNAc)_3(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$		0.1295	9.62	0.01 ± 0.00	0.10 ± 0.12
164		1225	2451	0.173	$(Hex)_2(HexNAc)_2(Deoxyhexose)_3(NeuAc)_1(Sulph)_1 + (Man)_3(GlcNAc)_2$	T.b.c.	0.4535	2.26	0.07 ± 0.15	0.16 ± 0.23
165		1257.5	2516	0.093	$(Hex)_2(HexNAc)_2(Deoxyhexose)_2(NeuAc)_2 + (Man)_3(GlcNAc)_2$	T.b.c.	0.4132	2.99	0.03 ± 0.06	0.10 ± 0.17
167		1258.5	2518	0.077	$(Hex)_4(HexNAc)_4(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		0.1334	4.02	0.04 ± 0.08	0.16 ± 0.14
170		1330.6	2662.2	0.235	$(Hex)_2(HexNAc)_2(Deoxyhexose)_3(NeuAc)_2 + (Man)_3(GlcNAc)_2$		0.3230	1.49	0.16 ± 0.33	0.24 ± 0.31
173		1367		0.018	$(Hex)_3(HexNAc)_3(Deoxyhexose)_1(NeuAc)_2 + (Man)_3(GlcNAc)_2$		0.3632	1.89	0.01 ± 0.00	0.02 ± 0.02

Supp Table 2 (cont).

Glycan #	[M − H]⁻_exp	[M − 2H]²⁻_exp	[M − H]⁻_calc	Δ[M − H]⁻	Composition	Proposed structure	P-value (T-test)	Fold change (T/N)	Normal (% Area)	Tumour (% Area)
177		1440	2881	-0.038	(Hex)$_3$ (HexNAc)$_3$ (Deoxyhexose)$_2$ (NeuAc)$_2$ + (Man)$_3$(GlcNAc)$_2$	T.b.c.	0.4295	-1.45	0.79 ± 1.10	0.55 ± 0.48
180		759.9	1520.8	0.226	(HexNAc)$_3$ + (Man)$_3$(GlcNAc)$_2$	(structure diagram)	0.2372	5.89	0.03 ± 0.04	0.16 ± 0.23
181		783.7	1568.4	-0.183	(Hex)$_1$ (HexNAc)$_1$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$	(structure diagram)	0.0836	5.13	0.04 ± 0.06	0.18 ± 0.21
182		1088.5		-0.021	Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_2$ + (Man)$_3$(GlcNAc)$_2$	T.b.c.	0.1963	73.53	0.01 ± 0.00	0.74 ± 1.19
183		934.4		0.088	(HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$ (GlcNAc)$_2$	(structure diagram)	0.0917	5.44	0.24 ± 0.35	1.33 ± 1.46
184		1015.4		0.035	(Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$	(structure diagram)	0.6987	-1.28	0.23 ± 0.34	0.18 ± 0.20
185		1015.4		0.035	(Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$	(structure diagram)	0.3369	3.90	0.07 ± 0.15	0.28 ± 0.46
186		1015.4		0.035	(Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$	(structure diagram)	0.2751	37.12	0.01 ± 0.00	0.37 ± 0.72
187		1161		-0.14	(Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	(structure diagram)	0.0823	5.90	0.05 ± 0.09	0.28 ± 0.28
188		1161		-0.14	(Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	T.b.c.	0.5860	1.89	0.04 ± 0.07	0.07 ± 0.11
189	1260.5		1260.5	0.026	(HexNAc)$_1$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$	(structure diagram)	0.0295	2.56	0.23 ± 0.41	0.59 ± 0.63
190		1286.6		0.251	(Hex)$_2$ (HexNAc)$_3$ (Deoxyhexose)$_3$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$	T.b.c.	0.7849	1.36	0.17 ± 0.30	0.23 ± 0.35

Supplementary Figure 1. Annotated MS/MS for *O*-glycans from tissue samples not previous observed in the cell line and tumour tissue comparison (Chapter 2). Glycans have been numbered according to Supplementary Table 1 (Chapter 3). Mass spectra corresponding to glycans with only compositional information is not shown. Glycan schemes were derived from GlycoWorkbench. Additional structural isomers cannot be excluded. Annotation was based on the presence of structural feature ions and common knowledge of known glycan synthetic pathways. Structural schemes not intended to show specific linkages (linkages not shown by linkage angles). Where ambiguity is indicated in parent structure, for example in the assignment of fucose, depiction of specific linkages are not intended for corresponding glycan fragments.

Glycan 44   *m/z* 813.2⁻
Composition: Hex)$_1$ (HexNAc)$_2$ (Deoxyhexose)$_1$ (Sulph)$_1$

Note: Two structures proposed since there were insufficient diagnostic fragments to determine whether core 1 or core 3.
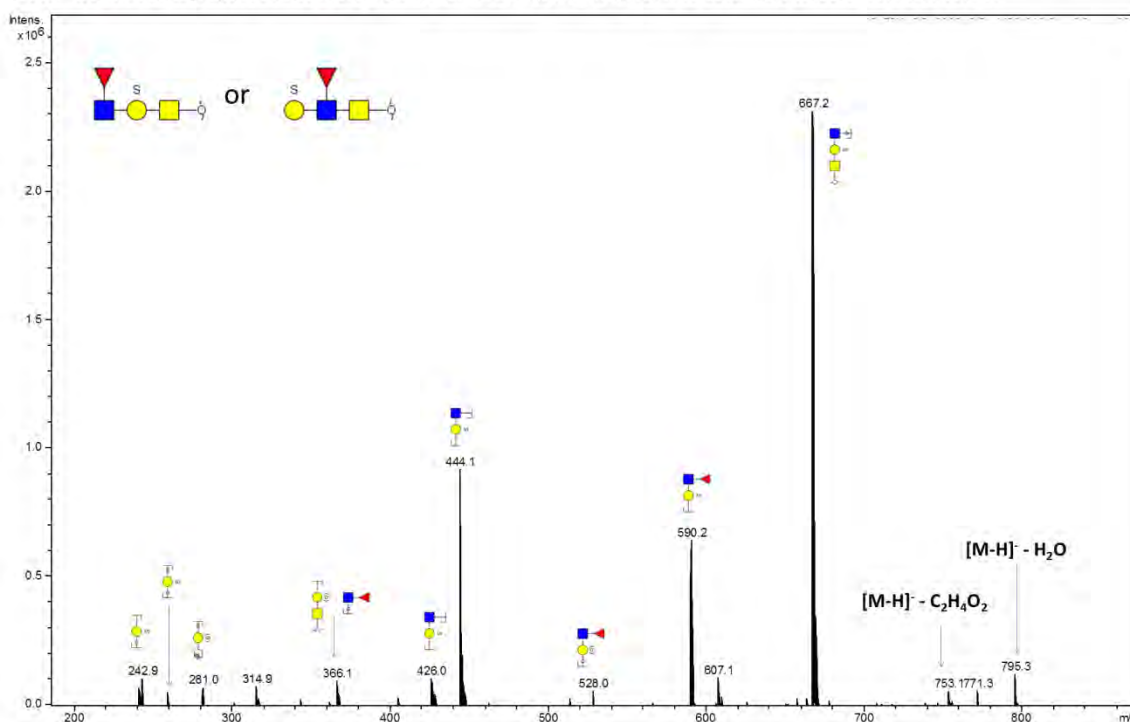
Glycan 45

m/z 879.5⁻
Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₂
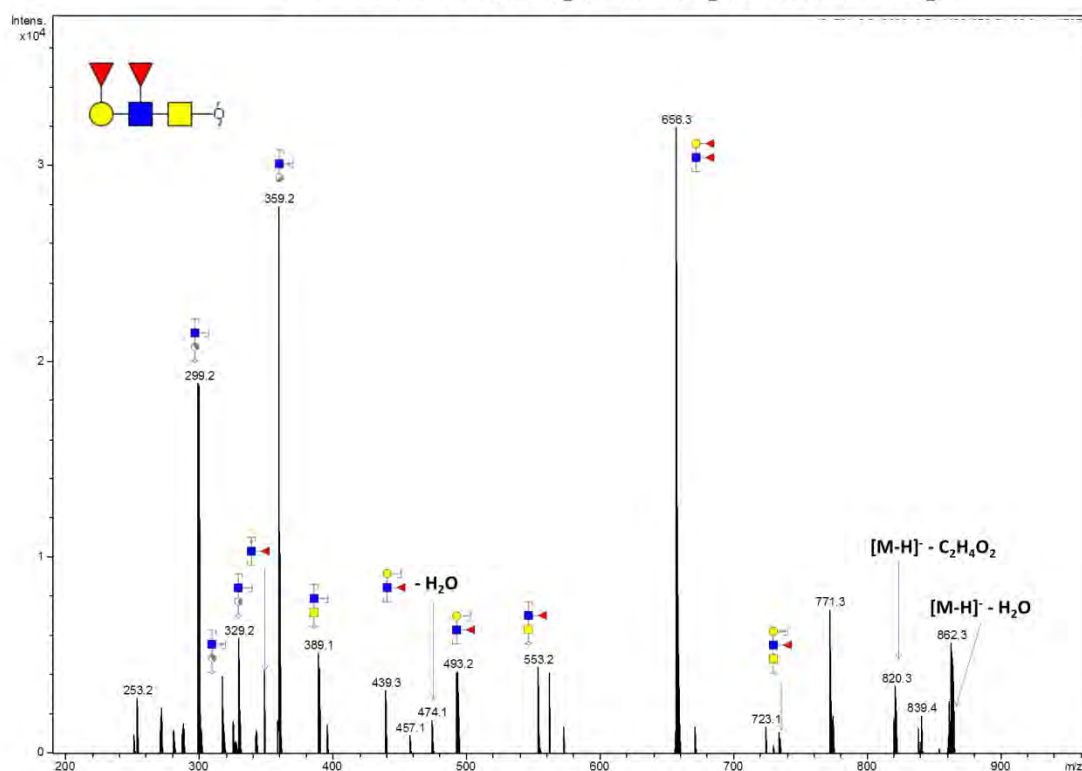
Glycan 46

m/z 879.5⁻
Composition: (Hex)₁ (HexNAc)₂ (Deoxyhexose)₂
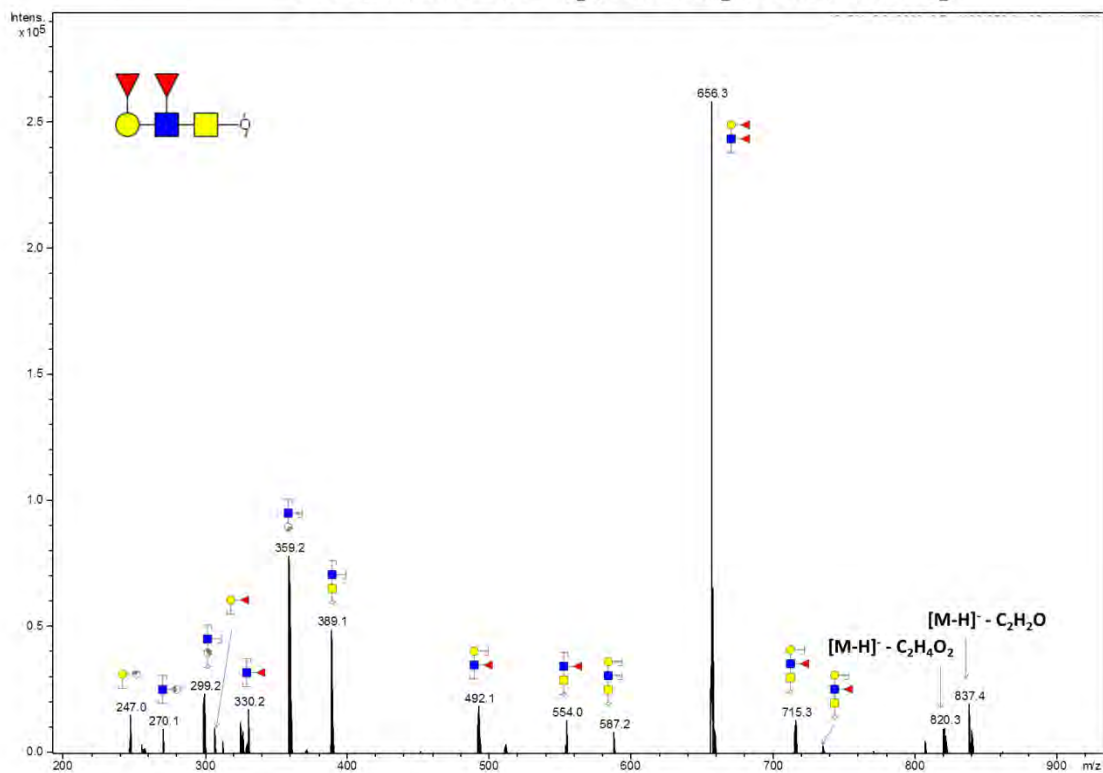
Glycan 49 *m/z* 667.0⁻

Composition: $(Hex)_1 (HexNAc)_2 (Sulph)_1$
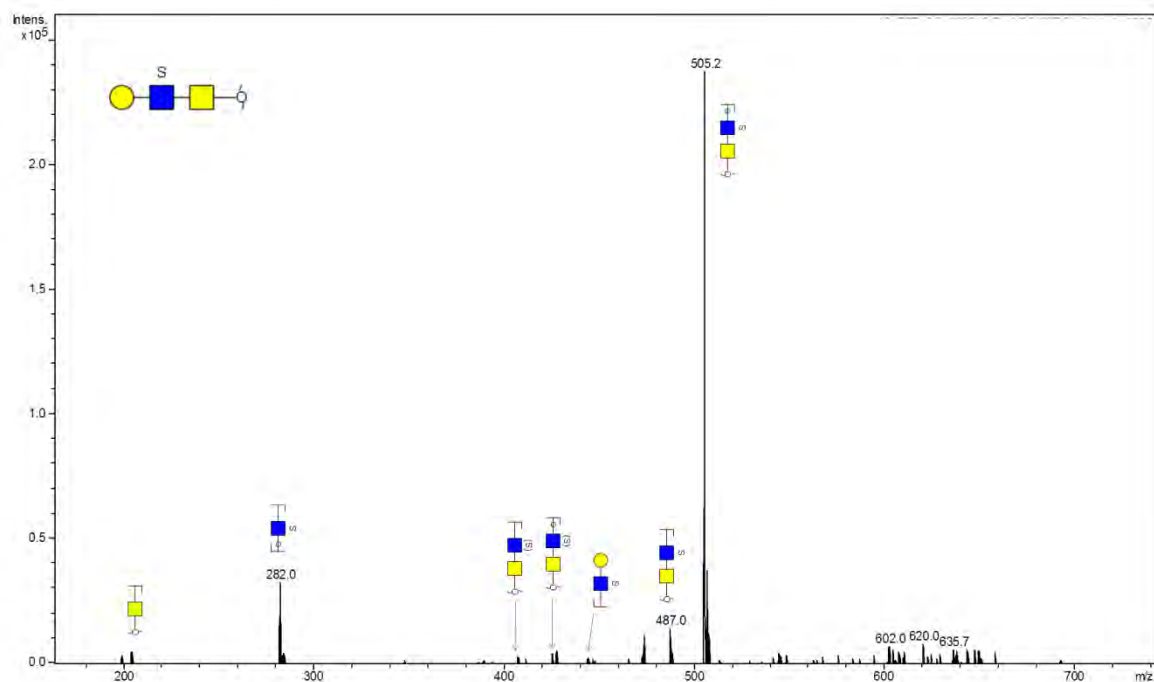
Glycan 52 *m/z* 1170.5⁻

Composition: $(Hex)_1 (HexNAc)_2 (Deoxyhexose)_2 (NeuAc)_1$

261

Supplementary Figure 2. **Annotated MS/MS for *N*-glycans from tissue samples not previous observed in the cell line and tumour tissue comparison (Chapter 2).** Glycans have been numbered according to Supplementary Table 2 (Chapter 3). Mass spectra corresponding to glycans with only compositional information is not shown. Glycan schemes were derived from GlycoWorkbench. Additional structural isomers cannot be excluded. Annotation was based on the presence of structural feature ions and common knowledge of known glycan synthetic pathways. Structural schemes not intended to show specific linkages (linkages not shown by linkage angles) or distinction between 3-arm and 6-arm. Where ambiguity is indicated in parent structure, for example in the assignment of fucose, depiction of specific linkages are not intended for corresponding glycan fragments. Where appropriate, diagnostic ion for bisecting GlcNAc has been boxed.

Glycan 180 *m/z* 759.9^{2-}
Composition: (HexNAc)$_3$ + (Man)$_3$(GlcNAc)$_2$
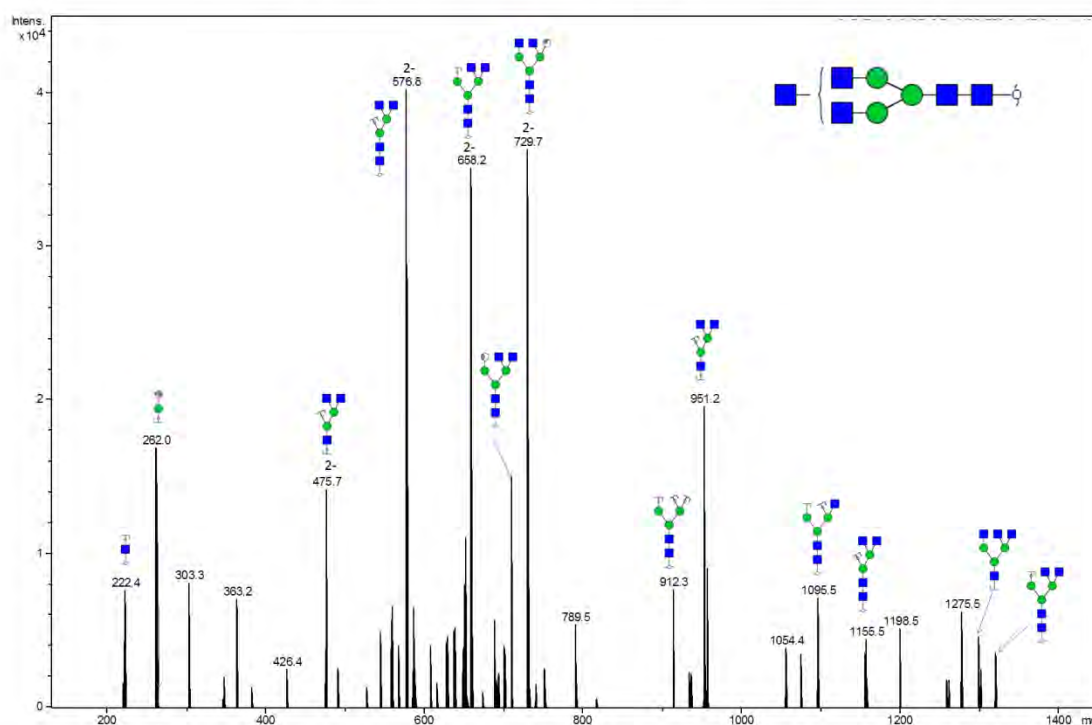
Glycan 181

m/z 783.8²⁻

Composition: (Hex)₁ (HexNAc)₁ (Deoxyhexose)₂ + (Man)₃(GlcNAc)₂
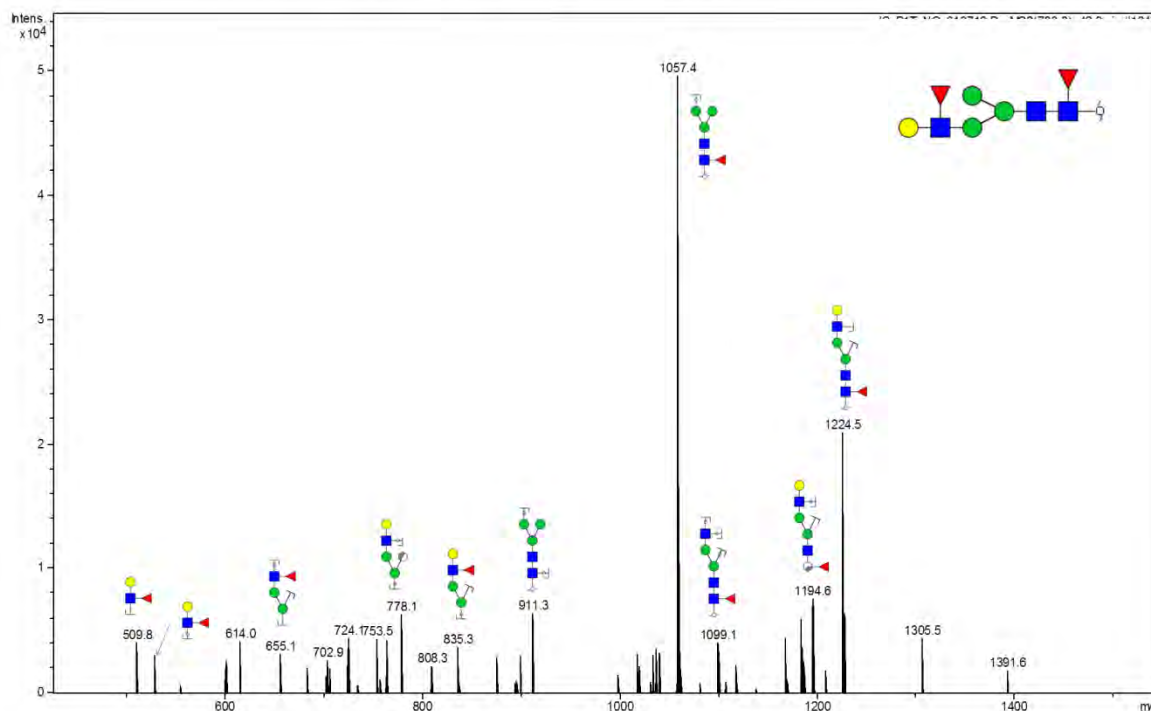
Glycan 183

m/z 934.4²⁻

Composition: (HexNAc)₄ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
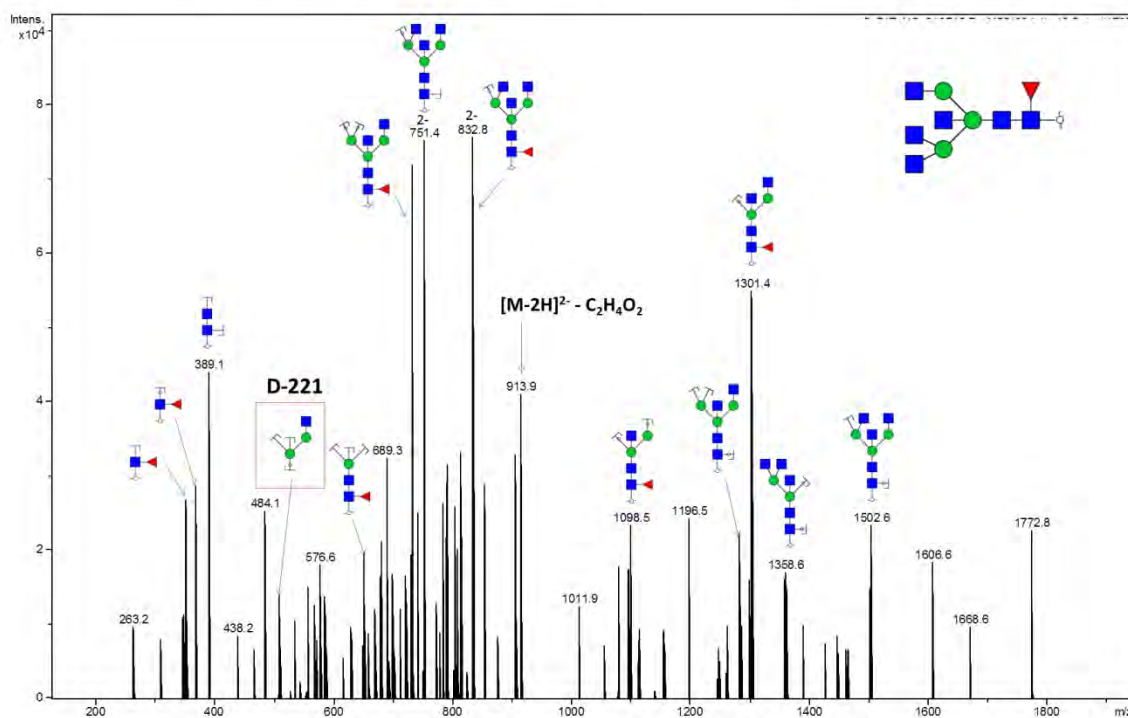
Glycan 184

m/z 1015.4^{2-}

Composition: (Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
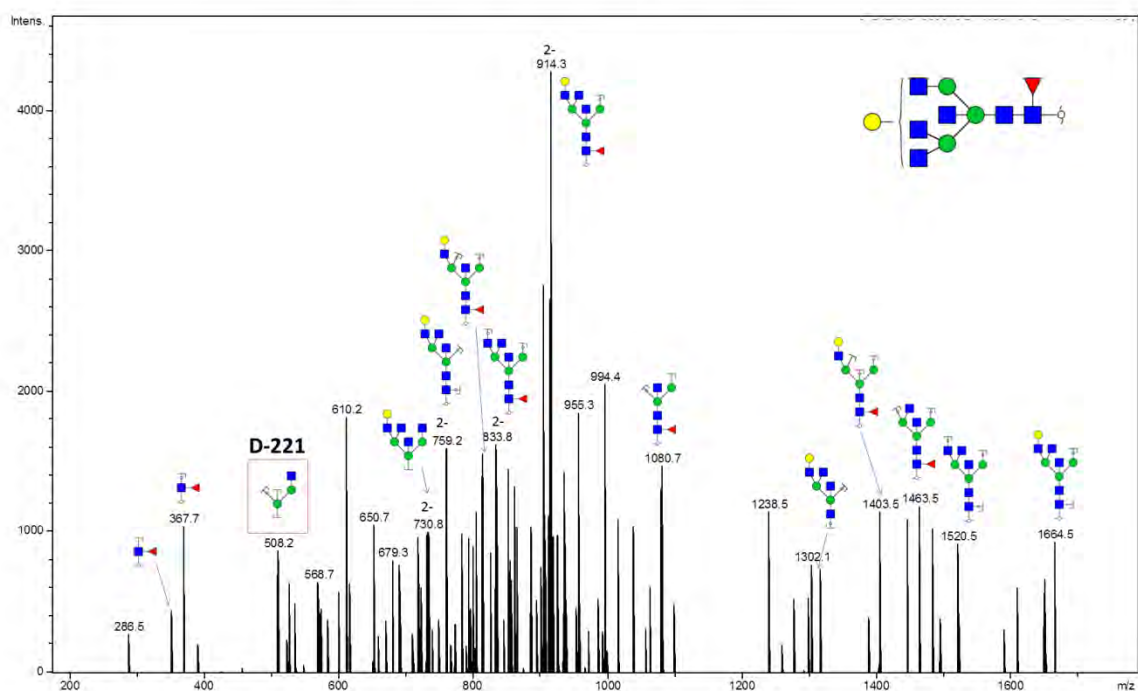
Glycan 185

m/z 1015.4^{2-}

Composition: (Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
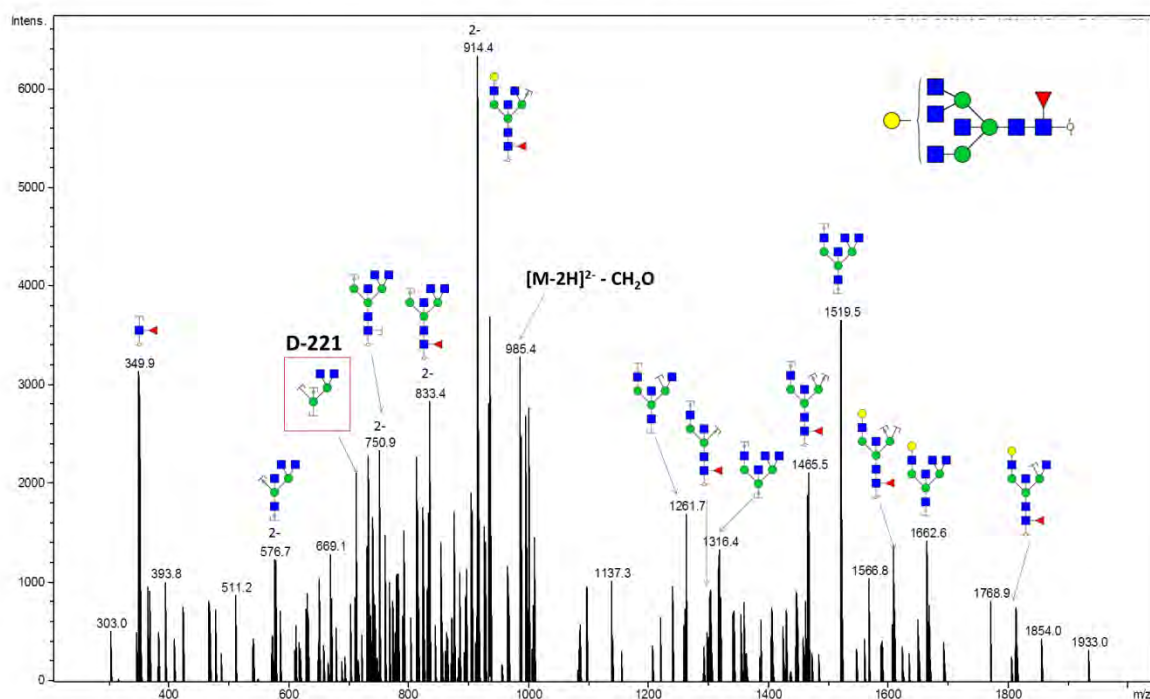
Glycan 186

m/z 1015.4^{2-}

Composition: (Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ + (Man)$_3$(GlcNAc)$_2$
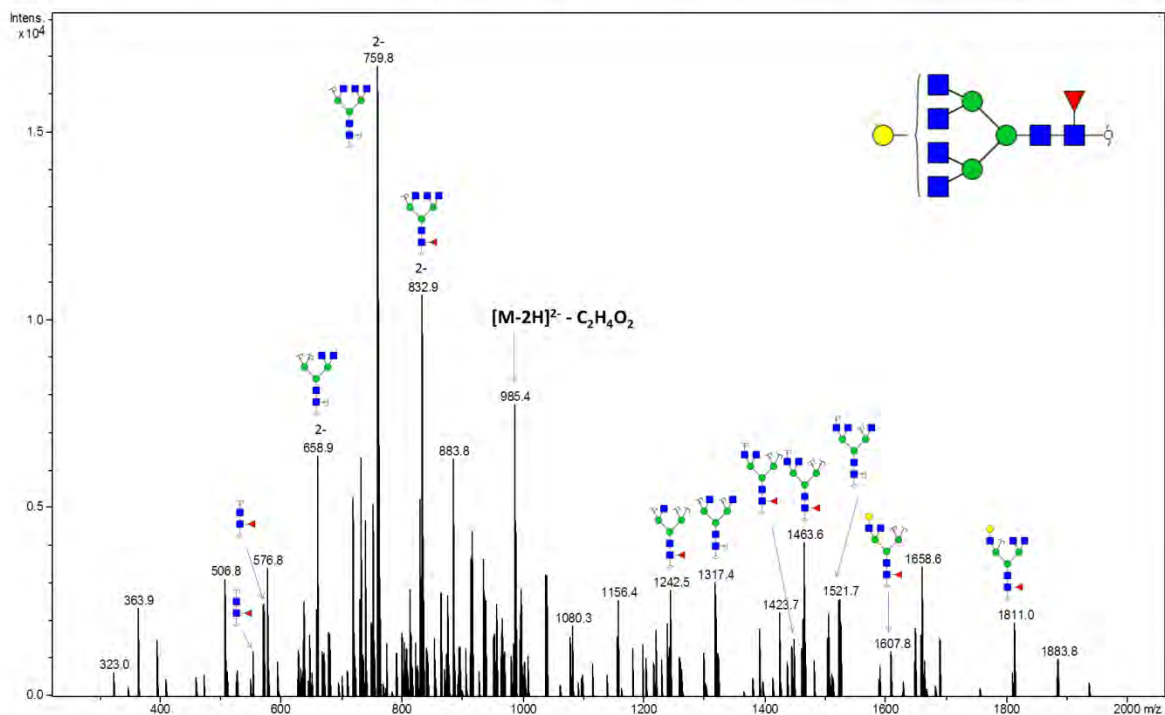
Glycan 187

m/z 2323^{2-}

Composition: (Hex)$_1$ (HexNAc)$_4$ (Deoxyhexose)$_1$ (NeuAc)$_1$ + (Man)$_3$(GlcNAc)$_2$
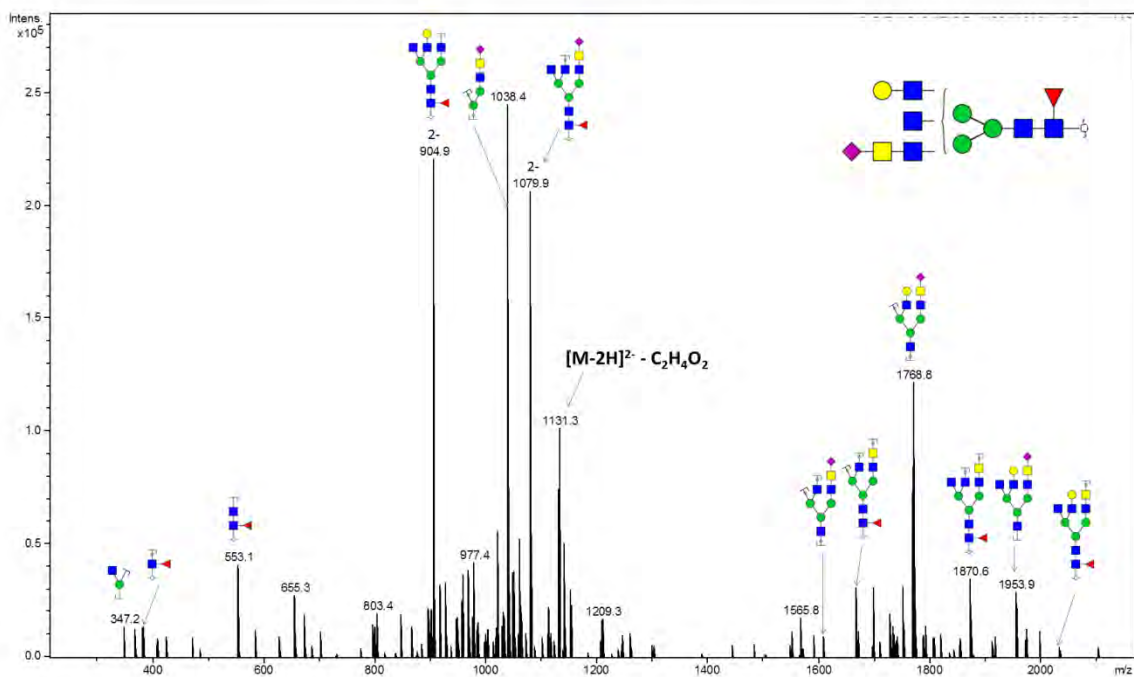
Glycan 189

m/z 1260.5⁻

Composition: (HexNAc)₁ (Deoxyhexose)₁ + (Man)₃(GlcNAc)₂
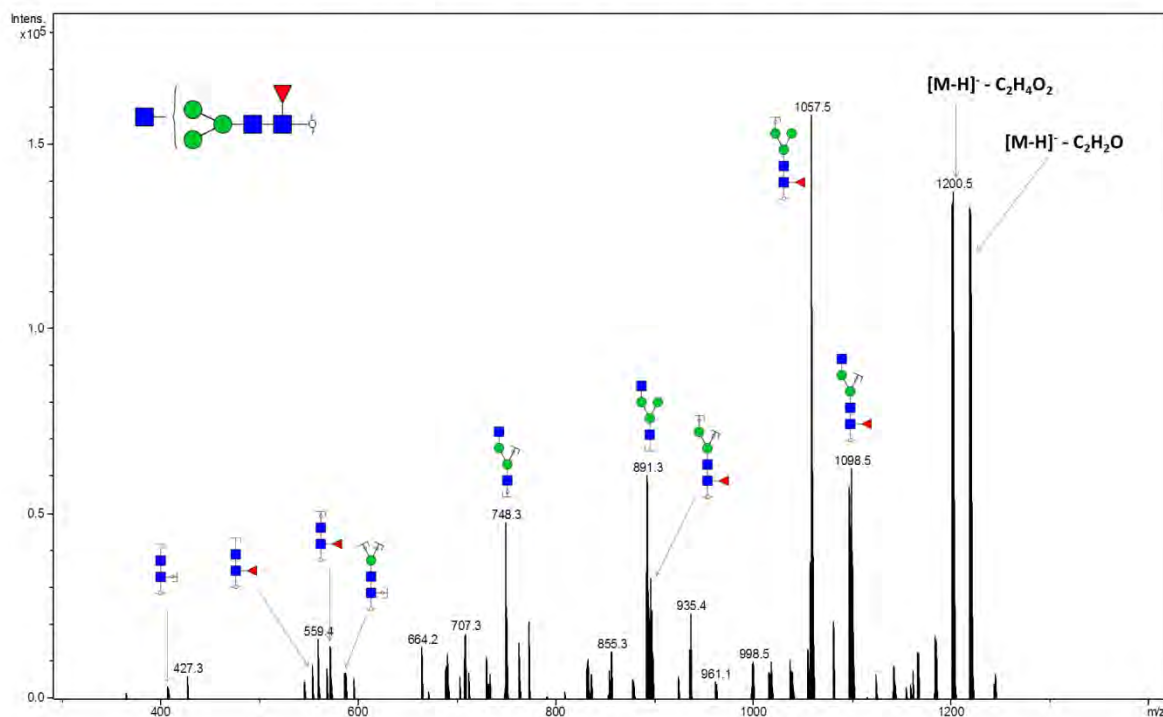
Chapter 4

Membrane proteomics and *N*-glycosylation site mapping of patient-matched normal and tumour colon tissues

4   Membrane proteomics and *N*-glycosylation site mapping of patient-matched normal and tumour colon tissues

4.1   Introduction

Frequently glycosylated, membrane proteins are known to mediate important biological functions such as cell-cell adhesion, signal transduction and ion transport [242, 243]. They represent approximately 30% of the human genome and constitute the majority (60%) of pharmaceutical targets [244]. Like glycans, membrane proteins are a potential source of cancer biomarkers. In fact, carcinoembryonic antigen (CEA), a "classical' biomarker which has been recommended for CRC surveillance [245], is a GPI-linked membrane glycoprotein [246]. But CEA has only limited sensitivity and specificity, so it is not well suited for early detection population based screening [214, 245, 247]. Therefore, there remains a clear need for new biomarker candidates for early CRC detection.

