

One sample, one pot and the whole glycome: towards a system to study disorders of protein glycosylation

Kirsty Leigh Skeene

PhD

University of York

Chemistry

September 2016

Abstract

Glycoprotein glycosylation is altered in many diseases, including congenital disorders of glycosylation, as well as in cell differentiation. Alterations to the presence and location of glycan biosynthetic enzymes can manifest in subtle changes to glycan profiles. To determine which of these changes are statistically significant, analysis of multiple biological samples in parallel is needed. Protein glycosylation is also a major consideration in the pharmaceutical industry, with an increase in engineering of glycoprotein-based therapeutics. In therapeutic protein development, glycan structures can influence efficacy and stability. It is therefore important to understand glycan biosynthesis and to have a convenient means to assess the structures of glycans in order to further understand disease and genetic disorders linked to errors in glycosylation, as well as to potentially contribute to disease diagnostics and to inform development of biotherapeutics.

A filter-aid *N*-glycan separation (FANGS) method was recently developed, to release and isolate *N*-glycans, from a small number of culture whole cell lysates, for mass spectrometric analysis. This thesis presents a FANGS-based approach, to carry out O-glycan release, using β -elimination. Through method optimisation, O-glycan release can be achieved with 20 minutes of sonication, without loss of labile sialic acid groups, and gives comparable results to those from well-accepted overnight incubation methods. In addition, the thesis presents a streamlined protocol for a one-sample-one-pot approach to release *N*- and O-glycans from the same sample, in the same pot, with the potential to subsequently analyse protein remaining in the filter. Analysing both *N*- and O-glycans from one sample is important since in disease it is not always only one type of glycosylation that is altered. The developed method has been applied to mesenchymal stromal cells (MSCs), before and after differentiation into adipocytes, as well as genetically manipulated MSCs, in addition to undifferentiated and differentiated normal human urothelial cells. In both cell systems, changes could be observed in both the *N*- and O-glycan profiles. The approach offers insight into potential functional role of glycans in cellular processes such as differentiation and disease.

List of Contents

Abstract.....	3
List of Contents	5
List of Tables	9
List of Figures	13
Acknowledgments	25
Author's declaration	27
Chapter 1: Introduction	29
1. Introduction:	31
1.1. Carbohydrates: monosaccharides, oligosaccharides, and glycans	31
1.2. Glycoprotein glycan biosynthesis	35
1.2.1. Biosynthesis of <i>N</i> -glycans	35
1.2.2. Biosynthesis of <i>O</i> -glycans	37
1.2.3. Glycan biosynthesis: the role of the Golgi	39
1.3. Glycans in disease and biological therapeutics	45
1.3.1. Congenital disorders of glycosylation.....	45
1.3.2. Biologics: glycoengineering, vaccines and drug targets	46
1.4. Methods for glycosylation analysis.....	50
1.4.1. Release of <i>O</i> -glycans from glycoproteins	53
1.5. Aims of the thesis	59
Chapter 2: Mass Spectrometry	61
2. Mass spectrometry	63
2.1. Introduction	63
2.2. Matrix-assisted laser desorption/ionisation (MALDI).....	63
2.5.1. Sample preparation	64
2.5.2. Ion formation in MALDI	64
2.6. Electrospray ionisation	68
2.7. Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer.....	70
2.7.1. Ion motion	71

2.7.2.	Excitation and detection	73
2.8.	Time of flight mass spectrometry	75
2.9.	Orbitrap	79
2.10.	Tandem mass spectrometry	82
2.10.1.	Fragmentation of carbohydrates	83
2.10.2.	Fragmentation of peptides	85
2.11.	Coupling high performance liquid chromatography (HPLC) to mass spectrometry (LC-MS)	86
2.12.	Mass spectrometric characterisation of glycans	88
Chapter 3: Experimental Procedures		91
3.	Experimental procedures	93
3.1.	Cell culture: Mesenchymal stromal cells	93
3.1.1.	Basal media	93
3.1.2.	Primary MSC culture	93
3.1.3.	Adipogenesis	93
3.2.	Histochemistry: Oil red O staining	93
3.3.	Extraction and collection of glycoproteins from CHO and MSC cells	94
3.4.	Extraction, collection of glycoproteins from porcine bladder urothelial tissue	94
3.5.	Solubilisation of standard glycoproteins in SDS (using FASP).	94
3.6.	Solubilisation of standard glycoproteins and urothelial cell samples, and preparation for O-glycan release, using an adapted enhanced filter aided sample preparation (eFASP) ¹⁷⁷	95
3.7.	Normal human urothelial cells	95
3.8.	Non-reductive β -elimination – using overnight incubation	96
3.8.1.	Non-reductive β -elimination – using sonication	96
3.8.2.	PNGase F treatment – N-glycan release	97
3.9.	Per-O-methylation	97
3.10.	MALDI-FT-ICR MS	98
3.11.	Glycan quantitation and statistical analysis	98
3.12.	Tandem mass spectrometry (MS/MS)	98
3.13.	Tryptic digestion	99
3.14.	Peptide analysis	99
3.15.	Protein identification	100
3.16.	Preparation of protein, glycopeptides and glycans for NMR analysis	100

3.16.1. “intact” urea-denatured glycoprotein (monoclonal antibody), IgG	100
3.16.2. Preparation of “large” <i>N</i> -glycan sample from IgG	100
3.16.3. Preparation of a “large” (glyco)peptide sample from GITR.....	101
3.17. Nuclear magnetic resonance spectroscopy	101
Chapter 4: Method Development.....	103
4. Method development	105
4.1. Method development for O-glycan release.....	105
4.2. Reducing incubation time for O-glycan release.....	110
4.3. One-sample, one-pot; the whole glycome	113
4.4. Optimum time for O-glycan release?	120
4.5. O-glycan release from Chinese hamster ovary cells (CHO)	126
4.6. Conclusions and future work	128
Chapter 5: Characterisation of antibody <i>N</i> -glycans, released by FANGS, by a combination of two-dimensional nuclear magnetic resonance spectroscopy experiments.	131
5. Characterisation of antibody <i>N</i> -glycans, released by FANGS, by a combination of two-dimensional nuclear magnetic resonance spectroscopy experiments.	133
5.1. Antibody research and glycome analysis.....	133
5.2. Nuclear magnetic resonance spectroscopy (NMR)	137
5.2.1. Instrumentation and experimental set-up	140
5.2.2. Chemical shift and spin-spin coupling	141
5.2.3. Two-dimensional NMR spectroscopy	142
5.2.3.1. Total correlation spectroscopy	142
5.2.3.2. Heteronuclear single quantum correlation	143
5.2.3.3. Heteronuclear multiple bond correlation	144
5.2.3.4. Heteronuclear multiple quantum correlation – correlation spectroscopy	144
5.2.3.5. Nuclear Overhauser effect spectroscopy	145
5.3. Results and discussion	145
5.4. Conclusions and future work	165
Chapter 6: Demonstration of the potential of the one-pot method to study glycosylation, and to monitor changes to glycosylation, in experimental cellular systems	167
6. Demonstration of the potential of the one-pot method to study glycosylation, and to monitor changes to glycosylation, in experimental cellular systems.	169
6.1. Mesenchymal stromal cells (MSCs)	169

6.3. O-glycans of Y201 and Y101 MSCs	171
6.3.1. Chemical inhibition of O-glycan biosynthesis	175
6.3.2. Genetic modification of MSCs: effect on glycan profile.....	180
6.3.3. Demonstration of the one-pot <i>N</i> - and O-glycan release method to study the <i>N</i> - and O-glycome of MSCs and to monitor the change to glycosylation between the wild- type and mutant cell lines.	183
6.3.4. Adipogenesis	188
6.4. Conclusions and future work	207
Chapter 7: Demonstration of the one-pot glycomics method applied to normal human urothelial cells.	209
7. Introduction.....	211
7.3. Data analysis.....	215
7.4. Application to normal human urothelial (NHU) cells.....	215
7.5. Conclusions and future work	235
Chapter 8: Summary of results, overall conclusions and future work.....	239
Abbreviations	245
References.....	247

List of Tables

Table 1: O-glycan types, classified in terms of the first monosaccharide attached to the protein chain, where R can be H, a monosaccharide or an oligosaccharide chain	37
Table 2: Glycan nomenclature for characterisation by mass spectrometry, including the symbols and masses for native and permethylated structures.	88
Table 3: Number of different glycan compositions detected at range of times treated with NH ₄ OH.....	123
Table 4: Number of different glycan compositions detected at range of times treated with NH ₄ OH.....	124
Table 5: number of glycans detected when treated with NH ₄ OH and a combination of sonication and rest periods at the incubation temperature	124
Table 6: Assignments of fucose ¹ H and ¹³ C resonances confirmed by matching cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	161
Table 7: Assignments of GlcNAc-1 ¹ H and ¹³ C resonances confirmed by matching cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	161
Table 8: Assignments of GlcNAc-2 ¹ H and ¹³ C resonances confirmed by matching cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	162
Table 9: Assignments of galactose ¹ H and ¹³ C resonances confirmed by matching cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	162
Table 10: Assignments of Man-z ¹ H and ¹³ C resonances confirmed by cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	163
Table 11: Assignments of Man-y ¹ H and ¹³ C resonances confirmed by cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	163
Table 12: Assignments of Man-x ¹ H and ¹³ C resonances confirmed by cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	164

Table 13: Assignments of GlcNAc-x/y (terminal? on Gal?) ¹ H and ¹³ C resonances confirmed by cross peak patterns in [¹ H ¹³ C]-HSQC, [¹ H ¹³ C]-HMBC, [¹ H ¹ H]-TOCSY, ¹ H ¹³ C-dept-HSQC, ¹ H ¹ H-NOESY, ¹ H ¹³ C-HMQC-COSY, ¹ H ¹ H-dqf-COSY.....	164
Table 14: List of O-glycans identified from Y201 and Y101 MSC cell lines, showing the mass accuracy	172
Table 15: P values for significant differences between Y201/Y101 WT O-glycan levels. Statistical analysis carried out by T test with Holm Sidak post-hoc test. ns = not significant	172
Table 16: Lectins used for lectin staining, by K. Wilson, for analysis of N- and O glycans.....	176
Table 17: List of permethylated O-glycans identified, with their mass accuracies, from Y201 and Y101 MSCs after 48 h pre-treatment with O-glycan biosynthesis inhibitor, BadGal.....	179
Table 18: List of P values for the significant differences between Y201 and Y101 WT O-glycans. Statistical analysis carried out using one-way ANOVA with Holm Sidak post-hoc test. ns = not significant	179
Table 19: List of permethylated O-glycans identified, along with their mass accuracies, in Cog4KD MSCs	181
Table 20: P values for significant differences between Y101 wild-type and Cog4KD O-glycans. Statistical analysis used a one-way ANOVA with Holm Sidak post-hoc test.....	182
Table 21: List of permethylated N-glycans identified, with their mass accuracies, in Cog4KD MSCs.....	185
Table 22: List of identified MSC permethylated O-glycans (Y201 cell line) and mass accuracies	193
Table 23: List of identified MSC permethylated O-glycans (Y101 cell line) and mass accuracies	193
Table 24: List of P values for the significant differences in O-glycan abundance between Y201 and Y101 MSCs cultured for 21 days in either basal medium or adipogenic medium. Statistical analysis carried out using one-way ANOVA with Holm Sidak post-hoc test.	194
Table 25: List of permethylated N-glycans identified in MSCs, along with their mass accuracies, after being cultured in either basal medium or adipogenic medium for 21 days.	198
Table 26: A list of permethylated O-glycans identified from Y201 MSCs after a 48 h pre-treatment with O-glycan biosynthesis inhibitor, BadGal, followed by 21 days in either basal or adipogenic medium	205

Table 27: A list of permethylated O-glycans identified in Y101 MSCs after a 48 h pre-treatment with O-glycan biosynthesis inhibitor, BadGal, followed by 21 days in either basal or adipogenic medium.....	206
Table 28: List of permethylated O-glycans identified from Y1419, and Y1860 NHU cell lines, showing mass accuracies.....	217
Table 29: Protein concentration extracted from NHU cell line, measured by Bradford assay.	218
Table 30: List of permethylated N-glycans identified from Y1419, and Y1860 NHU cell cultures, showing mass accuracies.....	221
Table 31: List of permethylated O-glycans identified from Y1419, and Y1860 NHU cell lines, showing mass accuracies.....	226
Table 32: concentration of protein extracted from cultures from each of the differentiated NHU cell lines, measured by Bradford assay.	227
Table 33: List of permethylated N-glycans identified from differentiated Y1419 and Y1860 NHU cell lines, showing mass accuracies. All glycan compositions listed were identified in the proliferating cell lines.	232

List of Figures

Figure 1: Ring closure of a hexose, galactose, to give α or β configuration at the anomeric carbon, in this case C1. Blue numbers designate the carbon atom number.	32
Figure 2: structures of common monosaccharide units that make up glycoprotein glycan structures.....	33
Figure 3: Structure of uridine diphosphate (UDP)-galactose, a nucleotide monosaccharide donor used in glycosyltransferase reactions	34
Figure 4: Formation of a β -1-3 glycosidic bond between galactose and <i>N</i> -acetylgalactosamine, catalysed by β -1-3 galactosyltransferase.....	34
Figure 5: Types of glycosidic linkage in a glycoprotein. a) <i>N</i> -glycosidic linkage via the amido nitrogen of Asn, b) O-glycosidic linkage via the oxygen in the hydroxyl group of Ser.	35
Figure 6: <i>N</i> -glycosylation. A lipid-linked precursor is synthesised in the endoplasmic reticulum (ER), followed by transfer of the resulting oligosaccharide to all asparagine side chains in the protein. Processing of the glycan core is initiated in the ER and further processing occurs as the glycoprotein migrates through the Golgi complex where monosaccharides are removed, and then others may be added by glycosyltransferase enzymes to produce that glycoprotein's <i>N</i> - glycan structures.....	36
Figure 7: O-glycosylation: mucin-type. Biosynthesis is initiated via addition of GalNAc to the hydroxyl group of serine or threonine side chains in the protein, forming the substrate for formation of either core 1 or core 3 O-glycans. Enzymes involved in synthesising each of the core structures are displayed. (C1GalT1, core 1 β 1,3-galactosyltransferase; β 3Gn-T6, core 3 β 1,3- <i>N</i> -acetylglucosaminyltransferase 6; core 2 or 4 β 6GlcNAc-T, β 1,6- <i>N</i> -acetylglucosaminyltransferases). The number of isoforms present for each enzyme is shown as either (1) or (3). Adapted from Tran et al ⁴³	39
Figure 8: retrograde trafficking in the Golgi, showing transport vesicles and how they are involved in the retrograde transport of enzymes in the Golgi. Adapted from Cottam et al ⁵²	41
Figure 9: Vesicle transport at the Golgi. 1) Vesicle formation: vesicles buds from the Golgi cisternae, initiated by coat proteins by interacting with membrane proteins to bend, shape and deform the membrane to form the vesicle, trapping inside the contents to be transported. 2) Rab proteins, e.g. Rab GTPase, bind to the transport vesicle. 3) Rab proteins bind to Rab effector proteins such as tethering factors (known as golgins in the Golgi) to aid tethering of the vesicle to the target membrane. 4) Once vesicle-bound Rab proteins are	

bound to a tethering factor, the protein coat around the vesicle is shed to prepare for fusion with the target membrane. Adapted from Ungar et al⁵⁹42

Figure 10: COG complex structure schematic, showing connections of each of the subunits. Subunits Cog1-Cog4 make up lobe A of the protein complex and Cog5-Cog8 make up lobe B. Cog2,3,4 and Cog5,6,7 are trimers, and these are connected via the dimer of Cog1,8. Adapted from Miller et al⁴⁶43

Figure 11: Two proposed mechanisms for COG–golgin interactions involved in vesicle tethering. Adapted from Ramirez et al⁶⁸44

Figure 12: Vesicle fusion with the target membrane catalysed by SNAREs. SNAREs on the vesicles (v-SNAREs) wrap around the target membrane SNAREs (t-SNAREs) securing the membranes in place and pulling them close enough to fuse and unload the vesicle contents⁴⁷44

Figure 13: Pentasaccharide of heparin, the anticoagulant used for the treatment of thrombosis; it contains a sulfated iduronic acid and sulfated glucosamine. Neomycin antibiotic and Relenza, an antiviral drug.....47

Figure 14: Filter-aided *N*-glycan separation (FANGS): 1) cell culture, 2) cell lysis, solubilisation in SDS, 3) transfer of solubilised cell lysate to filter, exchange of SDS for ammonium bicarbonate buffer, filter centrifuged to remove SDS, 4) PNGase F added to sample in the filter to release *N*-glycans from the glycoproteins, filter centrifuged to collect the released glycans in the filtrate. Modified from Rahman et al¹⁴¹, and from Millipore Amicon filter user guide.....52

Figure 15: Peeling reaction¹⁵²: monosaccharides are eliminated from the reducing residue of the released glycan.54

Figure 16: General mechanism for β -elimination and inclusion of NaBH₄ to prevent peeling by converting the reducing terminus of O-glycans to alditols.55

Figure 17: Non-reductive β -elimination, adapted from Rademaker et al¹⁵³56

Figure 18: Permethylation of carbohydrates.58

Figure 19: commonly used MALDI matrices for analysing carbohydrates, peptides and proteins.64

Figure 20 Matrix-assisted laser desorption/ionisation. Laser ablation of the matrix/analyte spot, causing excitation of the matrix and vaporisation into an expanding plume.....65

Figure 21: The lucky survivor model¹⁹³ a) ablated matrix/analyte clusters contain precharged analytes with counterions and so carry no net charge. Dissociation of those clusters leads to charge neutralisation, and so the analyte is not detected. b) charge separation upon

disruption of the matrix-analyte clusters gives a net charge and so the analyte ions are detected as “lucky survivors”. c) Matrix excess charges included in the clusters lead to generation of protonated analytes by counterion neutralisation and so the analyte ions are detected as “lucky survivors”.	66
Figure 22: Formation of a Taylor cone ¹⁹⁸	69
Figure 23: Charge residue model ¹⁹⁵ for gas-phase ion production in ESI.....	69
Figure 24: Ion evaporation model ²⁰⁰ for gas-phase ion formation in ESI.....	69
Figure 25: Schematic of ESI source.....	70
Figure 26: A cubic FT-ICR analyser cell. The cell consists of six plates made up of two trapping plates, which sit perpendicular to the magnetic field, two excitation plates and two detection plates, which sit parallel to the magnetic field.	71
Figure 27: Ion cyclotron motion. The direction of travel is the opposite for positively or negatively charged ions ²⁰¹	72
Figure 28: Ion motion as a result of trapping potential application in ICR cell. Ions oscillate back and forth, parallel to the magnetic field (B), between the trapping plates.....	72
Figure 29: The motion of an ion in the analyser cell has three oscillation components: cyclotron motion, trapping (axial) motion and magnetron motion, which results from a combination of the magnetic and electric fields.....	73
Figure 30: Excitation and detection of an ion in the cell of an ICR-MS: a) ions before excitation have a small cyclotron radius and are therefore not detected. b) During excitation, ions whose cyclotron frequency is in resonance with the applied electric field absorb energy and their cyclotron radius increases. c) When the applied electric field is turned off, the ions continue to orbit in the larger radius until a different electric field is applied, when they relax back.....	74
Figure 31: Fourier transform-ion cyclotron resonance (FT-ICR).....	74
Figure 32: Time of flight mass spectrometer. a) Linear. Poor mass resolution as ions the with the same m/z ratio but different kinetic energies reach the detector at different times. b) Reflectron. Improved mass resolution as ions with the same m/z ratio but different kinetic energies reach the detector simultaneously.....	77
Figure 33: a) delayed pulse extraction, using dual stage extraction. Ions are formed and allowed time to separate according to their initial velocities. For ions with the same m/z ratio but different velocities, ones with a higher velocity travel further towards the detector. On application of the delayed pulse more energy is supplied to those ions which had less energy meaning they join the ions that had more energy and reach the detector	

simultaneously with appropriate choice of voltage and delay time. b) In continuous extraction, ions with the same m/z ratio but different velocities reach the detector at different times as ions, upon their formation, are immediately extracted.79

Figure 34: Orbitrap Fusion hybrid Orbitrap instrument schematic. Reproduced from²¹³80

Figure 35: Orbitrap, to show how the outer electrodes are split in half along the z axis ($z=0$). Taken from Hu et al²¹¹81

Figure 36: Movement of analyte ions in an Orbitrap mass analyser. a) and b) as ions approach the opposite electrode, the electric field contracts the radius of the ion packet and decreases the trajectory close to the z-axis of the trap, giving time for further ion packets to arrive. c) Once all ions have entered the trap, the voltage of the central electrode stabilises, stabilising the ion trajectories. Frequencies of radial or rotational motion vary, as the frequency is dependent on ion velocity and radius of rotation when the ions enter the trap and so the ions dephase much more quickly than the axial motion. The dephasing causes the ion packet to form a ring made up of uniformly distributed ions of each m/z82

Figure 37: Fragmentation of glycans to produce B and Y ions. B ions are produced from cleavage of the protonated glycosidic bond, and Y ions are produced from cleavage of the protonated glycosidic bond followed by H-transfer²¹⁵.84

Figure 38: Fragmentation nomenclature of carbohydrates, adapted from Domon and Costello²¹⁵84

Figure 39: fragmentation and nomenclature of branched glycans by Domon and Costello²¹⁵. The branches that lead off the core structure are denoted as α or β and are then labelled in the same way as the main chain (e.g. B or Y ion and the number of monosaccharides in the fragment). If there is any further branching then labels such as α' and α'' are added.85

Figure 40: Peptide fragmentation and nomenclature86

Figure 41: Calculation of the mass of a glycan in both native and permethylated forms.89

Figure 42: MALDI-FT-ICR spectrum of permethylated O-glycans, ionised as $[M + Na]^+$, released from 20 pmol of porcine stomach mucin a) following FASP for SDS removal/exchange, followed by O-glycan release, in the filter, with NH_4OH . b) following eFASP for SDS removal/exchange, followed by O-glycan release, in the filter, with NH_4OH 106

Figure 43: MALDI-FT-ICR spectrum (m/z 1200-2220) of permethylated O-glycans, ionised as $[M + Na]^+$, released from 2 pmol mucin. a) alkylation using 4-vinylpyridine for 1 h, b) alkylation using iodoacetamide for 1 h, c) alkylation using 4-vinylpyridine for 15 min, d) alkylation using iodoacetamide for 15 min.....107

Figure 44 MALDI-FT-ICR mass spectrum of permethylated O-glycans released from porcine stomach mucin using overnight incubation, prepared using SDS solution following 1 in 10 dilution with exchange buffer a) containing DCA b) without DCA and 1 in 30 dilution with exchange buffer, c) containing DCA d) without DCA. The red box highlights that increasing the dilution and removing DCA gives cleaner glycan mass spectra. All glycans ionised as $[M + Na]^+$ 109

Figure 45 MALDI-FT-ICR mass spectrum of permethylated O-glycans, ionised as $[M + Na]^+$, released from porcine stomach mucin using 16 h incubation (top panel), 5 min sonication/10 min rest (x4) (middle panel) and 10 min microwave irradiation (bottom panel). 111

Figure 46: Collection of porcine bladder urothelium. a) Porcine bladder b) bladder cut open through the urethra c) inside the porcine bladder. d) surface of the bladder scraped with a scalpel e) scrapings collected for glycomic analyses. This method of harvesting porcine bladder urothelium has been developed and validated by Chung-Yao Wang in the group, in collaboration with Prof Jenny Southgate. 112

Figure 47 MALDI-FT-ICR mass spectrum of permethylated glycans released from porcine bladder urothelium, ionised as $[M + Na]^+$. a) O-glycans released using 16 h incubation b) O-glycans released using 20 min sonication at 45 °C 113

Figure 48 MALDI-FT-ICR mass spectrum of permethylated glycans released from porcine bladder urothelium, ionised as $[M + Na]^+$. a) N-glycans released before O-glycans b) O-glycans released after N-glycans in the same filter unit from the same sample 116

Figure 49: HCD spectra of precursor m/z 967, $[M + Na]^+$ ($Hex_2HexNAC_2$) from porcine urothelial N-glycans (top) and O-glycans (bottom). Glycans are permethylated..... 117

Figure 50: HCD spectra of precursor m/z 1141, $[M + Na]^+$ ($Fuc_1Hex_2HexNAC_2$) from porcine urothelial N-glycans (top) and O-glycans (bottom). Glycans are permethylated..... 118

Figure 51: HCD spectra of precursor m/z 1171, $[M + Na]^+$ ($Hex_3HexNAC_2$) from porcine urothelial N-glycans (top) and O-glycans (bottom). Glycans are permethylated..... 119

Figure 52: Relative abundance of permethylated O-glycans released from porcine stomach mucin over 2 - 24 h incubation with NH_4OH . Glycan relative abundance calculated as a percentage of the signal from a spiked internal standard. 122

Figure 53: Relative abundance of permethylated O-glycans released from porcine stomach mucin from 5 min sonication – 60 min (alternating 5 min sonication followed by 10 min rest period) with NH_4OH . Glycan relative abundance calculated as a percentage of the signal from a spiked internal standard. 125

Figure 54 MALDI-FT-ICR mass spectrum of permethylated O-glycans released and isolated from CHO cells. Glycans ionised as $[M + Na]^+$. A) CHO IdIC b) WT CHO.....	127
Figure 55: Comparison of the relative abundance (%) of O-glycans in WT CHO cells and IdIC CHO cells.	128
Figure 56: Generic structure of an antibody. An antibody consists of four polypeptide chains (two light chains and two heavy chains) each held together by disulfide bonds. The Fab region contains the antigen binding site. The Fc region contains glycans and can bind to cell surface receptors and activate the immune response.	134
Figure 57: Energy level diagram.....	138
Figure 58: NMR pulse to signal. a) RF field is applied, inducing B_1 magnetic field, perpendicular to B_0 . b) the induced magnetic field from the RF field creates sine waves at the detector as B_1 oscillates about the x,y axis. When the pulse ends the signal decays, known as free induction decay (FID) and the FID is converted to a frequency domain by Fourier transformation ²⁶¹	140
Figure 59: $[^1H^1H]$ -TOCSY. Example of the type of $^1H^1H$ correlations that could be observed in a $[^1H^1H]$ -TOCSY experiment for galactose, e.g. a) H1 with H3 and b) H1 with H4.....	143
Figure 60: $^1H^{13}C$ -HSQC. Example of a single bond correlation observed in a $[^1H^{13}C]$ -HSQC experiment of galactose, for the anomeric carbon with the anomeric proton.....	143
Figure 61: $^1H^{13}C$ -HMBC. Example of multiple bond correlation in a $^1H^{13}C$ -HMBC experiment of glycans. HMBC can provide linkage information from the anomeric proton to the carbon at the other side of the glycoside bond on the neighbouring monosaccharide unit. In addition, HMBC can provide multiple bond correlations of 1H and ^{13}C around a single ring system.	144
Figure 62: $^1H^{13}C$ -HMQC-COSY. Example of the type of correlation that could be observed in galactose, showing the correlation of the anomeric proton with C2 in the ring.....	144
Figure 63: $^1H^1H$ -NOESY. Example of the type of nOes that could be observed on a glycan. a) the nOe between the CH_3 in the amide acetyl group of a HexNAc, with the anomeric proton. b) through space interaction of the methyl group on a fucose with a 1H on an adjacent monosaccharide.	145
Figure 64: Primary amino acid sequence of the Fc region of the IgG around the glycosylation site. Red lines depict tryptic cleavage sites. The N, in orange, is the glycosylation site ²³⁴	146
Figure 65: N-Glycans found in the IgG ^{264,265,266}	146

Figure 66: $^1\text{H}^{13}\text{C}$ -HSQC NMR spectra of the denatured intact protein, tryptic digest, and released <i>N</i> -glycans. The anomeric region, depicting the $^1\text{H}^{13}\text{C}$ correlations between the anomeric carbon and the anomeric proton, is circled.	148
Figure 67: $^1\text{H}^{13}\text{C}$ -HSQC NMR spectra of the denatured intact protein, tryptic digest, and released <i>N</i> -glycans, expanded in the anomeric region.	149
Figure 68: $^1\text{H}^{13}\text{C}$ -HSQC NMR spectra, showing the methyl/acetyl group region, of the denatured intact protein, tryptic digest, and released <i>N</i> -glycans.	150
Figure 69: TOCSY short mixing time (13 ms) showing assignments of the anomeric protons of the released <i>N</i> -glycans with other protons around each monosaccharide ring system.	151
Figure 70: $^1\text{H}^1\text{H}$ -TOCSY long mixing time (120 ms) showing $^1\text{H}^1\text{H}$ correlations of anomeric protons, of released <i>N</i> -glycans, with protons around each monosaccharide ring system. ...	152
Figure 71: Evidence of connectivity between fucose H1C1 with fucose H2C2, using a combination of $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY. Cross peaks are indicated by the intersecting lines.	153
Figure 72: Evidence of connectivity of H1C1 of GlcNAc-1 to H2C2 of GlcNAc-1, using a combination of $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY.	153
Figure 73: Evidence of connectivity of H1C1 of GlcNAc-1 to H2C2 of GlcNAc-1, using a combination of $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY. Cross peaks are indicated by the intersecting lines.	154
Figure 74: Evidence of connectivity of H1C1 of GlcNAc-1 to H2C2 of GlcNAc-1, by $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY.	155
Figure 75: Comparison of $^1\text{H}^{13}\text{C}$ -HMBC, $^1\text{H}^{13}\text{C}$ -dept-HSQC, $^1\text{H}^1\text{H}$ -NOESY and $^1\text{H}^1\text{H}$ -TOCSY to show the Fuc(1-6)GlcNAc-1 linkage.	156
Figure 76: Comparison of $^1\text{H}^{13}\text{C}$ -HMBC, $^1\text{H}^{13}\text{C}$ -dept-HSQC, $^1\text{H}^1\text{H}$ -NOESY and $^1\text{H}^1\text{H}$ -TOCSY to show the Fuc(1-6)GlcNAc-1 linkage.	157
Figure 77: $^1\text{H}^1\text{H}$ -NOESY spectrum showing nOe cross peaks connecting GlcNAc-methyl resonances with anomeric proton resonances, from Fuc, Man and GlcNAc.	158
Figure 78: $^1\text{H}^1\text{H}$ -NOESY interactions connecting the GlcNAc-methyl with anomeric proton from a) Man, b) Fuc and c) GlcNAc.	158
Figure 79: $^1\text{H}^1\text{H}$ -TOCSY and $^1\text{H}^{13}\text{C}$ -dept-hsqc to show H6C6 connectivities with the anomeric proton.	160
Figure 80: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from a) Y101 and b) Y201 MSC cell lines. All glycans ionised as $[\text{M} + \text{Na}]^+$	171

Figure 81: Comparison of the O-glycan profiles of the Y201 and Y101 wild-type (WT) MSCs. Relative abundance calculated by calculating each glycan signal as a percentage of the spiked internal standard signal. P = * < 0.05, ** < 0.01, *** < 0.001.173

Figure 82: HCD product ion mass spectrum of permethylated O-glycan NeuAc₁Hex₁HexNAc₁ (precursor *m/z* 879).174

Figure 83: HCD product ion mass spectrum of permethylated O-glycan NeuAc₂Hex₁HexNAc₁ (precursor *m/z* 1240)174

Figure 84: CID product ion mass spectrum of permethylated O-glycan, NeuAc₁Hex₂HexNAc₂ (precursor *m/z* 1328)175

Figure 85 MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y201 and Y101 MSCs after 48 h pre-treatment with the O-glycan biosynthesis inhibitor, BadGal. All glycans ionised as [M + Na]⁺178

Figure 86: Comparison of relative O-glycan abundance between wild-type and BadGal treated MSCs from two cell lines (Y201 and Y101). Relative abundance calculated by calculating each glycan signal as a percentage of the spike internal standard. Blue asterisks are for statistically significant changes to the Y201 cell line and red asterisks are for the Y101 cell line, where P < 0.05 = *, P < 0.01 = ** and P < 0.001 = ***.180

Figure 87: MALDI-FT-ICR mass spectrum of permethylated O-glycans, released from Y101 Cog4KD MSCs. All glycans ionised as [M + Na]⁺181

Figure 88 Comparison of relative abundance of O-glycans between Y101 WT and Cog4KD MSCs. Relative abundance of glycans were quantified by comparing peak intensities with a spiked internal standard. Blue asterisks indicate statistically different changes to the relative glycans. P < 0.05 = *, < 0.01 = **182

Figure 89: MALDI-FT-ICR mass spectrum of permethylated *N*-glycans released from Y201 Cog4KD MSCs (top panel) and WT MSCs (bottom panel). All glycans ionised as [M + Na]⁺ ..184

Figure 90: Comparison of the relative abundance of permethylated *N*-glycans in Y101 WT and Cog4KD MSCs.187

Figure 91: Oilred O staining of Y201 and Y101 cell lines after being cultured in either basal or adipogenic medium for 21 days. Oilred O did not stain the cells in the basal medium. As cells differentiate into adipocytes lipid droplets form and the Oilred O stains the lipid droplets red. Y201 differentiate well and form many lipid droplets, Y101 although some staining is seen differentiation is limited.188