Despite the development of protocols for enrichment of membrane proteins, there are few recent studies of CRC membrane proteins [57, 248, 249]. These studies have generally employed differential centrifugation with sodium carbonate for membrane protein enrichment prior to proteomic analysis by mass spectrometry. Amongst the differentially expressed proteins, CEA has been shown to be upregulated in CRC [249]. However, the glycosylation status of CEA or other altered membrane proteins has not been widely investigated. While glycoproteins can be studied by membrane enrichment, there are also alternate strategies for their direct enrichment. Techniques include hydrophilic interaction liquid chromatography (HILIC), lectin affinity chromatography, hydrazide chemistry and titanium dioxide enrichment (for sialylated glycopeptides) [250] followed by mass spectrometry [251]. In addition, glycoprotein enrichment can be carried out on whole cell proteomes or after immunoprecipitation of a targeted glycoprotein. Mass spectrometry can be performed on the glycopeptides directly or following enzymatic removal of *N*-linked glycans by PNGase-F [252, 253].

In CRC, glycoproteomics has been performed on the colorectal cancer cell line HT29, which utilised a combination of membrane and glycoprotein enrichment by lectins [254]. In this study from 2008, only 20 proteins were identified, of which 65% were membrane associated and 45% were *N*-glycosylated. *N*-glycosylated proteins found include integrin and dipeptidyl peptidase IV, though their expression was not quantified. More recent work on nine pair-matched normal and tumour colon tissue identified 72 proteins with 17 detected exclusively and 14 upregulated in cancer after lectin enrichment and LC-MS/MS from tissue lysates [255]. While these studies identified altered glycoproteins, their glycosylation sites were not investigated. To this end, more comprehensive and targeted work has been performed with the acute phase glycoprotein haptoglobin. Though not a

membrane protein, it was found upregulated in the serum of CRC patients [256, 257]. The fucosylation of Asn 241 was reported to be higher in cancer than in patients with inflammatory bowel disease or healthy controls in this study.

However, it was only very recently that non-targeted approaches have been used to study *N*-glycosylation sites in cancer [258, 259]. For pancreatic cancer, the *N*-glycosylation sites of pancreatic tissue from healthy controls and patients with chronic pancreatitis or pancreatic cancer were compared [258]. A total of 649 *N*-glycosylation sites were found and the abundance of some *N*-glycopeptides was also altered between cancer and non-diseased pancreatic tissues. In addition, a study by Deeb *et. al.* [259] on 10 lymphoma cells lines identified 2383 *N*-glycosylation sites from 1321 protein groups. Differences in the *N*-glycosylation site expression between the 10 cell lines allowed their segregation into two major lymphoma subtypes, suggesting that glycopeptide differences may have potential for use as a cancer signature. Overall, these studies have highlighted the feasibility and importance of *N*-glycosylation site analysis using a combination of glycoprotein enrichment, enzymatic deglycosylation and mass spectrometry.

In CRC, there is a near absence of knowledge about glycoprotein alterations occurring concomitantly with glycosylation site changes in these tissues during carcinogenesis. Therefore, in this chapter, the peptides released from tissue membrane proteins after their deglycosylation by PNGase-F (i.e. previously *N*-linked glycopeptides) were analysed. The use of PNGase F releases the glycan from Asn residues, resulting in its deamidation and conversion to Asp (+ 0.98 Da) that can be readily detected by mass spectrometry. The corresponding global protein *N*-glycan profiles from the same patients (n=3) have been presented earlier in Chapter 3. The identification of *N*-glycosylation sites in these peptides was also of interest as little is known *N*-glycosylation site changes during CRC carcinogenesis. In order to distinguish nonenzymatic and enzymatic deamidation, peptides from proteins not treated with PNGase-F from the same samples were also analysed. Overall, this study aimed to investigate changes in the membrane glycoproteins and the identification of their *N*-glycosylation sites from normal and matched tumour colon tissue from CRC patients. Notably, this is also the first study to use epithelial cell enrichment prior to membrane protein preparation. The addition of the epithelial enrichment step may deepen coverage of the membrane subproteome and enhance identification of epithelial-specific biomarkers.

4.2 Materials and methods

A workflow of experimental procedures conducted for this Chapter is shown in Figure 1.
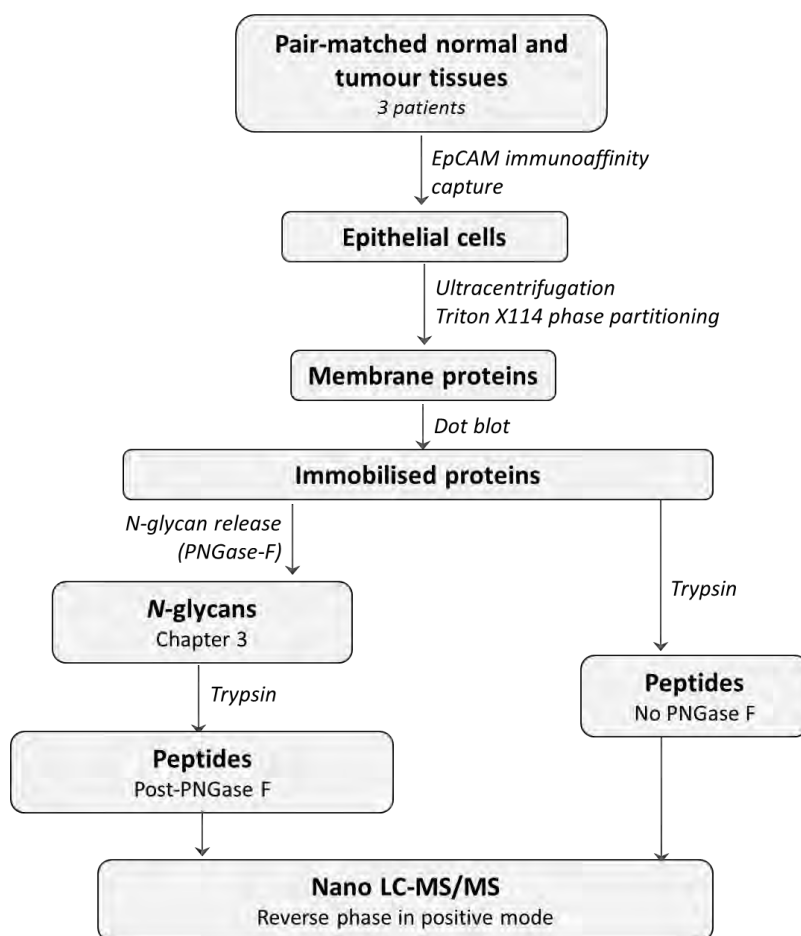
Figure 1. Proteomics analysis workflow for pair-matched normal and tumour colorectal tissue studied in this chapter. The preparation of membrane proteins was essentially the same as Chapter 2 and 3.

Tissue collection

Three samples consisting of normal mucosae and their patient-matched tumours were analysed in this study. The clinicopathological characteristics of the patient cohort (patients 4-6) are shown in Table 1 of Chapter 3. *N*-glycan profiles have also been obtained earlier from these sample patients and tissues have been collected as previously described in Chapter 3.

Immunoaffinity capture of epithelial cells and membrane protein enrichment

Both immunoaffinity capture of epithelial cells and membrane protein enrichment were performed essentially as described in the previous study [219] (Chapter 2). Immunoaffinity capture was carried out using magnetic beads conjugated to epithelial marker EpCAM (Cellection Epithelial Enrich Dynabeads, Invitrogen). Enrichment of epithelial cells was previously confirmed by Western Blotting [219] (Chapter 2). Membrane proteins were enriched using a combination of sodium carbonate stripping and Triton X114 phase partitioning as previously [221]. Similar to chapter 3, protein quantification was not performed due to limited sample recovery and thus, proteins were

normalised post-data acquisition using total protein abundance (as determined by label-free spectral counting).

Dot blotting and peptide release by tryptic digestion

Tryptic digestion of protein dot blots was performed essentially as detailed earlier [260]. It should be noted that in this workflow (Fig 1) trypsin digestion was performed following *N*-glycan release so that *N*-glycans could be collected for *N*-glycosylation profiling (Chapter 3) and to improve proteomic coverage[260]. Following enrichment, samples were resuspended in 10 µL of 8 M Urea. Proteins were dot blotted onto pre-activated PVDF membranes (0.2 µM, Bio-Rad) and stained with 0.1 % (v/v) Direct Blue 71 to enable protein visualisation. In order to distinguish between spontaneous chemical deamidation with enzymatic deamidation, each sample were dot blotted twice with one dot blot digested following PNGase-F glycan release while the other was digested without prior incubation with PNGase-F (*N*-glycan release as detailed in Chapter 3). Dried protein spots were cut and placed into 96-well plate. Spots were rewet with methanol, washed with water, and then incubated with 10 µL of trypsin (0.02 µg/µL, Promega) in 60 mM ammonium bicarbonate, at 37°C for 3 hours. Peptides were recovered by addition of 15 µL of 50% acetonitrile (v/v) with 1% formic acid (v/v) and sonication for 10 mins. Protein spots were also washed with 20 µL of water and wash fractions were pooled with recovered samples. Samples were dried and stored at -30°C until analysis by LC-MS/MS.

Nanoflow reverse phase chromatography (nanoLC-MS/MS) for peptide analysis

Dried peptides were resuspended in 10 µL in 2% acetonitrile (v/v) in 0.1% formic acid (v/v) for analysis by LC-MS/MS. Peptides were separated by an Eksigent LC system (Eksigent Technologies, USA) using a ProteCol C18 column (300 Å, 3 µm, 150 µm x 10 cm, SGE Analytical Sciences, Australia). Solvent A consisted of 2% acetonitrile with 0.1% formic acid (v/v) while Solvent B contained 90% (v/v) acetonitrile with 0.1% (v/v) formic acid. A 150 min linear gradient at a flow rate of 600 nl/min was used starting from 98:2 solvent A/solvent B at 0 mins changing to 60:40 solvent A/solvent B at 150 mins.

The LC eluent was analysed in positive ion mode using a hybrid quadruple TripleTof 5600 system (AB SCIEX, Concord, ON) [261], fitted with a Nanospray III source (AB SCIEX, Concord, ON) and a PicoTip fused silica emitter (360 µm od, 75 µm id, 15 µm diameter emitter orifice; New Objective, MA, USA). Ion spray voltage, heater interface temperature, curtain gas flow and nebulising gas flow was set 2.5 kV, 150°C, 20 and 5, respectively. Ions with *m/z* 350-1500 were acquired using an IDA (information dependent acquisition) approach where the top 15 ions of precursor scan were selected for MS/MS. Accumulation time was 0.25 sec in precursor scan and 0.1 sec in the product ion scan.

Membrane protein identification and quantification by normalised spectral abundance factors

The MS/MS of the raw data files (.Wiff format) were re-calibrated post-data acquisition to improve mass accuracy [262] using either Analyst TF (v 1.6, AB SCIEX) or Peakview (v 1.2.0.3, AB SCIEX). Re-calibrated wiff files were converted to Mascot generic format (MGF) using ProteinPilot Software (v 4.2, revision 1340) for searching in Mascot (v 2.4.0, Matrix Science, London, UK). The search was performed against homo sapiens assuming the enzyme trypsin allowing for up to 2 missed cleavages. Database selected was Swissprot (swissprot2014_04.fasta) which contained 20266 entries. There was no fixed peptide modification selected. Variable peptide modifications specified were deamidation at asparagine (+ 0.98 Da) and oxidation at methionine (+ 15.99 Da). Peptide tolerance was set at ± 5 ppm and MS/MS tolerance at ± 0.03 Da. Peptide charge states of 2+, 3+ and 4+ were considered.

Mascot results were further filtered using Scaffold (v 4.3.0, Proteome Software Inc., Portland, OR) for identification and quantification, which used the Peptide Prophet algorithm [263] for FDR estimation. Protein and peptide FDR were set at 1% and 0.1% respectively and proteins were identified using a minimum of 1 peptide. Proteins which contained similar peptides and could not be differentiated by MS/MS were grouped to satisfy the principles of parsimony. Proteins were also further categorised and cellular localisation mapped according to UniProt keywords such as "glycoprotein", "membrane" and "cytoplasm" similar to a previous study [259]. The annotation of "glycoproteins" contains both experimentally-validated and putative glycosylation sites (as predicted by NetNGlyc which is available at http://www.cbs.dtu.dk/services/NetNGlyc/) [264]. Prediction of transmembrane helices was carried out using TMHMM v 2.0 (available online at http://www.cbs.dtu.dk/services/TMHMM/) while prediction of signal peptides was performed using SignalP v 4.1 (available online at http://www.cbs.dtu.dk/services/SignalP/).

For protein quantification, a label-free spectral counting-based approach of normalised spectral abundance factors (NSAF) (calculated according to Zybailov *et. al.*[265]) was used. Peptides obtained from both PNGase-F treated and untreated fractions were combined for quantification. For statistical calculations, all spectral counts were adjusted by a spectral fraction of 0.5 in order to include proteins with zero spectral counts [266]. A p-value of smaller than 0.05 was required for significance when protein expression was compared in normal tissue tumours using Student's paired two-tailed t-test. Proteins were considered differentially expressed if it was present with a minimum of one peptide in all three patients with a p-value of smaller than 0.05 and a fold change greater than 1.5.

Identification and quantification of *N*-glycosylation sites using Scaffold PTM

272

For identification and quantification of *N*-glycosylation sites, peptides following PNGase-F digestion were analysed separately from peptides which were not treated previously with PNGase-F ('PNGase-untreated'). The small letter 'n' has been used to depict deamidation of asparagine. Deamidated sites with the *N*-glycosylation consensus sequence (NXS/T/C, where X can be any amino acid except proline (P)) were considered genuine if not found in the PNGase F-untreated fraction. For validation of deamidated sites, Mascot results were further analysed in Scaffold PTM (v 2.1.3, Proteome Software Inc., Portland, OR) and a localisation probability for each site was calculated based on the p-value of the Ascore [267]. A minimum localisation probability of 95% was required for analysis. Sequence logos (Weblogos) were generated online at http://weblogo.berkeley.edu/ [268] using sequences pre-aligned in Scaffold PTM. Additionally, motif analysis on pre-aligned sequences was carried out using Motif-x tool [269] which is available at http://motif-x.med.harvard.edu/motif-x.html. For this analysis, the width, number of occurrences and significance was set at 13, 5 and 0.000001 respectively. The background selected was the IPI human proteome (the FASTA file containing all human protein sequences).

4.3   Results and discussion

Membrane proteomics of epithelial-enriched normal and tumour colorectal tissue

A novel aspect in this proteomics study, as used in the glycomics studies reported earlier (Chapters 2 and 3), is the enrichment of epithelial cells from the tissue specimens using immunoaffinity capture. This enrichment has been previously confirmed using Western Blot (Chapter 2) using representative epithelial marker keratin 18 [270] and mesenchymal/fibroblast marker vimentin [271]. In this study, complementary proteomic profiles of epithelial cell-enriched fractions from colorectal tissue were also analysed, although the proteomic profile of the non-epithelial fractions were not obtained due to limited recovery. However, the abundance of a panel of markers for epithelial and non-epithelial cells types (Supp Table 1, with references) was analysed for the epithelial-fractions of the normal and cancer tissues (Fig 2). None of the cellular specific markers (Supp Table 1) were differentially expressed between normal and tumour tissues except CD44 which was increased in the tumours (p < 0.05). However, there were general observations which were notable in both normal and tumour tissues. As expected, the epithelial markers EpCAM and cytokeratin 8/18/19 were abundantly detected which was consistent with the enrichment of epithelial cells (Fig 2). In contrast, immune markers (such as CD45, CD4, CD8) and muscle cell markers (such as actin gamma, caldesmon and desmin) were generally low in abundance or absent (Fig 2) demonstrating there was little evidence for non-epithelial cell types. Interestingly, vimentin was also identified by MS, although it was not detected in the epithelial fraction by Western blotting (Chapter 2). As vimentin and CD44 have both been associated with epithelial-to-mesenchymal transition (EMT)[272-274], this suggests EMT may be occurring in these colon tissues. Alternatively, there may also be residual fibroblasts as vimentin is

also known as a fibroblast marker [271]. Despite this, the overall protein expression pattern is clearly supportive of epithelial cell enrichment by immunoaffinity capture, although the presence of some residual stromal cells cannot be completely excluded following enrichment.
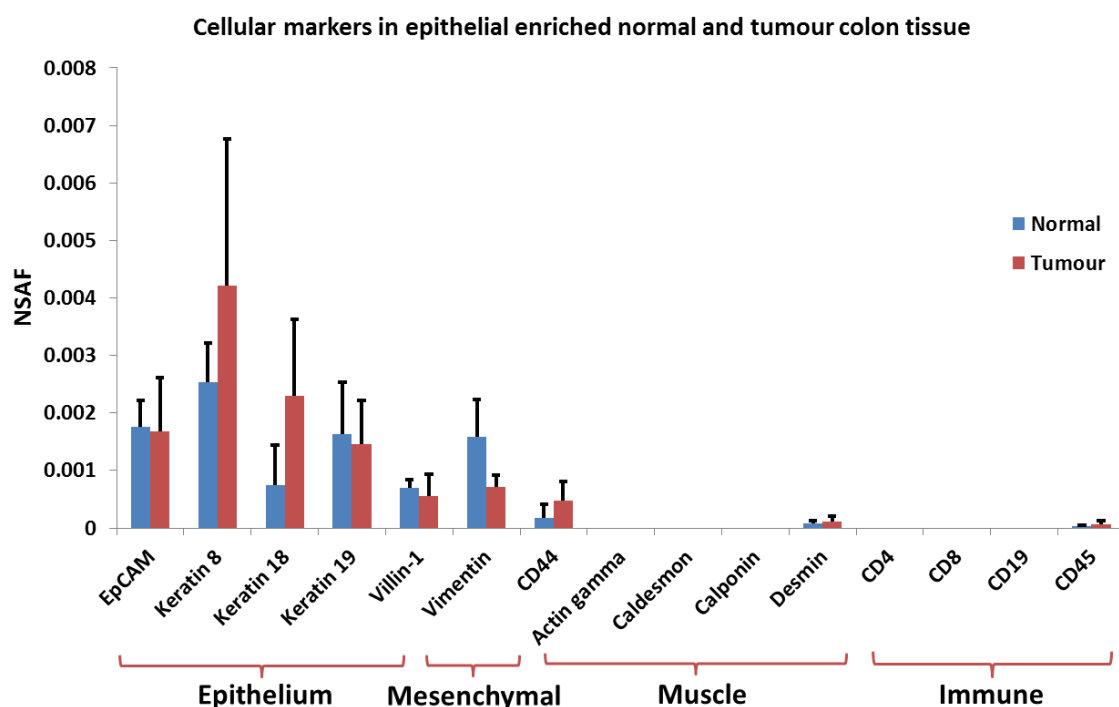
Figure 2. Expression of cellular markers in the epithelial fractions of colon tissue (normal and tumour) quantified by normalised spectral abundance factors (NSAF). None of the cellular markers analysed were differentially expressed between normal and tumour tissues except for CD44 which was higher in the tumours ($p < 0.05$). Expression profile obtained complemented the confirmation of enrichment previously shown by Western Blot (Chapter 2).

The combination of epithelial cell affinity enrichment and membrane protein enrichment enabled this study to identify 1648 proteins in total amongst all samples at 1% protein FDR (spectral counts of proteins and peptides provided in Supp Table 2 and 3). Of these, 392 (23.8%) were annotated as "glycoprotein" (predicted or experimentally confirmed) according to UniProt (Fig 3a). To give a general overview of the proteome, proteins were also classified according to their cellular localisation using the UniProt keywords "membrane" and "cytoplasm" (Fig 3b) which are mapped based on gene ontology (GO) annotation [275]. Following membrane enrichment, membrane-associated proteins form the largest proportion of identified proteins in the proteome at 51%, of which 8% were also associated with the cytoplasm. When protein abundance was also considered, a similar result was obtained (52% were membrane proteins, Supp Fig 1). Compared to a previous membrane proteomics study using essentially the same membrane enrichment procedure on rat liver [221], the proportion of membrane proteins was lower in this study (74% vs 51%). However, the

overall number of identified membrane proteins is much higher in this report (849 in this study vs 188 identified in [221]).
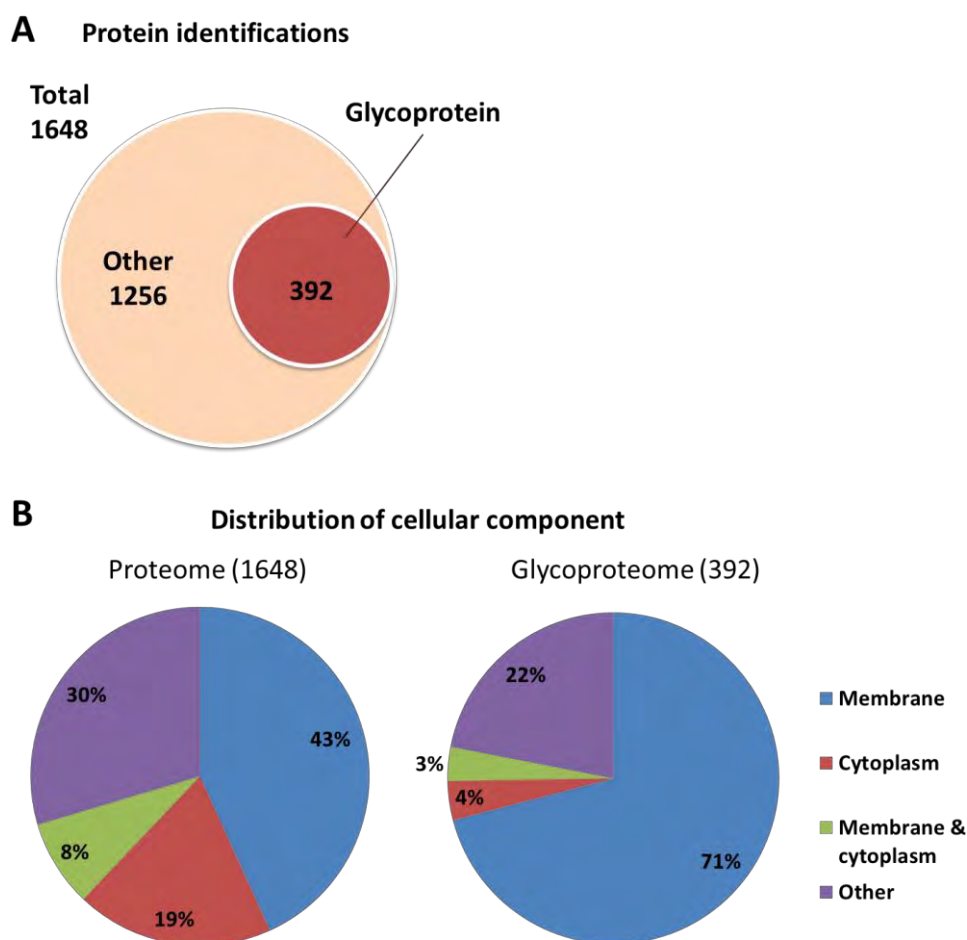
Figure 3. Membrane proteomics of pair-matched normal and cancerous colorectal tissues (n=3) following epithelial cell enrichment. Protein classification as 'glycoprotein' and cellular localisation were analysed using UniProt keywords. **A)** A total of 1648 proteins were identified, out of which 392 (23.8%) were glycoproteins. **B)** Comparison of cellular localisation between the proteome and glycoproteome suggests a higher percentage of membrane-associated proteins in the glycoproteome (43% vs 71%). Proteins annotated as 'other' have not been associated with either the membrane or the cytoplasm according to annotation by UniProt.

In terms of comparison with other membrane proteomics studies using human colon tissues, three major studies have been performed since 2010 with two studies using up to 28 patient-matched normal and tumours [57, 249] while the third study investigated matched polyps with tumours [248]. Similar to this study, all had employed sodium carbonate-based centrifugation methods for the enrichment of membrane proteins although Triton X114 phase partitioning was not performed. The proportion of membrane proteins (as a percentage of total number of proteins) reported varied from 51% [57] to 75% (642/856) [249] compared to 51% obtained in this study. Interestingly, although

workflow carried out in this study did not result in the highest enrichment of membrane proteins, the number of quantified membrane proteins was the highest (849) when contrasted with other proteomic studies comparing normal and tumour colonic tissues.

Out of the 1648 proteins identified, 52 proteins were found to be differentially expressed between normal and tumour colorectal tissues. The differentially expressed proteins (excluding the 9 glycoproteins located in Table 2) are reported in Table 1. Using DAVID bioinformatics tool for GO (gene ontology) analysis [276], the top 5 non-redundant GO terms for molecular function and biological process with the highest coverage are shown in Table 3, which indicates that the proteins are involved in the binding of various substrates and protein localisation. Pathway analysis was also performed using DAVID, although there was little overlap in pathways between various proteins, and no terms were significantly enriched. This is not surprising given that there were few differentially expressed proteins. However, a very small number of terms (≤3) such as "drug metabolism", "purine metabolism", "antigen processing and presentation" were shared by up to 2 proteins (maximum, Table 3).

Table 1. Differentially expressed proteins from 3 patient matched normal and tumour colonic tissues (minimum 1 peptide, 1.5 fold -change, p-value < 0.05). The p-value was calculated using two tailed paired student's t-test . Average fold change was determined by the mean of Tumour NSAF/Normal NSAF across all three patients. Glycoproteins have been excluded (these are presented in Table 2).

Cellular localisation	Uniprot ID	Protein name	Average fold change (T/N)	P-value (T-test)	NSAF Normal	± SD	Tumour	± SD
Membrane	DHB12_HUMAN	Estradiol 17-beta-dehydrogenase 12	14.461	0.0391	2.72E-04 ±	1.81E-04	2.93E-03 ±	8.10E-04
	CY24A_HUMAN	Cytochrome b-245 light chain	8.080	0.0192	1.99E-04 ±	1.24E-04	1.22E-03 ±	3.66E-04
	BRI3B_HUMAN	BRI3-binding protein	6.843	0.0499	1.85E-04 ±	4.56E-05	1.29E-03 ±	4.85E-04
	ENOA_HUMAN	Alpha-enolase	4.719	0.0391	3.41E-04 ±	2.07E-04	1.28E-03 ±	4.75E-04
	TMX2_HUMAN	Thioredoxin-related transmembrane protein 2	4.375	0.0455	1.83E-04 ±	1.97E-05	8.06E-04 ±	2.52E-04
	TOM40_HUMAN	Mitochondrial import receptor subunit TOM40 homolog	4.250	0.0167	2.35E-04 ±	1.56E-04	7.90E-04 ±	3.26E-05
	TMCO1_HUMAN	Transmembrane and coiled-coil domain-containing protein 1	4.136	0.0033	4.10E-04 ±	2.30E-04	1.51E-03 ±	3.10E-04
	S12A2_HUMAN	Solute carrier family 12 member 2	3.455	0.0474	3.19E-04 ±	1.12E-04	9.94E-04 ±	1.53E-04
	VDAC1_HUMAN	Voltage-dependent anion-selective channel protein 1	3.019	0.0002	3.18E-03 ±	2.33E-03	7.47E-03 ±	2.35E-03
	SPTC2_HUMAN	Serine palmitoyltransferase 2	2.804	0.0216	6.89E-05 ±	4.29E-05	1.35E-04 ±	2.87E-05
	GLYM_HUMAN	Serine hydroxymethyltransferase, mitochondrial	2.763	0.0426	3.67E-04 ±	1.75E-04	9.78E-04 ±	3.79E-04
	MYADM_HUMAN	Myeloid-associated differentiation marker	2.690	0.0498	3.11E-04 ±	2.01E-04	7.66E-04 ±	3.33E-04
	QCR8_HUMAN	Cytochrome b-c1 complex subunit 8	2.291	0.0279	8.47E-04 ±	3.66E-04	1.77E-03 ±	1.85E-04
	ATD3A_HUMAN	ATPase family AAA domain-containing protein 3A	2.172	0.0421	5.21E-04 ±	4.15E-04	8.89E-04 ±	2.80E-04
	CNIH4_HUMAN	Protein cornichon homolog 4	2.112	0.0100	2.79E-04 ±	1.73E-04	4.91E-04 ±	2.09E-04
	CALX_HUMAN	Calnexin	1.909	0.0157	1.38E-03 ±	4.34E-04	2.56E-03 ±	5.74E-04
	HYEP_HUMAN	Epoxide hydrolase 1	1.875	0.0491	2.70E-04 ±	1.46E-04	4.35E-04 ±	1.39E-04
	TRAP1_HUMAN	TNF receptor-associated protein	1.667	0.0432	1.11E-03 ±	3.47E-04	1.83E-03 ±	5.54E-04
	ESYT1_HUMAN	Extended synaptotagmin-1	1.502	0.0367	9.07E-05 ±	7.51E-05	1.27E-04 ±	8.73E-05
Cytoplasm	FLNA_HUMAN	Filamin-A	6.112	0.0031	1.38E-04 ±	1.01E-04	4.22E-04 ±	1.18E-04
	NUCL_HUMAN	Nucleolin	5.620	0.0186	5.46E-05 ±	3.40E-05	2.35E-04 ±	7.68E-05
	1433Z_HUMAN	14-3-3 protein zeta/delta	4.420	0.0350	4.71E-04 ±	4.47E-04	1.60E-03 ±	7.97E-04
	IMDH2_HUMAN	Inosine-5'-monophosphate dehydrogenase 2	3.520	0.0335	7.54E-05 ±	4.69E-05	2.21E-04 ±	9.41E-05
	PRS10_HUMAN	26S protease regulatory subunit 10B	2.804	0.0216	9.96E-05 ±	6.20E-05	1.96E-04 ±	4.15E-05
	NDKB_HUMAN	Nucleoside diphosphate kinase B	2.620	0.0225	6.59E-04 ±	5.45E-04	1.33E-03 ±	4.20E-04
	TCPQ_HUMAN	T-complex protein 1 subunit theta	2.600	0.0408	3.54E-04 ±	2.21E-04	6.41E-04 ±	1.20E-04
	HS90B_HUMAN	Heat shock protein HSP 90-beta	2.432	0.0423	9.67E-04 ±	4.87E-04	2.09E-03 ±	3.89E-04
	RLA0_HUMAN	60S acidic ribosomal protein P0	1.784	0.0376	9.56E-04 ±	4.01E-04	1.68E-03 ±	6.41E-04
	GRP75_HUMAN	Stress-70 protein (mitochondrial)	1.668	0.0069	2.04E-03 ±	6.59E-05	3.40E-03 ±	2.27E-04
	MYO1C_HUMAN	Unconventional myosin-Ic	-2.729	0.0148	4.12E-04 ±	2.50E-04	2.86E-04 ±	2.31E-04

Table 1. (cont).