Figure 92: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y201 MSCs after 21 days in either basal medium or adipogenic medium. All glycans ionised as $[M + Na]^+$	191
Figure 93: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y101 MSCs after 21 days cultured in either basal or adipogenic medium. All glycans ionised as $[M + Na]^+$	192
Figure 94: Comparison of the relative abundance of O-glycans in Y201 and Y101 MSC cell lines after being cultured in either basal or adipogenic medium for 21 days, alongside values for cells in basal medium for only approx 7 days (80% rather than complete confluence). Quantitation carried out by comparing the glycan peak intensities to the peak intensity of a spiked internal standard. Asterisks indicate a statistically significant change in the relative glycan abundance between the basal samples (21 days) and the adipogenic samples (green for Y201 and purple for Y101. P = * < 0.05, ** < 0.01, *** < 0.001.....	195
Figure 95: MALDI-FT-ICR mass spectrum of permethylated N-glycans released, after O-glycans, from Y201 MSCs cultured for 21 days in basal medium. All N-glycans ionised as $[M + Na]^+$	196
Figure 96: MALDI-FT-ICR mass spectrum of permethylated N-glycans (released after O-glycan release) from Y201 MSCs, following adipogenesis. All N-glycans ionised as $[M + Na]^+$	197
Figure 97: Comparison of the relative abundance of permethylated N-glycans identified in MSCs when cultured in either basal medium or adipogenic medium for 21 days.....	200
Figure 98: Oilred O staining of Y201 and Y101 MSC cell lines after pre-treatment with BadGal for 48 hours, followed by culture in basal or adipogenic medium for 21 days. Oilred O did not stain the cells in the basal medium. As cells differentiate into adipocytes lipid droplets form and the Oilred O stains the lipid droplets red. Y201 differentiate well and form many lipid droplets, Y101 although some staining is seen differentiation is limited.	201
Figure 99: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y201 MSCs after 48 h pre-treatment with BadGal, followed by 21 days in basal or adipogenic medium. All glycans ionised as $[M + Na]^+$	203
Figure 100: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y101 MSCs after 48 h pre-treatment with BadGal, followed by 21 days in basal or adipogenic medium. All glycans ionised as $[M + Na]^+$	204
Figure 101: Comparison of the O-glycan abundance in Y201 and Y101 MSC cell lines pre-treated with BadGal for 48 h, followed by culturing in basal or adipogenic medium for 21	

days. Relative abundance calculated by comparing peak intensity of spiked internal standard to each of the glycan peaks. P = * < 0.05, ** < 0.01, *** < 0.001.....207

Figure 102: A schematic of urinary bladder urothelium. Cells are arranged in distinct layers; basal cells, intermediate cells and the umbrella/superficial cells which are fully differentiated to perform their function as a protective barrier; the umbrella cells are connected via intercellular tight junctions. The apical surfaces of the superficial cells are covered in urothelial plaques. Adapted from Lewis²⁹⁴211

Figure 103: TER set-up. adapted from³³². To carry out the electrical measurements, one electrode is placed in the upper (apical) chamber of the membrane well, and the other in the lower basolateral chamber of the membrane well.214

Figure 104: NHU cultures of the three cell lines Y1419, Y1860 and Y1862.216

Figure 105: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from NHU cells (from three cell lines, Y1419, Y1860 and Y1862). All glycans ionised as $[M + Na]^+$ 216

Figure 106: Comparison of the O-glycan profiles of three NHU cell lines. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. The glycan profiles have been normalised to protein concentration.....218

Figure 107: MALDI-FT-ICR mass spectra of permethylated N-glycans released after O-glycan release from NHU cells (from three cell lines, Y1419, Y1860 and Y1862). All glycans ionised as $[M + Na]^+$ 220

Figure 108: Comparison of the N-glycan profiles of cultures from Y1419, Y1860 and Y1862 NHU cell lines. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. The glycan profiles have been normalised to protein concentration.....223

Figure 109: Schematic of the semi-permeable membrane well to which NHU cells were transferred to form a tight barrier, mimicking the barrier formed in the bladder.....224

Figure 110: TER readings over 0 to 8 days for each of the NHU cell lines (Y1419, Y1880, Y1862) to show formation of a tight barrier. (data provided by Ros Duke)225

Figure 111: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from NHU cells (from three cell lines, Y1419, Y1860 and Y1862). All glycans ionised as $[M + Na]^+$ 225

Figure 112: Comparison of the O-glycan profiles of the Y1419, Y1860 and Y1862 NHU cell cultures, differentiated on SNAP well membranes. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. The glycan profiles have been normalised to protein concentration.....227

Figure 113: Comparison of the O-glycan profiles of the, proliferating cells on 6 cm dishes and differentiated cells on semi-permeable membrane wells for Y1419 NHU cell line (top graph) and Y1860 (bottom graph). Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. Glycan profiles are normalised to protein concentration..... 229

Figure 114: MALDI-FT-ICR mass spectrum of permethylated N-glycans released from NHU cells (from three cell lines, Y1419, Y1860 and Y1862) after O-glycans. All glycans ionised as $[M + Na]^+$ 231

Figure 115: Comparison of the *N*-glycan profiles of the differentiated Y1419, Y1860 and Y1862 NHU cell lines. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. Glycan profiles are normalised to protein concentration. 234

Acknowledgments

Firstly, I would like to say thank you to my supervisors Jane Thomas-Oates, Ed Bergstrom and Graham Clarke (Bristol-Myers Squibb), for giving me the opportunity to carry out this PhD, and for supporting me throughout it. Jane offered continuous encouragement and positivity, which I definitely needed when I was convinced about 12 months into the PhD that I had no data and that nothing was working! Thanks to Ed for all his technical and practical advice. Thanks to Graham from Bristol-Myers Squibb for giving me the amazing opportunity to give a project update to BMS colleagues in the states – without that I would not have met Mike Reily, who then invited me to work in his lab for two months. The placement was an exciting opportunity to learn more about NMR, which I learned a great deal from Luciano Mueller and Nelly Aranibar (although I am sorry to say but I am still a mass spectrometrists at heart!) I made some great new friends there, including Bev Heravi who took me under wing and let me stay with her, in her home, for those two months.

Thanks to everyone I worked with In York, including Andy, Carwyn, Elena, Emily, Emma, Helen, Jo, Karl, Kirsty, Lance, Peter, Rachel, and Sam (a fantastic Mchem student) from the JTO group. In addition, special thanks goes to Daniel Ungar in Biology, along with PhD students Katie and Peter, for allowing me into their lab and teaching me the ropes for cell culture, and for providing help and advice throughout my PhD. Also thanks to Jenny Southgate and Ros Duke for their help with the urothelial cell study.

Finally, love and thanks goes to all my family and friends for their continued support and encouragement, and especially to Will who has had to put up with my moaning, tears and grumpiness on many occasions when I have come home having had a bad day in the lab or a stressful day of thesis writing. Without the support I may not have finished this PhD!

Oh! How could I forget Jeff?! The best fluffy house bunny in the world who has been with me throughout my thesis writing, and providing her approval by chasing off with chapter drafts or munching on the corners when I asked for her thoughts.

Author's declaration

I declare that the work presented in this thesis is my own, except where otherwise acknowledged and that it has not been submitted previously for a degree at this or any other university. All sources are acknowledged as references.

Chapter 1: Introduction

1. Introduction:

Glycans, by IUPAC definition, are “compounds consisting of a large number of monosaccharides linked glycosidically”. Perhaps more specifically, glycans are often glycoconjugates and are the carbohydrate chains that are glycosidically linked to proteins or lipids, biosynthesised through glycosylation of proteins and lipids. Glycans coat all cell surfaces and play a central role in cellular processes, disease and in the immune response. It has been observed that glycoprotein and lipid glycosylation is altered in cancer^{1,2,3,4}, Alzheimer’s disease^{5,6}, inflammation⁷, diabetes⁸ and in congenital disorders of glycosylation^{9,10,11,12}. Protein glycosylation is now a major consideration in the pharmaceutical industry with an increase in engineering of glycan/glycoprotein-based therapeutics (especially monoclonal antibodies)^{13,14}, and carbohydrate-based vaccines^{15,16,17,18} (against bacterial pathogens, HIV and cancer). In therapeutic protein development, glycan structures present can influence efficacy, stability^{19,20} and pharmacokinetics^{21,22}. Glycans are also important as drug targets²³. It is therefore important to understand glycan biosynthesis and to have a convenient means to assess the structures of glycans in order to further understand disease and genetic disorders linked to errors in glycosylation, as well as both to potentially contribute to disease diagnostics and to inform the development of biotherapeutics.

1.1. Carbohydrates: monosaccharides, oligosaccharides, and glycans

Monosaccharides are the basic carbohydrate monomer units. Hexoses, the most common building block of glycoprotein glycans, have chiral centres on four of their six carbon atoms. The substituents around each of those chiral carbons can be arranged in two stereochemically different ways, giving 16 possible different isomeric structures. An equilibrium exists between the open and closed ring forms of monosaccharides. The ring is formed by an intramolecular reaction of a hydroxyl group with a carbonyl resulting in a hemiacetal or hemiketal. For example, the intramolecular reaction of the hydroxyl group on carbon 5 of galactose with the C1 aldehyde yields the ring form in which carbon 1 becomes a chiral centre and can therefore exist in two stereochemical configurations. Carbon 1 in this structure is referred to as the anomeric carbon and the position of the OH group attached to

the anomeric carbon is designated α or β , Figure 1. The anomeric carbon is formed on the carbonyl carbon atom. For aldoses, this is carbon 1, while for ketoses such as *N*-acetylneuraminic acid, the anomeric carbon is carbon 2. Common monosaccharides found to make up glycoprotein glycan structures are: hexoses (e.g. mannose and galactose), deoxyhexoses (e.g. fucose), hexosamines (e.g. *N*-acetylglucosamine, *N*-acetylgalactosamine) and sialic acids (e.g. *N*-acetylneuraminic acid), Figure 2. Monosaccharides are also referred to as having D or L configuration, assignment of which is determined by the asymmetric carbon furthest away from the carbonyl.

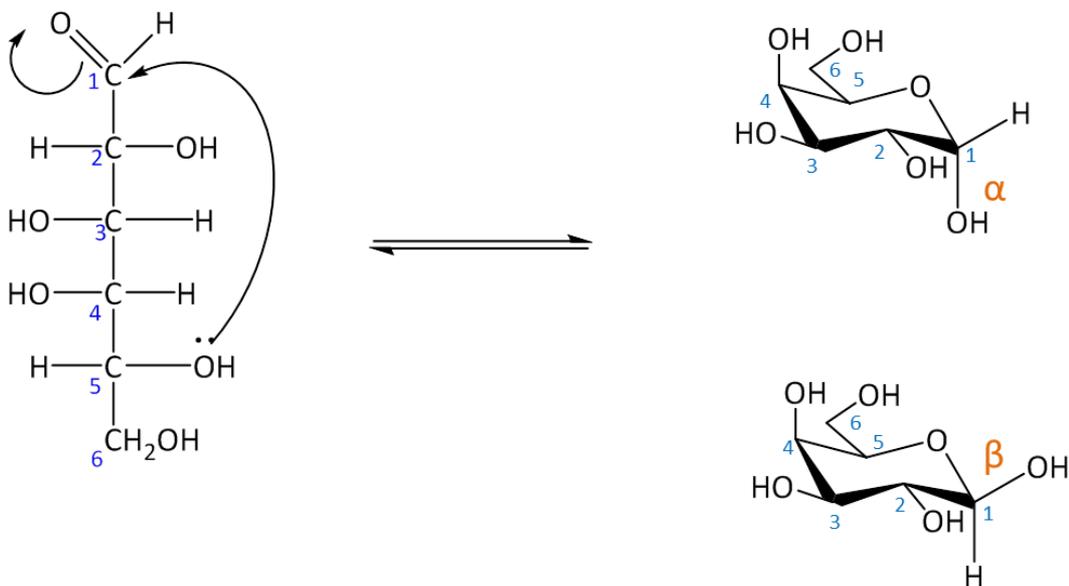


Figure 1: Ring closure of a hexose, galactose, to give α or β configuration at the anomeric carbon, in this case C1. Blue numbers designate the carbon atom number.

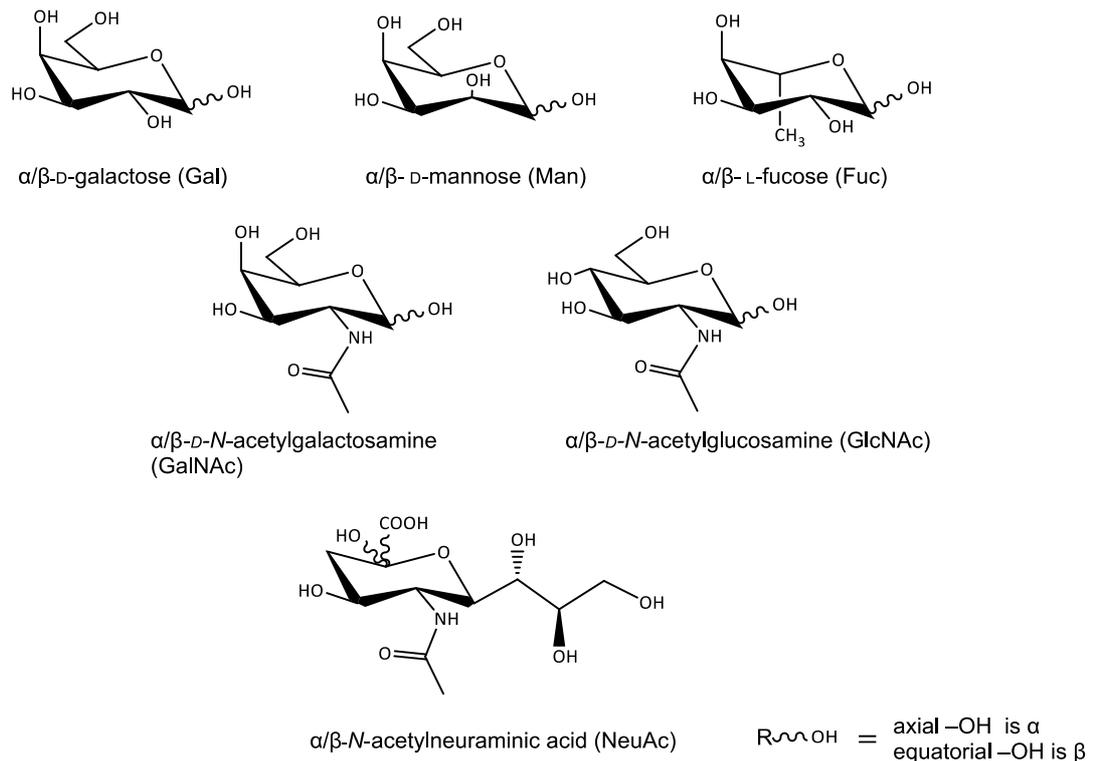
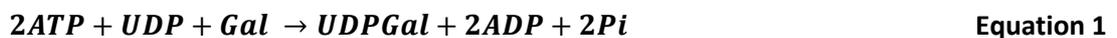


Figure 2: structures of common monosaccharide units that make up glycoprotein glycan structures.

Disaccharides are the simplest form of oligosaccharide, formed from two monosaccharides joined through a glycosidic bond. Bond formation involves a condensation reaction between the hemiacetal of a monosaccharide, and a hydroxyl group of another, eliminating a molecule of water. The formation of the glycosidic bond in cells is driven by the energy released from the hydrolysis of two phosphate bonds in a molecule of adenosine triphosphate (ATP). This takes place in two stages. Firstly, a nucleotide sugar donor such as uridine diphosphate (UDP)-galactose (Figure 3) is formed in the cytoplasm, as in equation 1:



UDP-Gal is then used to form a glycosidic linkage with another monosaccharide (the acceptor). Glycosyl transferase enzymes catalyse the transfer of a monosaccharide from the donor to an acceptor; the enzymes have specificity for donors and acceptors. For example, in the condensation reaction between galactose and N-acetylgalactosamine-Ser (GalNAc-Ser) in the growing O-glycan chain (Figure 4), a glycosidic linkage is formed between the C1

of galactose and the 3-hydroxyl group of GalNAc with UDP-galactose the donor and β -1-3-galactosyltransferase the enzyme that catalyses the reaction.

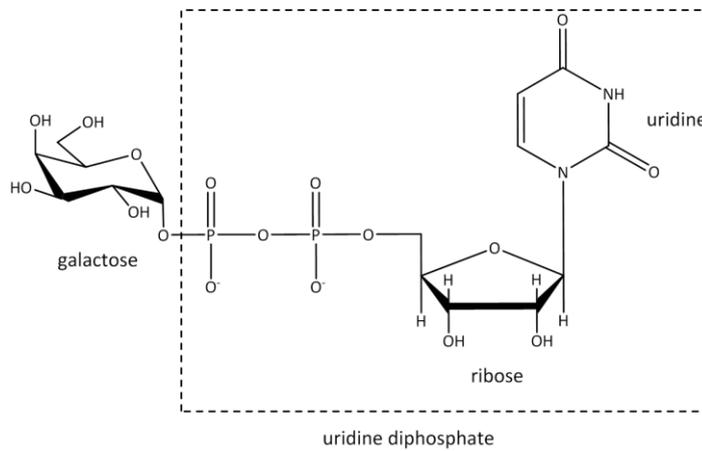


Figure 3: Structure of uridine diphosphate (UDP)-galactose, a nucleotide monosaccharide donor used in glycosyltransferase reactions

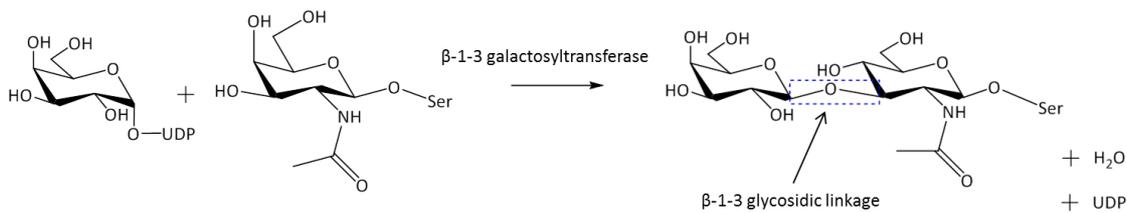


Figure 4: Formation of a β -1-3 glycosidic bond between galactose and *N*-acetylgalactosamine, catalysed by β -1-3 galactosyltransferase.