Cellular localisation	Uniprot ID	Protein name	Average fold change (T/N)	P-value (T-test)	NSAF Normal	± SD	Tumour	± SD
Secreted	THEM6_HUMAN	Protein THEM6	9.434	0.0219	1.86E-04 ±	1.16E-04	1.09E-03 ±	2.32E-04
Unknown[a]	IGKC_HUMAN	Ig kappa chain C region	-6.418	0.0163	2.15E-02 ±	8.29E-03	6.61E-03 ±	7.14E-03
	GCN1L_HUMAN	Translational activator GCN1	7.451	0.0259	3.46E-05 ±	2.61E-05	1.87E-04 ±	3.59E-05
	DDX21_HUMAN	Nucleolar RNA helicase 2	6.643	0.0327	4.95E-05 ±	3.08E-05	2.12E-04 ±	5.65E-05
	CLPX_HUMAN	ATP-dependent Clp protease ATP-binding clpX-like subunit	2.112	0.0100	6.12E-05 ±	3.81E-05	1.08E-04 ±	4.59E-05
	RS16_HUMAN	40S ribosomal protein S16	1.752	0.0195	1.43E-03 ±	9.61E-04	2.01E-03 ±	8.57E-04
	DECR_HUMAN	2,4-dienoyl-CoA reductase, mitochondrial	-2.124	0.0073	1.20E-03 ±	6.94E-05	5.70E-04 ±	6.09E-05
	ODO2_HUMAN	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	-2.403	0.0379	1.63E-03 ±	6.72E-04	7.28E-04 ±	4.29E-04
	LONM_HUMAN	Lon protease homolog (mitochondrial)	-2.912	0.0075	3.08E-04 ±	1.33E-04	1.71E-04 ±	1.32E-04
	ABHDB_HUMAN	Alpha/beta hydrolase domain-containing protein 11	-3.135	0.0076	9.34E-04 ±	1.08E-04	3.50E-04 ±	1.96E-04
	LMNA_HUMAN	Prelamin-A/C [Cleaved into: Lamin-A/C]	-6.146	0.0145	4.65E-04 ±	1.27E-04	8.90E-05 ±	5.00E-05
	ACSF2_HUMAN	Acyl-CoA synthetase family member 2, mitochondrial	-9.738	0.0433	3.52E-04 ±	1.33E-04	3.70E-05 ±	1.57E-05
	HMCS2_HUMAN	Hydroxymethylglutaryl-CoA synthase, mitochondrial	-25.532	0.0203	1.74E-03 ±	4.31E-04	7.61E-05 ±	3.62E-05

a) Unknown whether it is located in the cytoplasm or membrane or is secreted

Table 2. Differentially expressed glycoproteins from 3 patient matched normal and tumour colonic tissues (minimum 1 peptide, 1.5 fold-change, p-value < 0.05). Proteins were determined as "glycoprotein" from UniProt annotation. The p-value and average fold change were calculated according to Table 1.

Cellular localisation	Uniprot ID	Protein name	Average fold change (T/N)	P-value (T-test)	NSAF Normal	± SD	Tumour	± SD
Membrane	CCD47_HUMAN	Coiled-coil domain-containing protein 47	9.790	0.0395	8.02E-05 ±	4.99E-05	5.13E-04 ±	1.68E-04
	CD44_HUMAN	CD44 antigen	7.768	0.0411	1.80E-04 ±	2.33E-04	4.75E-04 ±	3.36E-04
	S39AE_HUMAN	Zinc transporter ZIP14	5.930	0.0272	1.56E-04 ±	1.71E-04	3.79E-04 ±	1.60E-04
	2B1G_HUMAN	HLA class II histocompatibility antigen, DRB1-16 beta chain	5.810	0.0301	9.44E-04 ±	1.19E-03	2.92E-03 ±	1.64E-03
	CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5	3.080	0.0437	2.29E-03 ±	2.40E-03	4.58E-03 ±	1.84E-03
	CAH12_HUMAN	Carbonic anhydrase 12	-8.529	0.0439	7.67E-04 ±	1.96E-04	1.09E-04 ±	5.20E-05
Secreted	IGJ_HUMAN	Immunoglobulin J chain	-3.481	0.0142	2.19E-03 ±	4.46E-04	6.36E-04 ±	1.49E-04
Unknown[a]	EST2_HUMAN	Cocaine esterase	-4.212	0.0071	3.86E-04 ±	1.84E-04	1.69E-04 ±	2.05E-04
Cytoplasm	SERPH_HUMAN	Serpin H1	2.967	0.0316	1.29E-04 ±	1.39E-05	3.86E-04 ±	9.12E-05

a) Unknown whether it is located in the cytoplasm or membrane or is secreted

Table 3. Top 5 non-redundant GO terms for biological process, molecular function and KEGG pathways with highest coverage for differentially expressed proteins using DAVID informatics tool.

Category	Term	P-Value	Fold Enrichment	Count	% of differentially expressed
Molecular function	Nucleoside and nucleotide binding	0.0319	1.93	14	27
	Unfolded protein binding	0.000001	18.82	7	13
	ATPase activity	0.0206	4.63	5	10
	Cofactor binding	0.0435	4.97	4	8
	Collagen binding	0.0057	25.76	3	6
Biological Process	Protein localisation	0.0434	2.62	7	13
	Protein folding	0.0002	11.18	6	12
	Response to organic substance	0.0598	2.75	6	12
	Protein transport	0.0723	2.60	6	12
	Transmembrane transport	0.0861	2.90	5	10
KEGG Pathway	Purine metabolism			2	4
	Drug metabolism			2	4
	Antigen processing and presentation			2	4

Due to the multifunctional nature of many proteins, it is not surprising that proteins can be associated with more than one GO term. One notable example in this dataset is nucleoside diphosphate kinase B (NDKB_HUMAN), a member of the nucleoside diphosphate kinase (NDK) family [277]. This protein was increased 2.6 fold in the tumours (Table 1) and associated with "nucleoside and nucleotide binding", "response to organic substance" and "purine metabolism". This protein functions to synthesise nucleoside triphosphates other than ATP and is known to have various roles in cell growth and differentiation [278]. In colorectal cancer, its expression has been observed to be increased in tumour tissues compared to normal epithelium as shown by immunochemistry [279]. In this study, unique peptides for NDK B were consistently observed in all three patients demonstrating elevated NDK B in CRC tumour cells. Although a related isoform, NDK A has been reported as upregulated in a previous proteomics study on colon cancer and other cancers [280] this NDK isoform was not detected in our study. Though isoform-specific functions require further investigation, the findings here provide further support for the upregulation of nucleoside diphosphate kinase expression between normal and primary colon tumours.

Other proteins highlighted by the GO analysis include the higher expression of heat shock proteins in tumours, such as mortalin (GRP75_HUMAN), and heat shock protein 90 (HS90B_HUMAN) which were increased 1.7 fold and 2.4 fold in the tumours, respectively (Table 1). These proteins were associated with the terms "protein localisation", "protein folding" and "protein transport" since they are known as molecular chaperones which are required for protein structural integrity [281] and

correct protein folding [282]. Interestingly, molecular chaperones have been associated with glycoproteins as aberrant protein *N*-glycosylation has been shown to impair protein folding [283]. As well, calnexin (CALX_HUMAN), another protein involved in protein folding, was also upregulated 1.9-fold in the tumours (Table 1). A possible explanation for the increased expression of molecular chaperones is the increased protein synthesis of tumour cells, which may in turn require the higher expression of molecular chaperones to facilitate correct protein folding and promote cell survival [284]. Overall, the increased expression of the above proteins in tumours is unsurprising since these findings are consistent with previous studies which widely report their upregulation in colorectal cancer [57, 285-290]. Given that these proteins have been found increased in tumours in multiple studies, it is not surprising that many of these have been identified as potential therapeutic targets for cancer therapy [291-295].

While the above proteins were highlighted due to the GO analysis, there were also other noteworthy proteins in our dataset such as BRI3 binding protein (BRI3B_HUMAN), filamin A (FLNA_HUMAN) and hydroxymethylglutaryl-CoA synthase (HMCS2_HUMAN). In the tumours, BRI3 binding protein (BRI3BP) was increased 6.8-fold, Filamin A was increased 6.1-fold while hydroxymethylglutaryl-CoA synthase (HMGCS2) was decreased 25.5 fold (Table 1). These proteins are known to be implicated in various cancer-related functions. For example, BRI3BP is known to play a role in tumorigenesis via functional disruption of tumour suppressor protein p53 [296]; filamin A mediates cell invasion, adhesion and metastasis [297-299]; while HMGCS2 has also been implicated in tumour growth [300]. These proteins have also been previously linked with CRC and on the whole, the changes in this study revealed similar trends to earlier reports as expected [296, 301, 302]. Thus, the general concordance between these results and the literature also provides evidence of the reliability of this proteomics workflow to capture representative protein changes despite the small patient group.

Besides confirming findings from other investigations, there were membrane proteins identified in this study, which to the author's best knowledge, have not been previously associated with CRC. Three of these were estradiol 17-beta-dehydrogenase 12 (DHB12_HUMAN), cytochrome b-245 light chain (CY24A_HUMAN) and ATPase family AAA domain-containing 3A (ATD3A_HUMAN), which will be discussed further below. Estradiol 17-beta dehydrogenase 12 (HSD17B12) was upregulated 14.5-fold in the tumours (Table 1). It has roles in the activation of estrogen (conversion of estrone (E1) to estradiol (E2)) and fatty acid metabolism [303]. HSD17B12 has also been associated with increased metastasis in hepatocellular carcinoma [304] and is considered a marker of poor prognosis in ovarian carcinoma [305]. However, the association of this isoform (of 14 isoforms [303, 306]) has not been linked to CRC before. Interestingly, a related isoform, HSD17B2, which inactivates estrogen (E2 to E1 conversion) has been observed as downregulated in CRC tumours [307, 308] which may also lead to

higher levels of active estrogen. Taken together, these findings implicate estrogen activation in tumorigenesis of CRC though the exact mechanism remains to be elucidated.

On a different note, cytochrome b-245 light chain (also known as p22phox) is a membrane bound subunit of the active NADPH oxidase (NOX), which generates reactive oxygen species involved in variety of functions such as microbe killing, angiogenesis [309] and cell signalling [310]. This protein was observed to be increased 8-fold in the tumours compared to the normal tissue (Table 1). The increased expression of p22phox has been linked with other cancers [311-313] but has not yet been correlated with CRC. Intriguingly, a related glycosylated subunit, gp91phox, known to form heterodimers with p22phox has been previously observed in a colon epithelial cell line and colon tissues [314]. This subunit (CY24B_HUMAN) was also detected in of this study although it was not differentially expressed between tumours and normal tissues. Though it has been reported that both subunits are vital for the fully functional NOX [310, 315, 316], it is not entirely clear how the change in p22phox:gp91phox ratio implied in this study will alter NOX function. However, as NOX enzymes are known to be involved in signal transduction related to angiogenesis [309], cell migration and growth [310], it is also speculated that these functions may be dysregulated by p22phox overexpression in tissue.

Besides the above proteins, the ATPase family AAA domain-containing 3A (ATAD3A) protein was also elevated in the tumours. It is a mitochondrial inner membrane protein that has been recently found highly expressed in cancers of the cervix [317], lung [318], brain [319] and prostate [318] using antibody-based approaches. However, to date it has not been identified in colorectal cancer and this study hence has shown this to be a differentially expressed protein between normal and tumour colon tissues. Similar to studies in other cancers, ATAD3A was upregulated in the colonic tumours (2.1-fold, Table 1) compared to the normal controls. Although additional verification is required in a larger patient cohort, the overexpression of ATAD3A may be a prognostic marker for poor survival in colorectal cancer, similar to lung [318] and cervical cancers [317]. Likewise, it may also be correlated with increased lymphovascular invasion and tumour stage as previously observed in other cancers [318, 320]. Overall, this is an interesting novel finding that warrants further investigation due to its strong association with tumour progression in other cancers.

Glycoproteins in CRC tumour tissues

Whilst membrane proteins are of importance, analysis of the glycoprotein subset is also highly relevant in the light of the previous glycomic analysis (detailed in Chapter 3). Thus, glycoproteins, especially altered glycoproteins from the membrane enriched samples, will be discussed in greater detail in the following sections. In this study, proteins were considered glycosylated if specified as a

"glycoprotein" according to UniProt keyword annotation. UniProt annotation identifies both experimentally verified and putative glycoproteins (proteins with computationally predicted glycosylation sites)[264]. Of 1648 proteins found, 392 (23.8%) were glycoproteins using this criteria (Fig 3a). Compared to the proteome, the percentage of membrane-associated proteins is higher in the glycoproteome (71% vs 43%) as expected, as glycoproteins are known to be membrane-associated or secreted [3].

Identification of N-glycosylation sites using PNGase-F

An aspect of interest from the glycoproteins identified is the localisation of their *N*-glycosylation site(s). Not only is this a step towards site-specific glycosylation profiling [321], but there is also little known about whether *N*-glycosylation sites are changed in altered glycoproteins in CRC. The *N*-glycosylation sites contain the consensus sequence NXS/T/C, where X represents any amino acid except proline although rarely serine or threonine may also be replaced by cysteine [322]. The use of PNGase-F releases the glycan from Asn residues, resulting in its deamidation and conversion to Asp that can be readily detected in the MS analysis of the tryptic peptides by searching the databases with a + 0.98 Da modification. In order to distinguish between genuine *N*-glycosylation sites resulting from enzymatic deamidation and spontaneous chemical deamidation [323, 324], two fractions were analysed from each tissue sample – the fractions were either PNGase-F treated or untreated prior to tryptic digestion. Deamidated peptides containing an *N*-glycosylation consensus sequence were therefore not included as an identified *N*-glycosylation site if they were present in the PNGase-F untreated fractions as these may have arisen from nonenzymatic chemical deamidation [323, 324]. This is especially likely when the asparagine is adjacent to a glycine (G) as the small glycine residue favours formation of a cyclic intermediate in the chemical asparagine (N) to aspartic acid (D) conversion [323, 325].

In total, 1628 proteins were identified in the PNGase-F treated fractions while 813 were identified in the untreated fractions (Table 4; 1648 proteins were identified only when PNGase-F treated and untreated fractions were analysed together).The lower number of protein identifications in the untreated fractions may be explained by lower protein coverage (due to glycan shielding of peptide backbone from tryptic digestion[260]) and/or possibly lower protein loading. The spectral counts and full list of deamidated peptides used in this analysis is provided in Supp Table 4 and 5. Overall, 281 deamidated Asn sites were identified in the PNGase-F treated samples whereas only 26 deamidated sites were identified in the untreated samples (Table 4), which could be attributed to nonenzymatic chemical transformation. However, it was noted that 13 of the 26 deamidated sites (50%) found in the untreated samples were also identified in the PNGase-F treated samples, indicating that a degree of nonenzymatic deamidation remained in the PNGase-F treated samples. A frequency plot of the

deamidated sites in PNGase-F treated and untreated samples is shown in Fig 4a. Analysis of these sequences showed that in the PNGase-F treated samples, the majority (70%, 197/281) contained the recognised *N*-linked glycosylation motif NXS/T (Fig 4b). In contrast, only 12%, (3/26) contained NXS/T in the untreated samples (Fig 4b). There were no deamidated peptide sequences found consisting of the NXC tripeptide in either PNGase-F treated or untreated samples. There were also 84/281 (30%, Fig 4b) deamidated sites in the PNGase-F treated samples which did not contain a known *N*-glycosylation consensus sequence (NXS/T/C). 43 of these sites (43/84, 51%) had the NG motif which was the only motif found to be statistically overrepresented at these sites (by motif analysis, Fig 4c). For the remaining 41 sites, the amino acid adjacent to the deamidated asparagine varied (Supp Fig 3) and the reason for these sites is unclear at present. However, the frequency of deamidation at any one particular amino acid was notably lower for these 41 sites (≤ 7 sites for any other amino acid) than for glycine. The overrepresentation of glycine and other amino acids adjacent to asparagine will be discussed further below.

As mentioned above, motif analysis was also carried out using the Motif-x tool [269, 326] to further determine whether there was a significant overrepresentation of NXS/T motifs in the PNGase-F treated samples. In this analysis, the background was selected as the IPI Human Proteome (the FASTA file containing all human protein sequences) while the foreground was all the sequences containing deamidated sites identified in this study (PNGase-F treated and untreated were analysed separately). The fold increase was calculated as (foreground matches/foreground size)/(background matches/background size)[269]. Using this tool, NXT and NXS motifs were significantly increased 6.4- and 5.4-fold (Fig 4c) in the PNGase-F treated fractions whereas in the untreated samples they were not significantly altered from the background. The presence of leucine residues upstream of deamidated asparagine (Fig 4a) was also remarkable although the motif was not significantly increased. The NG motif was found elevated relative to background in both PNGase-F treated and untreated fractions, although the majority of NG motifs found in untreated samples (9/11) were also identified in PNGase-F treated samples – again suggesting that a degree of spontaneous chemical deamidation remained in the PNGase-F treated samples. Nevertheless, the notable increase in deamidated peptides containing a NXS/T motif in PNGase-F treated compared to untreated samples indicates enzymatic deamidation occurred. More importantly, the enzymatically deamidated sites (i.e. formerly *N*-glycosylated) can be separated from nonenzymatically deamidated sites as the enzymatically deamidated sites would not be present in the untreated fractions. Only two deamidated peptides with NXS/T sites identified in untreated samples were also found in the PNGase-F treated samples (Fig 4d). These sites, along with the one site containing NPT sequence were removed from analysis and not considered as *N*-glycosylation sites. The latter was removed since it is not considered a canonical *N*-glycosylation motif.

Although there is a clear and significant number of deamidated sites with the canonical *N*-glycosylation NXS/T motif identified in this study (as mentioned above, Fig 4c), the overall percentage in the PNGase-F treated sample (70%, Fig 4b) was notably lower than other reports. Utilising a MS-based approach, typically >95% of deamidated sites identified contained an *N*-glycosylation consensus sequence [19, 259]. A likely explanation for the discrepancy with the current study is the noted use of ^{18}O-water during sample preparation which was not employed in this study. The use of ^{18}O-water during PNGase-F digestion (while ^{16}O-water was used for other processing steps) thus allowed ready discrimination between enzymatic and nonenzymatic deamidation. Therefore, one of the explanations for the number of deamidated sites in the PNGase-F treated samples which did not have an *N*-glycosylation site consensus sequence (84/281, 30%) found in this study is nonenzymatic deamidation at the asparagine residues. As mentioned previously, 51% (43/84) of these sites have the NG motif and this is the only motif that was statistically overrepresented at these sites (Fig 4c). Hence, nonenzymatic deamidation is not surprising at these sites, since it has been shown previously that the presence of glycine adjacent to the asparagine is especially prone to nonenzymatic deamidation [323]. This includes during conditions routinely utilised for trypsin digestion prior to MS like those employed in this study (pH 7.4-8.2, 37°C)[323, 325]. Indeed, the small size of the glycine may facilitate the formation of a cyclic intermediate which favours deamidation [323].

Nevertheless, the existence of noncanonical *N*-glycosylation consensus sequences has been suggested [322] although this has been controversial and some may only be experimental artefacts [324, 327] similar to that observed in this study. While the presence of noncanonical *N*-glycosylation sites cannot be completely excluded in this investigation, the overall evidence suggests nonenzymatic deamidation is the more likely explanation for the higher percentage of deamidated sites observed. Having mentioned this, it has also been noted that spontaneous deamidation may still occur during PNGase-F digestion in ^{18}O-water [324], although this was not always considered by the addition of untreated control samples in the previous studies [259]. Taken together, the comparison of the findings in this study with the literature indicates there may be a degree of nonenzymatic remaining in the treated samples. However, the inclusion of untreated controls accompanied by the stringent criteria for identification of deamidation sites [328] lends confidence that *N*-glycosylation sites identified in this investigation containing the NXS/T motif are genuine.

Table 4. Number of deamidated sites identified in PNGase-F treated and untreated fractions. As expected, the number of deamidated sites in untreated fractions was much lower than PNGase-F treated fractions.

Fraction	Protein No.	Deamidated sites	Unique sequences
PNGase-F treated	1628	281	266
No PNGase-F	813	26	22

A PNGase-F treated

No PNGase-F

B Distribution of deamidated sites

- ■ nXS
- ■ nXT
- ■ Other

C Motif analysis by Motif-X

	Motif	No.	% unique	Fold increase
Treated	……n.T….	110	41.4	6.42
	……n.S….	79	29.7	5.39
	……nG…..	38	14.3	7.60
Untreated	……nG…..	11	50.0	7.59

D Overlap of unique sequences with nXS/T

PNGase-F treated 187 — 2 — No PNGase-F 1
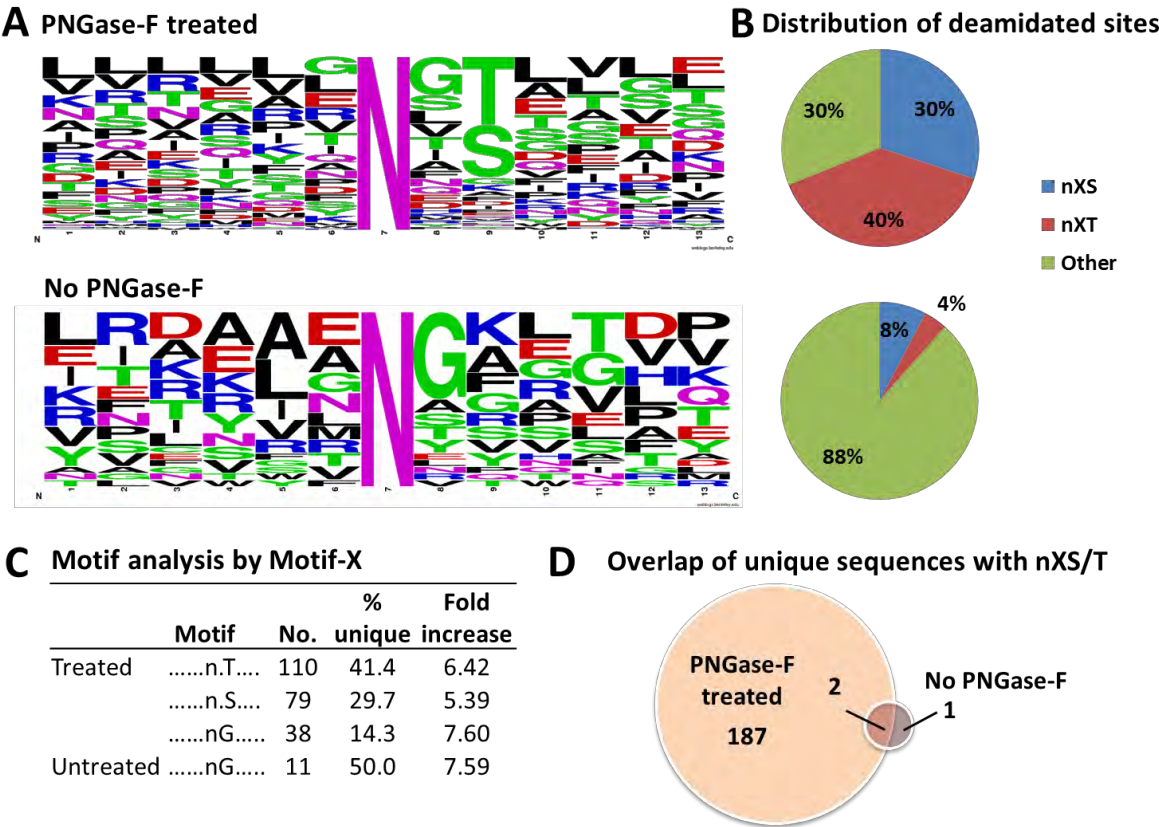
Figure 4. Motif analysis of deamidated sites indicates an increase in nXS/T motif from enzymatic deamidation (where n=Asp). A) Frequency plot of sequences surrounding the deamidated sites from PNGase-F treated and untreated samples shows an increase of serine and threonine in the third amino acid after asparagines, consistent with the well-established motif for *N*-glycosylation whereas the nonenzymatic sites showed a majority of nG motifs. **B)** Comparison of deamidation site distribution shows that the majority of deamidated sites (70%, 197/281) were in nXS/T motifs in PNGase-F treated samples, whereas they formed only a minority (12%, 3/26) in untreated fractions. One deamidated peptide in the PNGase-F treated sample had the motif nPT. No sequences containing nXC were found in either fraction. **C)** Motif analysis of unique sequences using motif-x tool indicated that the sequences nXT, nXS and nG were significantly overrepresented from background in PNGase-F treated samples. The selected background dataset was the "IPI Human Proteome" database. The foreground dataset was the sequences containing deamidated asparagines identified

this study. The fold increase was calculated as (foreground matches/foreground size)/(background matches/background size)[269]. **D)** Overlap of unique sequences with nXS/T between PNGase-F treated and untreated samples.

Table 5. Number of *N*-glycosylation sites identified and the proteins associated with these sites.

	No.
Sites with NXT	110
Sites with NXS	84
Total *N*-glycosylation sites	194
Proteins with *N*-glycosylation sites	128
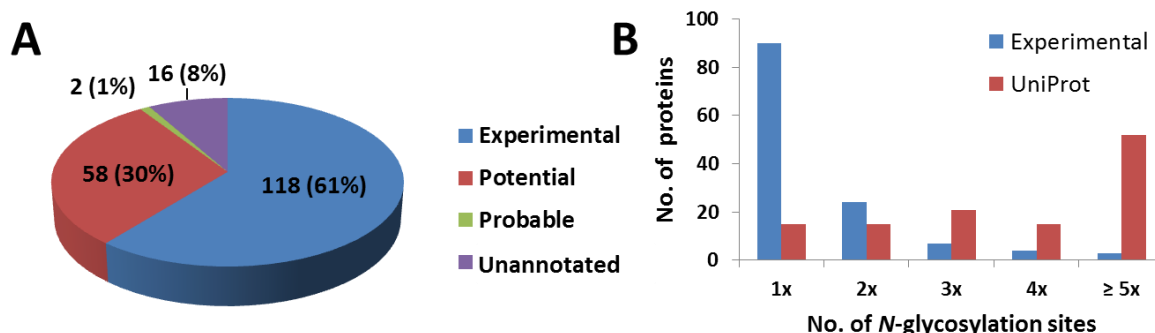
Figure 5. *N*-glycosylation sites identified in colon tissue and their UniProt annotation. A) UniProt annotation of 194 *N*-glycosylation sites from 128 proteins identified by mass spectrometry in this study. **B)** Number of *N*-glycosylation sites per protein (experimentally identified in this study vs. sites annotated by UniProt).

Overall, 194 *N*-glycosylation sites were identified which were mapped to 128 distinct proteins (Supp Table 6). Mass spectra of all *N*-glycosylation sites have been provided in Supp Fig 4. *N*-glycosylation sites with NXT motif (110 sites, Table 5) were more common than sites with NXS motif (84 sites, Table 5) and this is in agreement with previous reports of eukaryotic *N*-glycosylation including humans [19, 259]. When these were matched with sequence annotation from UniProt (release 2014_06), 61% (118 sites, Fig 5a) were reported previously with experimental verification. However, 60 sites (31%, Fig 5a) were annotated either as potential (30%, 58 sites) or probable sites (1%, 2 sites) and therefore, the findings in this study provided experimental confirmation for these *N*-glycosylation sites.

Table 6. Unannotated *N*-glycosylation sites identified and their corresponding Ascore and localisation probability (from Scaffold PTM).

Uniprot Accession	Protein name	Motif	Surrounding Sequence	Site	Best Ascore	Localization Probability
Cytoplasm						
DESP_HUMAN	Desmoplakin	......n.T....	TISSVRNLTIRSS	N2601	1000	100.00%
K6PL_HUMAN	6-phosphofructokinase, liver type	......n.S....	PPKEKSNFSLAIL	N399	45.54	100.00%
MYH9_HUMAN	Myosin-9	......n.T....	SHLLGINVTDFTR	N381	1000	100.00%
RL18_HUMAN	60S ribosomal protein L18	......n.T....	FLARRTNSTFNQV	N40	21.82	100.00%
VIME_HUMAN	Vimentin	......n.T....	RDGQVINETSQHH	N456	1000	100.00%
Membrane						
ARF1_HUMAN	ADP-ribosylation factor 1	......n.S....	ETVEYKNISFTVW	N60	1000	100.00%
AT131_HUMAN	Probable cation-transporting ATPase 13A1	......n.S....	VVPTPNNGSTELV	N150	26.2	100.00%
AT1A1_HUMAN	Sodium/potassium-transporting ATPase subunit alpha-1	......n.S....	EADTTENQSGVSF	N405	52.77	98.90%
CASD1_HUMAN	CAS1 domain-containing protein 1	......n.S....	AAVSILNSSTRNS	N241	136.26	100.00%
CC50B_HUMAN	Cell cycle control protein 50B	......n.S....	ARIRQGNYSAGLP	N273	1000	100.00%
LEG4_HUMAN	Galectin-4	......n.T....	INPRMGNGTVVRN	N243	1000	100.00%
SIA7F_HUMAN	Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 6	......n.T....	YSADVGNKTTYRV	N149	161.02	100.00%
Secreted						
FCGBP_HUMAN	IgGFc-binding protein	......n.T....	ITVQVANFTLRLE	N1317	1000	100.00%
OLFM4_HUMAN	Olfactomedin-4	......n.T....	GNIARVNLTTNTI	N352	23.78	100.00%
OLFM4_HUMAN	Olfactomedin-4	......n.T....	MVISKLNDTTLQV	N411	120.81	100.00%
Unknown[a]						
LEG9B_HUMAN	Galectin-9B	......n.S....	VRNTQINNSWGSE	N285	58.42	100.00%

a) Cellular localisation not annotated as either membrane, cytoplasm or secreted.

Table 7. *N*-glycosylation sites identified from carcinoembryonic antigen 5 and immunonglobulin J and their corresponding Ascore and localisation probability (from Scaffold PTM).

Uniprot Accession	Protein name	Motif	Surrounding Sequence	Site	Best Ascore	Localization Probability	Uniprot annotation
CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5	......n.T....	IQNIIQNDTGFYT	N115	27.01	99.80%	Potential
CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5	......n.S....	KPSISSNNSKPVE	N152	46.67	100.00%	Potential
CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5	......n.T....	LFNVTRNDTASYK	N208	1000	100.00%	Potential
CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5	......n.S....	PTISPLNTSYRSG	N246	341.31	100.00%	Experimental evidence
CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5	......n.T....	LQLSNDNRTLTLL	N375	41.7	100.00%	Potential
CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5	......n.S....	KPSISSNNSKPVE	N508	40.66	100.00%	Potential
IGJ_HUMAN	Immunoglobulin J chain	......n.S....	PLNNRENISDPTS	N71	34.27	100.00%	Experimental evidence

In addition, a further 16 sites (all containing NXS/T motif) that have not been previously annotated as potential by UniProt (Table 6) was also identified in this study. These 16 unannotated sites were mapped to 15 proteins (one protein had two sites). The majority of the corresponding proteins (60%, 9/15) were membrane-bound or secreted, 33% (5/15) were cytoplasmic and 7% (1/15) were not linked to either the cytoplasm or membrane. These unannotated sites from the membrane/secreted proteins (10 sites) were simply sequons (consensus sequences with NXS/T motif) identified as less likely to be *N*-glycosylated by NetNGlyc, the *N*-glycosylation site prediction tool utilised by UniProt [264]. The identified sites had NetNGlyc result scores of less than 9/9 or a threshold of less than 0.5 as determined by computational neural networks [329]. Hence, these sites were not considered as potential sites by the database as they were not predicted as sites with the highest confidence. Despite this, it is very plausible that these are genuine *N*-glycosylation sites since they are located on membrane-bound or secreted proteins. Thus, they can be *N*-glycosylated according to canonical glycosylation pathways through the endoplasmic reticulum (ER) and Golgi apparatus [25]. Therefore, this study has provided the first evidence for the *N*-glycosylation of these sites.

Intriguingly, there were also 5 other unannotated sites mapped to 5 cytoplasmic proteins (such as myosin 9, desmoplakin and vimentin) which are not classified as membrane proteins (Table 6). Moreover, none of the cytoplasmic proteins contained any predicted transmembrane helices or signal peptides (as determined by prediction tools, TMHMM v 2 and SignalP v 4.1) to indicate a potential alternative role as membrane or secreted protein. As cytoplasmic proteins are not generally known to be *N*-glycosylated, whether the deamidated peptides associated with these sites are exclusive or homologous to other proteins was also determined. All deamidated peptides were found to be exclusive to their corresponding identified proteins, further showing that they were not linked with the membrane. Additionally, the indicated *N*-glycosylation sites were examined to determine whether the asparagine was adjacent to glycine, which is known to be susceptible to nonenzymatic deamidation [323, 325]. None of the sites were located next to glycine and thus, there is little evidence to show that they were an experimental artefact. Taken together, the evidence appears to point to the existence of 5 novel *N*-glycosylation sites on proteins considered to be cytoplasmic, that have not been reported before. Having mentioned this, their existence is thought-provoking since *N*-glycosylation is only known to occur on membrane-bound/secreted proteins synthesised in the ER or Golgi. At present, alternative mechanisms for cytoplasmic *N*-glycosylation have not been shown. Given that the glycan moiety was not analysed in this study, more experimental validation, by identification of the actual glycosylated peptide(s) with the glycan attached, is needed to show that these proteins are indeed *N*-glycosylated. Regardless, it is a provocative finding that merits further experimental verification as it suggests the possibility of noncanonical *N*-glycosylation and alternative *N*-glycosylation pathways, at least in colon tissues.

288

Besides this, the number of *N*-glycosylation sites observed experimentally was also compared with the number of *N*-glycosylation sites annotated by UniProt. Experimentally, one *N*-glycosylation site was identified for the majority of proteins while fewer proteins were identified with multiple *N*-glycosylation sites (Fig 5b). This trend is consistent with previous study performed by Deeb *et. al.* where *N*-glycosylation sites from human lymphoma cell lines were also identified using a deglycosylation mass spectrometry-based approach [259]. Despite this, it was noted that the number of proteins annotated with multiple *N*-glycosylation sites (5+; only proteins identified in this study considered) was much higher in UniProt than experimentally identified (Fig 5b, red vs blue). This indicates that not all sites from highly *N*-glycosylated proteins were mapped in this experiment. Having observed this, an association between the number of experimentally identified sites and the abundance of the corresponding proteins was also considered. However, there was also no association between number of experimentally observed *N*-glycosylation sites and protein abundance (Supp Fig 2). On the whole, this suggests alternative strategies are required to target proteins with a high number of *N*-glycosylation sites regardless of the protein's abundance if all sites are to be identified. Nevertheless, the mass-spectrometry based method presented here still enabled high throughput confirmation of predicted *N*-glycosylation sites and identification of previously unreported *N*-glycosylation sites (Fig 5a). Therefore, this approach is likely to be complementary to others such as glycopeptide analysis where more detailed information about the attached glycan moiety is obtained, albeit also incomplete at this stage of technology development.