There are two major glycan types in a glycoprotein; *N*-glycans and *O*-glycans, named according to their linkage to the protein via an amino acid side chain. *N*-glycans form a glycosidic linkage with the amido nitrogen of an asparagine side chain in the specific consensus sequence -Asn-X-Ser- or -Asn-X-Thr- (where X is any amino acid except proline), where the term *consensus sequence* refers to commonly occurring sequences found amongst related protein sequences which are recognised by an enzyme. *O*-glycans form a glycosidic linkage via oxygen in the hydroxyl group of, most commonly, serine or threonine side chains, Figure 5.

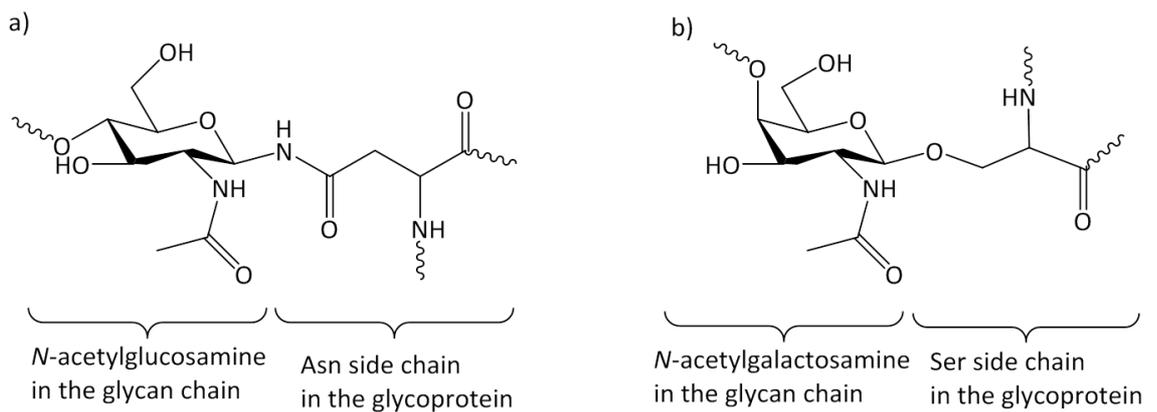


Figure 5: Types of glycosidic linkage in a glycoprotein. a) N-glycosidic linkage via the amido nitrogen of Asn, b) O-glycosidic linkage via the oxygen in the hydroxyl group of Ser.

1.2. Glycoprotein glycan biosynthesis

The biosynthetic code for glycans remains enigmatic because, unlike protein and nucleic acid biosynthesis, there is no template for glycan biosynthesis. Eukaryotic glycan biosynthesis is managed by a large number of glycosyltransferase enzymes (located in the endoplasmic reticulum (ER) and the Golgi apparatus) and is dependent on their expression levels, location and activity, that, because of the lack of a template, can change under the influence of local cellular environmental conditions^{24,25} such as availability of nutrients^{26,27}, changes in temperature²⁸ or pH²⁹. Consequently, the overall structure of a glycoprotein can vary quite significantly depending on the tissue in which it is expressed and local conditions, making analysis of glycoprotein structures both challenging and important; since structures cannot be predicted from a template, they have to be determined in each distinct case.

The enzymes and biosynthetic pathways leading to *N*- and *O*- protein glycosylation have been defined in a range of experimental systems, which enables an understanding of the range of possible glycan structures for both glycosylation types.

1.2.1. Biosynthesis of *N*-glycans

In *N*-glycosylation³⁰, a chain of monosaccharides ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) is biosynthesised on a lipid carrier (dolichol) in the ER. This lipid-linked precursor oligosaccharide is transferred to the amide nitrogen in asparagine's side chain. This transfer is catalysed by an oligosaccharyltransferase which has specificity for the *N*-glycosylation consensus sequence. Following transfer in the ER, the *N*-glycan undergoes trimming,

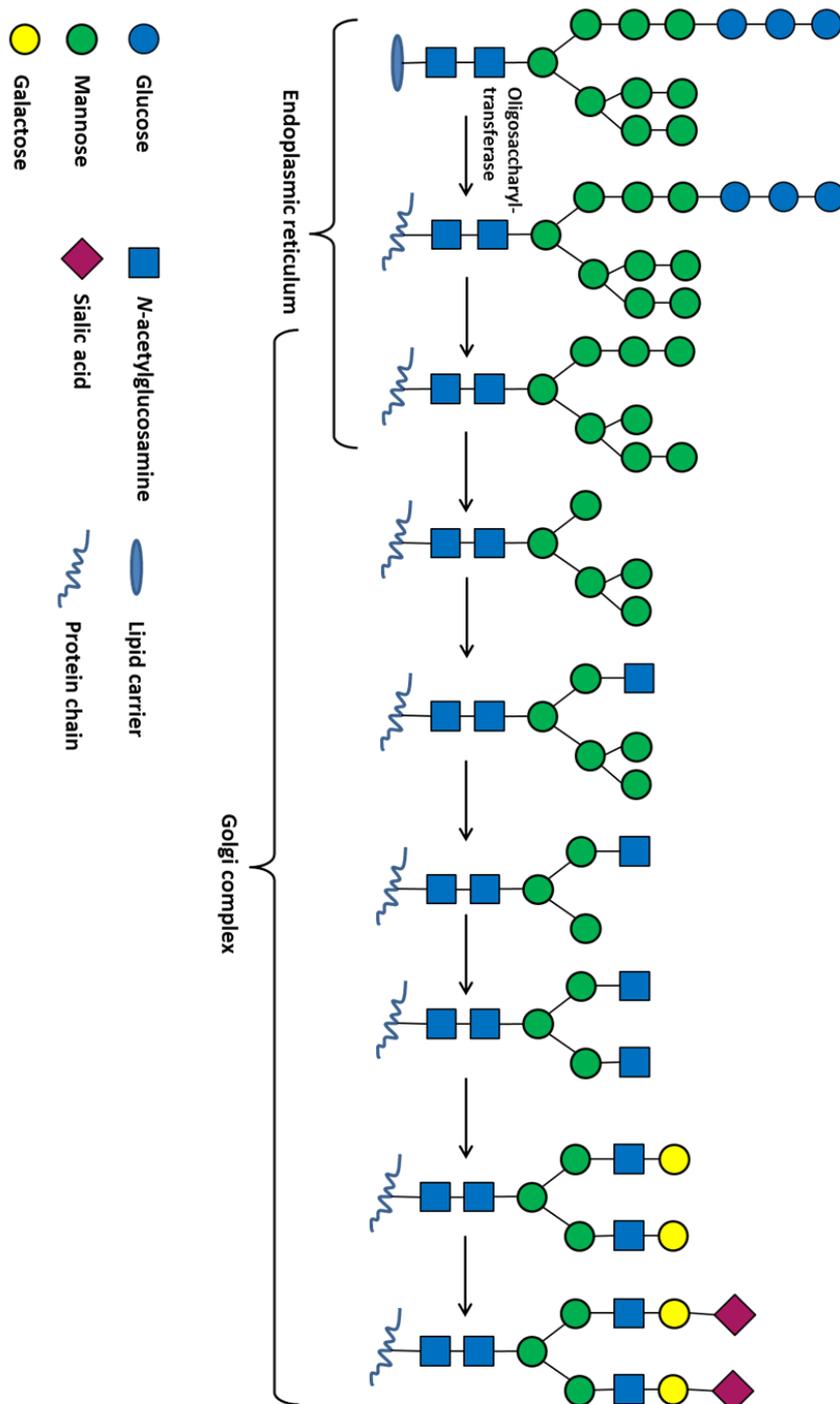


Figure 6: *N*-glycosylation. A lipid-linked precursor is synthesised in the endoplasmic reticulum (ER), followed by transfer of the resulting oligosaccharide to all asparagine side chains in the protein. Processing of the glycan core is initiated in the ER and further processing occurs as the glycoprotein migrates through the Golgi complex where monosaccharides are removed, and then others may be added by glycosyltransferase enzymes to produce that glycoprotein's *N*-glycan structures.

1.2.2. Biosynthesis of O-glycans

Biosynthesis of O-glycans is somewhat different from that of *N*-glycans. O-glycosylation occurs in the Golgi complex (while *N*-glycosylation is initiated in the endoplasmic reticulum). O-glycosylation is not in general directed to a consensus sequence; instead, O-glycans are built on a range of protein-glycan linkages, classified in terms of the first monosaccharide attached to the protein, Table 1.

Table 1: O-glycan types, classified in terms of the first monosaccharide attached to the protein chain, where R can be H, a monosaccharide or an oligosaccharide chain

Type of O-glycan	Structure	Location
Mucin-type ^{31,32,33}	R-GalNAc-Ser/Thr	Gastrointestinal, respiratory and genital tracts
Glycosaminoglycans (GAGs) ³⁴	GlcA β 1-3Gal β 1-3Gal β 1-4Xyl-Ser	Cartilage, skin, arteries
O-linked mannose ^{35,36}	R-Man-Ser/Thr	Nervous system, yeast, bacteria
O-linked fucose ³⁷	R-Fuc-Ser/Thr	Found in some secreted or cell surface proteins
O-linked <i>N</i> -acetylglucosamine ³⁸	R-GlcNAc-Ser/Thr	Nuclear and cytoplasmic proteins
O-linked galactose ^{39,40}	R-Gal-Ser/Thr	Plants, collagens

Mucins are large molecular weight glycoproteins (> 200 kDa) with 20 known human mucin genes expressed in mucosal sites such as the airways and gastrointestinal tract. Mucins can be found membrane bound or they are secreted. Mucins contain repeated amino acid sequences⁴¹, referred to as tandem repeats, which are rich in serine and threonine residues, the primary sites for O-glycosylation; mucins are heavily O-glycosylated. The main functions of mucins are to provide lubrication⁴¹ and to protect from micro-organisms. Mucins possess these functions because glycans are hydrophilic, with some sites on the proteins bearing

clusters of sialylated glycans giving regions with negative charge and they can therefore sequester large quantities of water to form mucus/gels which provide a physical barrier against bacterial pathogens and proteases⁴².

In O-glycosylation, monosaccharides are added to the protein-linked, growing glycan chain in stepwise reactions, where each transfer of a monosaccharide is catalysed by a different enzyme of the glycosyltransferase family⁴³. Similarly to N-glycosylation, as the glycoprotein migrates through the Golgi, monosaccharide moieties are added to the growing glycan chain by specific glycosyltransferases to give the mature glycan structure.

Biosynthesis of mucin type O-glycans involves addition of GalNAc from uridine diphospho-GalNAc to Ser/Thr residues, catalysed by a family of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases (pp-GalNAc-Ts)⁴⁴, each with its own specificity for amino acid sequences around the glycosylation target site⁴³. There are around 20 ppGalNAc-Ts in humans; some are expressed in a range of tissues, whilst others are restricted to a specific location. GalNAc-Ser/Thr is then the substrate for synthesis of core 1 (bearing a β 1-3 Gal linkage), catalysed by β 1,3-galactosyltransferase (C1GalT-1) and core 3 glycans (bearing a β 1-3 GlcNAc linkage), catalysed by β 1,3-*N*-acetylglucosaminyltransferase (β 3Gn-T6). Core 1 and core 3 glycans can then be modified to core 2 and core 4 glycans respectively, catalysed by β 1,6-*N*-acetylglucosaminyltransferases (β 6GlcNAc-T). There are three isoforms of β 6GlcNAc-T in mammals, two of which can catalyse core 2 formation and the third can either catalyse core 2 or core 4 production. Once the core structures are formed (Figure 7), they are further modified by numerous glycosyltransferases, such as fucosyltransferase and sialotransferase⁴³.

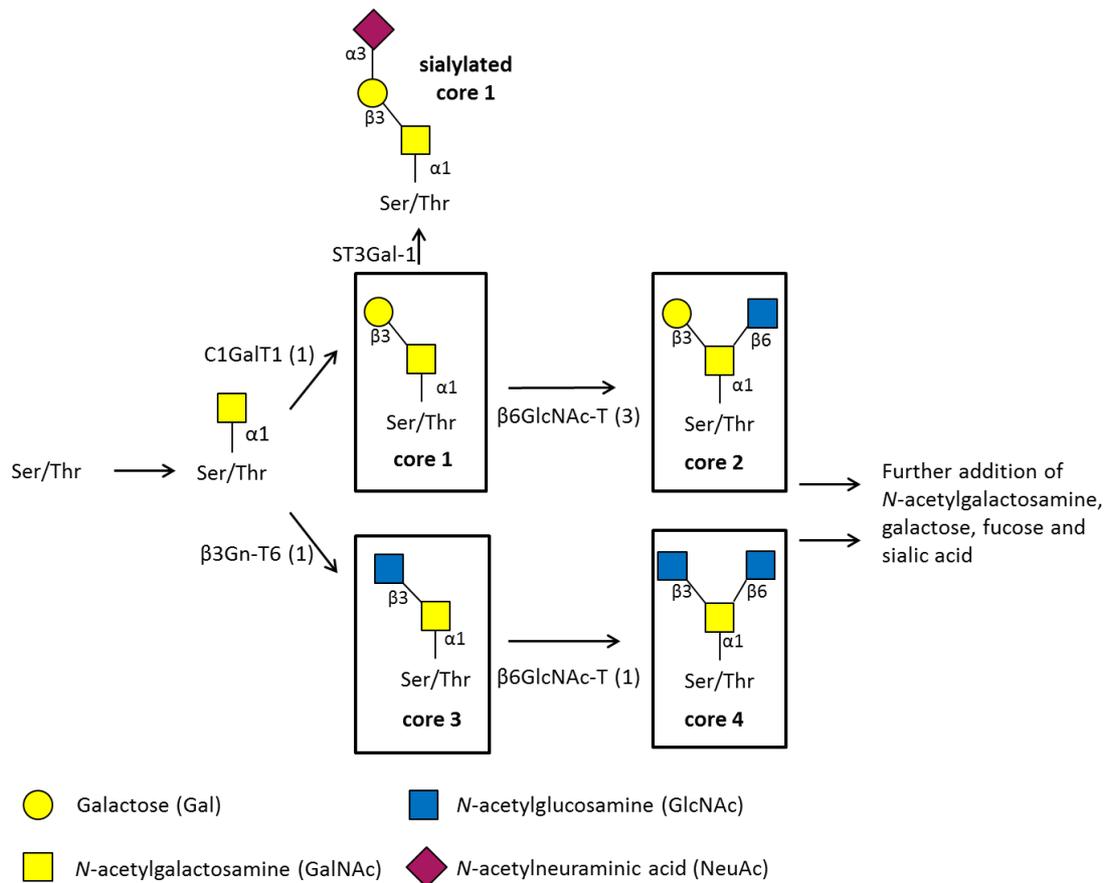


Figure 7: O-glycosylation: mucin-type. Biosynthesis is initiated via addition of GalNAc to the hydroxyl group of serine or threonine side chains in the protein, forming the substrate for formation of either core 1 or core 3 O-glycans. Enzymes involved in synthesising each of the core structures are displayed. (C1GalT1, core 1 β 1,3-galactosyltransferase; β 3Gn-T6, core 3 β 1,3-N-acetylglucosaminyltransferase 6; core 2 or 4 β 6GlcNAc-T, β 1,6-N-acetylglucosaminyltransferases). The number of isoforms present for each enzyme is shown as either (1) or (3). Adapted from Tran et al⁴³.

1.2.3. Glycan biosynthesis: the role of the Golgi

Whilst we know the repertoire of glycan structures of both protein glycosylation types, and the consequences for glycosylation of deleting the enzymes that directly synthesise the glycans are evident, it is very difficult to predict the consequences of enzymes for glycan biosynthesis not localizing correctly in the cell. There has been a range of research into how glycan processing in the Golgi occurs, in particular trying to understand the vesicle tethering process^{45,46}.

Most glycan processing occurs in the Golgi complex, a secretory organelle in the mammalian cell, made up of flattened stacks of cisternae. The cisternae are divided into three compartments; cis, medial and trans, containing many different enzymes all competing for the growing glycan chain. Enzyme distribution between cisternal sections is controlled by vesicle transport, a complex process involving multiple signals, and protein interactions between the vesicle transporting the growing glycan chain and enzymes, with the cisterna that the vesicle will fuse to. Vesicle budding (the initial stages of vesicle transport), how the vesicles migrate through the Golgi complex, and vesicle fusion with the target membrane have all been studied, but the exact mechanism underlying the process and how the Golgi manages to maintain the correct distribution of enzymes for correct glycosylation is not fully understood.

There are a number of models that suggest how the transport vesicles migrate through the Golgi complex. One model is cisternal maturation^{47,48,49}. This model states that cis cisternae continually form from clusters of vesicles that bud off and arrive at the Golgi, from the ER. Each cisterna matures through cis to medial to trans Golgi cisternae. Golgi enzymes are recycled through retrograde transport as vesicles bud off from Golgi cisternae and travel back down in the direction of trans cisternae through to cis, returning the enzymes to their correct Golgi compartments (Figure 8). A second model suggests that the Golgi cisternae do not move and instead vesicles travel from cis to trans Golgi cisternae⁵⁰. As in the cisternal maturation model, there are also vesicles travelling backwards (trans through to cis), taking the enzymes that were transported with the forward moving vesicles (cis through to trans) back to the correct Golgi compartments. A third model suggests that once the contents of the transport vesicle are delivered to the Golgi they are quickly distributed around the Golgi and sorted into those for processing and those to be removed⁵¹. Of the three models, cisternal maturation appears to be the most accepted.

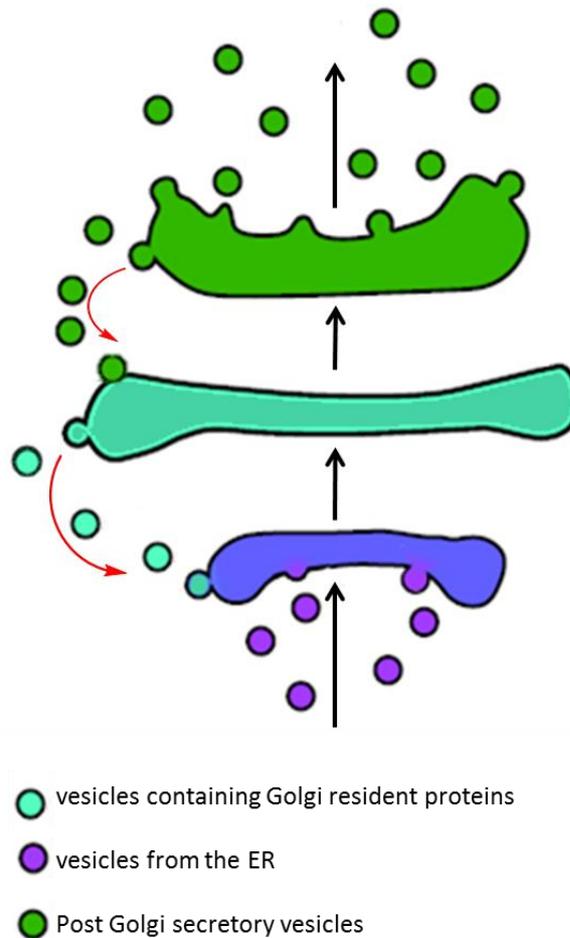


Figure 8: retrograde trafficking in the Golgi, showing transport vesicles and how they are involved in the retrograde transport of enzymes in the Golgi. Adapted from Cottam et al⁵².

In all of the described models, transport vesicles continuously bud from one membrane and fuse with another, carrying the maturing glycoprotein chain(s). Vesicle formation is initiated by coat proteins which select what is to go into the vesicle (e.g. glycoproteins or enzymes), by interaction with lipid and membrane proteins, and they also bend, shape and deform the membranes of the ER or Golgi to form a vesicle, trapping the contents inside^{53,54,55}, Figure 9 (1). Two types of protein-coated vesicles involved in transport of glycoproteins through the Golgi are coatomer protein complex-1, COPI, and coatomer protein complex-2, COPII, vesicles. COPII vesicles bud from the ER and COPI vesicles bud from the Golgi compartments.

During or soon after vesicle formation, Rab proteins, a subfamily of monomeric GTPases, are recruited and bind to the transport vesicle, Figure 9 (2). Rabs act as molecular switches and regulate the reversible assembly of protein complexes on the membrane. Once bound to a vesicle, Rab proteins can bind to other proteins, Rab effectors⁵⁶, such as motor or tethering

factors, Figure 9 (3), to help either move the vesicle along or to aid tethering of the vesicle with the target membrane^{57,58}. The vesicle moves by diffusion towards the target membrane. When the vesicle-bound Rab proteins bind to a tethering factor, the protein coat around the vesicle is shed in preparation for fusion with the target membrane, Figure 9 (4).

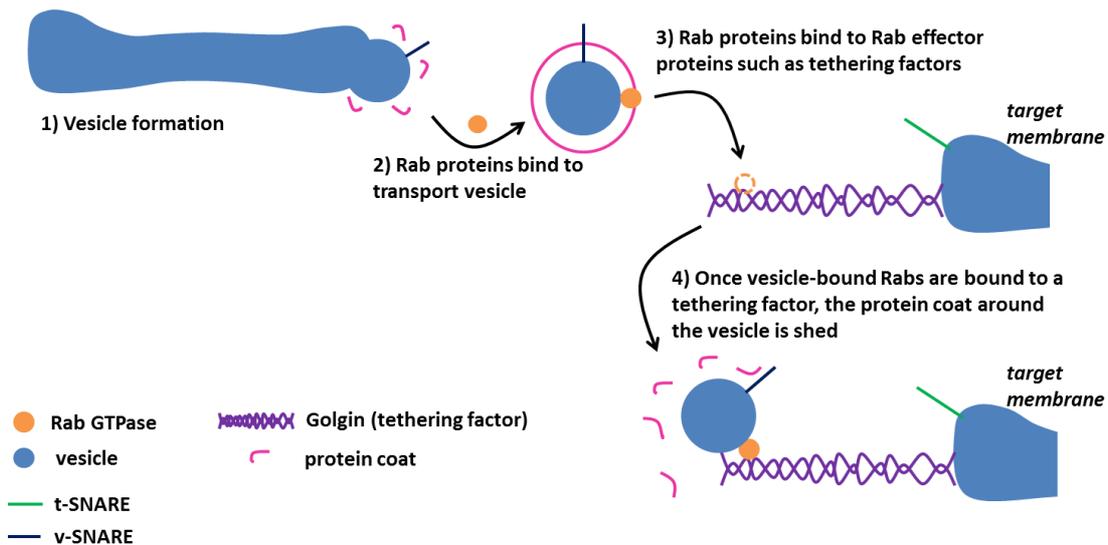


Figure 9: Vesicle transport at the Golgi. 1) Vesicle formation: vesicles buds from the Golgi cisternae, initiated by coat proteins by interacting with membrane proteins to bend, shape and deform the membrane to form the vesicle, trapping inside the contents to be transported. 2) Rab proteins, e.g. Rab GTPase, bind to the transport vesicle. 3) Rab proteins bind to Rab effector proteins such as tethering factors (known as golgins in the Golgi) to aid tethering of the vesicle to the target membrane. 4) Once vesicle-bound Rab proteins are bound to a tethering factor, the protein coat around the vesicle is shed to prepare for fusion with the target membrane. Adapted from Ungar et al⁵⁹.

Tethering factors such as coiled-coil proteins, (known as golgins at the Golgi) are long rod shaped molecules that project from membrane surfaces, opening up interaction sites on the Golgi cisternae⁶⁰. Golgins help to hold vesicles in position for fusion, a process interactions of tethering complexes, such as the conserved oligomeric Golgi (COG) complex, RabGTPases, and SNAREs⁵² between the transport vesicles and target membrane .

A protein complex, the conserved oligomeric Golgi (COG) complex, has been shown to play a central role in maintaining glycosylation stability in all eukaryotic cells^{45,46}. COG is thought to coordinate the retrograde transport of transport vesicles⁴⁶. COG has eight subunits, named Cog1-Cog8, arranged into two distinct lobes. Subunits Cog1-Cog4 make up lobe A whilst subunits Cog5-Cog8 make up lobe B. Subunits Cog2,3,4 and Cog5,6,7 form trimers connected by the dimer formed from Cog1 and Cog8, Figure 10.

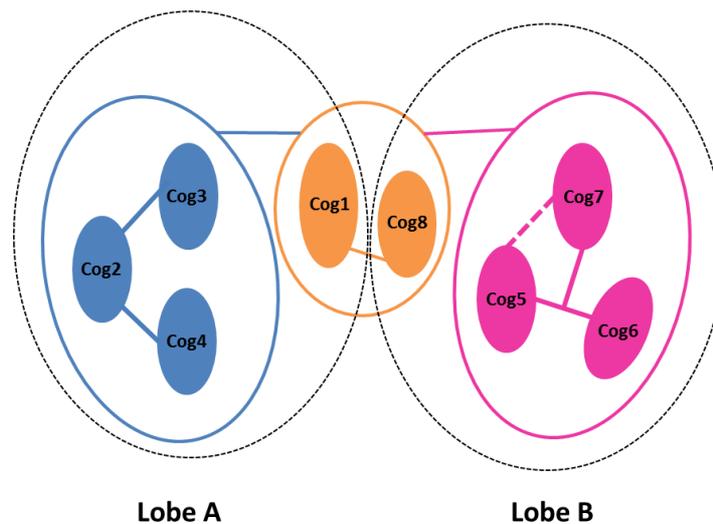


Figure 10: COG complex structure schematic, showing connections of each of the subunits. Subunits Cog1-Cog4 make up lobe A of the protein complex and Cog5-Cog8 make up lobe B. Cog2,3,4 and Cog5,6,7 are trimers, and these are connected via the dimer of Cog1,8. Adapted from Miller et al⁴⁶.

COG mutant cell lines have shown changes in glycosylation, which can be attributed to mislocalization of the Golgi-resident enzymes. What this does not suggest is exactly where and how COG functions. Protein interactions have been observed between COG⁴⁵ and coatamer protein complex-1 (COPI), the Rab GTPases^{61,62} and SNAREs^{63,64}, suggesting an involvement of COG in the final stages of tethering just before fusion of the vesicle and target membranes⁶⁵. However there is also evidence to suggest that COG is involved in the earlier stages, via interaction with golgins^{66,67,62,68}.

Although not fully understood, there are two mechanisms for how the COG-golgin interactions aid the vesicle tethering process. The first is that COG could be anchored to

both ends of the golgin⁶⁸, bending the golgin and pulling the vesicle and target membranes together. The second is that COG promotes movement of the vesicle along the length of the golgin chain using interactions with Rabs and SNAREs to help pull the vesicle towards the target membrane⁶⁸, Figure 11.

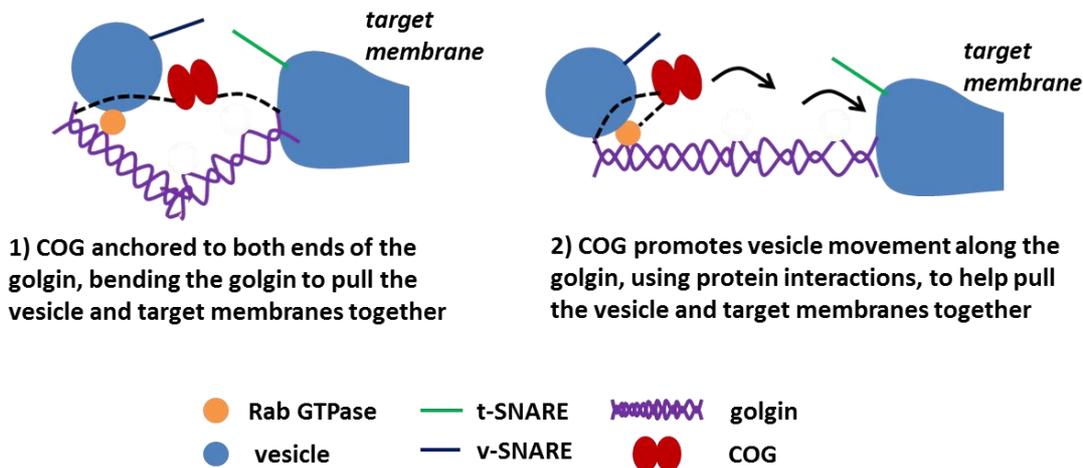


Figure 11: Two proposed mechanisms for COG–golgin interactions involved in vesicle tethering. Adapted from Ramirez et al⁶⁸.

Once a transport vesicle is tethered to the target membrane, SNARE proteins catalyse membrane fusion. SNAREs are present on the transport vesicle (v-SNARE) and the target membrane (t-SNARE). The two SNAREs wrap around each other⁶⁹, securing the membranes in place⁴⁷ as well as pulling the membranes close enough to fuse and unload the vesicle contents, Figure 12.

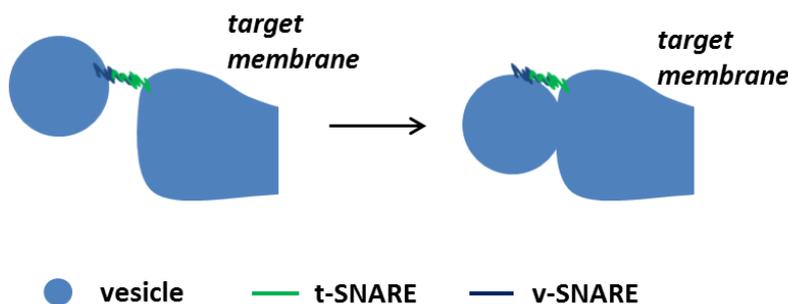


Figure 12: Vesicle fusion with the target membrane catalysed by SNAREs. SNAREs on the vesicles (v-SNAREs) wrap around the target membrane SNAREs (t-SNAREs) securing the membranes in place and pulling them close enough to fuse and unload the vesicle contents⁴⁷

In glycosylation, transport vesicles need to be accurate in targeting and fusing with the correct membrane to ensure the correct glycan is biosynthesised by unloading both the Golgi-resident enzymes and the glycoproteins in the correct place.

1.3. Glycans in disease and biological therapeutics

1.3.1. Congenital disorders of glycosylation

Congenital disorders of glycosylation, CDGs, first described by Jaak Jaeken in 1984⁷⁰, are a family of rare genetic disorders that present abnormalities in glycosylation⁷¹, and are not well diagnosed. Glycosylation defects may be due to mutations in the enzymes building the glycan, or those enzymes not localising correctly in the ER or the Golgi complex of the cell. While mutations in the enzymes directly synthesising the glycans have predictable outcomes, the outcome of enzyme mislocalisation is difficult to predict in terms of the structural consequences for glycosylation.

CDGs can be separated into four groups: those with errors in *N*-glycosylation, *O*-glycosylation, glycolipid and GPI-anchor protein glycosylation, and those which present errors in multiple glycosylation pathways. CDGs, although rare, make up around 100 genetic disorders known to be caused by glycosylation-related mutations¹⁰, and effects are expressed in most organ systems, particularly the central nervous system. The disorders present a vast array of abnormalities⁷² such as developmental delay in infancy and childhood, muscular dystrophy, strokes, seizures, immune dysfunction and cardiac failure. It is this breadth of symptoms, together with the rarity of the diseases that makes them slow and difficult to diagnose.

There are two main types of CDGs, CDG-I and CDG-II. The first type typically presents normal *N*-glycans attached to glycoproteins but the sites of attachment are not fully occupied. This is caused by defects in the assembly of the precursor oligosaccharide⁷³ and its transfer to the protein. The most common defect of this type, PMM2-CDG^{74,75,76} (previously known as CDG-1a⁷⁷), is in phosphomannomutase, the enzyme that catalyses isomerisation of mannose 6-phosphate to mannose 1-phosphate, required to build the lipid-linked precursor in *N*-glycosylation. A less common form, MPI-CDG (previously known as CDG-1b⁷⁷), results from

defects in the phosphomannose isomerase gene, the product of which generates mannose 6-phosphate from fructose 6-phosphate⁷⁸. Freeze et al⁷⁹ reported a treatment for this defect by administering high levels of mannose in the diet, which was shown to be sufficient to restore normal levels of the dolichol donor, preventing hypoglycosylation.

In CDG-II, almost all of the usual glycosylation sites are used but the glycans attached are truncated⁸⁰. This is a result of defects further along the glycosylation pathway during processing of the protein-bound glycans, and can therefore affect *N*-glycans, O-glycans or both simultaneously^{12,81,82}. CDG-IIs may also be caused by defects in the COG complex; CDGs have been reported with defects in Cog7^{83,84}, or Cog1^{85,86} where some also show reduced expression of Cog1-4 and Cog8⁸⁷. Such COG defects are considered to interfere with the accurate localisation of glycan processing enzymes, and thus to cause glycan truncation⁸⁷.