Next, it was determined whether these sites were identified in any of the altered glycoproteins found between tumour and normal tissue. In total, the expression of 9 identified glycoproteins were significantly changed (Table 2). This includes the well-known biomarker, carcinoembryonic antigen (CEAM5_HUMAN) where 6 of the 28 predicted *N*-glycosylation sites were found (N115, N152, N208, N246, N375, N508) and immunoglobulin J chain (IGJ_HUMAN) where the only known *N*-linked site (N71) was found as shown in Table 7. For example MS2 spectra from colon tumour tissues treated PNGase-F showing deamidation at N246 (parent peptide RSDSVILNVLYGPDAPTISPLnTSYR) and N508 for CEAM5 (parent peptide TITVSAELPKPSISSnNSKPVEDK), and at N71 on immunoglobulin J (parent peptide IIVPLNNREnISDPTSPLR), are shown in Fig 6a, c and d. Interestingly, the corresponding non-glycosylated peptides (intact or "native" Asn residue) from these sites was also found in the PNGase-F treated fractions, indicating less than 100% glycan occupancy of these protein sites in CRC tumours. For example, MS2 showing N246 of CEAM5 without deamidation (non-glycosylated) is shown in Fig 6b. However, neither the non-glycosylated or deamidated versions were found in the PNGase-F untreated fractions (normal or tumour), possibly due to low abundance. In terms of other proteins, the non-glycosylated "native" versions of the formerly *N*-glycosylated peptides were also generally

not found in PNGase-F treated fractions (normal or tumour). Only four other *N*-glycosylation sites (N60 of ARF1, N75 of FCGBP, N381 of MYH9 and N118 of TSN8) out of 194 were noted to have their corresponding non-glycosylated forms in the treated fractions.

Since limited information is reported for the alteration of glycosylation sites in colorectal cancer, the deamidated and native glycopeptides covering N246 and N508 from CEA5 were quantified in normal and tumour samples using NSAF (Fig 7). These two sites were chosen because they were consistently detected across all three cancer patients (min. one peptide per patient). This analysis indicated that there is an overall higher abundance of both deamidated and native peptides in tumour samples which reflects the higher general expression of CEA5 in tumours compared to normal tissues. However, the average fold-increase in deamidated peptides of N246 was still higher (6.7-fold, Fig 7) than the average increase in CEA5 (3-fold, Table 2) in tumours as determined by spectral counts of peptides with and without *N*-glycosylation sites. This suggests that the increase in deamidated peptides is not due only to an increase in protein abundance but also an increase the glycan occupancy of CEA5 at N246 in the tumours compared to normal tissues. Furthermore, there is a trend towards a higher ratio of deamidated:native N246 peptides (p = 0.065) compared to deamidated:native N508 peptides within tumours. Thus, while an increase in overall CEA5 contributes to the increased abundance of its corresponding deamidated (formerly *N*-glycosylated) peptides, it is also likely there are positional preferences for *N*-glycosylation along the polypeptide chain – a phenomenon which has been previously reported [330]. In general, *N*-glycosylated sites tend to be more common toward the *N*-terminus [331] and glycosylation efficiency appears reduced within 60 residues of the C-terminus [332]. Thus, the findings of CEA5 in this study support these previous observations with N246 trending a higher deamidated:native peptide ratio compared to N508 within tumours, indicating higher glycosylation at the site closer to *N*-terminus. Mechanistically, *N*-terminally located sites may spend more time being exposed to oligosaccharyltransferase (the sugar-adding enzyme) as they are translated and this may contribute to their increased likelihood of glycosylation.
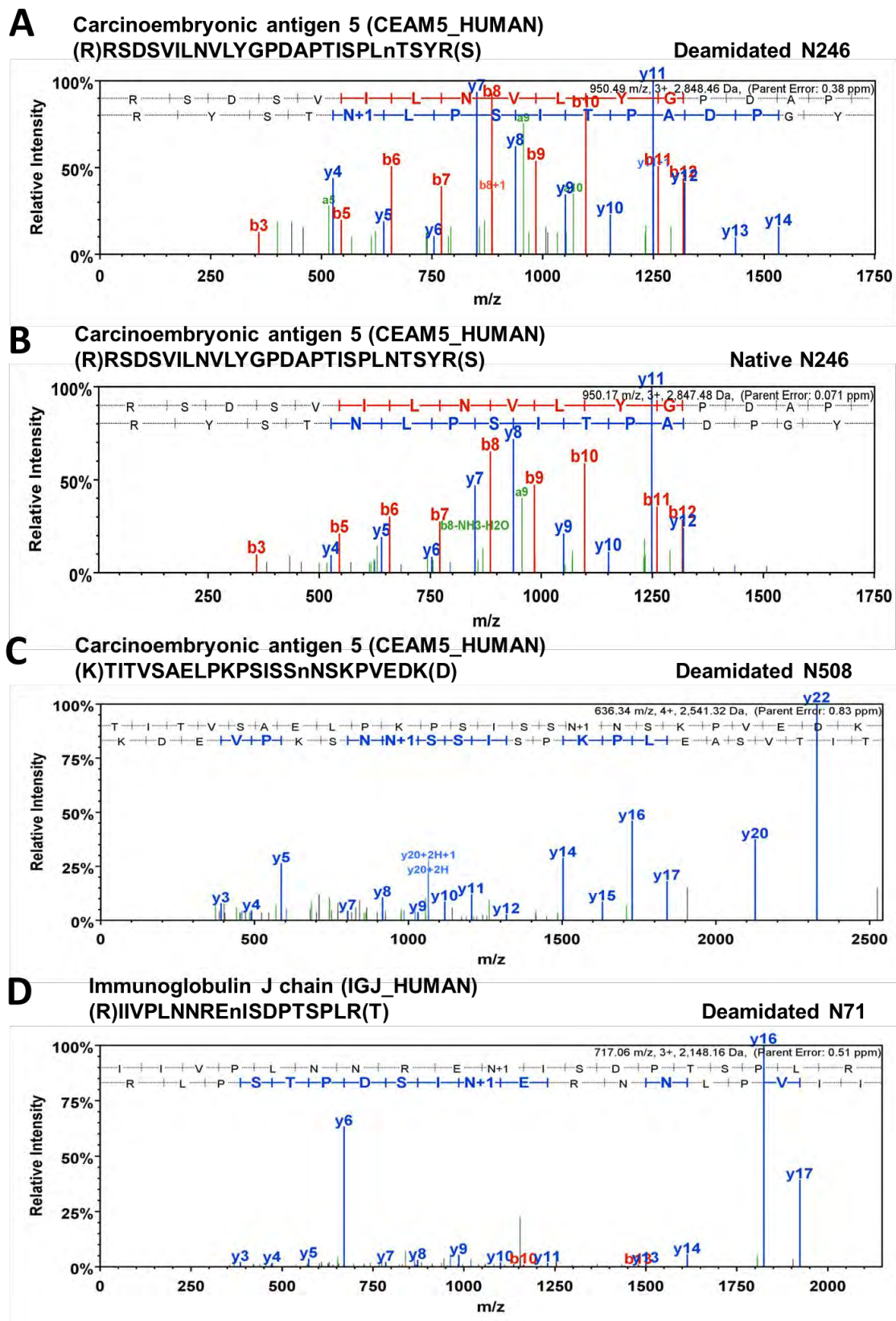
Figure 6. Example of tandem mass spectra from altered glycoproteins with deamidated (formerly *N*-glycosylated) sites. A-B) Mass spectra from carcinoembryonic antigen 5 (CEAM5_HUMAN). Asparagine 246 from this protein was found in deamidated (A) and native (non-glycosylated) forms (B). **C)** Deamidation of asparagine 508 from CEA5. **D)** Deamidation of asparagine 71 from immunoglobulin J.

N-glycosylation site status of CEA (CEAM5_HUMAN)
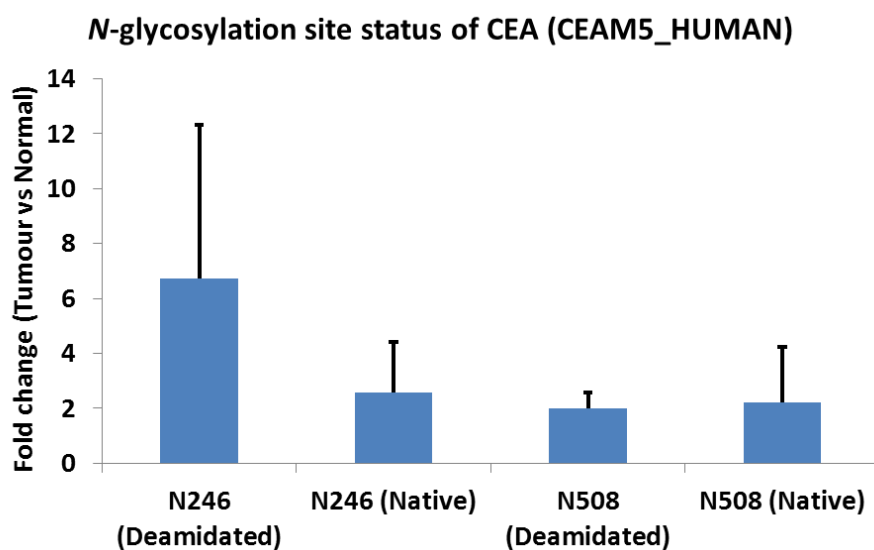
Figure 7. **Quantification of peptides with formerly *N*-glycosylated (deamidated) and non-glycosylated (native) asparagine 246 and asparagine 508 in carcinoembryonic antigen 5.** These sites were chosen for quantification since they were consistently detected (min. 1 peptide) in all three patients.

Additionally, for reasons not yet clear, there is also an increase in the ratio of deamidated:native N246 peptides between tumours and normal tissue, indicating a relative increase in glycosylation site occupancy in the tumours. Though the observations presented require additional samples for validation, this preliminary result suggests that glycosylation site occupancy may be correlated with carcinogenesis. This is not entirely surprising as differences in *N*-glycosylation site occupancy have been formerly associated with pathology. In a study by Hulsmeier *et. al*.[333], *N*-glycosylation occupancy of serum transferrin from individuals with congenital disorders of glycosylation (CDG) was compared with healthy controls samples. Of the two *N*-glycosylation sites (N413 and N611), the percentage occupancy of the second site at N611 was correlated with disease severity with highest occupancy for healthy controls and lowest observed for patients with severe form of CDG [333]. Moreover, *N*-attachment has been shown to influence biological properties of a glycoprotein such as its stability, solubility, half-life and enzymatic activity [330]. For example, *N*-glycan occupancy at N25 increased the resistance of interferon-gamma (IFN-γ) to protease digestion compared to wild-type IFN-γ [334]. Given that CEA is generally increased in CRC patients, increased glycosylation occupancy may be one potential mechanism which contributes to extending its serum half-life and therefore its increased expression in CRC. Hence, it would be interesting in future studies to confirm whether glycosylation site occupancy of N246 is increased in patient-derived tumour tissue or serum. Such changes in glycosylation site occupancy may also be exploited for assay development as biomarkers in combination with CEA protein expression for improvement of marker sensitivity and specificity.

Glycoprotein changes in CRC tumours compared to matched normal tissues

Apart from changes in CEA, other glycoproteins were also found to be differentially altered during carcinogenesis and many of these have been previously associated with CRC. Differences in *N*-glycosylation sites were not compared for these proteins (mentioned below) either because their corresponding deamidated peptides were not identified in the dataset; or in the case of immunoglobulin J, the deamidated site at N71 was identified in one patient only.

As examples of differentially expressed glycoproteins, CD44, zinc transporter ZIP14, serpin H1, carbonic anhydrase 12 (CA12) were increased while cocaine esterase was decreased (Table 2) in tumours compared to normal tissues in this study. Taken into account directionality of regulation (whether increased/decreased), the alteration of these proteins is in good agreement with the literature where CD44, ZIP14, serpin H1 are generally increased during CRC progression [335-339]. These proteins have a variety of biological functions which likely support malignancy. For example, CD44 overexpression in colon cancer cells increases cell migration/invasion and induces activation of the pro-growth Akt signalling pathways [273]. On the other hand, ZIP14 controls transport of divalent cations such as zinc and iron, which may have implications for DNA repair in tumour cells [340] while serpin H1 is a molecular chaperone that is involved in the synthesis of collagen, a component of the extracellular matrix which is involved in signalling networks regulating cell proliferation [341]. Working in a concerted manner, it is likely that the overexpression of these proteins contribute to carcinogenesis by affecting pathways involved in DNA damage, increased tumour growth and invasion.

Interestingly, though the downregulation of cocaine esterase (also known as carboxylesterase 2) in tumour tissue has also been reported in a variety of cancers, its expression in CRC was not markedly changed compared to normal tissue [342]. However, in this study, cocaine esterase was decreased 4.2-fold in the tumours compared to tissues (Table 2). While its precise role in carcinogenesis is unclear, one of its major functions is in the metabolism of drugs, including therapeutic prodrugs such as irinotecan which is actively used to treat colorectal cancer. Through hydrolysis, irinotecan is activated to generate SN-38 which is known to be 100 times more potent than irinotecan [342]. Thus, its decreased expression in colon tumour tissue may have implication for lower drug activation and overall tumour response to this drug [343]. While more experimental confirmation is required, this may ultimately have downstream effects on patient outcomes.

Although several glycoproteins have been found correlated with CRC, others were identified linked with CRC for the first time, namely immunoglobulin J chain and coiled-coil domain containing protein

47 (CCDC47) which was decreased 3.4 fold and increased 9.8-fold in this study, respectively (Table 2). Very limited information is known about CCDC47 though its coiled-coil domain may have functions in force production [344], cell motility [345] and membrane localisation of proteins [346]. Most importantly, coiled-coil domains of membrane proteins are believed to mediate protein-protein interactions [347]. This includes protein interactions of occludin which has been suggested to play a role in the maintenance of epithelial tight junctions which serve to bind adjacent epithelial cells [348]. Overall, it is unclear at present how elevated expression of CCDC47 influences cancer behaviour and further research is needed to elucidate its biological functions in colorectal cancer.

Immunoglobulin J chain was also shown to be downregulated in CRC tumour cells by 3.4-fold (Table 2). This immunoglobulin polypeptide plays a role in the initiation of IgA and IgM polymerisation [349]. The polymerisation of IgA into dimers and IgM into pentamers leads to highly valent antigen binding sites well-suited to binding foreign pathogens [350]. Its downregulation may indicate reduced immune surveillance in the tumour microenvironment, which may contribute to immune escape of the cancer cells. Interestingly, a component of the MHC class II molecule, the MHC class II histocompatibility antigen DRB1-16 beta chain, was upregulated (5.8-fold) in the same patients. Though it is unclear whether the MHC class II histocompatibility antigen is expressed by epithelial cells [351], antigen presenting cells or both, its presence suggests a degree of immune response in the tumour. Therefore, it appears that regulations of immune surveillance is highly intricate and require further investigation for deeper understanding of underlying molecular mechanisms. Nevertheless, the differential regulation of these two proteins strongly indicates alteration of immune-associated proteins, which may have implications in immune escape of tumours cells in colorectal cancer.

Limitations of current LC-MS/MS approach

Although spectral counting is commonly employed for label-free quantitation in proteomics due to its simplicity, many other approaches exist. In order to improve relative quantification, the use of isotopic labels may be employed to increase the robustness of quantification. The distinct advantage of using a labelled approach is that different samples may be mixed (following labelling) to minimise any differences in sample handling that may influence protein quantification. In addition, quantification of multiple samples can be performed simultaneously in the same MS run, further reducing variance associated with run-to-run spectral comparisons [352]. One labelled approach commonly utilised for human tissue samples is iTRAQ labelling where peptides from various samples are labelled with isobaric tags that are fragmented by MS/MS to give different reporter ions. The relative abundance of the reporter ions are used as a proxy to quantify proteins [353]. This approach has been successfully employed in an earlier study to investigate changes the in membrane

proteome of CRC tissues [57]. However, due to the limited recovery of peptide in this study, we chose not to utilise this method.

An alternative method is the use of SILAC (stable isotope labelling of amino acids in culture) labelling. As a form of metabolic labelling, SILAC has traditionally been used to quantify proteomes of cell lines since they are able to gradually incorporate 'light' or 'heavy' isotopes of amino acids over several generations. Nevertheless, a mix of SILAC-labelled cell lines (referred to as the super-SILAC mix [354]) can be spiked in as an internal standard for relative quantification of human tissue samples. A mix of different cell lines (of same cancer type) is used in order to provide internal standards to a representative number of proteins. Although this approach is likely to be more time-consuming than iTRAQ (as time is needed to generate several fully labelled SILAC cell lines), the limited sample recovery is less likely to be a concern because further sample processing was minimal compared to the label-free approach. Therefore, this method is promising as a future direction to improve the robustness of relative quantification of (formerly) *N*-glycosylated peptides investigated in this study.

4.4   Conclusions and future perspectives

In this study, membrane proteomic profiles from epithelial cells of patient-matched normal and tumour tissue was performed which resulted in the quantification of 1648 proteins – the highest number of proteins to date identified in membrane proteomics studies comparing normal and tumour samples. On the whole, 52 proteins including 9 glycoproteins were differentially expressed with altered protein abundances involved in nucleotide binding, protein transport, immunity and cell proliferation. The differential expression of known CRC associated proteins were validated and other proteins not previously implicated in CRC were identified.

As glycosylation is the main focus of this thesis, the glycoprotein subset of the proteome was investigated further. Specifically, the *N*-glycosylation sites were mapped after enzymatic deglycosylation as little is known about *N*-glycosylation site changes in colorectal cancer. 194 sites from 128 proteins were found which included confirmation of 60 potential sites and identification of a further 16 sites not annotated as being glycosylated in UniProt. This included the confirmation of 5 potential *N*-glycosylation sites on the currently used cancer marker, CEA. Quantification of formerly glycosylated and non-glycosylated peptides from CEA suggests *N*-glycosylation site differences may occur in CRC which warrant further investigation as these differences may be exploited to improve biomarker sensitivity and specificity.

Overall, this study has shown the viability of simultaneous membrane and *N*-glycosylation site profiling of proteins on a small sample set using high accuracy mass spectrometry. When combined

with *N*-glycosylation profiles obtained from the same samples (Chapter 3), there is new depth of information on the glycosylation and protein changes which occur during carcinogenesis. Though at this stage the global structural glycosylation changes observed cannot be directly attributed to specific proteins, these may be further investigated by immunopurification of target proteins followed by analysis of the intact glycopeptides. In addition, while the number of patients in this study is small, the experimental workflow was validated by the known biomarkers found and can be easily upscaled to a larger patient cohort. Targeted analysis of the carriers of the specific glycans found in this study (Chapter 3) may expedite throughput. Taken together, the application of this experimental approach provides a platform for in-depth exploration of glycosylation-related changes which may provide novel leads for biomarker discovery.

4.5 Supplemental information

Additional Supplemental information (Supp Tables 2-6, Supp Fig 4) is located in the provided CD (inserted in thesis).

Supplementary Table 1. Epithelial and non-epithelial cellular markers used for determination of epithelial enrichment by proteomics. EMT, epithelial-to-mesenchymal transition.

Cellular marker	Protein name	Protein short name	UniProt ID	Description	Ref(s)
Epithelium	Epithelial cell adhesion molecule	EpCAM	EPCAM_HUMAN	Cell surface antigen for epithelial cells	[355, 356]
	Keratin 8	Keratin 8	K2C8_HUMAN	Intermediate filament proteins expressed by epithelial cells	[270]
	Keratin 18	Keratin 18	K1C18_HUMAN	Intermediate filament proteins expressed by epithelial cells	[270]
	Keratin 19	Keratin 19	K1C19_HUMAN	Intermediate filament proteins expressed by epithelial cells	[270]
	Villin-1	Villin-1	VILI_HUMAN	Actin binding protein expressed on epithelial microvilli	[357-359]
Fibroblast/ mesenchymal	Vimentin	Vimentin	VIME_HUMAN	Intermediate filament protein expressed by fibroblast and mesenchymal cells	[271] [272]
Mesenchymal	CD44 antigen	CD44	CD44_HUMAN	Transmembrane receptor for hyaluronan associated with EMT	[273, 274]
Muscle	Actin gamma, smooth muscle	Actin gamma	ACTH_HUMAN	An actin isoform expressed by vascular/enteric smooth cells and myofibroblasts	[360]
	Caldesmon	Caldesmon	CALD1_HUMAN	Actin-associated protein marker for smooth muscle cells	[361]
	Calponin	Calponin	CNN1_HUMAN	Actin-binding protein specifically by smooth muscle	[362]
	Desmin	Desmin	DESM_HUMAN	Intermediate filament protein expressed by smooth muscle cells	[363]
Immune cells	T-cell surface glycoprotein CD4	CD4	CD4_HUMAN	Membrane receptor expressed by cytotoxic T-cells	[364]
	T-cell surface glycoprotein CD8	CD8	CD8_HUMAN	Antigen receptor expressed by predominantly by helper T-cells	[365]
	B-lymphocyte antigen CD19	CD19	CD19_HUMAN	B-cell surface receptor involved in cell signalling	[366]
	Receptor-type tyrosine protein phosphatase C (CD45)	CD45	PTPRC_HUMAN	Membrane glycoprotein found on all hematopoietic cells except erythrocytes	[367]

Supplementary Figure 1. Cellular localisation of membrane proteome by percentage NSAF.

Cellular localisation by abundance (% NSAF)

■ Membrane ■ Cytoplasm ■ Membrane & cytoplasm ■ Other
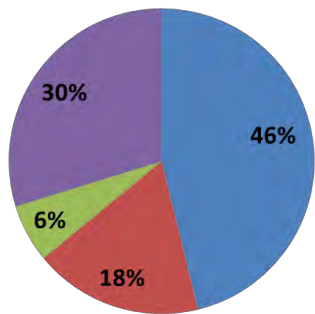
Supplementary Figure 2. Correlation between number of experimentally identified *N*-glycosylation sites and protein abundance determined by normalised spectral abundance factors (NSAF). Each data point represents the number of *N*-glycosylation sites and protein abundance of one protein.
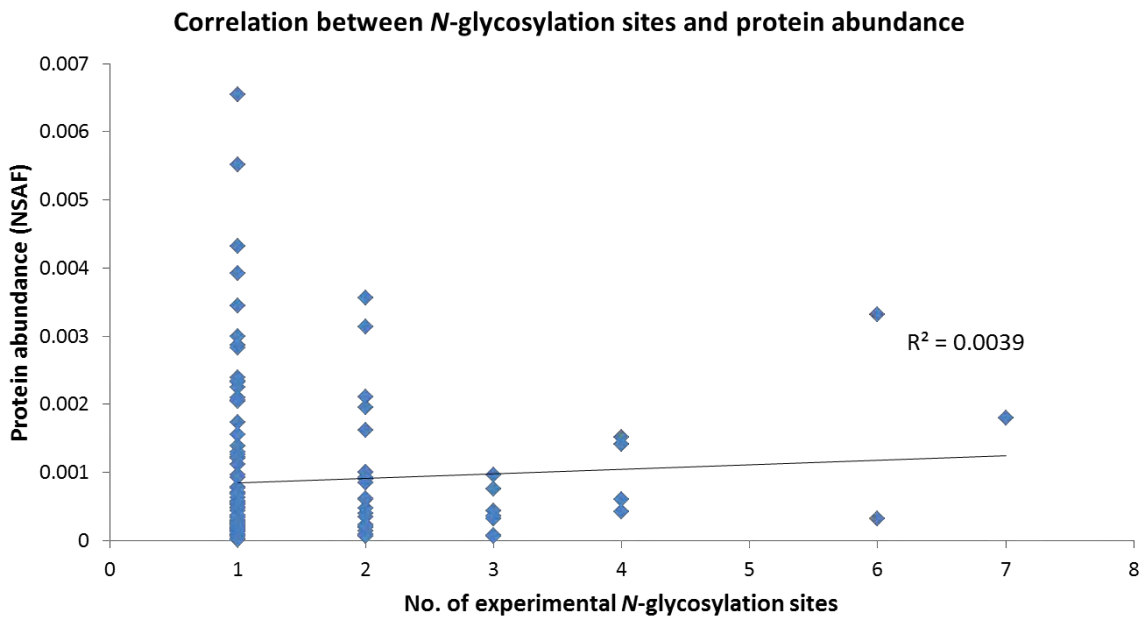
Supplementary Figure 3. Dependence of deamidation on the amino acid adjacent to the deamidated asparagine (n=84) at sites without the *N*-glycosylation consensus sequence (NXS/T).

Amino acid deamidation at sites without NXS/T
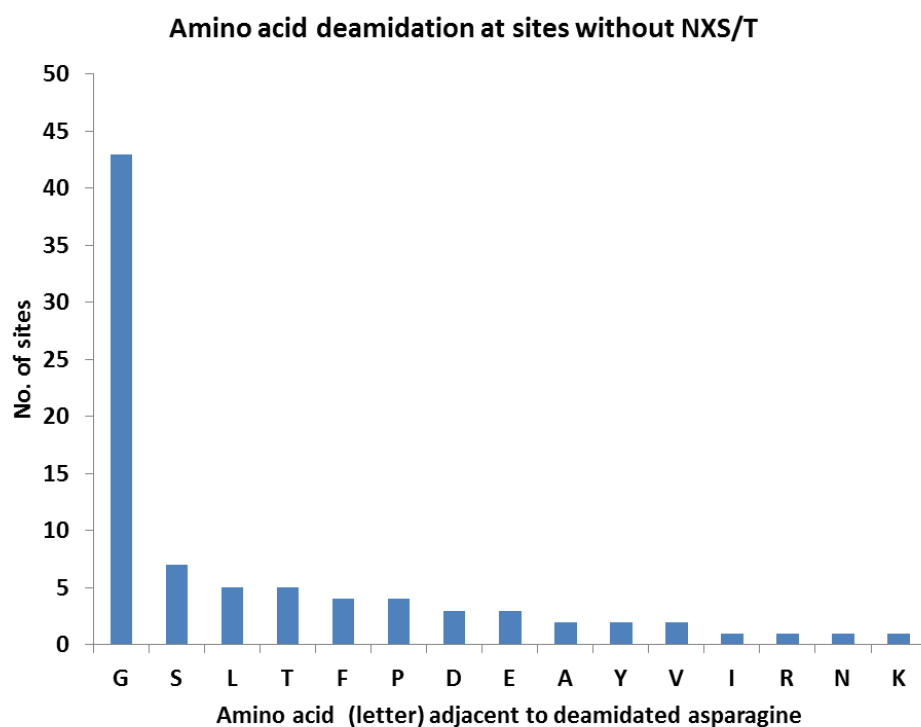
Chapter 5

Final discussion and conclusions

5 Final discussion and conclusions

In this final discussion and conclusions chapter, an overview of the findings presented in the thesis will be summarised. This will be followed by further discussion on the use of CRC cell lines as model systems and the utility of mass spectrometry in glycomics/glycoproteomics. Taking the above into account, various potential future directions following this thesis will be considered. Finally, the overall conclusions and perspectives will be presented.

5.1 Thesis overview

Colorectal cancer (CRC) is a prevalent disease that is newly diagnosed in approximately one million people worldwide each year [41]. There is still a need for novel biomarkers for this cancer since the currently approved FDA marker, carcinoembryonic antigen has limited sensitivity, which makes it unsuitable for early CRC detection [213]. The aim of this thesis was to investigate potentially novel sources of CRC biomarkers by examining membrane glycoproteins and *N*- and *O*-linked glycans released from these proteins in order to understand the biochemical processes underlying the production of these biomolecules in the context of CRC. The approach centred on using mass spectrometry for detection and comparison.

Several colon cancer cell lines were compared with colon tumour tissue to evaluate whether cell model systems provide a good representation of *in vivo* glycosylation observed in tumours. For the purpose of achieving a *bona fide* comparison with the epithelial origin of the cancer cell lines, a positive selection method based on EpCAM immunoaffinity was developed to enrich epithelial cells from tumour tissues prior to MS analysis (Chapter 2). The reduction of non-epithelial cell types allowed for a closer comparison with cell lines, which are a homogeneous epithelial population. The *N*- and *O*-glycan profiles of membrane proteins of the tumour tissue were compared with those from the cell lines using MS (detailed in Chapter 2). To date, this is first protein glycan analysis of colon cancer which has incorporated an additional enrichment step for epithelial cell selection. The results of *O*-glycan analysis were especially striking as the well-known glycan cancer marker sialyl-Tn, highly expressed in the EpCAM-enrich fractions from colon tumours, was undetected in the majority of cell lines. Moreover, a second potential glycan marker also detected in the tumour epithelial cells, which is a sialyl-Lewis X hexasaccharide, was also absent in the cell lines. The sialyl-Tn antigen was only produced in a mucinous cell line (LS174T). Further investigation suggested that sialyl-Tn formation in this cell line was associated with: 1) reciprocal regulation of the biosynthetic enzymes C1GalT1 and ST6GalNAc1, and 2) the expression of membrane-associated mucins including mucin-2, mucin-5B, and mucin-6. The association of this glycan with mucins is unsurprising given that sialyl-Tn has previously been identified as a pan-carcinoma cancer marker [368] carried by mucins [102, 369]. Thus, the findings presented here are consistent with mucins being the major protein carriers of

sialyl-Tn. More specifically, the evidence is consistent with previous reports that MUC2 carries sialyl-Tn [102, 122], although the data (from chapter 2) also implicates MUC5B and MUC6 as sialyl-Tn carrying glycoproteins. Thus, it is suggested that these mucins and sialyl-Tn are absent or lowly expressed in the non-mucinous cell lines examined.

In contrast, fewer overt changes were observed for the *N*-glycan profiles, where glycan abundances altered but no abundant glycans were expressed solely in either sample type. Nevertheless, when *N*-glycans were grouped into various classes, the expression of sialylated, fucosylated and bisecting *N*-glycans was found to vary between different cell lines. When investigated further, the expression of bisecting *N*-glycan structures corresponded well with the abundance of the relevant glycosyltransferase, GlcNAcT-III (encoded by the gene MGAT3).

Overall, this work (Chapter 2) showed that there are glycosylation differences between different cultured colon cancer cell lines and also between these cell lines and the epithelial cells of CRC tumours, especially in regards to *O*-glycans. These observations suggest that changes in the expression of glycan biosynthetic machinery and glycan protein carriers in different cellular environments can lead to the specific expression of particular membrane glycans. The findings also emphasise the need for clinical tissues to be analysed during the biomarker discovery phase and cautions against the sole use of cell models since known glycan markers have been shown in this study to be absent. This will be further discussed in the next section (5.2).

In the light of the above findings, a search for putative glycan biomarkers of CRC was conducted by using pair-matched normal and cancer colon tissues (Chapter 3). Though some previous work has been performed comparing mucin *O*-glycosylation [122, 177] and protein *N*-glycosylation in CRC [38], limited information is known about the glycosylation that is specifically characteristic of total membrane proteins from tumour epithelial cells. Comparison of the *N*- and *O*-glycans of membrane proteins from epithelial cells enriched by EpCAM immunoaffinity was carried out using LC-MS/MS. As an initial discovery approach, a small cohort (n=6 and n=3) of matched normal and cancerous tissue was analysed for their *N*- and *O*-glycosylation profiles respectively. For *O*-glycans, sialyl-Tn and the sialyl-Lewis X hexasaccharide were found to be increased in the tumours compared to normal tissue. Though these findings were not statistically significant because of large individual variations and limited numbers of measurements, the results presented agree with previous studies [78, 170]. Another previously unreported *O*-linked trisaccharide of composition $(HexNAc)_2 (NeuAc)_1$ with *m/z* 716 was also observed to be decreased in the tumours ($p < 0.05$). As mucins were implicated as the carriers of sialyl-Tn in the LS174T cell lines, mucin-2 and mucin-13 were quantified in a proteomics study (Chapter 4). These were quantified since they were consistently detected across all patients

although other mucins (such as mucin-1, mucin-4 and mucin-5B) were also detected. The quantification of mucin-2 and mucin-13 showed that mucin-2 showed a decrease in the tumour cells while mucin-13 was stable, which suggested that sialyl-Tn and mucin 2/13 expression are not positively correlated in the tissue environment. Given that mucin-2 is implicated as a major carrier of sialyl-Tn, this suggests that there may be an increase in sialyl-Tn being carried by mucin-2 in tumours concomitant with a degree of mucin-2 down expression. Alternatively, sialyl-Tn may be carried by a variety of other glycoproteins that have yet to be characterised. Moreover, the absence or inconsistent detection of mucin-5B and mucin-6 suggests these glycoproteins may be of lower abundance in the examined tumour tissues. Their heavily glycosylated nature may contribute to this observation as glycans tend to shield the polypeptide backbone from proteolytic cleavage.

In addition, analysis of the abundance of *O*-glycan cores showed that the core 1 *O*-glycan (Galβ1-3GalNAc or the T-antigen) were differentially increased in tumours ($p < 0.05$) compared to the matching normal tissue, which was mirrored by a downward trend in core 3 (GlcNAcβ1-3GalNAc) glycans. Again reinforcing the difference in cell lines and tumour tissue, the increased expression of core 1 glycans was associated with increased expression of sialyl-Tn (NeuAcα2-6GalNAc) in colon tissues contrasting with the decreased expression of core 1 glycans correlating with increased sialyl-Tn in the LS174T cell line (Chapter 2). This observation may be explained by a concomitant increase in the activity of C1GalT1 and ST6GalNAc1 glycosyltransferases in the tumours. As this occurs, there may be a reciprocal decrease in core 3 synthase activity, which may explain the decreased core 3 glycans in the tumours. Having mentioned the above, the activities of the relevant glycosyltransferases are regulated by many different factors including nucleotide sugar transporters, substrate availability (such as CMP-sialic acid, UDP-GalNAc) and its own protein expression. Thus, a complex interplay of various components may contribute to the tissue glycan changes observed. Nevertheless, changes in glycosyltransferase expression may be determined as one of the many factors to be investigated during further hypothesis testing. This will be discussed further in section 5.5.

On the other hand, for the *N*-glycans, bisecting and paucimannosidic type *N*-glycans were found altered in CRC tumours. Bisecting GlcNAc *N*-glycans were found decreased ($p = 0.017$) while paucimannosidic type *N*-glycans were found to be increased in the tumours ($p = 0.022$). These changes in bisecting and paucimannose *N*-glycans between normal and tumour tissues are in agreement with the only other study on the detailed glycan structures in CRC published by Balog *et. al.* [38], that analysed the protein glycosylation of pair-matched normal and tumour colon tissues from 13 patients. However, when individual *N*-glycans were quantified, some *N*-glycans were reported as differentially expressed for the first time in this thesis. In total, 12 differentially

expressed *N*-glycans were observed in this study but only one structure (bisecting *N*-glycan 65, Table 2, Chapter 3) was observed by Balog *et. al.* as differentially expressed. Of the 12 glycans, 5 were bisected *N*-glycans that were decreased more than 2-fold in the tumours, indicating that multiple structures were synthesised by the differential activity of GlcNAcT-III, the enzyme response for addition of bisecting GlcNAc to *N*-glycans. As for paucimannose, the differing activities of glycosidases such as hexosaminidase and α-mannosidase II may explain its increased abundance in the tumours. The changes in the bisecting and paucimannose expression during CRC carcinogenesis observed in this study suggests that these glycans may play important biological roles in malignancy that is yet to be fully explained.