A range of analytical tools is needed to help improve diagnosis, understanding, and treatment of these complex multisystem disorders. The main test carried out in the clinic for *N*-glycosylation defects is monitoring of serum transferrin by isoelectric focussing. This allows charge separation of transferrin isoforms that differ in the number of sialic acids⁸⁸. There is an increase in the use of mass spectrometry to monitor glycosylation defects, with some of the earlier applications using electrospray ionisation in 1992⁸⁹ and MALDI in 1994⁹⁰, where changes in glycan profiles between normal and CDG-I samples could be observed in mass spectra of serum transferrin. There is now a range of mass spectrometric applications for monitoring changes to glycan profiles in serum^{91,92,93} for CDG studies, including analysis of glycopeptides in serum to look at site specific glycoforms⁹⁴.

1.3.2. Biologics: glycoengineering, vaccines and drug targets

Glycans have a major significance in the pharmaceutical and medical industries because of the importance of glycan/glycoprotein/carbohydrate-based therapeutics. There has been a significant rise in the engineering of glycan components for creation of new therapeutics.

A major landmark in glycobiology and medicine was the discovery that glycans were the structurally-defining constituents of the A B O blood group substances^{95,96,97,98}. The main component of H antigen (blood type O) is fucose, to which the addition of GalNAc produces

the A antigen or the addition of Gal produces B antigen. This has been of major importance in matching blood donors for blood transfusions.

Glycan-based drug molecules have been around for many years. Heparin, the carbohydrate-based natural anticoagulant, was first discovered in 1916 by J. Mclean. Isolated from animal sources, mainly from porcine mucosa, has been used in the clinic to treat thrombosis since the 1940s¹⁷. Heparin has its effect through a structurally specific pentasaccharide constituent^{99,100,101,102} of heparin (identified in the 1980s) binding to antithrombin to stop blood clotting^{103,104}, Figure 13. Since then there have been synthetic forms of this pentasaccharide unit produced, such as Arixtra. Relenza¹⁰⁵ is an antiviral drug, based on a monosaccharide structure (Figure 13); the drug works by binding to the viral neuraminidase¹⁰⁶ preventing neuraminidase from binding to the sialylated glycan on the host cells¹⁰⁷ and therefore deactivating the virus and preventing entry into cells. In addition, antibiotics such as neomycin¹⁰⁸, an aminoglycoside, binds the 30S ribosome, so interfering with protein synthesis.

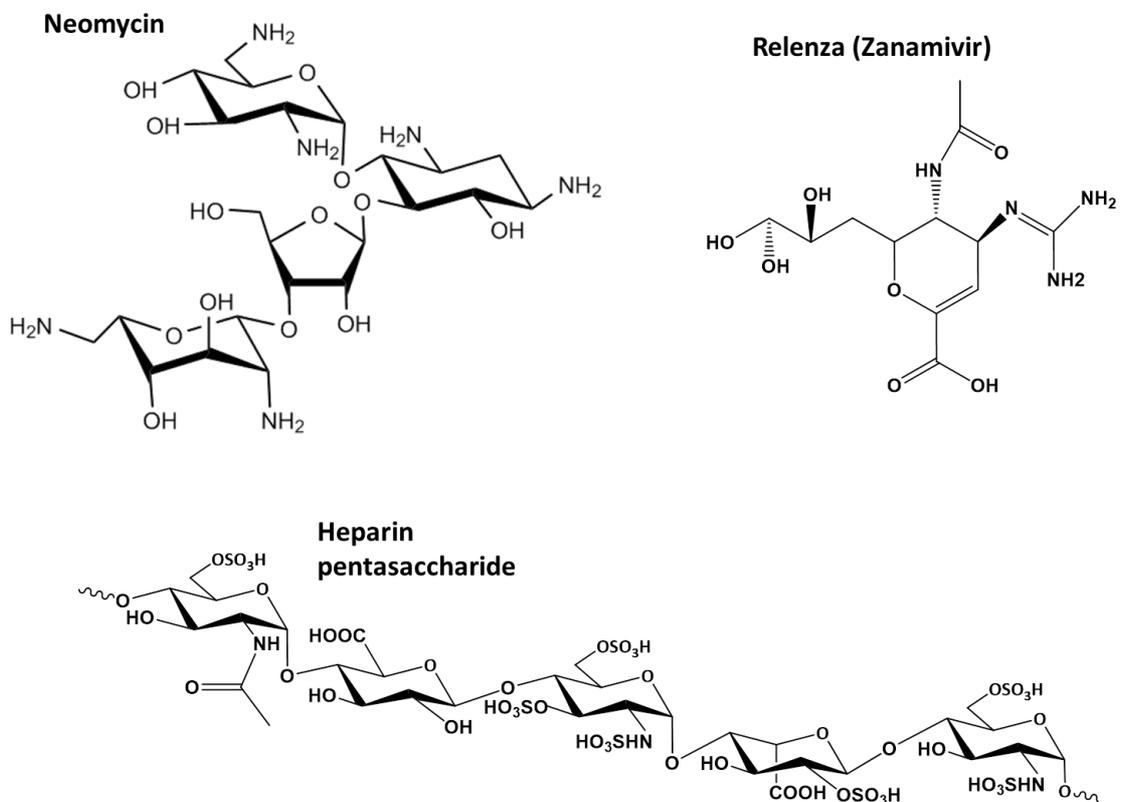


Figure 13: Pentasaccharide of heparin, the anticoagulant used for the treatment of thrombosis; it contains a sulfated iduronic acid and sulfated glucosamine. Neomycin antibiotic and Relenza, an antiviral drug.

Biopharmaceuticals, therapeutics often produced through genetic engineering of animal, plant or bacterial cells, is a major and still growing area in the pharmaceutical industry. The first biopharmaceutical, approved in 1982, was Humulin, recombinant human insulin. Since then over 200 biopharmaceuticals have been approved¹⁰⁹, with monoclonal antibodies (which are glycoproteins) dominating the biotherapeutics area, alongside hormones, enzymes, and other (glyco)proteins. Humira (adalimumab) is a monoclonal antibody used to treat rheumatoid arthritis, Crohn's disease and chronic psoriasis¹⁰⁹. Gazyva (Obinutuzumab), approved for treatment of chronic lymphatic leukaemia¹¹⁰, is one of the first approved glycoengineered monoclonal antibodies, and targets the surface protein, CD20, associated with B lymphocytes. In clinical trials Gazyva was found to be more effective than the wild-type, non-glycoengineered form, Retuximab¹¹¹, illustrating the importance of glycans in the success of biotherapeutics. In contrast, Taliglucerase (Elelyso) is a recombinant glucocerebrosidase, produced from an engineered carrot root cell line, used to treat Gaucher disease, a lysosomal storage disorder. Elelyso contains terminal mannose residues on its glycans, helping uptake to its target macrophages¹¹².

Surfaces of bacteria, viruses and parasites are covered with glycans, which are distinct from those on the host and so carbohydrate-specific antibodies are responsible for protection by being able to provide an immune response against carbohydrate antigens, resulting in target-cell death. Certain individuals lack such antibodies and are therefore more vulnerable to infections and so need vaccinating. For example, infants are immunized against meningitis and the elderly against *Streptococcus pneumoniae*.

Carbohydrate vaccines were first derived from natural sources containing glycans however there is increasing work on synthesising carbohydrate-based vaccines. For example, the fully synthesised haemophilus influenza type b (Hib) vaccine based on the capsular pentasaccharide antigen structure from Hib. The polysaccharide is conjugated to tetanus toxin¹¹³. In addition a vaccine has been developed by Huang et al¹⁶ against a breast cancer-associated carbohydrate antigen, Globo-H. The vaccine uses Globo-H conjugated to diphtheria toxoid, co-administered with α -galactosylceramide analog adjuvant, and has been shown to induce a good antibody response against Globo-H on breast cancer.

In addition to carbohydrate-based vaccines, glycans are also important as drug/vaccine targets. There is often aberrant glycosylation found on cancer cells. The protein MUC1 is

found to be over expressed in cancer cells and MUC1 glycosylation shows increased branching and sialylation¹¹⁴. In addition MUC1 on tumour cells contain core 2 O-glycans that promote metastasis¹¹⁵. These factors make MUC1 and the tumour-associated glycan antigens a good target for cancer vaccines^{15,16}. Developing vaccines for cancer is useful because they can raise antibodies against tumour-associated carbohydrate antigens from tumours that remain after a primary treatment, such as chemotherapy, to try to prolong survival free from the disease.

Therapeutic glycoproteins are often produced as a mixture of glycoforms, each of which will have different properties, effecting solubility, stability and efficacy. Isolating individual glycosylated products and determining their in vivo and in vitro functions is an almost impossible task. The human antibody IgG (discussed in more detail in chapter 5) contains one glycosylation site but can display up to 500 glycoforms¹¹⁶. To try to simplify and create a more homogenous display of glycoforms in therapeutics, cell lines (expression hosts) are engineered to produce specific glycosylation structures¹¹⁷ and through studying these engineered systems a greater understanding of the glycan functions can be established. Chinese hamster ovary cells have been engineered to express β -1,4-N-acetylglucosaminyltransferase III which adds a bisecting GlcNAc to the N-glycan core. This result lead to increased antibody dependant cellular cytotoxicity^{118,119} which was later found to be caused by the GlcNAc blocking core fucosylation and it was the absence of the fucose causing the desired effect. This has led to development of more de-fucosylated protein therapeutics.

In addition to engineering cell lines to produce controlled glycan structures, glycans can be altered after expression in the cell lines. This is used for treatment of lysosomal storage disorder where a glucocerebrosidase enzyme is used that is deglycosylated in vitro using α -neuraminidase, β -galactosidase and β -N-acetylglucosaminidase to expose mannose¹²⁰, improving targeting and internalization.

Cell surface glycan remodelling is used to create non-natural sugars where a reactive functional group is added to the glycans to help to target antibodies or drug delivery vehicles, such as nanoparticles. For example, Gartner and Bertozzi¹²¹ used azide-containing glycans to add complementary DNA strands to different cell types to assemble microtissues, useful for studying organ development and disease progression. In addition, the Bertozzi

group found that immunizing rabbits with a ketone-containing sialic acid bound to a protein carrier led to production of antibodies capable of binding to cancer cells to promote cell lysis¹²².

1.4. Methods for glycosylation analysis

Early work in glycomics sought to define the major glycan structures present on single, isolated glycoproteins, often from large quantities of samples, where the information gathered provided detail on composition of the glycans, the linkages, anomericity (α or β) and configuration (D or L). In 1967, Carlson was the first to report an oligosaccharide to contain both a fucose and a sialic acid in porcine submaxillary mucin¹²³. In 1982 Spik et al¹²⁴ reported the primary structure for glycans obtained from human lactotransferrin structure was determined by methanolysis, methylation analysis, hydrazinolysis (to release glycans from the protein), enzymatic cleavage, such as desialylation with neuraminidase, GC-MS and ¹H-NMR analysis. The combination of all the different sample preparation and analytical methods allowed distinction of four different oligosaccharides all fully characterised in terms of monosacchirdes unit present, their linkage position, branching and anomericity.

As the knowledge of glycan biosynthesis increased^{30,125} with further structural investigation, a knowledge of the possible repertoire of glycan structures also increased, leading into work which focussed on glycan profiling using mass spectrometry^{126,127,128,129} to determine glycans present in tissues and whole systems. If further structural analysis was required to determine particular isomeric forms present, especially for studies investigating disease, then methods previously developed for composition, linkage analysis etc could be used. For example, glycans from different populations of porcine stomach mucins^{130,131} were investigated where there was found to be regional differences in the glycan structure^{132,133}.

Cell culture is useful to study changes to glycosylation and the effects of these changes through disease, mutations, cell differentiation, and drug treatments, under controlled conditions. Cell culture offers the advantage over having to use human/animal tissues, serum or biopsies which are harder to manipulate and less readily available. However, a lot of glycomic studies using cell culture carry out deep glycan profiling which requires large numbers of cells (10 – 100 million)^{134,117}, limiting the feasibility of replication. Since glycosylation changes during cell differentiation and disease, batch-to-batch variations in biotherapeutics, including variation between cell types and growth conditions used, the lack

of replicates hinders the ability to report such differences to glycosylation. To study glycan biosynthesis and to try to understand the effects of biosynthetic enzyme mislocalisation that can manifest in subtle changes to glycan profiles, methods are needed to detect differences that are biologically relevant against a background of normal biological variation between individuals. To determine which of these glycan profile changes are significant, and which are just low-level biological variation, analysis of multiple biological samples in parallel is needed. It is not necessary to carry out ultimate deep glycan profiling to make this assessment, but rather to profile reproducibly enough to uncover overall shifts in relative glycan proportions. This makes it possible to avoid the use of very large numbers of cells that would be necessary for very deep profiling, but that at the same time limit the feasibility of analysing sufficient numbers of replicates.

There are reports of studies that use fewer cells such as Nakeno et al¹³⁵ who reported *N*- and *O*-glycan analysis on isolated membranes from 10^7 cells of a leukaemia cell line. This however, was for a detailed glycome analysis involving fluorescent labelling and LC-MS, not applicable for medium through-put analysis with primary cell cultures and expensive reagents. Other studies also involve complex sample preparation for the goal of deep glycan profiling even though the sample amounts were reduced from using three 10 cm dishes per sample¹³⁶ to one 10 cm dish (1-2 million cells)¹³⁷. A study was also reported to analyse *N*-glycans from human embryonic stem cells¹³⁸ which, although it reported only using 100 000 cells, again had multistep sample preparation not suited for a medium through-put, glycomic profiling that would enable generation of biological replicates.

Glycomic workflows frequently use separate protocols and samples for looking at *N*-glycans, *O*-glycans and for proteomic studies. A study reported release of *N*- and *O*-glycans from mesenchymal stem cells¹³⁹, and although the *O*-glycans were released from the de-*N*-glycosylated protein, after recovering the protein through a precipitation-extraction step, the *O*-glycans had to be pooled together before *O*-glycan analysis and so again, replicates were not possible. To understand the effects on total glycan profiles, it would be ideal to analyse both *N*-glycans and *O*-glycans from the same sample in an integrated protocol, i.e. a one-sample-one-pot method, making the best use of samples. Since some genetic diseases can occur as a result of a genetic defect that affects both *N*- and *O*-glycosylation simultaneously, whereas some defects only affect one glycan type, analysis of both *N*- and *O*-glycans from the same sample is ideal. A one-sample-one-pot pot method also has the

advantage of reducing the amount of material needed to carry out such studies, and would offer the potential, after *N*- and *O*-glycan release, enabling a complementary proteomic comparison between samples.

Recently, a paper was published describing a method to amplify the cellular *O*-glycome for in-depth glycan profiling using around 80 000 cells¹⁴⁰. This study addresses the challenge of cutting down on sample size, however this protocol is exclusively for studying *O*-glycan, and although *O*-glycans are important, it does not allow for simultaneous study of both *N*- and *O*-glycan from the same sample.

Methods are therefore needed to solubilise and extract glycoproteins from small numbers of cultured cells and to release and isolate their glycans for mass spectrometric analysis. Recently, a method has been developed for the release and isolation of *N*-glycans from glycoproteins solubilised from whole cell lysates, using a centrifugal filter. The method has been named filter-aided *N*-glycan separation, or FANGS¹⁴¹. The simple protocol makes use of a filter-aided sample preparation (FASP) approach for SDS removal, first described by Manza et al¹⁴², and then by Wiśniewski et al¹⁴³ for exchange of SDS with urea. In FANGS, following SDS exchange, the glycoprotein extract is then incubated with PNGase F to release the *N*-glycans. *N*-glycans are collected after centrifugation, Figure 14. FANGS has been successful using only 3.5×10^5 cells, equivalent to one well of a six well culture plate and so this method is very useful, being applicable to low cell numbers and therefore can be used to analyse multiple samples in parallel. In the literature there is no equivalent method for the release of *O*-linked glycans.

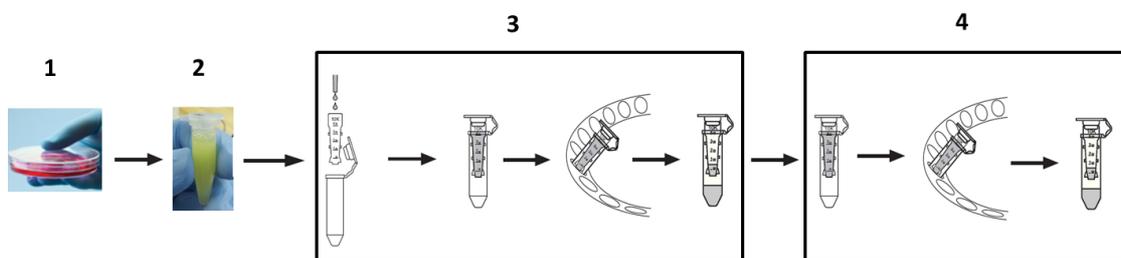


Figure 14: Filter-aided *N*-glycan separation (FANGS): 1) cell culture, 2) cell lysis, solubilisation in SDS, 3) transfer of solubilised cell lysate to filter, exchange of SDS for ammonium bicarbonate buffer, filter centrifuged to remove SDS, 4) PNGase F added to sample in the filter to release *N*-glycans from the glycoproteins, filter centrifuged to collect the released glycans in the filtrate. Modified from Rahman et al¹⁴¹, and from Millipore Amicon filter user guide.