Four monoantennary glycan structures were also found to be increased in tumours compared to non-neoplastic colonic mucosa. Additionally, there was also an overall increase in monoantennary structures in the cancer tissues. Concomitantly, there was a decrease in biantennary and an upward trend in triantennary structures in tumour tissues. These observations highlight the intricate regulation of *N*-glycan branching which is altered in CRC. A possible explanation for these changes is alterations in the activity of α-mannosidase II and GlcNAcT-IV/GlcNAcT-V enzymes as these are required for the trimming of residual mannose residues for biantennary extension and the addition of GlcNAc to the biantennary *N*-glycan to form triantennary *N*-glycans, respectively. Interestingly, this complex dynamic of dysregulated *N*-glycan branching has not been previously mentioned in earlier studies in CRC. However, increased *N*-glycan branching is consistent with trends reported in other cancers [107, 236]. Though the above proposed mechanisms for the regulation of *N*-glycan branching require further validation (discussed further in section 5.5), the differential expression of the monoantennary glycans suggests they might be candidate glycan markers of CRC.

Given that the glycosylation changes in this study occurs on membrane proteins, it was relevant to consider whether membrane protein expression was also changed between normal and tumour colonic tissues. In this regard, label-free membrane proteomics was performed on the epithelial cells of normal and tumour tissues from three CRC patients. The results, presented in chapter 4, revealed that 52 proteins were differentially expressed (40 upregulated and 12 downregulated), including 9 glycoproteins (5 upregulated and 4 downregulated). The proteins were associated with the broad processes of nucleotide binding, protein transport, immunity and cell proliferation. The differential expression of proteins previously correlated with CRC such as carcinoembryonic antigen 5 (CEA5) [57] and CD44 [273] were observed. However, proteins not previously implicated in CRC were also identified such as coiled-coiled domain containing protein 47 (CCDC47) and immunoglobulin J (IGJ) chain. In addition to membrane protein profiling, the *N*-glycosylation sites of the glycoproteome subset were mapped on the basis of the change in peptide mass accompanying enzymatic

deglycosylation that converts Asn to Asp. Overall, 194 sites on 128 proteins were found which allowed confirmation of 60 sites annotated in UniProt either as "potential" or "probable" sites and the identification of another 16 unannotated sites. Remarkably, 5 potential *N*-glycosylation sites were confirmed in the CRC biomarker protein CEA. Most importantly, changes in formerly glycosylated and non-glycosylated peptides of CEA indicated that there are *N*-glycosylation site differences between normal and tumour tissues. Though the glycan moiety previously residing on these glycosylation sites was not determined, the observation that there are differences in *N*-glycosylation sites warrant further investigation as the glycosylation site changes combined with the biomarker protein, may provide utility for improving biomarker sensitivity and specificity.

5.2   The use of CRC cellular models as an investigative approach in glycomics

The marked changes observed in the *O*-glycosylation profiles of cell lines in comparison to CRC tumours is one of the most critical elements in this thesis. Within the context of biomarker discovery, the observed changes imply that some glycan based cancer markers may be underexpressed in many commonly utilized colon cancer cell line models, and this heralds caution against relying solely upon cell lines for biomarker discovery studies. Interestingly, this finding also suggests that cell lines have some inherent differences. For example, the mucinous cell line LS174T expressed sialyl-Tn, while it was not detected in four other non-mucinous cell lines, which is consistent with the trend found in another study [369]. In addition, the presence of mucins in LS174T cell lines results in more heterogeneous glycosylation than the membrane *O*-glycosylation observed of the other cell lines. Given that mucins are also found on the membrane epithelia of CRC patient tumours, LS174T (and other mucinous cell lines) may therefore be more suitable than non-mucinous cell lines as models of membrane protein *O*-glycosylation. The expression of sialyl-Tn in LS174T correlates with their mucin producing capacity that mirrors the complex *O*-glycosylation found in CRC tumours. In contrast, the differences in the *N*-glycosylation profiles of CRC tumours and cell lines were more subtle and glycan changes that occurred appeared quantitative (due to differences in relative abundance) rather than qualitative (presence/absence). As *N*-glycans are carried by membrane proteins other than mucins (though mucins cannot be excluded as potential carriers) different cell models of *N*-glycosylation may be used.

Taking this into account, one aspect of *N*-glycan change which merits further investigation is the increase of paucimannosidic *N*-glycans in CRC tumour tissues compared to the normal colonic epithelia. Since the regulation of bisecting *N*-glycans was already observed in CRC cell lines (Chapter 2), it would be interesting to use cell lines to investigate the biosynthetic pathways regulating paucimannose expression as it has not been previously studied in CRC. Consequently, it is relevant to

evaluate whether any cell lines studied may be used to model paucimannose changes which may yield insights into the regulation of paucimannose synthesis. Interestingly, SW480 and SW620 cells may be potential models since the expression of paucimannose as a percentage of the total membrane *N*-glycome of SW480 cells is similar to the normal tissue epithelial cells (1.2% ± 1.2% in normal tissue vs 1.2% ± 0.1% in SW480 cells, Fig 1a) while the paucimannose expression in SW620 cells is increased and comparable to the tumours (2.7% ± 1.8% in the tumours vs 2.8% ± 0.2% in SW620 cells, Fig 1a). Notably, the SW480 and SW620 colon cancer cells were derived from the same patient with the SW480 cells isolated from the primary tumour and SW620 cells isolated from a lymph node metastasis respectively [217]. Thus, these cell lines may be considered *in vitro* models of CRC progression. It also indicates that while mucinous colon cancer cell lines may be better models of membrane protein *O*-glycosylation, some aspects of *in vivo N*-glycosylation, such as paucimannosylation, in non-mucinous cell models may be comparable with tissue changes. In this regard, a variety of techniques may be employed more easily in cell lines to elucidate the underlying molecular mechanisms, such as qRT-PCR, Western Blotting, immunofluorescence microscopy or selected reaction monitoring (SRM) targeted towards the possible altered glycosyltransferases/glycosidases.

The use of cell lines however is still restricted to the modelling of malignant cell states as colon cell lines representing normal healthy tissues are not available except as primary cultures. This is important in the case of bisecting *N*-glycans where their expression was increased in SW620 compared to the less metastatic SW480 cell line, indicating an increase in bisecting *N*-glycans during cancer progression (Fig 1b). In contrast, the converse was observed between normal and tumour tissues (Fig 1b). Indeed, there is a stark difference in the expression of bisecting *N*-glycans observed between SW480 colon cancer cells (0.1% ± 0.01%) and normal colonic tissues (36.1% ± 18.3%; Fig 1b). It is possible that the expression of bisecting *N*-glycans is decreased during CRC carcinogenesis only to increase later during CRC metastasis. However, extended subculture on two-dimensional plates and patient variability cannot be excluded as factors that may also lead to differences in bisecting *N*-glycan expression. Regardless, it suggests that normal healthy tissues at present remain irreplaceable to investigate alterations in bisecting *N*-glycans as there is a lack of 'normal' non-cancerous colon cell lines for the profiling of glycosylation and other glycan-related changes in the healthy state.
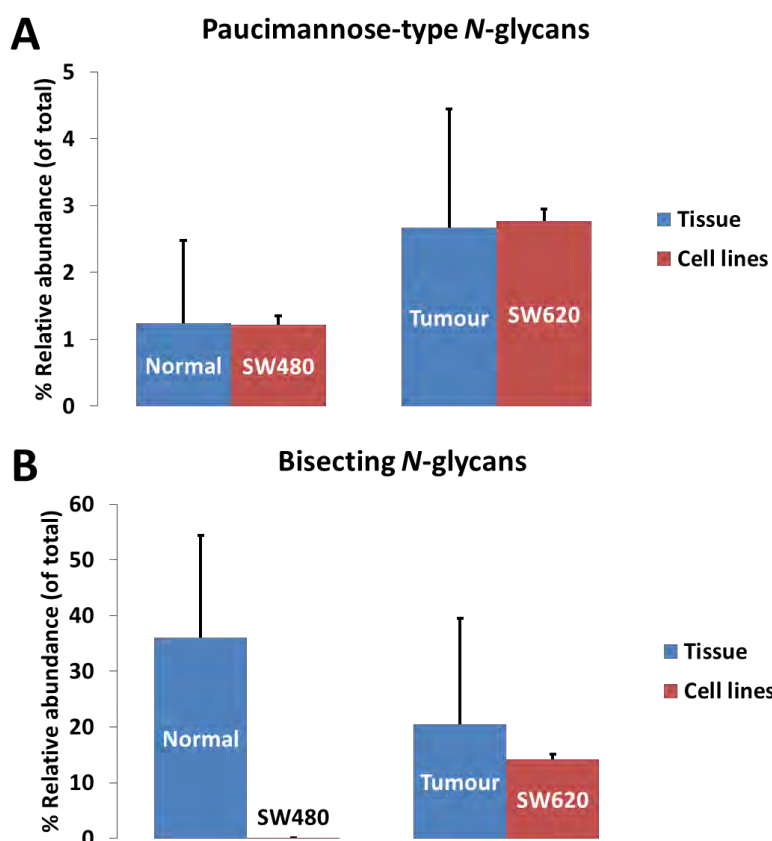
Figure 1. Expression of paucimannose-type (A) and bisecting *N*-glycans (B) in normal/tumour colonic tissues and the colon cancer cell lines, SW480 and SW620. Data shown as mean ± SD for tissues (n=6) and mean ± max/mean for cell lines (biological duplicates). The *N*-glycans were quantified as described in Chapter 2 and 3.

While cell line models clearly do not encompass the full complexity of *in vivo* tumours, they remain a valuable tool for elucidating molecular mechanisms underlying observed glycan changes that occur during cancer progression. Having mentioned this, it is prudent to validate findings from CRC cell lines in clinical tissues. However, their judicious use still gives insights which can be combined with other approaches to provide a cohesive view of glycosylation changes and their regulation in CRC.

5.3   The use of mass spectrometry for in-depth glycomic and glycoproteomic analysis

Historically, the use of monoclonal antibodies has been more common to detect glycan changes in CRC compared to mass spectrometry [78, 169, 170]. Even now, there are few mass spectrometry-based studies on glycan changes in CRC during carcinogenesis and/or cancer progression [38, 102]. Unsurprisingly, proteomic studies on CRC have been much more common (recently reviewed in [76]). However, to date, there are no known studies on CRC which combine *N*-glycosylation profiling with a parallel proteomics analysis on the same patient samples, as performed in this thesis (Chapter 3 and

4). Not only was membrane protein expression profiled, but the *N*-glycosylation sites were mapped on a large number of proteins, revealing novel information on *N*-glycosylation site localisation. Therefore, in-depth information on protein glycosylation, protein identification and *N*-glycosylation sites may be obtained from the same patient samples. Although the chosen method does not allow analysis of the glycan moiety on the formerly glycosylated peptides, it could be the first step in a multi-level analysis that goes on to utilise other glycoprotein/glycopeptide analysis strategies, as shown by Parker *et. al.* [321], where glycan and deglycosylated peptide information was combined to create a database for further glycopeptide analysis in rat brain samples.

Additionally, the overall experimental workflow may be upscaled and applied to other cancer types. In this regard, the most likely hurdle in the upscaling process is the low throughput of glycan structural analysis, which remains largely manual and is a noted drawback of the non-targeted approach presented in this thesis. To overcome this bottleneck, there have been efforts made towards development of bioinformatics tools which facilitate automated glycan analysis [145, 370]. The creation of curated databases which house glycan spectral information such as UniCarb-DB [146], provides a basis for automated spectral matching algorithms. The development and application of these algorithms is likely to increase glycomics analysis throughput similar to those now routinely employed in proteomics [371]. Although various obstacles still need to be overcome prior to high-throughput glycomics analysis, this study has nevertheless taken the initial steps to demonstrate the wealth of information which may be mined by harnessing the power of mass spectrometry, making it a critical component in the growing arsenal used to discover glycosylation changes to be exploited in the fight against cancer. In the following section, glycan alterations identified in CRC and presented in this thesis, will be compared with other cancer types.

5.4 Major glycan changes in CRC compared with other cancers

Using a mass spectrometry-based approach, altered glycosylation patterns have been discovered in the epithelial cells of colonic tumours compared to their matched normal mucosa in this thesis (Chapter 3). Table 1 summarises the major *N*- and *O*-glycan changes found or confirmed in CRC (including findings from other studies) compared with other cancers. While not intended to be comprehensive (for a more complete review, see authored publication in Chapter 1), the comparisons provide a glimpse into whether the trends observed in CRC differ or mirror those seen amongst different cancer types, particularly for the glycosylation of tissues or cell lines.

For *N*-glycans, bisecting *N*-glycans on proteins is emerging as an altered glycan phenotype of cancer although the directionality of the change (increase/decrease) is dependent on the cancer type. Thus far, bisecting *N*-glycans have been observed to be decreased during carcinogenesis of CRC [38] and

progression of breast cancer [372] (Table 1). In contrast, bisecting *N*-glycans were found increased during carcinogenesis of ovarian cancer [373](Table 1). Moreover, when expression of the corresponding enzyme (GlcNAcT-III) was investigated, there was generally good correlation between enzyme expression (as determined by mRNA expression) and bisecting glycan expression (as determined by LC-MS/MS, Chapter 2) [219, 373]. However, it is not yet clear whether the global changes in bisecting *N*-glycans observed in cancer are reflected in the overall glycoproteome or only on a subset of glycoproteins. Nonetheless, it is likely that this glycan type plays an important role given its general association with cancer. For example, bisecting *N*-glycans have been shown to modify specific proteins implicated in tumour invasion and migration, such as E-cadherin in CRC [209] and integrin α5β1 in cervical cancer [374] which may have impact on tumour metastasis. Despite this, there may be a number of other glycoproteins modified by bisecting *N*-glycans not yet identified which may also be functionally modulated. Thus, further work is required to empirically demonstrate the concerted effect of altered bisecting *N*-glycans on its carrier proteins in CRC progression.

While there have been some cancer studies showing the changes in bisecting *N*-glycans from tissues/cell lines, there have been limited reports on the expression of paucimannose and monoantennary glycans in cancer in general. Paucimannose changes in colon cancer have only been reported in this thesis and in the LC-MS/MS study by Balog *et. al.* [38]. Both these studies observed an increase in paucimannose *N*-glycans on the proteins of CRC tumours compared to normal colonic mucosa. Changes in paucimannose glycans in other cancer types have not yet been reported. Having mentioned this, alterations in paucimannose glycan expression have been shown in other non-cancer diseased conditions associated with inflammation such as cystic fibrosis [375] and systemic lupus erythematosus [376]. In fact, our laboratory has shown that paucimannose glycans are a dominant feature in the sputum of cystic fibrosis patients with average expression of 43.4% (± 4.1%) of the total *N*-glycans [375]. This is notably higher than the paucimannose expression that was observed in CRC (<3 %, detailed in Chapter 2 and 3; percentage expression not reported by Balog *et. al.* [38]). Overall, this suggests that paucimannose glycans are not unique to cancer but may be associated with other inflammation-related conditions. Therefore there is a need to profile patients with other, non-cancerous colonic diseases such as inflammatory bowel disease to determine basal expression of this glycan type in non-cancer diseased states and its cancer-specificity. This is discussed further in section 5.6.

Table 1. Major glycan changes identified or confirmed in this thesis compared with other cancer types.
Information regarding tissue/cell line glycan changes for ovarian and breast cancer have been taken from the reviewed publication (presented at end of Chapter 1)[36]. Where limited information is available for other cancers, glycan changes from other non-cancer conditions have also been included. Green arrows indicate trends identified or confirmed in this thesis. ↑: Increased, ↓: decreased, d: detected, H: human, M: mouse, S: serum, Sp: sputum, MAb: monoclonal antibody. [a]Ion corresponds to second eluted isomer with composition $(HexNAc)_2 (NeuAc)_1$. [b]Ion corresponds to sialyl Lewis X hexasaccharide previously identified as a potential cancer marker in CRC.

Glycan change	Regulation	Tissue	Cell line	Other	Cancer vs non-cancer	Early/late/ metastatic	MAb/ Lectin	HPLC/MS	Ref(s)
N-glycans									
Bisecting									
Colorectal cancer	↓	H			X			X	[38]
Ovarian cancer	↑		H		X			X	[373]
Breast cancer	↓	M				X	X		[372]
Paucimannose									
Colorectal cancer	↑	H			X			X	[38]
Lupus	↑	M						X	[376]
Cystic fibrosis	↑			Sp				X	[375]
Monoantennary									
Colorectal cancer	↑	H			X			X	This thesis
Lung cancer	↑			S	X			X	[377]
Triantennary									
Colorectal cancer	↑	H			X			X	This thesis
Ovarian cancer	↑			S	X	X		X	[378]
O-glycans									
Sialyl-Tn									
Colorectal cancer	↑	H				X	X	X	[37, 78, 80, 102, 169, 170]
Breast cancer	↑	H			X	X	X	X	[379-382]
Ovarian cancer	↑,d	H			X	X	X		[383-387]
m/z 716b[a]									
Colorectal cancer	↑, d	H			X			X	This thesis, [102, 177]
Ulcerative colitis	d	H						X	[388]
m/z 1315[b]									
Colorectal cancer	↑	H			X			X	This thesis, [102]
Ulcerative colitis	d	H						X	[388]

On a different note, changes in monoantennary *N*-glycans have only also been reported in lung cancer, although these changes were observed in the serum glycoproteins of lung cancer patients [377] rather than in lung cancer tissues or cell lines. In that study, monoantennary glycans were found to be significantly increased in the serum glycome of lung cancer patients compared to non-cancer healthy controls, although no change was significantly observed in biantennary *N*-glycans using LC-MS/MS [377]. In comparison, in this study a significant increase in monoantennary *N*-glycans (3.2-fold, p = 0.0083) as well as a significant decrease in biantennary *N*-glycans (1.5-fold, p = 0.0041) was shown in the CRC tumours compared to normal mucosa (Chapter 3). Nevertheless, a significant increase in the expression of triantennary *N*-glycans (p = 0.00184) have also been observed in ovarian cancer serum compared to healthy female controls by LC-MS/MS [378] (Table 1). Taken together, the overall findings in both of these studies suggest that an increase in mono-/triantennary glycans and dysregulated glycan branching is not specific to CRC but can occur during carcinogenesis across different cancer types. Similar to the other major *N*-glycan changes mentioned above, the biological functions of this glycan change on the overall cancer behaviour remains to be fully explored. For monoantennary *N*-glycans, a potential option to study this would be *N*-glycan remodelling by gene transfection or silencing of relevant enzymes (such as knockdown of the enzyme which extends the second antenna of biantennary glycan, GlcNAcT-II) prior to additional cell functional assays. The use of *in vitro* models for functional glycan analysis is also discussed further in section 5.6.

For *O*-glycans, the trend of increased sialyl-Tn expression in CRC tumours is consistent with observations from other epithelial cancers such ovarian [386], breast [382] and pancreatic cancers [389]. However, the decreased expression of an isomer of the trisaccharide (with *m/z* 716) and sialyl-Lewis X carrying hexasaccharide (with *m/z* 1315) observed in this study in CRC tumour tissues have not been reported in other cancer studies. However both these *O*-glycans have been reported to be expressed in patients with active ulcerative colitis (a form of inflammatory bowel disease) [388] although there was no significant change of either glycan found between patients with ulcerative colitis (n=15) and patient controls (n=25) who did not have colonic inflammation. Thus, the decreased expression of these two glycans may possibly be specific to CRC, but given the small number of *O*-glycan patient profiles (n=3) analysed in this study, requires further validation in a larger patient group. Intriguingly, sialyl-Tn was also found to be significantly increased in ulcerative colitis patients in the same study [388] which demonstrates that its overexpression may also be linked to inflammation. This further underscores the complexities of glycan changes in disease as the same glycans may be altered in different pathological conditions. While these findings indicate that identification of CRC-specific glycan markers is likely to be a challenge, the overall observations in this thesis have been in good agreement with the reported glycan changes in other cancers (where

available). The most notable were in the expression of bisecting *N*-glycans where directionality of change is dependent on cancer type. Other changes such as increases in paucimannosidic and sialyl-Tn were similar between different cancers as well as observed in other non-cancer conditions. The implications of these findings on future work will be detailed more below.

5.5   Future directions
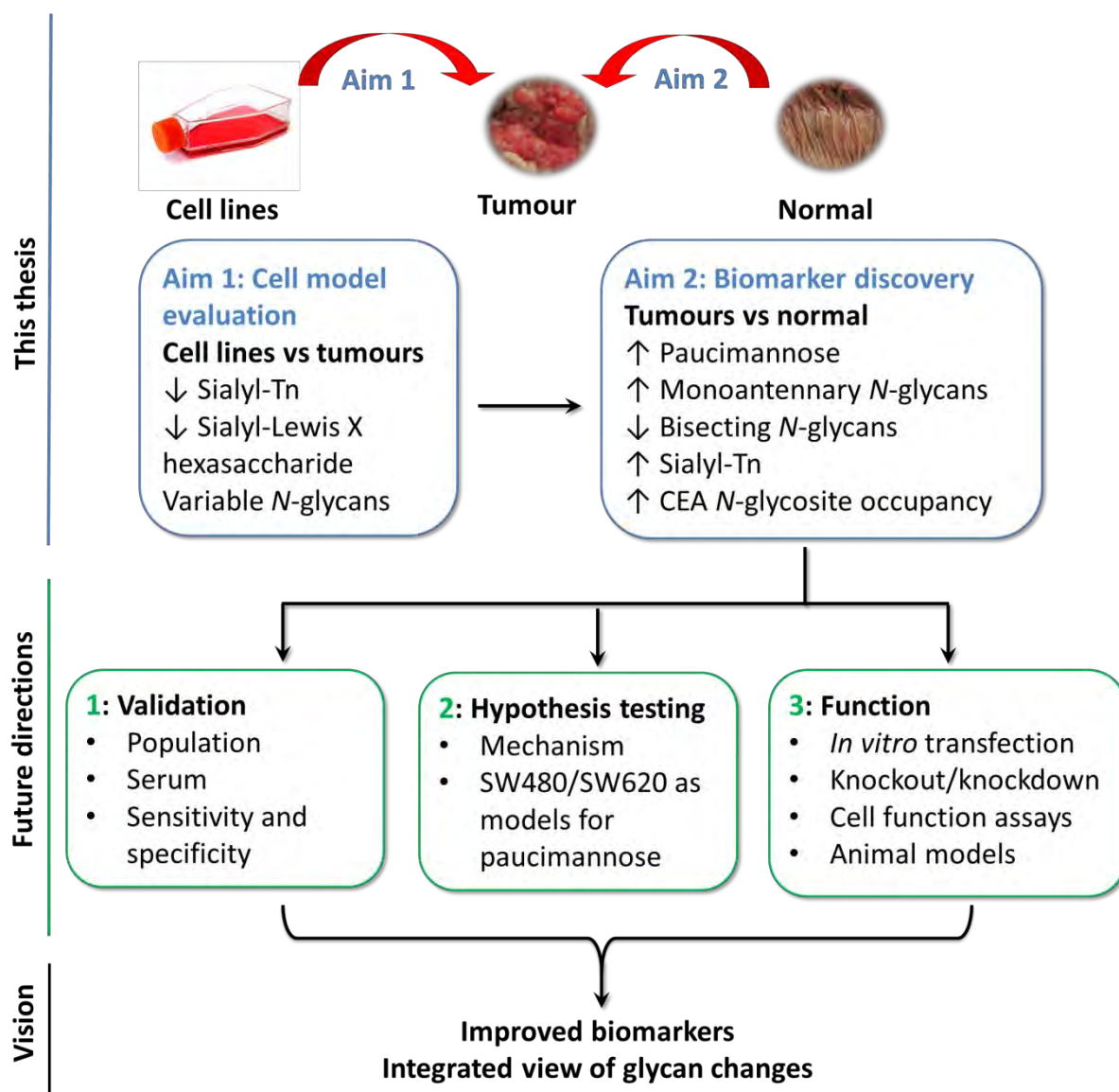
Figure 2. Possible future directions following findings reported in this thesis.

There are many directions that can be undertaken following the findings presented in this thesis and these have been outlined in Fig 2. In terms of biomarkers, although the strengths of using LC-MS/MS approaches have been highlighted above (section 5.4), it is clear that further validation of novel biomarker candidates in a larger patient population (step 1, boxed in green, Fig 2) is extremely important given the small patient number analysed in this thesis (n=3 - 6) as mentioned above. This is especially the case for differentially expressed *N*-glycans where there have only been two published

glycomics studies (the study presented in this thesis and Balog *et. al.* [38]) so far on protein glycosylation changes in CRC – each with patient cohorts of less than 15. In addition, though some colonic tumours (n= 4 - 5) were classified in stage C and D in the other study, neither of the above glycomics investigations has examined whether there are stage-specific glycan patterns which may be potentially useful as molecular signatures to aid pathological classification. Moreover, little is known about *N*-glycan structural diversity along the colon in the healthy and tumour states although mucin *O*-glycans of normal human colon have been shown to distribute regio-specifically [177]. Interestingly, membrane proteomic profiling along different segments of the normal colon has also revealed distinct regional differences [390]. Since mucin *O*-glycans and proteomic differences have been reported along different regions of the normal colon, there may also be *N*-glycan changes yet to be profiled in the normal (as well as tumour) colonic epithelia along the different segments from the proximal to distal colon.

Additionally, there have been limited reports on glycosylation differences in the non-cancer diseased colon (such as inflammatory bowel disease) compared with either normal or tumour colonic tissues. Only changes in mucin *O*-glycosylation of patients with active and inactive ulcerative colitis (one of the two major forms of inflammatory bowel disease) have been reported [388] as discussed earlier in section 5.4. However, *N*-glycosylation changes have not been previously investigated in inflammatory bowel conditions. Given that there is a well-known association between CRC and colonic inflammation [391, 392], it would be advantageous to determine whether differentially expressed *N*-glycans alter in other non-cancer diseases of the colon and not only cancer. Furthermore, it would be useful to determine whether there are correlations between glycan-related changes and other clinicopathological parameters such as lymphatic invasion, histological differentiation and overall survival. Thus, additional work in a larger patient cohort will provide more in-depth information on the potential of these glycan candidate markers as diagnostic or prognostic indicators of CRC. As manual data analysis of samples is time-consuming, targeted glycan analysis with quantification against an internal standard (which could be a labelled or non-mammalian glycan) is a potential approach to increase accurate analysis at high throughput.

Having discussed possible future directions, given the very small number of samples available for glycan tissue analysis, it is clear that the comparison of glycan profiles between epithelial cells of normal and tumour colon tissues in this thesis was only intended to be a proof-of-concept study. Higher sample numbers are clearly required to confirm or refute the conclusions reached in this thesis. Another aspect that is difficult to determine is whether having an epithelial enrichment provides a superior approach to identifying glycan markers compared to glycan profiling of mixed cell tissues. For instance, we identified 11 novel *N*-glycans as differentially expressed between normal

and tissue epithelial cells which were not previously observed in an earlier study [38]. Since epithelial enrichments were not performed in that study, the epithelial enrichment of the samples analysed in this thesis may have enabled these novel identifications. However, differences in sample preparation (such as total vs membrane protein preparation in this thesis) and LC-MS/MS approach (HILIC vs PGC in this study) cannot be excluded as contributing factors. In order to clearly elucidate the basis for these differences, more samples are required to perform complementary LC-MS/MS analyses with different sample preparations. As the availability of tissue samples was highly limited in this study, these comparisons were not performed. However, a more thorough comparison between epithelial-enriched and mixed-cell glycan profiles is acknowledged as a future direction to be pursued.

Besides additional glycan profiling using LC-MS/MS, other mass spectrometry-based glycoprotein/glycopeptide strategies not used in this thesis could also be employed in the near-future to provide a connection between glycan and protein changes observed. In particular, the information obtained in the tissue *N*-glycome and deglycosylated peptide analysis can be concatenated to construct a potential glycopeptide library for further *N*-glycopeptide analysis as demonstrated previously in rat brain [321]. It would be interesting to use this approach to determine the site-specific *N*-glycosylation of differential proteins of interest. For example, for carcinoembryonic antigen 5 (CEA5), it would be relevant to analyse whether differentially expressed *N*-glycans found in this study are carried by specific *N*-glycosylation sites on this CRC-associated glycoprotein. Additionally, differences in glycosylation site occupancy of CEA between normal and cancer colonic tissues (proposed in Chapter 4) can be confirmed. While CEA is known to be present in the blood, further analysis in patient serum will also reveal whether there are other tissue membrane glycoproteins that are shed and/or secreted into the blood (step 1, boxed in green, Fig 2). This would allow them to be profiled in the serum which will improve ease of CRC screening. With a patient cohort of adequate size, one can also determine whether the glycan and glycoprotein changes found in this and other studies could be combined to form novel biomarker panels with increased sensitivity and specificity (step 1, boxed in green, Fig 2), which can lead to improved early detection and ultimately, patient outcomes in CRC.

Even though validation of novel glycan marker candidates (step 1, boxed in green, Fig 2) are a crucial future step in the translation from bench to bedside, some of the molecular mechanisms underpinning the observed glycan changes such as altered paucimannose expression and the *O*-linked trisaccharide isomer remain to be fully elucidated (section 5.2). There are many factors that contribute towards altered glycosylation such as nucleotide sugar donors, nucleotide sugar transporters and glycan enzymes (glycosyltransferases/glycosidases). To add to this complexity, the mRNA stability, protein expression and turnover of each of these components may also impact on glycosylation. As a starting point, changes in the expression of glycan enzymes may be investigated in

in vitro models (step 2, boxed in green, Fig 2). For instance, the following changes proposed to correspond to paucimannose expression (Chapter 3) may be studied further in suitable cell line(s): 1) Increased activity of α-mannosidase II, which trims the last remaining mannose residues prior to biantennary extension (Fig 6a, Chapter 3), and 2) increased activity of hexosaminidase, which removes GlcNAc residue for the formation of paucimannose (Fig 6a, Chapter 3). As discussed previously (section 5.2), SW480 and SW620 colon cell lines may be used for these experiments due to differences in their inherent paucimannose expression. Thus, cell model systems may be employed to provide mechanistic insights for greater holistic understanding of glycan change, especially in regards to changes in enzyme protein expression. Many of the relevant enzymes (glycosyltransferases and glycosidases) are likely to be low in abundance and therefore the capacity of subculturing of cell lines allows the generation of abundant sample material for further enrichment/fractionation as required [393, 394] prior to MS-based techniques such as SRM. Alternatively, other techniques such as qRT-PCR [81, 373], immunofluorescence microscopy [395] or Western blotting may be also used to validate change in glycan enzyme expression. Indeed, some studies have shown that protein expression of glycosyltranferases on human tissues can be determined using Western Blotting [155] or immunohistochemistry [156]. Therefore, while cell lines are a useful initial starting tool to probe underlying mechanistic changes, the protein changes of the enzymes can be validated further in human tissue samples using efficacious antibodies to demonstrate their relevance to human *in vivo* changes.

Having mentioned this, model systems (*in vitro* and animal *in vivo*) could also be employed to further explore the biological functions of observed glycan alterations (if any) through experimental manipulation of glycan synthesis (step 3, boxed in green, Fig 2). In particular, in this thesis, the findings of qRT-PCR (Chapter 2) indicated that glycosyltransferases such as GlcNAcT-III (encoded by MGAT3), core 1 synthase (encoded by C1GALT1) and ST6GalNAc1 (encoded by ST6GALNAC1) are associated with CRC-associated glycan changes, namely the expression of bisecting *N*-glycans and sialyl-Tn. However, it appears that it is more common to alter only one enzyme during experimental manipulation of a synthetic pathway. For instance, fucosyltransferase 8 (Fut8) or GlcNAcT-III have been transfected into WiDr colon cancer cells previously [155, 211] but both enzymes have not been transfected and altered together. Given the findings regarding bisecting *N*-glycans/sialyl-Tn expression in this thesis, it would be interesting transfect both GlcNAcT-III and ST6GalNAc1 into SW480 (or other cancer cells) which may have low endogenous expression of both enzymes (Chapter 2) and subsequently monitor whether sialyl-Tn and bisecting *N*-glycan expression are altered. Since a subset of mucins have been proposed as major sialyl-Tn carriers, the expression of mucins (such as mucin-2/5B/6) may also be required as their expression is also low in SW480 cells. To this end, co-transfection with a mucin transcription factor such as GATA-4 [396] or Forkhead box transcription

factors [397] may be needed as the expression of these regulators have been shown to induce mucin-2 expression. The direction transfection of the above-mentioned mucins has also been considered. However, this may be difficult given their relatively large sizes (>250,000 kDa according to UniProt). Overall, the aim would be to establish a cell model where bisecting and sialyl-Tn can be altered singly or in combination to examine the synergistic effect of these glycan changes on cell function by *in vitro* analyses such as cell invasion and cell proliferation assays (step 3, boxed in green, Fig 2). As well, these cells may be injected into nude mice to generate animal models to determine changes in metastatic potential and cell growth (step 3, boxed in green, Fig 2).

Taken together, the findings presented in this thesis may be further extended in various ways towards the discovery of improved glycan and glycoprotein biomarkers in CRC. Clearly, the validation of novel potential glycan markers found in larger patient cohort is an important priority. The correlation of the glycan changes and other tumour location, TNM stage and inflammation should also be considered. As well, alterations in *N*-glycosylation site occupancy and site-specific *N*-glycosylation of differentially expressed glycoproteins (such as CEA) could be further profiled using MS-based glycopeptide analysis in patient colonic tissues and serum. Finally, model systems may be utilised to give mechanistic insights to the observed glycan changes – especially in regards to synergistic effects of multiple glycan changes and their biological functions during malignancy. Overall, the validation of candidate markers in clinical patient samples combined with a strategic use of model systems is predicted to lead to a cohesive and well-rounded understanding of identified glycan changes. It is envisaged that this will result in improved biomarkers, therapeutic surveillance and overall treatment outcomes for CRC patients.

5.6   Conclusions and perspectives

In summary, this thesis has examined the alterations in membrane glycosylation and glycoproteins in the context of colorectal cancer. In particular, this is the first known study to perform a comprehensive glycomics comparison of several cell line models and colonic tumours which also accounted for the heterogeneity of cell types found in tumour tissue. The findings empirically reveal glycosylation differences which imply that analysis of primary tumours are indispensable for glycan biomarker discovery. However, cell lines may still provide insights into molecular mechanisms governing membrane glycan expression, such as the variable sialyl-Tn expression investigated in this thesis.

Further work in patient-matched normal and cancer colon tissues in this study showed glycan and membrane protein alterations as well as intriguing *N*-glycosylation site changes that warrant further investigation. Overall, this study has demonstrated that in-depth analysis can be performed using

only a limited number, and quantity, of clinical samples (6 patients) using mass spectrometry. The strength of this study is that profiling of *N*-glycans, proteins and their corresponding *N*-glycosylation sites can be performed on the same tissues, which to the author's best knowledge, has not been previously demonstrated in human cancer tissues. Additionally, the abundant information obtained revealed changes in the glycans and proteins which may provide novel leads in cancer biomarker discovery. As some of the changes agreed and extended alterations that have been previously reported, there is potential for the novel candidate markers to be used in combination with other known markers to form a new and improved biomarker panel with increased specificity and sensitivity for early detection of CRC.

Moreover, although the patient cohort in this study was relatively small, the experimental workflow utilised may be conveniently upscaled. The low throughput nature of glycomics analysis may be overcome in the near-future by the current development of bioinformatics tools. Furthermore, the glycosylation changes found in this study may have functional roles in carcinogenesis, such as cell invasion, proliferation and metastasis that remain to be fully explored. This study therefore represents an initial but important foray into the mining of glycan and glycan-related changes which may be ultimately exploited in the battle against CRC. It is envisioned that mass spectrometric analysis of clinical tissues and judicious use of cell lines will provide complementary approaches which can be integrated to give an overall "bird's eye" view on the intricate regulation of glycan expression towards improved therapeutic treatment, monitoring and patient outcomes in CRC.

6   References

[1] Adamczyk, B., Tharmalingam, T., Rudd, P. M., Glycans as cancer biomarkers. *Biochim Biophys Acta* 2012, *1820*, 1347-1353.

[2] Mechref, Y., Hu, Y., Garcia, A., Hussein, A., Identifying cancer biomarkers by mass spectrometry-based glycomics. *Electrophoresis* 2012, *33*, 1755-1767.

[3] Varki, A., Sharon, N., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[4] Ali, N. A., Molloy, M. P., Quantitative phosphoproteomics of transforming growth factor-beta signaling in colon cancer cells. *Proteomics* 2011, *11*, 3390-3401.

[5] Leroy, C., Fialin, C., Sirvent, A., Simon, V., Urbach, S., Poncet, J., Robert, B., Jouin, P., Roche, S., Quantitative Phosphoproteomics Reveals a Cluster of Tyrosine Kinases That Mediates Src Invasive Activity in Advanced Colon Carcinoma Cells. *Cancer Research* 2009, *69*, 2279-2286.

[6] Minchin, R. F., Kadlubar, F. F., Ilett, K. F., Role of acetylation in colorectal cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1993, *290*, 35-42.

[7] Zhao, Y., Fang, X., Wang, Y., Zhang, J., Jiang, S., Liu, Z., Ma, Z., Xu, L., Li, E., Zhang, K., Comprehensive Analysis for Histone Acetylation of Human Colon Cancer Cells Treated with a novel HDAC Inhibitor. *Curr Pharm Des* 2013.

[8] Kondo, Y., Shen, L., Issa, J.-P. J., Critical Role of Histone Methylation in Tumor Suppressor Gene Silencing in Colorectal Cancer. *Molecular and Cellular Biology* 2003, *23*, 206-215.

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

[10] Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth, J. D., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Symbol nomenclature for glycan representation. *Proteomics* 2009, *9*, 5398-5399.

[11] Harvey, D. J., Merry, A. H., Royle, L., P. Campbell, M., Dwek, R. A., Rudd, P. M., Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics* 2009, *9*, 3796-3801.

[12] Bertozzi, C. R., Rabuka, D., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[13] McNaught, A. D., Nomenclature of Carbohydrates. *Pure and Applied Chemistry* 1996, *68*, 1919-2008.

[14] Hua, S., An, H. J., Ozcan, S., Ro, G. S., Soares, S., DeVere-White, R., Lebrilla, C. B., Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. *Analyst* 2011, *136*, 3663-3671.

[15] Koizumi, K., High-performance liquid chromatographic separation of carbohydrates on graphitized carbon columns. *Journal of chromatography. A* 1996, *720*, 119-126.

[16] Krishnamoorthy, L., Mahal, L. K., Glycomic analysis: an array of technologies. *ACS Chem Biol* 2009, *4*, 715-732.

[17] Schaller, C. IB4. Carbohydrates. [cited 2014 September 3]; Available from: http://chemwiki.ucdavis.edu/?title=Under_Construction/Schaller/Part_I:__Structure_in_Organic_Biological_%26_Inorganic_Chemistry/IB._Introduction_to_Biomolecules/IB4._Carbohydrates.

[18] Zambon, M., Lessons from the 1918 influenza. *Nat Biotech* 2007, *25*, 433-434.

[19] Zielinska, Dorota F., Gnad, F., Schropp, K., Wiśniewski, Jacek R., Mann, M., Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. *Mol Cell* 2012, *46*, 542-548.

[20] Freeze, H. H., Haltiwanger, R. S., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[21] Dell, A., Morris, H. R., Glycoprotein structure determination by mass spectrometry. *Science* 2001, *291*, 2351-2356.

[22] Brockhausen, I., Schachter, H., Stanley, P., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[23] Rini, J., Esko, J., Varki, A., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[24] Brooks, S. A., Carter, T. M., Royle, L., Harvey, D. J., Fry, S. A., Kinch, C., Dwek, R. A., Rudd, P. M., Altered Glycosylation of Proteins in Cancer: What Is the Potential for New Anti-Tumour Strategies. *Anti-Cancer Agents in Medicinal Chemistry* 2008, *8*, 2-21.

[25] Stanley, P., Schachter, H., Taniguchi, N., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[26] Varki, A., Esko, J., Colley, K., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, New York 2009.

[27] El-Battari, A., Prorok, M., Angata, K., Mathieu, S., Zerfaoui, M., Ong, E., Suzuki, M., Lombardo, D., Fukuda, M., Different glycosyltransferases are differentially processed for secretion, dimerization, and autoglycosylation. *Glycobiology* 2003, *13*, 941-953.

[28] Freeze, H. H., Elbein, A. D., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[29] Geisberg, Joseph V., Moqtaderi, Z., Fan, X., Ozsolak, F., Struhl, K., Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell*, *156*, 812-824.

[30] Haynes, C. M., Titus, E. A., Cooper, A. A., Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death. *Mol Cell* 2004, *15*, 767-776.

[31] Brooks, S. A., Carter, T. M., Royle, L., Harvey, D. J., Fry, S. A., Kinch, C., Dwek, R. A., Rudd, P. M., Altered glycosylation of proteins in cancer: what is the potential for new anti-tumour strategies. *Anticancer Agents Med Chem* 2008, *8*, 2-21.

[32] Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., Darnell, J., *Molecular Cell Biology*, W. H. Freeman, New York 2000.

[33] Stanley, P., Cummings, R. D., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H*., et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[34] Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M. A., Samyn-Petit, B., Julien, S., Delannoy, P., The human sialyltransferase family. *Biochimie* 2001, *83*, 727-737.

[35] Marcos, N. T., Pinho, S., Grandela, C., Cruz, A., Samyn-Petit, B., Harduin-Lepers, A., Almeida, R., Silva, F., Morais, V., Costa, J., Kihlberg, J., Clausen, H., Reis, C. A., Role of the Human ST6GalNAc-I and ST6GalNAc-II in the Synthesis of the Cancer-Associated Sialyl-Tn Antigen. *Cancer Research* 2004, *64*, 7050-7057.

[36] Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., Packer, N. H., Cell surface protein glycosylation in cancer. *Proteomics* 2014, *14*, 525-546.

[37] Itzkowitz, S. H., Yuan, M., Montgomery, C. K., Kjeldsen, T., Takahashi, H. K., Bigbee, W. L., Kim, Y. S., Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer Res* 1989, *49*, 197-204.

[38] Balog, C. I. A., Stavenhagen, K., Fung, W. L. J., Koeleman, C. A., McDonnell, L. A., Verhoeven, A., Mesker, W. E., Tollenaar, R. A. E. M., Deelder, A. M., Wuhrer, M., N-glycosylation of Colorectal Cancer Tissues. *Molecular & Cellular Proteomics* 2012, *11*, 571-585.

[39] Misonou, Y., Shida, K., Korekane, H., Seki, Y., Noura, S., Ohue, M., Miyamoto, Y., Comprehensive clinico-glycomic study of 16 colorectal cancer specimens: elucidation of aberrant glycosylation and its mechanistic causes in colorectal cancer cells. *J Proteome Res* 2009, *8*, 2990-3005.

[40] Zeller, J. L., Lynm, C., Glass, R. M., Colon cancer. *JAMA* 2008, *300*, 2816-2816.

[41] Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., Forman, D., Global cancer statistics. *CA: A Cancer Journal for Clinicians* 2011, *61*, 69-90.

[42] AIHW. Australia's Health. 2012; Available from: http://www.aihw.gov.au/publication-detail/?id=10737422172.

[43] Levine, D. S., Haggitt, R. C., Normal Histology of the Colon. *The American Journal of Surgical Pathology* 1989, *13*, 966-984.

[44] Geibel, J. P., Secretion and absorption by colonic crypts. *Annu Rev Physiol* 2005, *67*, 471-490.

[45] Carey, W. D., Colon physiology. A review. *Cleve Clin Q* 1977, *44*, 73-81.

[46] Eroschenko, V., *diFiore's Atlas of Histology with Functional Correlations*, Lippincott Williams & Wilkins, Baltimore 2013.

[47] Specian, R. D., Oliver, M. G., Functional biology of intestinal goblet cells. *Am J Physiol* 1991, *260*, C183-193.

[48] Owens, B. M. J., Simmons, A., Intestinal stromal cells in mucosal immunity and homeostasis. *Mucosal Immunol* 2013, *6*, 224-234.

[49] Fleming, M., Ravula, S., Tatishchev, S. F., Wang, H. L., Colorectal carcinoma: Pathologic aspects. *Journal of Gastrointestinal Oncology* 2012, *3*, 153-173.

[50] Aaltonen, L., Hamilton, S. R., *Pathology and genetics of tumours of the digestive system*, IARC Press 2000.

[51] Jankova, L., Chan, C., Fung, C. L., Song, X., Kwun, S. Y., Cowley, M. J., Kaplan, W., Dent, O. F., Bokey, E. L., Chapuis, P. H., Baker, M. S., Robertson, G. R., Clarke, S. J., Molloy, M. P., Proteomic comparison of colorectal tumours and non-neoplastic mucosa from paired patient samples using iTRAQ mass spectrometry. *Mol Biosyst* 2011, *7*, 2997-3005.

[52] Fielding, L. P., Arsenault, P. A., Chapuis, P. H., Dent, O., Gathright, B., Hardcastle, J. D., Hermanek, P., Jass, J. R., Newland, R. C., Clinicopathological staging for colorectal cancer: an International Documentation System (IDS) and an International Comprehensive Anatomical Terminology (ICAT). *J Gastroenterol Hepatol* 1991, *6*, 325-344.

[53] Davis, N. C., Newland, R. C., Terminology and classification of colorectal adenocarcinoma: the Australian clinico-pathological staging system. *Aust N Z J Surg* 1983, *53*, 211-221.

[54] American Cancer Society. How is colorectal cancer staged? 2014 [cited 2014 February 7]; Available from: http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-staged.

[55] Ogino, S., Nosho, K., Kirkner, G. J., Kawasaki, T., Meyerhardt, J. A., Loda, M., Giovannucci, E. L., Fuchs, C. S., CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 2009, *58*, 90-96.

[56] Shen, L., Toyota, M., Kondo, Y., Lin, E., Zhang, L., Guo, Y., Hernandez, N. S., Chen, X., Ahmed, S., Konishi, K., Hamilton, S. R., Issa, J.-P. J., Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proceedings of the National Academy of Sciences* 2007, *104*, 18654-18659.

[57] Chen, J. S., Chen, K. T., Fan, C. W., Han, C. L., Chen, Y. J., Yu, J. S., Chang, Y. S., Chien, C. W., Wu, C. P., Hung, R. P., Chan, E. C., Comparison of membrane fraction proteomic profiles of normal and cancerous human colorectal tissues with gel-assisted digestion and iTRAQ labeling mass spectrometry. *FEBS J* 2010, *277*, 3028-3038.

[58] Ghosh, D., Yu, H., Tan, X. F., Lim, T. K., Zubaidah, R. M., Tan, H. T., Chung, M. C., Lin, Q., Identification of key players for colorectal cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. *J Proteome Res* 2011, *10*, 4373-4387.

[59] Fang, J. Y., Richardson, B. C., The MAPK signalling pathways and colorectal cancer. *Lancet Oncol* 2005, *6*, 322-327.

[60] Slattery, M. L., Lundgreen, A., Wolff, R. K., MAP kinase genes and colon and rectal cancer. *Carcinogenesis* 2012, *33*, 2398-2408.

[61] Segditsas, S., Tomlinson, I., Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* 2006, *25*, 7531-7537.

[62] Xie, W., Rimm, D. L., Lin, Y., Shih, W. J., Reiss, M., Loss of Smad signaling in human colorectal cancer is associated with advanced disease and poor prognosis. *Cancer J* 2003, *9*, 302-312.

[63] Vogelstein, B., Kinzler, K. W., The multistep nature of cancer. *Trends in Genetics* 1993, *9*, 138-141.

[64] Fearon, E. R., Vogelstein, B., A genetic model for colorectal tumorigenesis. *Cell* 1990, *61*, 759-767.

[65] Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Kinzler, K. W., Cancer Genome Landscapes. *Science* 2013, *339*, 1546-1558.

[66] Frank, S. A., Princeton (NJ) 2007.

[67] Powell, S. M., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., Vogelstein, B., Kinzler, K. W., APC mutations occur early during colorectal tumorigenesis. *Nature* 1992, *359*, 235-237.

[68] Kinzler, K., Nilbert, M., Su, L., Vogelstein, B., Bryan, T., Levy, D., Smith, K., Preisinger, A., Hedge, P., McKechnie, D., et, a., Identification of FAP locus genes from chromosome 5q21. *Science* 1991, *253*, 661-665.

[69] Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K. W., Vogelstein, B., Nowak, M. A., Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences* 2010, *107*, 18545-18550.

[70] Whitehall, V., Leggett, B., Microsatellite instability: Detection and management in sporadic colorectal cancer. *Journal of Gastroenterology and Hepatology* 2011, *26*, 1697-1699.

[71] Dietmaier, W., Wallinger, S., Bocker, T., Kullmann, F., Fishel, R., Rüschoff, J., Diagnostic Microsatellite Instability: Definition and Correlation with Mismatch Repair Protein Expression. *Cancer Research* 1997, *57*, 4749-4756.

[72] Jass, J. R., Whitehall, V. L. J., Young, J., Leggett, B. A., Emerging concepts in colorectal neoplasia. *Gastroenterology* 2002, *123*, 862-876.

[73] Esteller, M., Dormant hypermethylated tumour suppressor genes: questions and answers. *The Journal of Pathology* 2005, *205*, 172-180.

[74] Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., Issa, J.-P. J., CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences* 1999, *96*, 8681-8686.

[75] Jimenez, C. R., Knol, J. C., Meijer, G. A., Fijneman, R. J., Proteomics of colorectal cancer: Overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *J Proteomics* 2010.

[76] de Wit, M., Fijneman, R. J. A., Verheul, H. M. W., Meijer, G. A., Jimenez, C. R., Proteomics in colorectal cancer translational research: Biomarker discovery for clinical applications. *Clinical Biochemistry* 2013, *46*, 466-479.

[77] Tishkoff, S. A., Verrelli, B. C., Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 2003, *4*, 293-340.

[78] Itzkowitz, S. H., Bloom, E. J., Kokal, W. A., Modin, G., Hakomori, S.-I., Kim, Y. S., Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients. *Cancer* 1990, *66*, 1960-1966.

[79] Elzagheid, A., Emaetig, F., Buhmeida, A., Laato, M., El-Faitori, O., Syrjänen, K., Collan, Y., Pyrhönen, S., Loss of MUC2 expression predicts disease recurrence and poor outcome in colorectal carcinoma. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 2013, *34*, 621-628.

[80] Imada, T., Rino, Y., Hatori, S., Takahashi, M., Amano, T., Kondo, J., Suda, T., Sialyl Tn antigen expression is associated with the prognosis of patients with advanced colorectal cancer. *Hepato-gastroenterology* 1999, *46*, 208-214.

[81] Petretti, T., Kemmner, W., Schulze, B., Schlag, P. M., Altered mRNA expression of glycosyltransferases in human colorectal carcinomas and liver metastases. *Gut* 2000, *46*, 359-366.

[82] Rangarajan, A., Weinberg, R. A., Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat Rev Cancer* 2003, *3*, 952-959.

[83] Mestas, J., Hughes, C. C., Of mice and not men: differences between mouse and human immunology. *J Immunol* 2004, *172*, 2731-2738.

[84] Brockhausen, I., Yang, J., Lehotay, M., Ogata, S., Itzkowitz, S., Pathways of mucin O-glycosylation in normal and malignant rat colonic epithelial cells reveal a mechanism for cancer-associated Sialyl-Tn antigen expression. *Biol Chem* 2001, *382*, 219-232.

[85] Johnson, R. L., Fleet, J. C., Animal models of colorectal cancer. *Cancer and Metastasis Reviews* 2013, *32*, 39-61.

[86] Flatmark, K., Maelandsmo, G. M., Martinsen, M., Rasmussen, H., Fodstad, O., Twelve colorectal cancer cell lines exhibit highly variable growth and metastatic capacities in an orthotopic model in nude mice. *Eur J Cancer* 2004, *40*, 1593-1598.

[87] Bu, X., Li, L., Li, N., Tian, X., Huang, P., Suppression of mucin 2 enhances the proliferation and invasion of LS174T human colorectal cancer cells. *Cell Biol Int* 2011, *35*, 1121-1129.

[88] Kuan, S. F., Byrd, J. C., Basbaum, C. B., Kim, Y. S., Characterization of quantitative mucin variants from a human colon cancer cell line. *Cancer Res* 1987, *47*, 5715-5724.

324

[89] Moriwaki, K., Noda, K., Furukawa, Y., Ohshima, K., Uchiyama, A., Nakagawa, T., Taniguchi, N., Daigo, Y., Nakamura, Y., Hayashi, N., Miyoshi, E., Deficiency of GMDS leads to escape from NK cell-mediated tumor surveillance through modulation of TRAIL signaling. *Gastroenterology* 2009, *137*, 188-198.

[90] Bresalier, R. S., Niv, Y., Byrd, J. C., Duh, Q. Y., Toribara, N. W., Rockwell, R. W., Dahiya, R., Kim, Y. S., Mucin production by human colonic carcinoma cells correlates with their metastatic potential in animal models of colon cancer metastasis. *J Clin Invest* 1991, *87*, 1037-1045.

[91] Ji, H., Greening, D. W., Barnes, T. W., Lim, J. W., Tauro, B. J., Rai, A., Xu, R., Adda, C., Mathivanan, S., Zhao, W., Xue, Y., Xu, T., Zhu, H.-J., Simpson, R. J., Proteome profiling of exosomes derived from human primary and metastatic colorectal cancer cells reveal differential expression of key metastatic factors and signal transduction components. *Proteomics* 2013, *13*, 1672-1686.

[92] Bresalier, R. S., Byrd, J. C., Brodt, P., Ogata, S., Itzkowitz, S. H., Yunker, C. K., Liver metastasis and adhesion to the sinusoidal endothelium by human colon cancer cells is related to mucin carbohydrate chain length. *Int J Cancer* 1998, *76*, 556-562.

[93] Pan, C., Kumar, C., Bohl, S., Klingmueller, U., Mann, M., Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol Cell Proteomics* 2009, *8*, 443-450.

[94] Ogata, S., Chen, A., Itzkowitz, S. H., Use of model cell lines to study the biosynthesis and biological role of cancer-associated sialosyl-Tn antigen. *Cancer Res* 1994, *54*, 4036-4044.

[95] Brockhausen, I., Yang, J., Dickinson, N., Ogata, S., Itzkowitz, S. H., Enzymatic basis for sialyl-Tn expression in human colon cancer cells. *Glycoconjugate journal* 1998, *15*, 595-603.

[96] Vavasseur, F., Dole, K., Yang, J., Matta, K. L., Myerscough, N., Corfield, A., Paraskeva, C., Brockhausen, I., O-glycan biosynthesis in human colorectal adenoma cells during progression to cancer. *Eur J Biochem* 1994, *222*, 415-424.

[97] Dall'Olio, F., Chiricolo, M., Mariani, E., Facchini, A., Biosynthesis of the cancer-related sialyl-alpha 2,6-lactosaminyl epitope in colon cancer cell lines expressing beta-galactoside alpha 2,6-sialyltransferase under a constitutive promoter. *Eur J Biochem* 2001, *268*, 5876-5884.

[98] Hanski, C., Klussmann, E., Wang, J., Bohm, C., Ogorek, D., Hanski, M. L., Kruger-Krasagakes, S., Eberle, J., Schmitt-Graff, A., Riecken, E. O., Fucosyltransferase III and sialyl-Le(x) expression correlate in cultured colon carcinoma cells but not in colon carcinoma tissue. *Glycoconjugate journal* 1996, *13*, 727-733.

[99] Yue, L., Yu, H., Zhang, C., Liu, J., The effect of overexpression of α1,3-fucosyltransferase VII on adhesive capability of human colon carcinoma HT-29 cells to HUVECs. *Advanced Materials Research* 2012, *345*, 250-256.

[100] Nakagoe, T., Fukushima, K., Nanashima, A., Sawai, T., Tsuji, T., Jibiki, M., Yamaguchi, H., Yasutake, T., Ayabe, H., Matuo, T., Tagawa, Y., Arisawa, K., Expression of Lewis(a), sialyl Lewis(a), Lewis(x) and sialyl Lewis(x) antigens as prognostic factors in patients with colorectal cancer. *Can J Gastroenterol* 2000, *14*, 753-760.

[101] Nakamori, S., Kameyama, M., Imaoka, S., Furukawa, H., Ishikawa, O., Sasaki, Y., Kabuto, T., Iwanaga, T., Matsushita, Y., Irimura, T., Increased Expression of Sialyl Lewisx Antigen Correlates with Poor Survival in Patients with Colorectal Carcinoma: Clinicopathological and Immunohistochemical Study. *Cancer Res* 1993, *53*, 3632-3637.

[102] Robbe-Masselot, C., Herrmann, A., Maes, E., Carlstedt, I., Michalski, J. C., Capon, C., Expression of a core 3 disialyl-Le(x) hexasaccharide in human colorectal cancers: a potential marker of malignant transformation in colon. *J Proteome Res* 2009, *8*, 702-711.

[103] Fong, B., Ma, K., McJarrow, P., Quantification of Bovine Milk Oligosaccharides Using Liquid Chromatography–Selected Reaction Monitoring–Mass Spectrometry. *Journal of Agricultural and Food Chemistry* 2011, *59*, 9788-9795.

[104] Flowers, S. A., Ali, L., Lane, C. S., Olin, M., Karlsson, N. G., Selected Reaction Monitoring to Differentiate and Relatively Quantitate Isomers of Sulfated and Unsulfated Core 1 O-Glycans from Salivary MUC7 Protein in Rheumatoid Arthritis. *Molecular & Cellular Proteomics* 2013, *12*, 921-931.

[105] Zhang, H., Wang, Z., Stupak, J., Ghribi, O., Geiger, J. D., Liu, Q. Y., Li, J., Targeted glycomics by selected reaction monitoring for highly sensitive glycan compositional analysis. *Proteomics* 2012, *12*, 2510-2522.

[106] Fukasawa, T., Asao, T., Yamauchi, H., Ide, M., Tabe, Y., Fujii, T., Yamaguchi, S., Tsutsumi, S., Yazawa, S., Kuwano, H., Associated expression of alpha2,3sialylated type 2 chain structures with lymph node metastasis in distal colorectal cancer. *Surg Today* 2013, *43*, 155-162.

[107] Handerson, T., Camp, R., Harigopal, M., Rimm, D., Pawelek, J., β1,6-Branched Oligosaccharides Are Increased in Lymph Node Metastases and Predict Poor Outcome in Breast Carcinoma. *Clinical Cancer Research* 2005, *11*, 2969-2973.

[108] Fernandes, B., Sagman, U., Auger, M., Demetrio, M., Dennis, J. W., β1–6 Branched Oligosaccharides as a Marker of Tumor Progression in Human Breast and Colon Neoplasia. *Cancer Research* 1991, *51*, 718-723.

[109] Drake, P. M., Schilling, B., Niles, R. K., Prakobphol, A., Li, B., Jung, K., Cho, W., Braten, M., Inerowicz, H. D., Williams, K., Albertolle, M., Held, J. M., Iacovides, D., Sorensen, D. J., Griffith, O. L., Johansen, E., Zawadzka, A. M., Cusack, M. P., Allen, S., Gormley, M., Hall, S. C., Witkowska, H. E., Gray, J. W., Regnier, F., Gibson, B. W., Fisher, S. J., Lectin Chromatography/Mass Spectrometry Discovery Workflow Identifies Putative Biomarkers of Aggressive Breast Cancers. *Journal of Proteome Research* 2012, *11*, 2508-2520.

[110] Abbott, K. L., Lim, J. M., Wells, L., Benigno, B. B., McDonald, J. F., Pierce, M., Identification of candidate biomarkers with cancer-specific glycosylation in the tissue and serum of endometrioid ovarian cancer patients by glycoproteomic analysis. *Proteomics* 2010, *10*, 470-481.

[111] Chen, S., Zheng, T., Shortreed, M. R., Alexander, C., Smith, L. M., Analysis of Cell Surface Carbohydrate Expression Patterns in Normal and Tumorigenic Human Breast Cell Lines Using Lectin Arrays. *Analytical Chemistry* 2007, *79*, 5698-5702.

[112] Kim, E. H., Misek, D. E., Glycoproteomics-based identification of cancer biomarkers. *Int J Proteomics* 2011, *2011*, 601937.

[113] Brisson, J. R., Vinogradov, E., McNally, D. J., Khieu, N. H., Schoenhofen, I. C., Logan, S. M., Jarrell, H., The application of NMR spectroscopy to functional glycomics. *Methods in molecular biology* 2010, *600*, 155-173.

[114] Duus, J., Gotfredsen, C. H., Bock, K., Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem Rev* 2000, *100*, 4589-4614.

[115] Mulloy, B., Hart, G. W., Stanley, P., in: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H.*, et al.* (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY) 2009.

[116] Knox, J. H., Kaur, B., Millward, G. R., Structure and performance of porous graphitic carbon in liquid chromatography. *Journal of Chromatography A* 1986, *352*, 3-25.

[117] Alpert, A. J., Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A* 1990, *499*, 177-196.

[118] Melmer, M., Stangler, T., Premstaller, A., Lindner, W., Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis. *Journal of chromatography. A* 2011, *1218*, 118-123.

[119] Wuhrer, M., Koeleman, C. A., Deelder, A. M., Two-dimensional HPLC separation with reverse-phase-nano-LC-MS/MS for the characterization of glycan pools after labeling with 2-aminobenzamide. *Methods in molecular biology* 2009, *534*, 79-91.

[120] Chen, X., Flynn, G. C., Analysis of N-glycans from recombinant immunoglobulin G by on-line reversed-phase high-performance liquid chromatography/mass spectrometry. *Analytical Biochemistry* 2007, *370*, 147-161.

[121] Jensen, P. H., Karlsson, N. G., Kolarich, D., Packer, N. H., Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* 2012, *7*, 1299-1310.

[122] Larsson, J. M., Karlsson, H., Sjovall, H., Hansson, G. C., A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* 2009, *19*, 756-766.

[123] Davies, M. J., Smith, K. D., Carruthers, R. A., Chai, W., Lawson, A. M., Hounsell, E. F., Use of a porous graphitised carbon column for the high-performance liquid chromatography of oligosaccharides, alditols and glycopeptides with subsequent mass spectrometry analysis. *Journal of Chromatography A* 1993, *646*, 317-326.

[124] Ahn, J., Bones, J., Yu, Y. Q., Rudd, P. M., Gilar, M., Separation of 2-aminobenzamide labeled glycans using hydrophilic interaction chromatography columns packed with 1.7 μm sorbent. *Journal of Chromatography B* 2010, *878*, 403-408.

[125] Wuhrer, M., de Boer, A. R., Deelder, A. M., Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrom Rev* 2009, *28*, 192-206.

[126] Thomsson, K. A., Karlsson, H., Hansson, G. C., Sequencing of Sulfated Oligosaccharides from Mucins by Liquid Chromatography and Electrospray Ionization Tandem Mass Spectrometry. *Analytical Chemistry* 2000, *72*, 4543-4549.

[127] Pabst, M., Bondili, J. S., Stadlmann, J., Mach, L., Altmann, F., Mass + Retention Time = Structure: A Strategy for the Analysis of N-Glycans by Carbon LC-ESI-MS and Its Application to Fibrin N-Glycans. *Analytical Chemistry* 2007, *79*, 5051-5057.

[128] Ali, L., Kenny, D. T., Hayes, C. A., Karlsson, N. G., Structural Identification of O-Linked Oligosaccharides Using Exoglycosidases and MSn Together with UniCarb-DB Fragment Spectra Comparison. *Metabolites* 2012, *2*, 648-666.

[129] Marino, K., Bones, J., Kattla, J. J., Rudd, P. M., A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* 2010, *6*, 713-723.

[130] Fenn, J., Mann, M., Meng, C., Wong, S., Whitehouse, C., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, *246*, 64-71.

[131] Dell, A., Carman, N. H., Tiller, P. R., Thomas-Oates, J. E., Fast atom bombardment mass spectrometric strategies for characterizing carbohydrate-containing biopolymers. *Biological Mass Spectrometry* 1988, *16*, 19-24.

[132] Khoo, K. H., From mass spectrometry-based glycosylation analysis to glycomics and glycoproteomics. *Advances in neurobiology* 2014, *9*, 129-164.

[133] Zaia, J., Mass Spectrometry and the Emerging Field of Glycomics. *Chemistry & Biology* 2008, *15*, 881-892.

[134] Leymarie, N., Zaia, J., Effective use of mass spectrometry for glycan and glycopeptide structural analysis. *Anal Chem* 2012, *84*, 3040-3048.

[135] Douglas, D. J., Frank, A. J., Mao, D., Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews* 2005, *24*, 1-29.

[136] Yates, J. R., Mass Spectral Analysis in Proteomics. *Annual Review of Biophysics and Biomolecular Structure* 2004, *33*, 297-316.

[137] Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., Graham Cooks, R., The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry : JMS* 2005, *40*, 430-443.

[138] Yates, J. R., Ruse, C. I., Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* 2009, *11*, 49-79.

[139] Everest-Dass, A., Abrahams, J., Kolarich, D., Packer, N., Campbell, M., Structural Feature Ions for Distinguishing N- and O-Linked Glycan Isomers by LC-ESI-IT MS/MS. *Journal of the American Society for Mass Spectrometry* 2013, *24*, 895-906.

[140] Han, L., Costello, C. E., Mass spectrometry of glycans. *Biochemistry. Biokhimiia* 2013, *78*, 710-720.

328

[141] Harvey, D. J., Ionization and collision-induced fragmentation of N-linked and related carbohydrates using divalent cations. *Journal of the American Society for Mass Spectrometry* 2001, *12*, 926-937.

[142] Harvey, D. J., Royle, L., Radcliffe, C. M., Rudd, P. M., Dwek, R. A., Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Anal Biochem* 2008, *376*, 44-60.

[143] Zhang, S., Williamson, B. L., Characterization of protein glycosylation using chip-based nanoelectrospray with precursor ion scanning quadrupole linear ion trap mass spectrometry. *J Biomol Tech* 2005, *16*, 209-219.

[144] Deshpande, N., Jensen, P. H., Packer, N. H., Kolarich, D., GlycoSpectrumScan: Fishing Glycopeptides from MS Spectra of Protease Digests of Human Colostrum sIgA. *Journal of Proteome Research* 2009, *9*, 1063-1075.

[145] Apte, A., Meitei, N., in: Li, J. (Ed.), *Functional Glycomics*, Humana Press 2010, pp. 269-281.

[146] Hayes, C. A., Karlsson, N. G., Struwe, W. B., Lisacek, F., Rudd, P. M., Packer, N. H., Campbell, M. P., UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics* 2011, *27*, 1343-1344.

[147] Campbell, M. P., Nguyen-Khuong, T., Hayes, C. A., Flowers, S. A., Alagesan, K., Kolarich, D., Packer, N. H., Karlsson, N. G., Validation of the curation pipeline of UniCarb-DB: Building a global glycan reference MS/MS repository. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2014, *1844*, 108-116.

[148] MCP. Guidelines for the publication of glycomic studies. 2014; Available from: http://www.mcponline.org/site/misc/glycomic.xhtml.

[149] Uemura, T., Shiozaki, K., Yamaguchi, K., Miyazaki, S., Satomi, S., Kato, K., Sakuraba, H., Miyagi, T., Contribution of sialidase NEU1 to suppression of metastasis of human colon cancer cells through desialylation of integrin beta4. *Oncogene* 2009, *28*, 1218-1229.

[150] Miyagi, T., Takahashi, K., Hata, K., Shiozaki, K., Yamaguchi, K., Sialidase significance for cancer progression. *Glycoconjugate journal* 2012, *29*, 567-577.

[151] Ito, H., Hiraiwa, N., Sawada-Kasugai, M., Akamatsu, S., Tachikawa, T., Kasai, Y., Akiyama, S., Ito, K., Takagi, H., Kannagi, R., Altered mRNA expression of specific molecular species of fucosyl- and sialyl-transferases in human colorectal cancer tissues. *Int J Cancer* 1997, *71*, 556-564.

[152] Iwai, T., Kudo, T., Kawamoto, R., Kubota, T., Togayachi, A., Hiruma, T., Okada, T., Kawamoto, T., Morozumi, K., Narimatsu, H., Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc Natl Acad Sci U S A* 2005, *102*, 4572-4577.

[153] Kudo, T., Ikehara, Y., Togayachi, A., Morozumi, K., Watanabe, M., Nakamura, M., Nishihara, S., Narimatsu, H., Up-regulation of a set of glycosyltransferase genes in human colorectal cancer. *Lab Invest* 1998, *78*, 797-811.

[154] Kemmner, W., Roefzaad, C., Haensch, W., Schlag, P. M., Glycosyltransferase expression in human colonic tissue examined by oligonucleotide arrays. *Biochimica et Biophysica Acta (BBA) - General Subjects* 2003, *1621*, 272-279.

[155] Osumi, D., Takahashi, M., Miyoshi, E., Yokoe, S., Lee, S. H., Noda, K., Nakamori, S., Gu, J., Ikeda, Y., Kuroki, Y., Sengoku, K., Ishikawa, M., Taniguchi, N., Core fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells. *Cancer Sci* 2009, *100*, 888-895.

[156] Jia, N., Barclay, W. S., Roberts, K., Yen, H. L., Chan, R. W., Lam, A. K., Air, G., Peiris, J. M., Dell, A., Nicholls, J. M., Haslam, S. M., Glycomic characterisation of respiratory tract tissues of ferrets: implications for its use in influenza virus infection studies. *The Journal of biological chemistry* 2014.

[157] Trinchera, M., Malagolini, N., Chiricolo, M., Santini, D., Minni, F., Caretti, A., Dall'Olio, F., The biosynthesis of the selectin-ligand sialyl Lewis x in colorectal cancer tissues is regulated by fucosyltransferase VI and can be inhibited by an RNA interference-based approach. *The International Journal of Biochemistry & Cell Biology* 2011, *43*, 130-139.