1.4.1. Release of O-glycans from glycoproteins

For O-glycosylation, there is no enzyme that shows a broad specificity for O-glycan release to compare with PNGase F for *N*-glycan release, so that chemical methods are the most appropriate for pan-specific O-glycan release. PNGase F, first reported by Plummer et al in 1984, has broad substrate specificity and cleaves the *N*-glycans from the asparagine side chain at the glycosylamine linkage, leaving the *N*-glycan structure intact¹⁴⁴. It can cleave high-mannose, hybrid, bi- tri- and tetra-antenary *N*-glycans.

Hydrazinolysis, originally described by Matsushima and Fuji¹⁴⁵ and later by Bayard¹⁴⁶, releases O-glycans as hydrazones which are then converted to reducing glycans by exchange with acetone. However, these procedures use high temperatures (60 – 90 °C), which can also lead to peeling (where monosaccharides are successively cleaved from the reducing terminus of the released O-glycan, Figure 15). Although recently peeling has been found to be significantly reduced by adding ethylenediaminetetraacetic acid (EDTA) directly to the hydrazine^{147,148}. Hydrazinolysis is also inconvenient as it requires anhydrous conditions, there can be numerous by-products formed and hydrazine is toxic. Hydrazinolysis can also release *N*-glycans and through controlling the condition either *N*-glycans, O-glycans or both *N*- and O-glycans can be released¹⁴⁹. *N*-glycan release requires stronger conditions, 85-100 °C¹⁴⁹, 5-16 h, compared to milder conditions for O-glycan release (60 °C, 6 h)¹⁵⁰. The exact mechanism for hydrazinolysis of glycans is not fully understood. There is suggestion that the reaction occurs by an initial β -elimination type reaction followed directly by reaction of the glycan with hydrazine to form a hydrazone¹⁵⁰.

β -elimination involving treatment of the glycoprotein with NaOH can be used to release O-glycans from glycoproteins^{123,151}. However, the strongly alkaline conditions can result in peeling reactions¹⁵². Inclusion in the reaction of NaBH₄ (reductive β -elimination) helps to prevent peeling by converting the O-glycan's reducing terminal residue to an alditol, Figure 16.

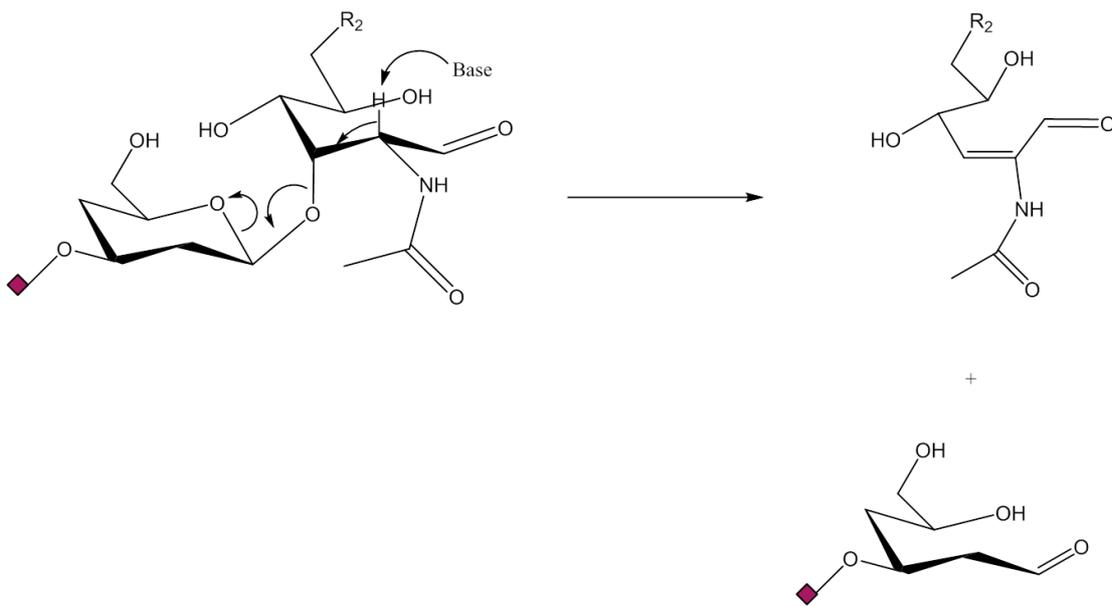


Figure 15: Peeling reaction¹⁵²: monosaccharides are eliminated from the reducing residue of the released glycan.

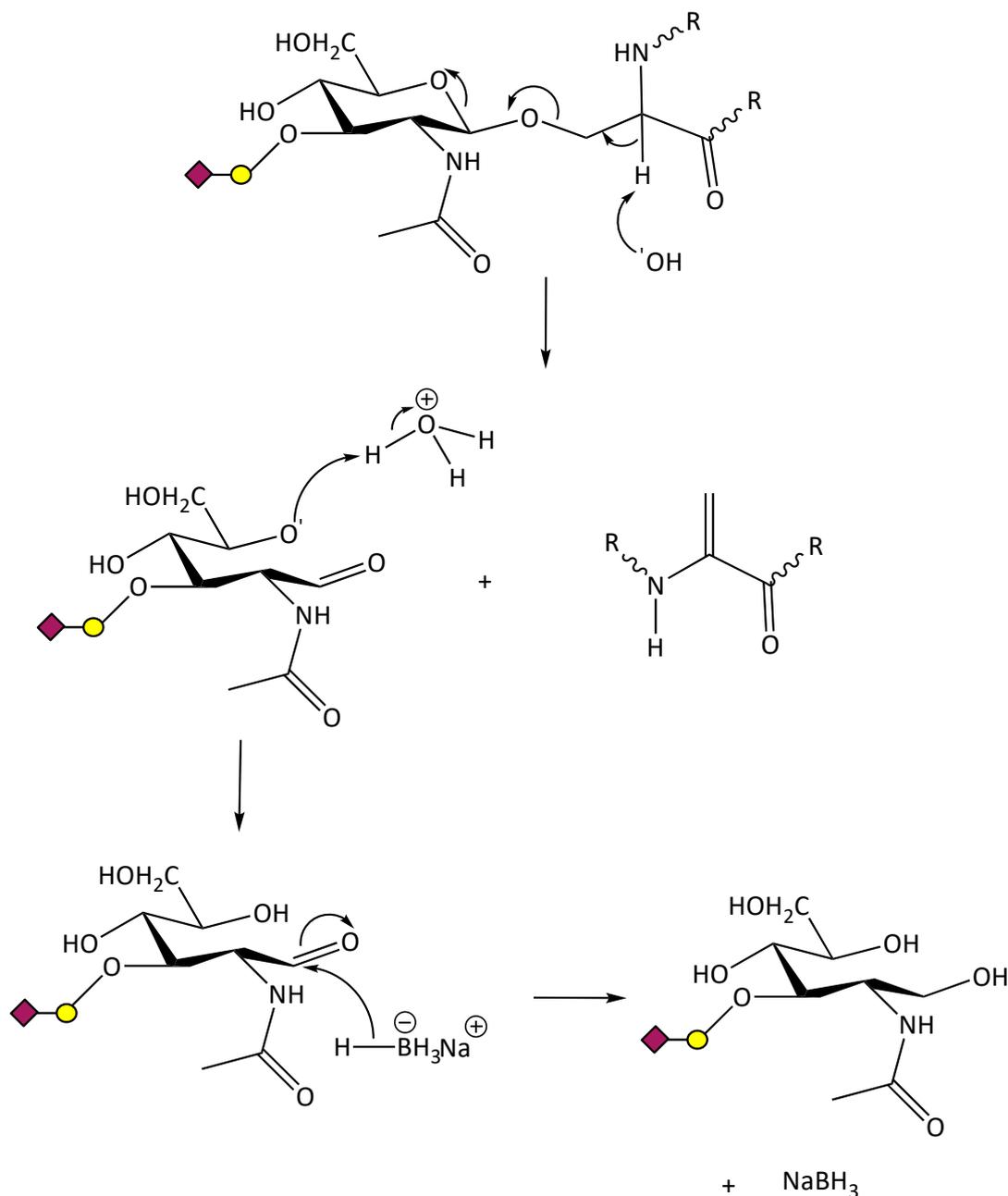


Figure 16: General mechanism for β -elimination and inclusion of NaBH_4 to prevent peeling by converting the reducing terminus of O-glycans to alditols.

Alternatively, glycoproteins can be treated under similar conditions with NH_4OH , releasing O-glycans and adding NH_3 across the double bond of the amino acid residue left behind, in a Michael-type addition reaction (Figure 17). This method uses milder conditions than reductive and other non-reductive elimination methods, and it does not require a desalting step prior to MS analysis, as NH_4OH is volatile and so is easily removed under reduced pressure¹⁵³, making it a convenient method to use.

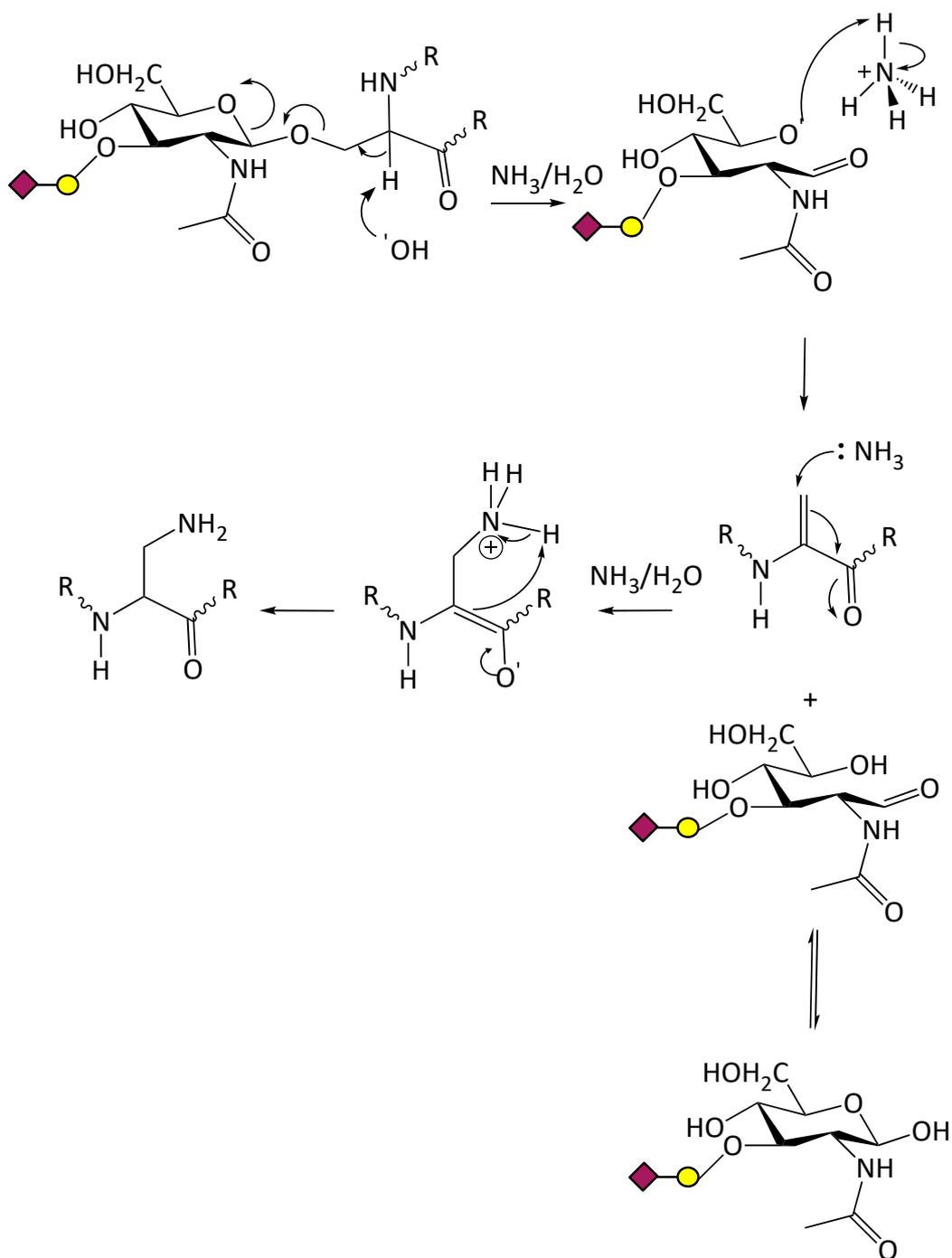


Figure 17: Non-reductive β-elimination, adapted from Rademaker et al¹⁵³

Novotny et al¹⁵⁴ have reported a β-elimination protocol based on the Rademaker method, that uses NH_4OH and ammonium carbonate rather than just NH_4OH . These reagents used together are stated to help to prevent peeling by blocking the reducing terminus of the released O-glycans; the ammonium carbonate converts the reducing glycans to glycosylamines which can then react with boric acid, yielding reducing glycans after removal of the base. Rademaker does not make use of ammonium carbonate nor does the study

suggest that peeling is observed when using NH_4OH . Since NH_4OH is a weaker base than NaOH commonly used in the β -elimination of O-glycans, it is tempting to assume that if peeling is observed it would be to a lesser extent, further so as the Rademaker method also uses a lower temperature. The study by Novotny did not investigate whether using NH_4OH resulted in peeling before they came up with a solution. Instead, they just made the assumption that peeling was observed. Likewise it is not necessary to treat with boric acid.

Conditions for releasing glycans from glycoproteins/peptides generally involve overnight or longer treatment. We are unaware of reports in the literature describing faster release of O-glycans. A shorter incubation time for glycan release would be very convenient for enhancing the through-put of analyses. There are reports of using microwave assistance to speed up incubation times in a range of chemical and enzymatic reactions. Domestic microwaves have been used to assist in trypsin digestion of proteins, and recently a method has been described¹⁵⁵ using a domestic microwave to release *N*-glycans from purified glycoproteins such as bovine fetuin and complex glycoprotein mixtures, using PNGase F, in 20 minutes (using 20% of the maximum power of the microwave). This method was reported to be just as effective as overnight incubation and microwave irradiation at low power was found not to affect the released glycans, leaving labile sialic acid residues intact. Sonication, used widely in industry and research, is also used to accelerate reactions in organic synthesis, stimulate enzymes in enzymatic synthesis, sonication has been used to extract proteins out of SDS-PAGE gels¹⁵⁶, and is also used to disrupt cell membranes without damaging the cell contents¹⁵⁷. However, there appears to be nothing in the literature to suggest sonication has been assessed for use in the release of glycans from glycoproteins.

Sonication is ultrasound, sound energy of a frequency beyond what the human ear can hear. Sonication works by a series of compression and rarefaction waves induced by the molecules of the medium through which the sonication passes. Bubbles form, which are unstable from interference of other forming bubbles close in space, causing sudden expansion and collapse of the bubbles, releasing energy as a result¹⁵⁷. In a water bath sonicator, samples are placed in a reaction vessel and the reaction vessel is placed into the water bath and so the energy is transferred from the water, through the reaction vessel walls and so the energy, which reaches the sample, is relatively low ($1\text{-}5\text{ W cm}^{-2}$)¹⁵⁷. This is in contrast to an ultrasonic probe, which is placed directly into the reaction vessel inducing energy of several hundred W cm^{-2} . Temperature control of the water bath is generally poor.

1.4.2. Derivatisation and quantitation of glycans

Carbohydrates tend to produce a poor signal response in mass spectrometry because they do not have many sites for protonation and so they do not ionise efficiently, unlike other biological molecules such as protein/peptides. Carbohydrates are therefore frequently derivatised^{158,159,160} by permethylation^{161,162,163,164}, peracetylation^{165,164} or reducing terminal tagging¹⁶⁰ to help improve the sensitivity of detection, the rate of change of response, by enhancing the analyte's surface activity in the matrix in MALDI, and in ESI, improving mass spectrometric response.

Permethylation produces methyl ethers on the carbohydrate hydroxyl groups, converting the glycan to a more hydrophobic species. The resulting decrease in intermolecular hydrogen bonding increases the surface activity of the glycan sitting on the MALDI matrix surface, or in an electrospray droplet, leading to an increase in the ion signal intensity¹⁶⁶. In addition, permethylation can stabilise sialic acid linkages¹⁶⁷ by eliminating negative charge associated with acidic groups. The permethylation method that is now routinely used is catalysed by sodium hydroxide through reaction with slow addition of iodomethane^{161,163,162} (Figure 18). This method has been shown to produce higher yields of permethylated products in a shorter time than the original Hakomori¹⁶⁸ permethylation method reported in 1964, catalysed by methylsulfinyl carbanion in dimethyl sulfoxide. As a result of permethylation, tandem mass spectrometry can yield diagnostic fragmentation patterns to help determine the glycan structure (discussed in Chapter 2).