[158] Yang, J. M., Byrd, J. C., Siddiki, B. B., Chung, Y. S., Okuno, M., Sowa, M., Kim, Y. S., Matta, K. L., Brockhausen, I., Alterations of O-glycan biosynthesis in human colon cancer tissues. *Glycobiology* 1994, *4*, 873-884.

[159] Hundt, S., Haug, U., Brenner, H., Blood markers for early detection of colorectal cancer: a systematic review. *Cancer Epidemiol Biomarkers Prev* 2007, *16*, 1935-1953.

[160] Young, G. P., Bosch, L. J., Fecal Tests: From Blood to Molecular Markers. *Curr Colorectal Cancer Rep* 2011, *7*, 62-70.

[161] Ang, C.-S., Rothacker, J., Patsiouras, H., Gibbs, P., Burgess, A. W., Nice, E. C., Use of multiple reaction monitoring for multiplex analysis of colorectal cancer-associated proteins in human feces. *Electrophoresis* 2011, *32*, 1926-1938.

[162] Kwong, G. A., von Maltzahn, G., Murugappan, G., Abudayyeh, O., Mo, S., Papayannopoulos, I. A., Sverdlov, D. Y., Liu, S. B., Warren, A. D., Popov, Y., Schuppan, D., Bhatia, S. N., Mass-encoded synthetic biomarkers for multiplexed urinary monitoring of disease. *Nat Biotech* 2013, *31*, 63-70.

[163] Torpy, J. M., Lynm, C., Glass, R. M., Colon cancer screening. *JAMA* 2006, *295*, 1208-1208.

[164] Akamine, S., Nakagoe, T., Sawai, T., Tsuji, T., Tanaka, K., Hidaka, S., Shibasaki, S., Nanashima, A., Yamaguchi, H., Nagayasu, T., Yasutake, T., Differences in prognosis of colorectal cancer patients based on the expression of sialyl Lewisa, sialyl Lewisx and sialyl Tn antigens in serum and tumor tissue. *Anticancer Res* 2004, *24*, 2541-2546.

[165] Nakamori, S., Kameyama, M., Imaoka, S., Furukawa, H., Ishikawa, O., Sasaki, Y., Izumi, Y., Irimura, T., Involvement of carbohydrate antigen sialyl Lewis X in colorectal cancer metastasis. *Diseases of the Colon & Rectum* 1997, *40*, 420-431.

[166] Bresalier, R. S., Ho, S. B., Schoeppner, H. L., Kim, Y. S., Sleisenger, M. H., Brodt, P., Byrd, J. C., Enhanced sialylation of mucin-associated carbohydrate structures in human colon cancer metastasis. *Gastroenterology* 1996, *110*, 1354-1367.

[167] Nakamori, S., Kameyama, M., Imaoka, S., Furukawa, H., Ishikawa, O., Sasaki, Y., Kabuto, T., Iwanaga, T., Matsushita, Y., Irimura, T., Increased Expression of Sialyl Lewisx Antigen Correlates with Poor Survival in Patients with Colorectal Carcinoma: Clinicopathological and Immunohistochemical Study. *Cancer Research* 1993, *53*, 3632-3637.

[168] Campbell, B. J., Finnie, I. A., Hounsell, E. F., Rhodes, J. M., Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin. *J Clin Invest* 1995, *95*, 571-576.

[169] Thor, A., Ohuchi, N., Szpak, C. A., Johnston, W. W., Schlom, J., Distribution of Oncofetal Antigen Tumor-associated Glycoprotein-72 Defined by Monoclonal Antibody B72.3. *Cancer Res* 1986, *46*, 3118-3124.

[170] Ogata, S., Koganty, R., Reddish, M., Michael Longenecker, B., Chen, A., Perez, C., Itzkowitz, S., Different modes of sialyl-Tn expression during malignant transformation of human colonic mucosa. *Glycoconjugate journal* 1998, *15*, 29-35.

[171] Malagolini, N., Santini, D., Chiricolo, M., Dall'Olio, F., Biosynthesis and expression of the Sda and sialyl Lewis x antigens in normal and cancer colon. *Glycobiology* 2007, *17*, 688-697.

[172] Shibuya, N., Tazaki, K., Song, Z., Tarr, G. E., Goldstein, I. J., Peumans, W. J., A Comparative Study of Bark Lectins from Three Elderberry (Sambucus) Species. *Journal of Biochemistry* 1989, *106*, 1098-1103.

[173] Beauchemin, N., Huot, J., Arabzadeh, A., in: Ablin, R. J., Jiang, W. G., Dvorak, H. F., Gold, P*., et al.* (Eds.), *Metastasis of Colorectal Cancer*, Springer Netherlands 2010, pp. 173-203.

[174] Ohshio, G., Yoshioka, H., Manabe, T., Sakahara, H., Yamabe, H., Imamura, M., Inoue, M., Tanaka, N., Nakada, H., Yamashina, I., Expression of sialosyl-Tn antigen (monoclonal antibody MLS102 reactive) in normal tissues and malignant tumors of the digestive tract. *Journal of Cancer Research and Clinical Oncology* 1994, *120*, 325-330.

[175] Wheelock, M. J., Johnson, K. R., Cadherins as modulators of cellular phenotype. *Annu Rev Cell Dev Biol* 2003, *19*, 207-235.

[176] Christofori, G., Semb, H., The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends in Biochemical Sciences* 1999, *24*, 73-76.

[177] Robbe, C., Capon, C., Coddeville, B., Michalski, J. C., Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *The Biochemical journal* 2004, *384*, 307-316.

[178] King, M. J., Chan, A., Roe, R., Warren, B. F., Dell, A., Morris, H. R., Bartolo, D. C., Durdey, P., Corfield, A. P., Two different glycosyltransferase defects that result in GalNAc alpha-O-peptide (Tn) expression. *Glycobiology* 1994, *4*, 267-279.

[179] Vavasseur, F., Yang, J. M., Dole, K., Paulsen, H., Brockhausen, I., Synthesis of O-glycan core 3: characterization of UDP-GlcNAc: GalNAc-R beta 3-N-acetyl-glucosaminyltransferase activity from

colonic mucosal tissues and lack of the activity in human cancer cell lines. *Glycobiology* 1995, *5*, 351-357.

[180] Akama, R., Sato, Y., Kariya, Y., Isaji, T., Fukuda, T., Lu, L., Taniguchi, N., Ozawa, M., Gu, J., N-acetylglucosaminyltransferase III expression is regulated by cell-cell adhesion via the E-cadherin-catenin-actin complex. *Proteomics* 2008, *8*, 3221-3228.

[181] Shirazi, T., Longman, R. J., Corfield, A. P., Probert, C. S., Mucins and inflammatory bowel disease. *Postgrad Med J* 2000, *76*, 473-478.

[182] McGuckin, M. A., Lindén, S. K., Sutton, P., Florin, T. H., Mucin dynamics and enteric pathogens. *Nat Rev Micro* 2011, *9*, 265-278.

[183] Einerhand, A. W., Renes, I. B., Makkink, M. K., van der Sluis, M., Buller, H. A., Dekker, J., Role of mucins in inflammatory bowel disease: important lessons from experimental models. *Eur J Gastroenterol Hepatol* 2002, *14*, 757-765.

[184] Johansson, M. E., Gustafsson, J. K., Holmen-Larsson, J., Jabbar, K. S., Xia, L., Xu, H., Ghishan, F. K., Carvalho, F. A., Gewirtz, A. T., Sjovall, H., Hansson, G. C., Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* 2014, *63*, 281-291.

[185] Atuma, C., Strugala, V., Allen, A., Holm, L., The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *Am J Physiol Gastrointest Liver Physiol* 2001, *280*, G922-929.

[186] Ermund, A., Schutte, A., Johansson, M. E., Gustafsson, J. K., Hansson, G. C., Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer's patches. *Am J Physiol Gastrointest Liver Physiol* 2013, *305*, G341-347.

[187] Bergstrom, K. S., Xia, L., Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* 2013, *23*, 1026-1037.

[188] Gupta, B., Maher, D., Ebeling, M., Stephenson, P., Puumala, S., Koch, M., Aburatani, H., Jaggi, M., Chauhan, S., Functions and regulation of MUC13 mucin in colon cancer cells. *Journal of Gastroenterology* 2013, 1-14.

[189] Williams, S. J., Wreschner, D. H., Tran, M., Eyre, H. J., Sutherland, G. R., McGuckin, M. A., Muc13, a novel human cell surface mucin expressed by epithelial and hemopoietic cells. *The Journal of biological chemistry* 2001, *276*, 18327-18336.

[190] Pinto, R., Carvalho, A. S., Conze, T., Magalhaes, A., Picco, G., Burchell, J. M., Taylor-Papadimitriou, J., Reis, C. A., Almeida, R., Mandel, U., Clausen, H., Soderberg, O., David, L., Identification of new cancer biomarkers based on aberrant mucin glycoforms by in situ proximity ligation. *J Cell Mol Med* 2012, *16*, 1474-1484.

[191] Sewell, R., Bäckström, M., Dalziel, M., Gschmeissner, S., Karlsson, H., Noll, T., Gätgens, J., Clausen, H., Hansson, G. C., Burchell, J., Taylor-Papadimitriou, J., The ST6GalNAc-I Sialyltransferase

Localizes throughout the Golgi and Is Responsible for the Synthesis of the Tumor-associated Sialyl-Tn O-Glycan in Human Breast Cancer. *Journal of Biological Chemistry* 2006, *281*, 3586-3594.

[192] Conze, T., Carvalho, A. S., Landegren, U., Almeida, R., Reis, C. A., David, L., Söderberg, O., MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. *Glycobiology* 2010, *20*, 199-206.

[193] Sasaki, M., Yamato, T., Nakanuma, Y., Expression of sialyl-Tn, Tn and T antigens in primary liver cancer. *Pathol Int* 1999, *49*, 325-331.

[194] Thomsson, K. A., Prakobphol, A., Leffler, H., Reddy, M. S., Levine, M. J., Fisher, S. J., Hansson, G. C., The salivary mucin MG1 (MUC5B) carries a repertoire of unique oligosaccharides that is large and diverse. *Glycobiology* 2002, *12*, 1-14.

[195] Desgrosellier, J. S., Cheresh, D. A., Integrins in cancer: biological implications and therapeutic opportunities. *Nat Rev Cancer* 2010, *10*, 9-22.

[196] Humphries, M. J., Integrin structure. *Biochem Soc Trans* 2000, *28*, 311-339.

[197] Bates, R. C., Bellovin, D. I., Brown, C., Maynard, E., Wu, B., Kawakatsu, H., Sheppard, D., Oettgen, P., Mercurio, A. M., Transcriptional activation of integrin beta6 during the epithelial-mesenchymal transition defines a novel prognostic indicator of aggressive colon carcinoma. *J Clin Invest* 2005, *115*, 339-347.

[198] Hazelbag, S., Kenter, G. G., Gorter, A., Dreef, E. J., Koopman, L. A., Violette, S. M., Weinreb, P. H., Fleuren, G. J., Overexpression of the αvβ6 integrin in cervical squamous cell carcinoma is a prognostic factor for decreased survival. *The Journal of Pathology* 2007, *212*, 316-324.

[199] Seales, E. C., Jurado, G. A., Brunson, B. A., Wakefield, J. K., Frost, A. R., Bellis, S. L., Hypersialylation of beta1 integrins, observed in colon adenocarcinoma, may contribute to cancer progression by up-regulating cell motility. *Cancer Res* 2005, *65*, 4645-4652.

[200] Shaikh, F. M., Seales, E. C., Clem, W. C., Hennessy, K. M., Zhuo, Y., Bellis, S. L., Tumor cell migration and invasion are regulated by expression of variant integrin glycoforms. *Exp Cell Res* 2008, *314*, 2941-2950.

[201] Zhao, Y., Itoh, S., Wang, X., Isaji, T., Miyoshi, E., Kariya, Y., Miyazaki, K., Kawasaki, N., Taniguchi, N., Gu, J., Deletion of Core Fucosylation on α3β1 Integrin Down-regulates Its Functions. *Journal of Biological Chemistry* 2006, *281*, 38343-38350.

[202] Cavallaro, U., Christofori, G., Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nat Rev Cancer* 2004, *4*, 118-132.

[203] Yap, A. S., Kovacs, E. M., Direct cadherin-activated cell signaling: a view from the plasma membrane. *The Journal of Cell Biology* 2003, *160*, 11-16.

[204] Takeichi, M., Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* 1991, *251*, 1451-1455.

[205] Paschos, K. A., Canovas, D., Bird, N. C., The role of cell adhesion molecules in the progression of colorectal cancer and the development of liver metastasis. *Cell Signal* 2009, *21*, 665-674.

[206] Mohri, Y., Prognostic significance of E-cadherin expression in human colorectal cancer tissue. *Surg Today* 1997, *27*, 606-612.

[207] Kwak, J. M., Min, B. W., Lee, J. H., Choi, J. S., Lee, S. I., Park, S. S., Kim, J., Um, J. W., Kim, S. H., Moon, H. Y., The prognostic significance of E-cadherin and liver intestine-cadherin expression in colorectal cancer. *Dis Colon Rectum* 2007, *50*, 1873-1880.

[208] Geng, F., Shi, B. Z., Yuan, Y. F., Wu, X. Z., The expression of core fucosylated E-cadherin in cancer cells and lung cancer patients: prognostic implications. *Cell Res* 2004, *14*, 423-433.

[209] Pinho, S. S., Reis, C. A., Paredes, J., Magalhaes, A. M., Ferreira, A. C., Figueiredo, J., Xiaogang, W., Carneiro, F., Gartner, F., Seruca, R., The role of N-acetylglucosaminyltransferase III and V in the post-transcriptional modifications of E-cadherin. *Hum Mol Genet* 2009, *18*, 2599-2608.

[210] Iijima, J., Zhao, Y., Isaji, T., Kameyama, A., Nakaya, S., Wang, X., Ihara, H., Cheng, X., Nakagawa, T., Miyoshi, E., Kondo, A., Narimatsu, H., Taniguchi, N., Gu, J., Cell-Cell Interaction-dependent Regulation of N-Acetylglucosaminyltransferase III and the Bisected N-Glycans in GE11 Epithelial Cells: INVOLVEMENT OF E-CADHERIN-MEDIATED CELL ADHESION. *Journal of Biological Chemistry* 2006, *281*, 13038-13046.

[211] Kitada, T., Miyoshi, E., Noda, K., Higashiyama, S., Ihara, H., Matsuura, N., Hayashi, N., Kawata, S., Matsuzawa, Y., Taniguchi, N., The addition of bisecting N-acetylglucosamine residues to E-cadherin down-regulates the tyrosine phosphorylation of beta-catenin. *The Journal of biological chemistry* 2001, *276*, 475-480.

[212] Lalkhen, A. G., McCluskey, A., Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain* 2008, *8*, 221-223.

[213] Fakih, M. G., Padmanabhan, A., CEA monitoring in colorectal cancer. What you should know. *Oncology (Williston Park)* 2006, *20*, 579-587; discussion 588, 594, 596 passim.

[214] Duffy, M. J., Lamerz, R., Haglund, C., Nicolini, A., Kalousová, M., Holubec, L., Sturgeon, C., Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update. *International Journal of Cancer* 2013, n/a-n/a.

[215] Meyerhardt, J. A., Mangu, P. B., Flynn, P. J., Korde, L., Loprinzi, C. L., Minsky, B. D., Petrelli, N. J., Ryan, K., Schrag, D. H., Wong, S. L., Benson, A. B., Follow-Up Care, Surveillance Protocol, and Secondary Prevention Measures for Survivors of Colorectal Cancer: American Society of Clinical Oncology Clinical Practice Guideline Endorsement. *Journal of Clinical Oncology* 2013.

[216] McLeish, J. A., Thursfield, V. J., Giles, G. G., Survival from colorectal cancer in Victoria: 10-year follow up of the 1987 management survey. *ANZ J Surg* 2002, *72*, 352-356.

[217] Leibovitz, A., Stinson, J. C., McCombs, W. B., 3rd, McCoy, C. E., Mazur, K. C., Mabry, N. D., Classification of human colorectal adenocarcinoma cell lines. *Cancer Res* 1976, *36*, 4562-4569.

[218] Tom, B., Rutzky, L., Jakstys, M., Oyasu, R., Kaye, C., Kahan, B., Human colonic adenocarcinoma cells. *In Vitro* 1976, *12*, 180-191.

[219] Chik, J. H. L., Zhou, J., Moh, E. S. X., Christopherson, R., Clarke, S. J., Molloy, M. P., Packer, N. H., Comprehensive glycomics comparison between colon cancer cell cultures and tumours: Implications for biomarker studies. *Journal of Proteomics* 2014, *108*, 146-162.

[220] Zhou, J., Belov, L., Huang, P. Y., Shin, J.-S., Solomon, M. J., Chapuis, P. H., Bokey, L., Chan, C., Clarke, C., Clarke, S. J., Christopherson, R. I., Surface antigen profiling of colorectal cancer using antibody microarrays with fluorescence multiplexing. *Journal of Immunological Methods* 2010, *355*, 40-51.

[221] Lee, A., Kolarich, D., Haynes, P. A., Jensen, P. H., Baker, M. S., Packer, N. H., Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *J Proteome Res* 2009, *8*, 770-781.

[222] Everest-Dass, A. V., Jin, D., Thaysen-Andersen, M., Nevalainen, H., Kolarich, D., Packer, N. H., Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by Candida albicans. *Glycobiology* 2012, *22*, 1465-1479.

[223] Hounsell, E. F., Lawson, A. M., Feeney, J., Gooi, H. C., Pickering, N. J., Stoll, M. S., Lui, S. C., Feizi, T., Structural analysis of the O-glycosidically linked core-region oligosaccharides of human meconium glycoproteins which express oncofoetal antigens. *Eur J Biochem* 1985, *148*, 367-377.

[224] Smith, P. L., Myers, J. T., Rogers, C. E., Zhou, L., Petryniak, B., Becker, D. J., Homeister, J. W., Lowe, J. B., Conditional control of selectin ligand expression and global fucosylation events in mice with a targeted mutation at the FX locus. *The Journal of Cell Biology* 2002, *158*, 801-815.

[225] Lübke, T., Marquardt, T., von Figura, K., Körner, C., A New Type of Carbohydrate-deficient Glycoprotein Syndrome Due to a Decreased Import of GDP-fucose into the Golgi. *Journal of Biological Chemistry* 1999, *274*, 25986-25989.

[226] Kurimoto, A., Kitazume, S., Kizuka, Y., Nakajima, K., Oka, R., Fujinawa, R., Korekane, H., Yamaguchi, Y., Wada, Y., Taniguchi, N., The absence of core fucose up-regulates GnT-III and Wnt target genes: a possible mechanism for an adaptive response in terms of glycan function. *The Journal of biological chemistry* 2014, *289*, 11704-11714.

[227] Wang, X., Inoue, S., Gu, J., Miyoshi, E., Noda, K., Li, W., Mizuno-Horikawa, Y., Nakano, M., Asahi, M., Takahashi, M., Uozumi, N., Ihara, S., Lee, S. H., Ikeda, Y., Yamaguchi, Y., Aze, Y., Tomiyama, Y., Fujii, J., Suzuki, K., Kondo, A., Shapiro, S. D., Lopez-Otin, C., Kuwaki, T., Okabe, M., Honke, K., Taniguchi, N., Dysregulation of TGF-β1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* 2005, *102*, 15791-15796.

[228] Barrow, H., Tam, B., Duckworth, C. A., Rhodes, J. M., Yu, L. G., Suppression of core 1 Gal-transferase is associated with reduction of TF and reciprocal increase of Tn, sialyl-Tn and Core 3 glycans in human colon cancer cells. *PLoS One* 2013, *8*, e59792.

[229] Zhang, W., Cao, P., Chen, S., Spence, A. M., Zhu, S., Staudacher, E., Schachter, H., Synthesis of paucimannose N-glycans by Caenorhabditis elegans requires prior actions of UDP-N-acetyl-D-glucosamine:alpha-3-D-mannoside beta1,2-N-acetylglucosaminyltransferase I, alpha3,6-mannosidase II and a specific membrane-bound beta-N-acetylglucosaminidase. *The Biochemical journal* 2003, *372*, 53-64.

[230] Tulsiani, D. R., Hubbard, S. C., Robbins, P. W., Touster, O., alpha-D-Mannosidases of rat liver Golgi membranes. Mannosidase II is the GlcNAcMAN5-cleaving enzyme in glycoprotein biosynthesis and mannosidases Ia and IB are the enzymes converting Man9 precursors to Man5 intermediates. *Journal of Biological Chemistry* 1982, *257*, 3660-3668.

[231] Altmann, F., Marz, L., Processing of asparagine-linked oligosaccharides in insect cells: evidence for alpha-mannosidase II. *Glycoconjugate journal* 1995, *12*, 150-155.

[232] Altmann, F., Schwihla, H., Staudacher, E., Glössl, J., März, L., Insect Cells Contain an Unusual, Membrane-bound -N-Acetylglucosaminidase Probably Involved in the Processing of Protein N-Glycans. *Journal of Biological Chemistry* 1995, *270*, 17344-17349.

[233] Gil-Martín, E., Rodríguez-Berrocal, J., De La Cadena, M. P., Fernández-Briera, A., Alterations of glycosidases in human colonic adenocarcinoma. *Clinical Biochemistry* 1997, *30*, 17-25.

[234] Kirkegaard, T., Jäättelä, M., Lysosomal involvement in cell death and cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2009, *1793*, 746-754.

[235] Thaysen-Andersen, M., Venkatakrishnan, V., Loke, I., Laurini, C., Diestel, S., Parker, B. L., Packer, N. H., Human neutrophils secrete bioactive paucimannosidic proteins from azurophilic granules into pathogen-infected sputum. *The Journal of biological chemistry* 2015, *290*, 8789-8802.

[236] Mehta, A., Norton, P., Liang, H., Comunale, M. A., Wang, M., Rodemich-Betesh, L., Koszycki, A., Noda, K., Miyoshi, E., Block, T., Increased Levels of Tetra-antennary N-Linked Glycan but Not Core Fucosylation Are Associated with Hepatocellular Carcinoma Tissue. *Cancer Epidemiology Biomarkers & Prevention* 2012, *21*, 925-933.

[237] Millioni, R., Tolin, S., Puricelli, L., Sbrignadello, S., Fadini, G. P., Tessari, P., Arrigoni, G., High Abundance Proteins Depletion *vs* Low Abundance Proteins Enrichment: Comparison of Methods to Reduce the Plasma Proteome Complexity. *PLoS ONE* 2011, *6*, e19603.

[238] Labianca, R., Nordlinger, B., Beretta, G. D., Mosconi, S., Mandalà, M., Cervantes, A., Arnold, D., Group, o. b. o. t. E. G. W., Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2013, *24*, vi64-vi72.

[239] Thomsson, K. A., Karlsson, N. G., Hansson, G. C., Liquid chromatography-electrospray mass spectrometry as a tool for the analysis of sulfated oligosaccharides from mucin glycoproteins. *Journal of chromatography. A* 1999, *854*, 131-139.

[240] Karlsson, N., Karlsson, H., Hansson, G., Strategy for the investigation of O-linked oligosaccharides from mucins based on the separation into neutral, sialic acid- and sulfate-containing species. *Glycoconjugate journal* 1995, *12*, 69-76.

[241] Griffin, T. J., Xie, H., Bandhakavi, S., Popko, J., Mohan, A., Carlis, J. V., Higgins, L., iTRAQ Reagent-Based Quantitative Proteomic Analysis on a Linear Ion Trap Mass Spectrometer. *Journal of Proteome Research* 2007, *6*, 4200-4209.

[242] Hynes, R. O., Integrins: Versatility, modulation, and signaling in cell adhesion. *Cell* 1992, *69*, 11-25.

[243] Wickström, S. A., Fässler, R., Regulation of membrane traffic by integrin signaling. *Trends in Cell Biology* 2011, *21*, 266-273.

[244] Overington, J. P., Al-Lazikani, B., Hopkins, A. L., How many drug targets are there? *Nat Rev Drug Discov* 2006, *5*, 993-996.

[245] Locker, G. Y., Hamilton, S., Harris, J., Jessup, J. M., Kemeny, N., Macdonald, J. S., Somerfield, M. R., Hayes, D. F., Bast, R. C., Jr., ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol* 2006, *24*, 5313-5327.

[246] Chandrasekaran, E. V., Davila, M., Nixon, D. W., Goldfarb, M., Mendicino, J., Isolation and structures of the oligosaccharide units of carcinoembryonic antigen. *The Journal of biological chemistry* 1983, *258*, 7213-7222.

[247] Labianca, R., Nordlinger, B., Beretta, G. D., Brouquet, A., Cervantes, A., Group, O. b. o. t. E. G. W., Primary colon cancer: ESMO Clinical Practice Guidelines for diagnosis, adjuvant treatment and follow-up. *Annals of Oncology* 2010, *21*, v70-v77.

[248] Kume, H., Muraoka, S., Kuga, T., Adachi, J., Narumi, R., Watanabe, S., Kuwano, M., Kodera, Y., Matsushita, K., Fukuoka, J., Masuda, T., Ishihama, Y., Matsubara, H., Nomura, F., Tomonaga, T., Discovery of Colorectal Cancer Biomarker Candidates by Membrane Proteomic Analysis and Subsequent Verification using Selected Reaction Monitoring (SRM) and Tissue Microarray (TMA) Analysis. *Molecular & Cellular Proteomics* 2014, *13*, 1471-1484.

[249] Han, C.-L., Chen, J.-S., Chan, E.-C., Wu, C.-P., Yu, K.-H., Chen, K.-T., Tsou, C.-C., Tsai, C.-F., Chien, C.-W., Kuo, Y.-B., Lin, P.-Y., Yu, J.-S., Hsueh, C., Chen, M.-C., Chan, C.-C., Chang, Y.-S., Chen, Y.-J., An Informatics-assisted Label-free Approach for Personalized Tissue Membrane Proteomics: Case Study on Colorectal Cancer. *Molecular & Cellular Proteomics* 2011, *10*.

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

[251] Zhang, Y., Jiao, J., Yang, P., Lu, H., Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. *Clin Proteomics* 2014, *11*, 18.

[252] Yin, X., Bern, M., Xing, Q., Ho, J., Viner, R., Mayr, M., Glycoproteomic analysis of the secretome of human endothelial cells. *Mol Cell Proteomics* 2013, *12*, 956-978.

[253] Kaji, H., Yamauchi, Y., Takahashi, N., Isobe, T., Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat Protoc* 2006, *1*, 3019-3027.

[254] Vercoutter-Edouart, A.-S., Slomianny, M.-C., Dekeyzer-Beseme, O., Haeuw, J.-F., Michalski, J.-C., Glycoproteomics and glycomics investigation of membrane N-glycosylproteins from human colon carcinoma cells. *PROTEOMICS* 2008, *8*, 3236-3256.

[255] Li, Y., Wen, T., Zhu, M., Li, L., Wei, J., Wu, X., Guo, M., Liu, S., Zhao, H., Xia, S., Huang, W., Wang, P., Wu, Z., Zhao, L., Shui, W., Li, Z., Yin, Z., Glycoproteomic analysis of tissues from patients with colon cancer using lectin microarrays and nanoLC-MS/MS. *Molecular BioSystems* 2013, *9*, 1877-1887.

[256] Park, S.-Y., Lee, S.-H., Kawasaki, N., Itoh, S., Kang, K., Hee Ryu, S., Hashii, N., Kim, J.-M., Kim, J.-Y., Hoe Kim, J., α1-3/4 fucosylation at Asn 241 of β-haptoglobin is a novel marker for colon cancer: A combinatorial approach for development of glycan biomarkers. *International Journal of Cancer* 2012, *130*, 2366-2376.

[257] Park, S. Y., Yoon, S. J., Jeong, Y. T., Kim, J. M., Kim, J. Y., Bernert, B., Ullman, T., Itzkowitz, S. H., Kim, J. H., Hakomori, S. I., N-glycosylation status of beta-haptoglobin in sera of patients with colon cancer, chronic inflammatory diseases and normal subjects. *Int J Cancer* 2010, *126*, 142-155.

[258] Pan, S., Chen, R., Tamura, Y., Crispin, D. A., Lai, L. A., May, D. H., McIntosh, M. W., Goodlett, D. R., Brentnall, T. A., Quantitative glycoproteomics analysis reveals changes in N-glycosylation level associated with pancreatic ductal adenocarcinoma. *J Proteome Res* 2014, *13*, 1293-1306.

[259] Deeb, S. J., Cox, J., Schmidt-Supprian, M., Mann, M., N-linked Glycosylation Enrichment for In-depth Cell Surface Proteomics of Diffuse Large B-cell Lymphoma Subtypes. *Molecular & Cellular Proteomics* 2014, *13*, 240-251.

[260] Wilson, N. L., Schulz, B. L., Karlsson, N. G., Packer, N. H., Sequential analysis of N- and O-linked glycosylation of 2D-PAGE separated glycoproteins. *J Proteome Res* 2002, *1*, 521-529.

[261] Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., Muddiman, D. C., Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). *Analytical Chemistry* 2011, *83*, 5442-5446.

[262] Stravs, M. A., Schymanski, E. L., Singer, H. P., Hollender, J., Automatic recalibration and processing of tandem mass spectra using formula annotation. *Journal of Mass Spectrometry* 2013, *48*, 89-99.

[263] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Analytical Chemistry* 2002, *74*, 5383-5392.

[264] Uniprot. Glycosylation. 2014 [cited 2014 October 7]; Available from: http://www.uniprot.org/manual/carbohyd.

[265] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., Washburn, M. P., Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces cerevisiae. *Journal of Proteome Research* 2006, *5*, 2339-2347.

[266] Mirzaei, M., Soltani, N., Sarhadi, E., Pascovici, D., Keighley, T., Salekdeh, G. H., Haynes, P. A., Atwell, B. J., Shotgun Proteomic Analysis of Long-distance Drought Signaling in Rice Roots. *Journal of Proteome Research* 2011, *11*, 348-358.

[267] Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., Gygi, S. P., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotech* 2006, *24*, 1285-1292.

[268] Crooks, G. E., Hon, G., Chandonia, J. M., Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res* 2004, *14*, 1188-1190.

[269] Chou, M. F., Schwartz, D., Biological sequence motif discovery using motif-x. *Curr Protoc Bioinformatics* 2011, *Chapter 13*, Unit 13 15-24.

[270] Moll, R., Franke, W. W., Schiller, D. L., Geiger, B., Krepler, R., The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell* 1982, *31*, 11-24.

[271] Franke, W. W., Schmid, E., Osborn, M., Weber, K., Different intermediate-sized filaments distinguished by immunofluorescence microscopy. *Proceedings of the National Academy of Sciences* 1978, *75*, 5034-5038.

[272] Christiansen, J. J., Rajasekaran, A. K., Reassessing Epithelial to Mesenchymal Transition as a Prerequisite for Carcinoma Invasion and Metastasis. *Cancer Res* 2006, *66*, 8319-8326.

[273] Cho, S. H., Park, Y. S., Kim, H. J., Kim, C. H., Lim, S. W., Huh, J. W., Lee, J. H., Kim, H. R., CD44 enhances the epithelial-mesenchymal transition in association with colon cancer invasion. *Int J Oncol* 2012, *41*, 211-218.

[274] Zhang, Y., Wei, J., Wang, H., Xue, X., An, Y., Tang, D., Yuan, Z., Wang, F., Wu, J., Zhang, J., Miao, Y., Epithelial mesenchymal transition correlates with CD24+CD44+ and CD133+ cells in pancreatic cancer. *Oncol Rep* 2012, *27*, 1599-1605.

[275] Uniprot. What are the differences between UniprotKB keywords and the GO terms? 2007 [cited 2014 July 30]; Available from: http://www.uniprot.org/faq/23.

[276] Huang da, W., Sherman, B. T., Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009, *4*, 44-57.

[277] Lacombe, M.-L., Munier, A., Mehus, J., Lambeth, D., The Human Nm23/Nucleoside Diphosphate Kinases. *Journal of Bioenergetics and Biomembranes* 2000, *32*, 247-258.

[278] Kimura, N., Shimada, N., Fukuda, M., Ishijima, Y., Miyazaki, H., Ishii, A., Takagi, Y., Ishikawa, N., Regulation of Cellular Functions by Nucleoside Diphosphate Kinases in Mammals. *Journal of Bioenergetics and Biomembranes* 2000, *32*, 309-315.

[279] Sarris, M., Lee, C. S., nm23 protein expression in colorectal carcinoma metastasis in regional lymph nodes and the liver. *Eur J Surg Oncol* 2001, *27*, 170-174.

[280] Kuramitsu, Y., Nakamura, K., Proteomic analysis of cancer tissues: shedding light on carcinogenesis and possible biomarkers. *Proteomics* 2006, *6*, 5650-5661.

[281] Picard, D., Heat-shock protein 90, a chaperone for folding and regulation. *Cell Mol Life Sci* 2002, *59*, 1640-1648.

[282] Walter, S., Buchner, J., Molecular chaperones--cellular machines for protein folding. *Angew Chem Int Ed Engl* 2002, *41*, 1098-1113.

[283] McCormick, L. M., Urade, R., Arakaki, Y., Schwartz, A. L., Bu, G., Independent and Cooperative Roles of N-Glycans and Molecular Chaperones in the Folding and Disulfide Bond Formation of the Low-Density Lipoprotein (LDL) Receptor-Related Protein†. *Biochemistry* 2005, *44*, 5794-5803.

[284] Jolly, C., Morimoto, R. I., Role of the Heat Shock Response and Molecular Chaperones in Oncogenesis and Cell Death. *Journal of the National Cancer Institute* 2000, *92*, 1564-1572.

[285] Dundas, S. R., Lawrie, L. C., Rooney, P. H., Murray, G. I., Mortalin is over-expressed by colorectal adenocarcinomas and correlates with poor survival. *The Journal of Pathology* 2005, *205*, 74-81.

[286] Wadhwa, R., Takano, S., Kaur, K., Deocaris, C. C., Pereira-Smith, O. M., Reddel, R. R., Kaul, S. C., Upregulation of mortalin/mthsp70/Grp75 contributes to human carcinogenesis. *Int J Cancer* 2006, *118*, 2973-2980.

[287] Drecoll, E., Nitsche, U., Bauer, K., Berezowska, S., Slotta-Huspenina, J., Rosenberg, R., Langer, R., Expression analysis of heat shock protein 90 (HSP90) and Her2 in colon carcinoma. *International Journal of Colorectal Disease* 2014, *29*, 663-671.

[288] Moser, C., Lang, S. A., Kainz, S., Gaumann, A., Fichtner-Feigl, S., Koehl, G. E., Schlitt, H. J., Geissler, E. K., Stoeltzing, O., Blocking heat shock protein-90 inhibits the invasive properties and hepatic growth of human colon cancer cells and improves the efficacy of oxaliplatin in p53-deficient colon cancer tumors in vivo. *Mol Cancer Ther* 2007, *6*, 2868-2878.

[289] Alfonso, P., Canamero, M., Fernandez-Carbonie, F., Nunez, A., Casal, J. I., Proteome analysis of membrane fractions in colorectal carcinomas by using 2D-DIGE saturation labeling. *J Proteome Res* 2008, *7*, 4247-4255.

[290] Kwong, K. Y., Bloom, G. C., Yang, I., Boulware, D., Coppola, D., Haseman, J., Chen, E., McGrath, A., Makusky, A. J., Taylor, J., Steiner, S., Zhou, J., Yeatman, T. J., Quackenbush, J., Synchronous global assessment of gene and protein expression in colorectal cancer progression. *Genomics* 2005, *86*, 142-158.

[291] Lu, X., Xiao, L., Wang, L., Ruden, D. M., Hsp90 inhibitors and drug resistance in cancer: The potential benefits of combination therapies of Hsp90 inhibitors and other anti-cancer drugs. *Biochemical Pharmacology* 2012, *83*, 995-1004.

[292] Workman, P., Burrows, F., Neckers, L. E. N., Rosen, N., Drugging the Cancer Chaperone HSP90. *Annals of the New York Academy of Sciences* 2007, *1113*, 202-216.

[293] Mathupala, S. P., Pedersen, P. L., Voltage dependent anion channel-1 (VDAC-1) as an anti-cancer target. *Cancer Biol Ther* 2010, *9*, 1053-1056.

340

[294] Evans, C. G., Chang, L., Gestwicki, J. E., Heat shock protein 70 (hsp70) as an emerging drug target. *J Med Chem* 2010, *53*, 4585-4602.

[295] Guzhova, I. V., Shevtsov, M. A., Abkin, S. V., Pankratova, K. M., Margulis, B. A., Intracellular and extracellular Hsp70 chaperone as a target for cancer therapy. *International Journal of Hyperthermia* 2013, *29*, 399-408.

[296] Ha, S.-A., Shin, S. M., Lee, Y. J., Kim, S., Kim, H. K., Namkoong, H., Lee, H., Lee, Y. S., Cho, Y.-S., Park, Y. G., Jeon, H. M., Oh, C., Kim, J. W., HCCRBP-1 directly interacting with HCCR-1 induces tumorigenesis through P53 stabilization. *International Journal of Cancer* 2008, *122*, 501-508.

[297] Baldassarre, M., Razinia, Z., N. Brahme, N., Buccione, R., Calderwood, D. A., Filamin A controls matrix metalloproteinase activity and regulates cell invasion in human fibrosarcoma cells. *Journal of Cell Science* 2012, *125*, 3858-3869.

[298] Kim, H., Nakamura, F., Lee, W., Shifrin, Y., Arora, P., McCulloch, C. A., Filamin A is required for vimentin-mediated cell adhesion and spreading. *Am J Physiol Cell Physiol* 2010, *298*, C221-236.

[299] Fiori, J. L., Zhu, T. N., O'Connell, M. P., Hoek, K. S., Indig, F. E., Frank, B. P., Morris, C., Kole, S., Hasskamp, J., Elias, G., Weeraratna, A. T., Bernier, M., Filamin A modulates kinase activation and intracellular trafficking of epidermal growth factor receptors in human melanoma cells. *Endocrinology* 2009, *150*, 2551-2560.

[300] Martinez-Outschoorn, U. E., Lin, Z., Whitaker-Menezes, D., Howell, A., Sotgia, F., Lisanti, M. P., Ketone body utilization drives tumor growth and metastasis. *Cell Cycle* 2012, *11*, 3964-3971.

[301] Ang, C.-S., Nice, E. C., Targeted In-Gel MRM: A Hypothesis Driven Approach for Colorectal Cancer Biomarker Discovery in Human Feces. *J Proteome Res* 2010, *9*, 4346-4355.

[302] Camarero, N., Mascaró, C., Mayordomo, C., Vilardell, F., Haro, D., Marrero, P. F., Ketogenic HMGCS2 Is a c-Myc Target Gene Expressed in Differentiated Cells of Human Colonic Epithelium and Down-Regulated in Colon Cancer. *Molecular Cancer Research* 2006, *4*, 645-653.

[303] Moeller, G., Adamski, J., Integrated view on 17beta-hydroxysteroid dehydrogenases. *Molecular and Cellular Endocrinology* 2009, *301*, 7-19.

[304] Chen, N., Sun, W., Deng, X., Hao, Y., Chen, X., Xing, B., Jia, W., Ma, J., Wei, H., Zhu, Y., Qian, X., Jiang, Y., He, F., Quantitative proteome analysis of HCC cell lines with different metastatic potentials by SILAC. *Proteomics* 2008, *8*, 5108-5118.

[305] Szajnik, M., Szczepanski, M. J., Elishaev, E., Visus, C., Lenzner, D., Zabel, M., Glura, M., DeLeo, A. B., Whiteside, T. L., 17beta Hydroxysteroid dehydrogenase type 12 (HSD17B12) is a marker of poor prognosis in ovarian carcinoma. *Gynecol Oncol* 2012, *127*, 587-594.

[306] Marchais-Oberwinkler, S., Henn, C., Möller, G., Klein, T., Negri, M., Oster, A., Spadaro, A., Werth, R., Wetzel, M., Xu, K., Frotscher, M., Hartmann, R. W., Adamski, J., 17β-Hydroxysteroid dehydrogenases (17β-HSDs) as therapeutic targets: Protein structures, functions, and recent progress in inhibitor development. *The Journal of Steroid Biochemistry and Molecular Biology* 2011, *125*, 66-82.

[307] English, M. A., Hughes, S. V., Kane, K. F., Langman, M. J., Stewart, P. M., Hewison, M., Oestrogen inactivation in the colon: analysis of the expression and regulation of 17beta-hydroxysteroid dehydrogenase isozymes in normal colon and colonic cancer. *Br J Cancer* 2000, *83*, 550-558.

[308] English, M. A., Kane, K. F., Cruickshank, N., Langman, M. J., Stewart, P. M., Hewison, M., Loss of estrogen inactivation in colonic cancer. *J Clin Endocrinol Metab* 1999, *84*, 2080-2085.

[309] Ushio-Fukai, M., Nakamura, Y., Reactive oxygen species and angiogenesis: NADPH oxidase as target for cancer therapy. *Cancer Letters* 2008, *266*, 37-52.

[310] Brown, D. I., Griendling, K. K., Nox proteins in signal transduction. *Free Radic Biol Med* 2009, *47*, 1239-1253.

[311] Li, Q., Fu, G.-B., Zheng, J.-T., He, J., Niu, X.-B., Chen, Q.-D., Yin, Y., Qian, X., Xu, Q., Wang, M., Sun, A.-F., Shu, Y., Rui, H., Liu, L.-Z., Jiang, B.-H., NADPH oxidase subunit p22phox-mediated reactive oxygen species contribute to angiogenesis and tumor growth through AKT and ERK1/2 signaling pathways in prostate cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2013, *1833*, 3375-3385.

[312] Edderkaoui, M., Nitsche, C., Zheng, L., Pandol, S. J., Gukovsky, I., Gukovskaya, A. S., NADPH Oxidase Activation in Pancreatic Cancer Cells Is Mediated through Akt-dependent Up-regulation of p22phox. *Journal of Biological Chemistry* 2011, *286*, 7779-7787.

[313] Block, K., Gorin, Y., New, D. D., Eid, A., Chelmicki, T., Reed, A., Choudhury, G. G., Parekh, D. J., Abboud, H. E., The NADPH oxidase subunit p22phox inhibits the function of the tumor suppressor protein tuberin. *Am J Pathol* 2010, *176*, 2447-2455.

[314] Kikuchi, H., Hikage, M., Miyashita, H., Fukumoto, M., NADPH oxidase subunit, gp91phox homologue, preferentially expressed in human colon epithelial cells. *Gene* 2000, *254*, 237-243.

[315] Ambasta, R. K., Kumar, P., Griendling, K. K., Schmidt, H. H. H. W., Busse, R., Brandes, R. P., Direct Interaction of the Novel Nox Proteins with p22phox Is Required for the Formation of a Functionally Active NADPH Oxidase. *Journal of Biological Chemistry* 2004, *279*, 45935-45941.

[316] Bedard, K., Krause, K. H., The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol Rev* 2007, *87*, 245-313.

[317] Chen, T. C., Hung, Y. C., Lin, T. Y., Chang, H. W., Chiang, I. P., Chen, Y. Y., Chow, K. C., Human papillomavirus infection and expression of ATPase family AAA domain containing 3A, a novel anti-autophagy factor, in uterine cervical cancer. *Int J Mol Med* 2011, *28*, 689-696.

[318] Fang, H.-Y., Chang, C.-L., Hsu, S.-H., Huang, C.-Y., Chiang, S.-F., Chiou, S.-H., Huang, C.-H., Hsiao, Y.-T., Lin, T.-Y., Chiang, I.-P., Hsu, W.-H., Sugano, S., Chen, C.-Y., Lin, C.-Y., Ko, W.-J., Chow, K.-C., ATPase family AAA domain-containing 3A is a novel anti-apoptotic factor in lung adenocarcinoma cells. *Journal of Cell Science* 2010, *123*, 1171-1180.

[319] Li, S., Lamarche, F., Charton, R., Delphin, C., Gires, O., Hubstenberger, A., Schlattner, U., Rousseau, D., Expression analysis of ATAD3 isoforms in rodent and human cell lines and tissues. *Gene* 2014, *535*, 60-69.

[320] Huang, K. H., Chow, K. C., Chang, H. W., Lin, T. Y., Lee, M. C., ATPase family AAA domain containing 3A is an anti-apoptotic factor and a secretion regulator of PSA in prostate cancer. *Int J Mol Med* 2011, *28*, 9-15.

[321] Parker, B. L., Thaysen-Andersen, M., Solis, N., Scott, N. E., Larsen, M. R., Graham, M. E., Packer, N. H., Cordwell, S. J., Site-Specific Glycan-Peptide Analysis for Determination of N-Glycoproteome Heterogeneity. *J Proteome Res* 2013, *12*, 5791-5800.

[322] Zielinska, D. F., Gnad, F., Wiśniewski, J. R., Mann, M., Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* 2010, *141*, 897-907.

[323] Krokhin, O. V., Antonovici, M., Ens, W., Wilkins, J. A., Standing, K. G., Deamidation of -Asn-Gly-sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem* 2006, *78*, 6645-6650.

[324] Palmisano, G., Melo-Braga, M. N., Engholm-Keller, K., Parker, B. L., Larsen, M. R., Chemical Deamidation: A Common Pitfall in Large-Scale N-Linked Glycoproteomic Mass Spectrometry-Based Analyses. *J Proteome Res* 2012, *11*, 1949-1957.

[325] Robinson, N. E., Robinson, Z. W., Robinson, B. R., Robinson, A. L., Robinson, J. A., Robinson, M. L., Robinson, A. B., Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J Pept Res* 2004, *63*, 426-436.

[326] Schwartz, D., Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotech* 2005, *23*, 1391-1398.

[327] Hao, P., Ren, Y., Alpert, A. J., Sze, S. K., Detection, Evaluation and Minimization of Nonenzymatic Deamidation in Proteomic Sample Preparation. *Molecular & Cellular Proteomics* 2011, *10*.

[328] Nepomuceno, A. I., Gibson, R. J., Randall, S. M., Muddiman, D. C., Accurate Identification of Deamidated Peptides in Global Proteomics Using a Quadrupole Orbitrap Mass Spectrometer. *Journal of Proteome Research* 2013, *13*, 777-785.

[329] Gupta, R., Brunak, S., Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 2002, 310-322.

[330] Jones, J., Krag, S. S., Betenbaugh, M. J., Controlling N-linked glycan site occupancy. *Biochimica et Biophysica Acta (BBA) - General Subjects* 2005, *1726*, 121-137.

[331] Nilsson, I. M., von Heijne, G., Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *Journal of Biological Chemistry* 1993, *268*, 5798-5801.

[332] Gavel, Y., von Heijne, G., Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* 1990, *3*, 433-442.

[333] Hülsmeier, A. J., Paesold-Burda, P., Hennet, T., N-Glycosylation Site Occupancy in Serum Glycoproteins Using Multiple Reaction Monitoring Liquid Chromatography-Mass Spectrometry. *Molecular & Cellular Proteomics* 2007, *6*, 2132-2138.

[334] Sareneva, T., Pirhonen, J., Cantell, K., Julkunen, I., N-glycosylation of human interferon-gamma: glycans at Asn-25 are critical for protease resistance. *The Biochemical journal* 1995, *308 (Pt 1)*, 9-14.

[335] Wielenga, V. J., Heider, K. H., Offerhaus, G. J., Adolf, G. R., van den Berg, F. M., Ponta, H., Herrlich, P., Pals, S. T., Expression of CD44 variant proteins in human colorectal cancer is related to tumor progression. *Cancer Res* 1993, *53*, 4754-4756.

[336] Mulder, J. W. R., Sewnath, M., Offerhaus, G., Pals, S., Oosting, J., Kruyt, P., Weidema, W., Seldenrijk, C., Colorectal cancer prognosis and expression of exon-v6-containing CD44 proteins. *The Lancet* 1994, *344*, 1470-1472.

[337] Thorsen, K., Mansilla, F., Schepeler, T., Oster, B., Rasmussen, M. H., Dyrskjot, L., Karni, R., Akerman, M., Krainer, A. R., Laurberg, S., Andersen, C. L., Orntoft, T. F., Alternative splicing of SLC39A14 in colorectal cancer is regulated by the Wnt pathway. *Mol Cell Proteomics* 2011, *10*, M110 002998.

[338] Sveen, A., Bakken, A. C., Ågesen, T. H., Lind, G. E., Nesbakken, A., Nordgård, O., Brackmann, S., Rognum, T. O., Lothe, R. A., Skotheim, R. I., The exon-level biomarker SLC39A14 has organ-confined cancer-specificity in colorectal cancer. *International Journal of Cancer* 2012, *131*, 1479-1485.

[339] Araki, K., Mikami, T., Yoshida, T., Kikuchi, M., Sato, Y., Oh-ishi, M., Kodera, Y., Maeda, T., Okayasu, I., High expression of HSP47 in ulcerative colitis-associated carcinomas: proteomic approach. *Br J Cancer* 2009, *101*, 492-497.

[340] Song, Y., Leonard, S. W., Traber, M. G., Ho, E., Zinc Deficiency Affects DNA Damage, Oxidative Stress, Antioxidant Defenses, and DNA Repair in Rats. *The Journal of Nutrition* 2009, *139*, 1626-1631.

[341] Iwai, L. K., Payne, L. S., Luczynski, M. T., Chang, F., Xu, H., Clinton, R. W., Paul, A., Esposito, E. A., Gridley, S., Leitinger, B., Naegle, K. M., Huang, P. H., Phosphoproteomics of collagen receptor networks reveals SHP-2 phosphorylation downstream of wild-type DDR2 and its lung cancer mutants. *The Biochemical journal* 2013, *454*, 501-513.

[342] Xu, G., Zhang, W., Ma, M. K., McLeod, H. L., Human carboxylesterase 2 is commonly expressed in tumor tissue and is correlated with activation of irinotecan. *Clin Cancer Res* 2002, *8*, 2605-2611.

[343] Jones, R. P., Sutton, P., Greensmith, R. M., Santoyo-Castelazo, A., Carr, D. F., Jenkins, R., Rowe, C., Hamlett, J., Park, B. K., Terlizzo, M., O'Grady, E., Ghaneh, P., Fenwick, S. W., Malik, H. Z., Poston, G. J., Kitteringham, N. R., Hepatic activation of irinotecan predicts tumour response in patients with colorectal liver metastases treated with DEBIRI: exploratory findings from a phase II study. *Cancer Chemother Pharmacol* 2013, *72*, 359-368.

[344] Burkhard, P., Stetefeld, J., Strelkov, S. V., Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol* 2001, *11*, 82-88.

[345] Becker-Heck, A., Zohn, I. E., Okabe, N., Pollock, A., Lenhart, K. B., Sullivan-Brown, J., McSheene, J., Loges, N. T., Olbrich, H., Haeffner, K., Fliegauf, M., Horvath, J., Reinhardt, R., Nielsen, K. G., Marthin, J. K., Baktai, G., Anderson, K. V., Geisler, R., Niswander, L., Omran, H., Burdine, R. D., The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nat Genet* 2011, *43*, 79-84.

[346] Knodler, L. A., Ibarra, J. A., Perez-Rueda, E., Yip, C. K., Steele-Mortimer, O., Coiled-coil domains enhance the membrane association of Salmonella type III effectors. *Cell Microbiol* 2011, *13*, 1497-1517.

[347] Qian, F., Germino, F. J., Cai, Y., Zhang, X., Somlo, S., Germino, G. G., PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nat Genet* 1997, *16*, 179-183.

[348] Nusrat, A., Chen, J. A., Foley, C. S., Liang, T. W., Tom, J., Cromwell, M., Quan, C., Mrsny, R. J., The Coiled-coil Domain of Occludin Can Act to Organize Structural and Functional Elements of the Epithelial Tight Junction. *Journal of Biological Chemistry* 2000, *275*, 29816-29822.

[349] Koshland, M. E., The coming of age of the immunoglobulin J chain. *Annu Rev Immunol* 1985, *3*, 425-453.

[350] Johansen, F. E., Braathen, R., Brandtzaeg, P., Role of J chain in secretory immunoglobulin formation. *Scand J Immunol* 2000, *52*, 240-248.

[351] Thelemann, C., Eren, R. O., Coutaz, M., Brasseit, J., Bouzourene, H., Rosa, M., Duval, A., Lavanchy, C., Mack, V., Mueller, C., Reith, W., Acha-Orbea, H., Interferon-γ Induces Expression of MHC Class II on Intestinal Epithelial Cells and Protects Mice from Colitis. *PLoS ONE* 2014, *9*, e86844.

[352] Ong, S.-E., Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protocols* 2007, *1*, 2650-2660.

[353] Martinez-Aguilar, J., Chik, J., Nicholson, J., Semaan, C., McKay, M. J., Molloy, M. P., Quantitative mass spectrometry for colorectal cancer proteomics. *Proteomics Clin Appl* 2013, *7*, 42-54.

[354] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Meth* 2010, *7*, 383-385.

[355] Momburg, F., Moldenhauer, G., Hammerling, G. J., Moller, P., Immunohistochemical study of the expression of a Mr 34,000 human epithelium-specific surface glycoprotein in normal and malignant tissues. *Cancer Res* 1987, *47*, 2883-2891.

[356] Moldenhauer, G., Momburg, F., Moller, P., Schwartz, R., Hammerling, G. J., Epithelium-specific surface glycoprotein of Mr 34,000 is a widely distributed human carcinoma marker. *Br J Cancer* 1987, *56*, 714-721.

[357] Friederich, E., Vancompernolle, K., Louvard, D., Vandekerckhove, J., Villin function in the organization of the actin cytoskeleton. Correlation of in vivo effects to its biochemical activities in vitro. *The Journal of biological chemistry* 1999, *274*, 26751-26760.

[358] Bretscher, A., Weber, K., Villin: the major microfilament-associated protein of the intestinal microvillus. *Proc Natl Acad Sci U S A* 1979, *76*, 2321-2325.

[359] West, A. B., Isaac, C. A., Carboni, J. M., Morrow, J. S., Mooseker, M. S., Barwick, K. W., Localization of villin, a cytoskeletal protein specific to microvilli, in human ileum and colon and in colonic neoplasms. *Gastroenterology* 1988, *94*, 343-352.

[360] Chaponnier, C., Gabbiani, G., Pathological situations characterized by altered actin isoform expression. *J Pathol* 2004, *204*, 386-395.

[361] Watanabe, K., Kusakabe, T., Hoshi, N., Saito, A., Suzuki, T., h-Caldesmon in leiomyosarcoma and tumors with smooth muscle cell-like differentiation: its specific expression in the smooth muscle cell tumor. *Hum Pathol* 1999, *30*, 392-396.

[362] Winder, S. J., Walsh, M. P., Calponin: thin filament-linked regulation of smooth muscle contraction. *Cell Signal* 1993, *5*, 677-686.

[363] Rangdaeng, S., Truong, L. D., Comparative immunohistochemical staining for desmin and muscle-specific actin. A study of 576 cases. *Am J Clin Pathol* 1991, *96*, 32-45.

[364] Salgame, P., Abrams, J. S., Clayberger, C., Goldstein, H., Convit, J., Modlin, R. L., Bloom, B. R., Differing lymphokine profiles of functional subsets of human CD4 and CD8 T cell clones. *Science* 1991, *254*, 279-282.

[365] Rahemtulla, A., Fung-Leung, W. P., Schilham, M. W., Kundig, T. M., Sambhara, S. R., Narendran, A., Arabian, A., Wakeham, A., Paige, C. J., Zinkernagel, R. M., Miller, R. G., Mak, T. W., Normal development and function of CD8+ cells but markedly decreased helper cell activity in mice lacking CD4. *Nature* 1991, *353*, 180-184.

[366] Pesando, J. M., Bouchard, L. S., McMaster, B. E., CD19 is functionally and physically associated with surface immunoglobulin. *J Exp Med* 1989, *170*, 2159-2164.

[367] Clark, E. A., Ledbetter, J. A., Leukocyte cell surface enzymology: CD45 (LCA, T200) is a protein tyrosine phosphatase. *Immunol Today* 1989, *10*, 225-228.

[368] Julien, S., Videira, P. A., Delannoy, P., Sialyl-Tn in Cancer: (How) Did We Miss the Target? *Biomolecules* 2012, *2*, 435-466.

[369] Dahiya, R., Itzkowitz, S. H., Byrd, J. C., Kim, Y. S., Mucin oligosaccharide biosynthesis in human colonic cancerous tissues and cell lines. *Cancer* 1992, *70*, 1467-1476.

[370] Kronewitter, S. R., De Leoz, M. L. A., Strum, J. S., An, H. J., Dimapasoc, L. M., Guerrero, A., Miyamoto, S., Lebrilla, C. B., Leiserowitz, G. S., The glycolyzer: Automated glycan annotation software for high performance mass spectrometry and its application to ovarian cancer glycan biomarker discovery. *Proteomics* 2012, *12*, 2523-2538.

[371] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.

[372] Miwa, H. E., Koba, W. R., Fine, E. J., Giricz, O., Kenny, P. A., Stanley, P., Bisected, complex N-glycans and galectins in mouse mammary tumor progression and human breast cancer. *Glycobiology* 2013, *23*, 1477-1490.

[373] Anugraham, M., Jacob, F., Nixdorf, S., Everest-Dass, A. V., Heinzelmann-Schwarz, V., Packer, N. H., Specific glycosylation of membrane proteins in epithelial ovarian cancer cell lines: glycan structures reflect gene expression and DNA methylation status. *Molecular & Cellular Proteomics* 2014.

[374] Isaji, T., Gu, J., Nishiuchi, R., Zhao, Y., Takahashi, M., Miyoshi, E., Honke, K., Sekiguchi, K., Taniguchi, N., Introduction of Bisecting GlcNAc into Integrin α5β1 Reduces Ligand Binding and Down-regulates Cell Adhesion and Cell Migration. *Journal of Biological Chemistry* 2004, *279*, 19747-19754.

[375] Venkatakrishnan, V., Thaysen-Andersen, M., Chen, S. C. A., Nevalainen, H., Packer, N. H., Cystic fibrosis and bacterial colonisation define the sputum N-glycosylation phenotype. *Glycobiology* 2014.

[376] Hashii, N., Kawasaki, N., Itoh, S., Nakajima, Y., Kawanishi, T., Yamaguchi, T., Alteration of N-glycosylation in the kidney in a mouse model of systemic lupus erythematosus: relative quantification of N-glycans using an isotope-tagging method. *Immunology* 2009, *126*, 336-345.

[377] Arnold, J. N., Saldova, R., Galligan, M. C., Murphy, T. B., Mimura-Kimura, Y., Telford, J. E., Godwin, A. K., Rudd, P. M., Novel Glycan Biomarkers for the Detection of Lung Cancer. *J Proteome Res* 2011, *10*, 1755-1764.

[378] Alley, W. R., Vasseur, J. A., Goetz, J. A., Svoboda, M., Mann, B. F., Matei, D. E., Menning, N., Hussein, A., Mechref, Y., Novotny, M. V., N-linked Glycan Structures and Their Expressions Change in the Blood Sera of Ovarian Cancer Patients. *J Proteome Res* 2012, *11*, 2282-2300.

[379] Imai, J., Ghazizadeh, M., Naito, Z., Asano, G., Immunohistochemical expression of T, Tn and sialyl-Tn antigens and clinical outcome in human breast carcinoma. *Anticancer Res* 2001, *21*, 1327-1334.

[380] Lloyd, K. O., Burchell, J., Kudryashov, V., Yin, B. W., Taylor-Papadimitriou, J., Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. Demonstration of simpler and fewer glycan chains in tumor cells. *The Journal of biological chemistry* 1996, *271*, 33325-33334.

[381] Lavrsen, K., Madsen, C. B., Rasch, M. G., Woetmann, A., Ødum, N., Mandel, U., Clausen, H., Pedersen, A. E., Wandall, H. H., Aberrantly glycosylated MUC1 is expressed on the surface of breast cancer cells and a target for antibody-dependent cell-mediated cytotoxicity. *Glycoconjugate journal* 2013, *30*, 227-236.

[382] Soares, R., Marinho, A., Schmitt, F., Expression of sialyl-Tn in breast cancer. Correlation with prognostic parameters. *Pathol Res Pract* 1996, *192*, 1181-1186.

[383] Tashiro, Y., Yonezawa, S., Kim, Y. S., Sato, E., Immunohistochemical study of mucin carbohydrates and core proteins in human ovarian tumors. *Hum Pathol* 1994, *25*, 364-372.

[384] Ryuko, K., Iwanari, O., Kitao, M., Moriwaki, S., Immunohistochemical evaluation of sialosyl-Tn antigens in various ovarian carcinomas. *Gynecol Oncol* 1993, *49*, 215-224.

[385] Davidson, B., Berner, A., Nesland, J. M., Risberg, B., Kristensen, G. B., Trope, C. G., Bryne, M., Carbohydrate antigen expression in primary tumors, metastatic lesions, and serous effusions from

patients diagnosed with epithelial ovarian carcinoma: evidence of up-regulated Tn and Sialyl Tn antigen expression in effusions. *Hum Pathol* 2000, *31*, 1081-1087.

[386] Ogawa, H., Ghazizadeh, M., Araki, T., Tn and sialyl-Tn antigens as potential prognostic markers in human ovarian carcinoma. *Gynecol Obstet Invest* 1996, *41*, 278-283.

[387] Ohno, S., Ohno, Y., Nakada, H., Suzuki, N., Soma, G., Inoue, M., Expression of Tn and sialyl-Tn antigens in endometrial cancer: its relationship with tumor-produced cyclooxygenase-2, tumor-infiltrated lymphocytes and patient prognosis. *Anticancer Res* 2006, *26*, 4047-4053.

[388] Larsson, J. M., Karlsson, H., Crespo, J. G., Johansson, M. E., Eklund, L., Sjovall, H., Hansson, G. C., Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflamm Bowel Dis* 2011, *17*, 2299-2307.

[389] Itzkowitz, S., Kjeldsen, T., Friera, A., Hakomori, S., Yang, U. S., Kim, Y. S., Expression of Tn, sialosyl Tn, and T antigens in human pancreas. *Gastroenterology* 1991, *100*, 1691-1700.

[390] van der Post, S., Hansson, G. C., MEMBRANE PROTEIN PROFILING OF HUMAN COLON REVEALS DISTINCT REGIONAL DIFFERENCES. *Molecular & Cellular Proteomics* 2014.

[391] Campbell, B. J., Yu, L. G., Rhodes, J. M., Altered glycosylation in inflammatory bowel disease: a possible role in cancer development. *Glycoconjugate journal* 2001, *18*, 851-858.

[392] Ullman, T. A., Itzkowitz, S. H., Intestinal Inflammation and Cancer. *Gastroenterology* 2011, *140*, 1807-1816.e1801.

[393] Bell, A. W., Ward, M. A., Blackstock, W. P., Freeman, H. N., Choudhary, J. S., Lewis, A. P., Chotai, D., Fazel, A., Gushue, J. N., Paiement, J., Palcy, S., Chevet, E., Lafreniere-Roula, M., Solari, R., Thomas, D. Y., Rowley, A., Bergeron, J. J., Proteomics characterization of abundant Golgi membrane proteins. *The Journal of biological chemistry* 2001, *276*, 5152-5165.

[394] Bergeron, J. J., Paiement, J., Khan, M. N., Smith, C. E., Terminal glycosylation in rat hepatic Golgi fractions: heterogeneous locations for sialic acid and galactose acceptors and their transferases. *Biochim Biophys Acta* 1985, *821*, 393-403.

[395] Murugaesu, N., Iravani, M., van Weverwijk, A., Ivetic, A., Johnson, D. A., Antonopoulos, A., Fearns, A., Jamal-Hanjani, M., Sims, D., Fenwick, K., Mitsopoulos, C., Gao, Q., Orr, N., Zvelebil, M., Haslam, S. M., Dell, A., Yarwood, H., Lord, C. J., Ashworth, A., Isacke, C. M., An In Vivo Functional Screen Identifies ST6GalNAc2 Sialyltransferase as a Breast Cancer Metastasis Suppressor. *Cancer Discovery* 2014, *4*, 304-317.

[396] van der Sluis, M., Melis, M. H. M., Jonckheere, N., Ducourouble, M.-P., Büller, H. A., Renes, I., Einerhand, A. W. C., Van Seuningen, I., The murine Muc2 mucin gene is transcriptionally regulated by the zinc-finger GATA-4 transcription factor in intestinal cells. *Biochemical and Biophysical Research Communications* 2004, *325*, 952-960.

[397] van der Sluis, M., Vincent, A., Bouma, J., Male, A. K.-V., van Goudoever, J. B., Renes, I. B., Van Seuningen, I., Forkhead box transcription factors Foxa1 and Foxa2 are important regulators of Muc2

mucin expression in intestinal epithelial cells. *Biochemical and Biophysical Research Communications* 2008, *369*, 1108-1113.

MACQUARIE
UNIVERSITY

External Approval Noted- Molloy (Ref: 5201200399)

Ethics Secretariat <ethics.secretariat@mq.edu.au>   Wed, May 16, 2012 at 1:43 PM
To: A/Prof Mark Molloy <mark.molloy@mq.edu.au>
Cc: jenny.chik@students.mq.edu.au

Dear A/Prof Molloy

Re: "Determination of extensive immunophenotypes of colorectal cancers
using a CD antibody micorarray"

The above application was considered by the Executive of the Human Research
Ethics Committee. In accordance with section 5.5 of the National Statement
on Ethical Conduct in Human Research (2007) the Executive noted the final
approval from the Sydney South West Area Health Service and your right to
proceed under their authority.

Please do not hesitate to contact the Ethics Secretariat if you have any
questions or concerns.

Please retain a copy of this email as this is your official notification of
external approval being noted.

Yours sincerely

Dr Karolyn White
Director of Research Ethics
Chair, Human Research Ethics Committee

350

MACQUARIE UNIVERSITY

Biosafety approval

JENNY CHIK <jenny.chik@students.mq.edu.au>   Wed, Oct 1, 2014 at 1:55 PM
To: JENNY CHIK <jenny.chik@students.mq.edu.au>

---------- Forwarded message ----------
From: "Bio Safety" <biosafety@mq.edu.au>
To: "Mark Molloy" <MMolloy@proteome.org.au>
Cc:
Date: Fri, 27 Aug 2010 13:23:07 +1100
Subject: Biosafety Application REF 5201000885

Dear Associate Professor Molloy

RE: Sugars in the real world: Are cultured cancer cells a good model
system for studying protein glycosylation? (REF: 5201000885)

The above application was considered by the Biosafety Committee and
Final Approval of the above application is granted, effective 27
August 2010.

You must inform the Committee of your willingness to accept and comply
with any conditions by signing the Agreement Statement attached to
this email, and returning one copy of this Agreement to the Committee
Secretary, Ms Nicola Myton, in the Research Office at Macquarie
University.

The following personnel are authorised to conduct this research:

Associate Professor Mark Molloy – Chief Investigator/Supervisor
Ms Jenny Chik – Co-Investigator

Please note the following standard requirements of approval:

1. Approval will be for a period of twelve (12) months. At the end of
this period, if the project is continuing then a progress report must
be submitted. If, at the end of this period the project has been
completed, abandoned, discontinued or not commenced for any reason,
you are required to submit a Final Report. If you complete the work
earlier than you had planned you must submit a Final Report as soon as
the work is completed. These reports are located at the following
address:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/biosafety_
research_ethics/forms

A Progress/Final Report for this study will be due on: 1 September 2011.

2. Please remember the Committee must be notified of any alteration to
the project.

3. If you will be applying for or have applied for internal or
external funding for the above project it is your responsibility to
provide the Macquarie University's Research Grants Management
Assistant with a copy of this email as soon as possible. Internal and
External funding agencies will not be informed that you have final
approval for your project and funds will not be released until the
Research Grants Management Assistant has received a copy of this
email.

If you need to provide a hard copy letter of Final Approval to an
external organisation as evidence that you have Final Approval, please
do not hesitate to contact the Committee Secretary at the address
below.

Please retain a copy of this email as this is your formal notification
of final Biosafety approval.

Yours Sincerely

Dr Sinan Ali
Chair, Macquarie University Biosafety Committee
--
Office of the Deputy Vice Chancellor (Research)

Macquarie University Biosafety Secretariat

Research Office
Level 3, Research HUB, Building C5C
Macquarie University
NSW 2109

Ph: +61 2 9850 4194
Fax: +61 2 9850 4465

Email: biosafety@mq.edu.au

--
Jenny Chik

PhD student
CBMS Macquarie University
Room: E8A 303
Email: jenny.chik@students.mq.edu.au
Tel: +61 2 9850 8270
Fax: +61 2 9850 8313

--
Jenny Chik

PhD student
CBMS Macquarie University
Room: E8A 303
Email: jenny.chik@students.mq.edu.au
Tel: +61 2 9850 8270

AgreementStatement.doc
208K

352

MACQUARIE UNIVERSITY

Biosafety Workshop

This is to certify that

Jenny Chik

has successfully completed the above workshop which was conducted by the
Macquarie University Biosafety Committee on 31 July 2009

Dr Sinan Ali,
Chair, Biosafety Committee
July 2009

Expiry Date:
July 2012

MACQUARIE
UNIVERSITY

Biosafety and
Biohazards Workshop

This is to certify that

Jenny Chik

has successfully completed the above workshop which was conducted by the Macquarie University Institutional Biosafety Committee and Biohazards Safety Committee on 10 April 2012

Dr Subra Vemulpad
Chair, Institutional Biosafety Committee
April 2012

Expiry Date:
April 2015