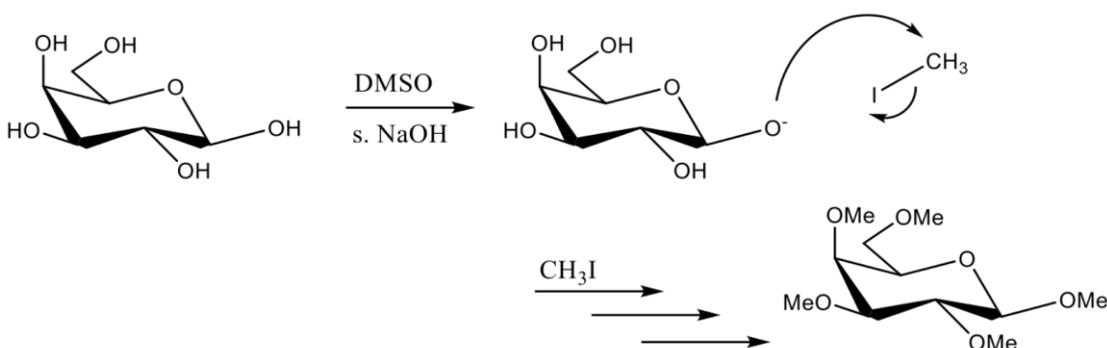


Figure 18: Permethylation of carbohydrates.

An important aspect of glycan profiling is obtaining reliable quantitative information about relative glycan abundance to compare across sample replicates and across different samples

such as wild-type cells and diseased or differentiated cells. Quantitative methods using mass spectrometry include spiking an internal standard, such as maltoheptaose^{169,170}, into samples. This method is useful since sample and the standard are simultaneously analysed. Internal standards such as maltoheptaose are useful for relative quantitation of glycans since they are similar chemically and are around the correct mass range. Stable isotope labelling is also used for relative quantitative glycomic analyses. This involves addition of different species of isotopes onto chemically similar analytes, which will generate a measurable mass difference on the mass spectrometer. In addition, relative quantitation of glycans has been carried out using ¹³CH₃I and ¹²CH₂DI to permethylate the glycans prior to analysis by mass spectrometry and can be used for both *N*- and *O*-glycan analysis^{171,172,173,174}. Other methods for glycan quantification include ¹⁸O labelling of the reducing end of *N*-glycans released by PNGase F in heavy water^{172,175}. Relative quantitation of glycans has also been carried out by peak height measurements in mass spectra where peak smoothing has been carried out to remove isotopic resolution, or by summing the peak heights of all the isotopes¹⁷⁶.

1.5. Aims of the thesis

The aim of the work in this thesis was to develop a method to release and isolate *O*-glycans from glycoproteins, from a small number of cultured whole cell lysates for analysis by mass spectrometry.

This thesis presents a FANGS-based approach, making use of a centrifugal filter, to carry out *O*-glycan release using non-reductive β -elimination, as well as exploiting the advantages of an optimised FASP^{142,143} approach, eFASP¹⁷⁷, which uses alternative reagents to FASP. Method optimisation reveals how *O*-glycan release can be achieved with a total of 20 minutes of sonication, without loss of labile sialic acid groups, and giving comparable results to the well-accepted overnight incubation methods. In addition, the thesis presents a streamlined protocol that has been developed for a one-sample-one-pot approach to release *N*- and *O*-glycans from the same sample, in the same sample pot, with the potential to subsequently analyse the protein remaining in the filter. The developed method has been demonstrated using porcine stomach mucin as a standard soluble glycoprotein, and has then been applied to porcine bladder urothelial cells, mesenchymal stromal cells and normal human urothelial cells.

Chapter 4 describes the method development work for the release and isolation of O-glycans from glycoproteins of whole cell lysates, in a centrifugal filter, for analysis by mass spectrometry. The chapter continues to show how the established *N*-glycan release method can be combined with the O-glycan release method to offer a simple, and convenient one-sample-one-pot method to enable analysis of the whole glycome. In addition, after glycan removal it is possible to analyse the protein remaining in the filter unit.

Chapter 5 describes work carried out on a seven week placement at Bristol-Myers Squibb, Lawrenceville, New Jersey, USA. The project involved investigation of the FANGS method for larger scale *N*-glycan isolation, to enable analysis of glycans from monoclonal antibodies using a suite of 2D NMR techniques, in order to establish whether NMR could be used as a glycan fingerprinting tool to quickly and effectively monitor and analyse batches of biologics.

Chapter 6 describes application of the O-glycan release method to mesenchymal stem cells and shows how the method can be effectively used to compare changes to the glycan profiles between genetically modified MSC cells, chemically treated cells and those which have differentiated into adipocytes, to offer insight into the roles of glycans on adipogenesis. The chapter also shows that the one-pot method can be applied to 'real' cultured cell samples. The approach focuses on determining the often-subtle changes in relative levels of the majority of the glycans in a glycan profile in a reproducible and statistically robust way, without the need to focus on detecting the most minor components of the glycome.

Chapter 7 describes application of the one-sample-one-pot method to normal urothelial cells cultured from human donor samples, and enables comparison of both *N*- and O-glycans between donors and between proliferating and differentiated cells.

Chapter 2: Mass Spectrometry

2. Mass spectrometry

2.1. Introduction

Since mass spectrometry is the main analytical technique used in this thesis this section offers a description of each of the mass spectrometry techniques used.

Mass spectrometry, quite simply, is a means of weighing molecules and their fragments to gain insight into the molecular weight and structural of an analyte. The general set-up of a mass spectrometer involves an ion source, whereby analytes are ionised. The resulting ions are transferred to a mass analyser where they are separated according to their mass to charge ratio (m/z) and then they are detected.

2.2. Matrix-assisted laser desorption/ionisation (MALDI)

Introduced by Karas and Hillenkamp^{178,179} as a technique to generate intact gas phase ions of large, thermally-labile and involatile molecules such as proteins, carbohydrates, oligonucleotides and synthetic polymers, MALDI has become very widely adopted and is now a core technique for the analysis of all types of involatile analyte above around 700 Da.

MALDI was developed from an earlier ionisation technique, laser desorption ionisation (LDI) which involves firing laser pulses onto a sample surface. The laser pulses ablate material from the sample surface, resulting in formation of gaseous ions and molecules. LDI has an upper limit for the size of the molecules that can be desorbed as intact gaseous ions of around 500 Da. This is because the analyte molecules absorb energy from the intense laser pulses which can cause sample damage and so analytes above approx 500 Da are usually observed as fragmentation products of the analyte¹⁸⁰. The development of MALDI allowed observation intact analyte ions in excess of 100 000 Da.

In MALDI, a matrix compound is added in excess over the analyte to ensure isolation of the analyte molecules from each other, as well as providing insulation from the laser energy on co-crystallisation of the analyte with the matrix. This insulation is provided because the matrix, frequently an organic acid, absorbs strongly at the laser wavelength and is therefore a mediator for energy transfer between the laser beam and the sample, allowing for softer

ionisation. Commonly used matrices for analysing carbohydrates, peptides and proteins are 2,5-dihydroxybenzoic acid (DHB)¹⁸¹ and α -cyano-4-hydroxycinnamic acid (CHCA)^{182,183}, Figure 19.

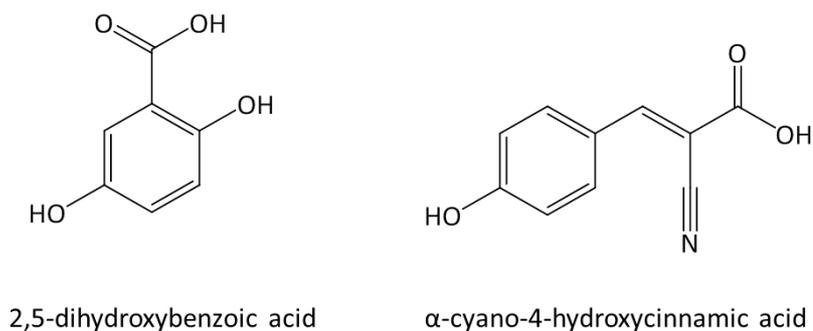


Figure 19: commonly used MALDI matrices for analysing carbohydrates, peptides and proteins.

2.5.1. Sample preparation

Preparing a sample for MALDI is achieved by dissolving the matrix in a solvent, usually a water/organic mix, using solvents such as methanol, ethanol or acetonitrile. The analyte sample is dissolved in a similar solvent. An aliquot of the sample solution is then mixed with an aliquot of the matrix solution and around 1 μ L of the mixture solution is applied to a stainless steel target and dried under ambient conditions. This method is generally referred to as the dried-droplet method. The matrix solution can alternatively be applied to the target and allowed to dry before applying the sample solution on top; the spot, however constituted, can also be dried under vacuum. The quality of MALDI data depends on homogeneity of the sample/matrix spot¹⁸⁴. Sometimes the sample/matrix material will accumulate around the edges of the spot upon drying. DHB can form needle shaped crystals around the edge of the sample area and the best spectra are acquired where more homogeneous crystals are visible, usually away from the needles. In these cases, better data are often acquired around the edge of the spot.

2.5.2. Ion formation in MALDI

In MALDI, the matrix/analyte co-crystals are irradiated with intense laser pulses. The matrix, chosen for its chromophore, absorbs the energy of the laser light, exciting the matrix, causing it to vaporise into an expanding plume, pulling the analyte molecules with it (Figure 20). During this laser ablation/desorption step, many primary ion formation pathways^{185,186}

such as pre-charged analytes in the matrix clusters¹⁸⁷, charge separation¹⁸⁸, photoionisation, thermal ionisation, disproportionation reactions and excited state proton transfer, have been proposed. In addition there is a range of secondary ion formation pathways^{186,189,190} such as ion-molecule reactions¹⁹¹, gas-phase proton transfer, cation transfer, electron transfer, matrix-matrix reactions and matrix-analyte reactions that have been discussed.

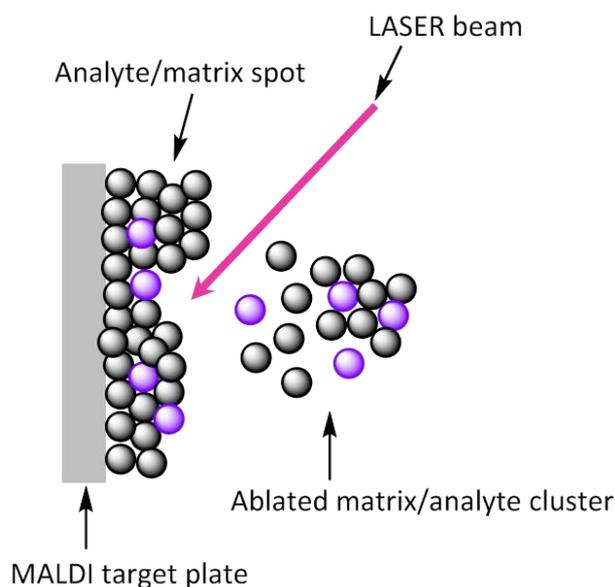


Figure 20 Matrix-assisted laser desorption/ionisation. Laser ablation of the matrix/analyte spot, causing excitation of the matrix and vaporisation into an expanding plume.

The exact mechanism by which MALDI works has been a popular debate since its introduction, yet it is still not fully understood. Understanding the ionisation mechanism is important to better control fragmentation, charge states, and to improve ion yields. However it is a difficult task because there are multiple experimental conditions that affect the appearance of a mass spectrum such as laser wavelengths, pulse energies/lengths, sample preparation methods, whether the instrument is run in positive or negative mode, as well as the use of different matrices and classes of analyte. It is most probable that no single mechanism can explain all ions observed in a MALDI mass spectrum.

Two models most discussed to date are the lucky survivor theory¹⁸⁸ and the gas phase protonation model^{192,186}. The original lucky survivor theory stated analytes are incorporated in the matrix crystals with their respective solution charge states preserved and suggests two possible outcomes of laser irradiation. One possibility is that ablated clusters contain precharged analytes with counterions and so carry no net charge. Dissociation of the clusters (break down or separate into smaller clusters) leads to overall charge neutralisation,

and so would not be detected mass spectrometrically (Figure 21a). Alternatively, upon disruption of the matrix-analyte crystals there is charge separation and so the clusters carry a net charge. Provided that the ions are not then neutralised by absorption of electrons or photoelectrons from the target, they are detected as “lucky survivors” (Figure 21b). There is now a refined lucky survivor model¹⁹³ which states that matrix molecules with a charge, described by the authors as excess charges, that become included in the clusters lead to the generation of protonated analytes following counterion neutralisation to neutralise the matrix ion, and leave the analyte ion with a surviving charge (Figure 21c).

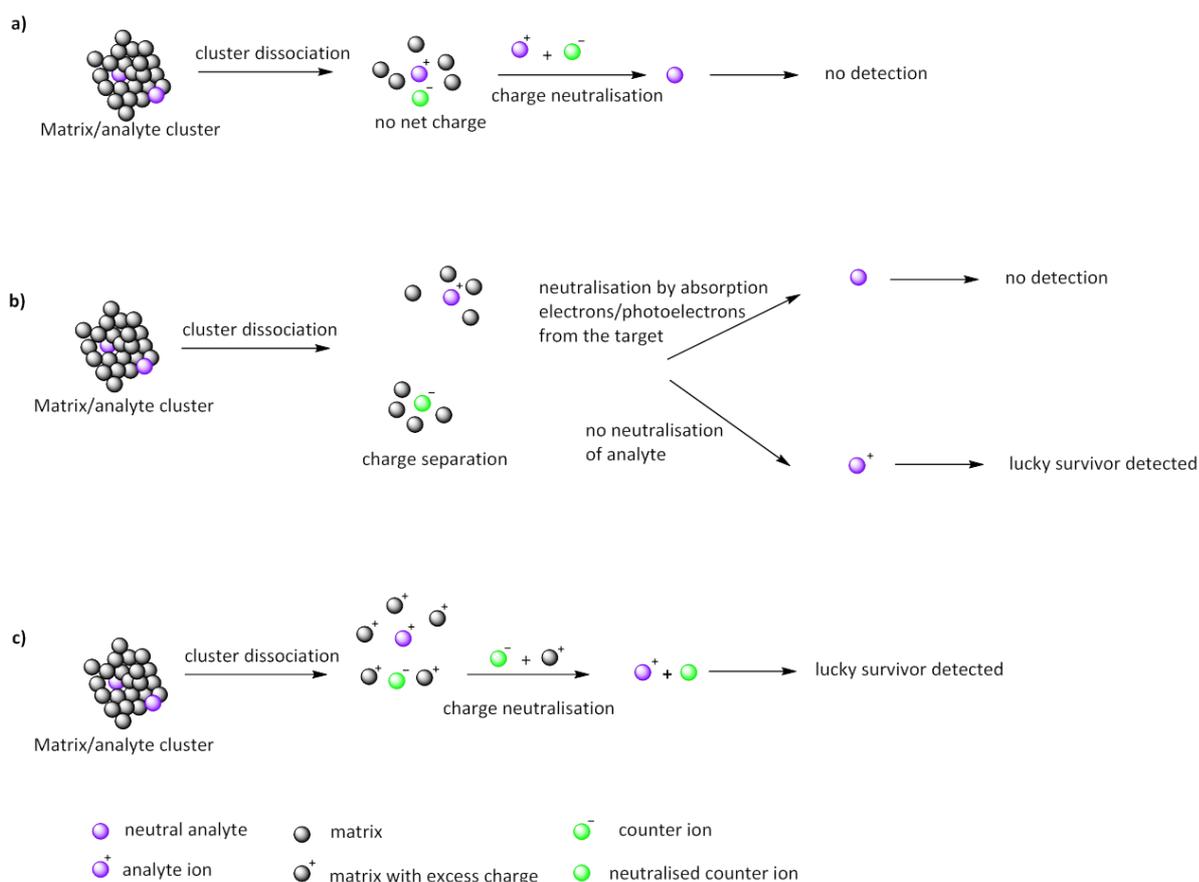
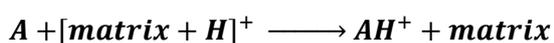


Figure 21: The lucky survivor model¹⁹³ a) ablated matrix/analyte clusters contain precharged analytes with counterions and so carry no net charge. Dissociation of those clusters leads to charge neutralisation, and so the analyte is not detected. b) charge separation upon disruption of the matrix-analyte clusters gives a net charge and so the analyte ions are detected as “lucky survivors”. c) Matrix excess charges included in the clusters lead to generation of protonated analytes by counterion neutralisation and so the analyte ions are detected as “lucky survivors”.

The gas phase protonation model suggests there are neutral analytes in the gas phase from either incorporation into the matrix crystals as uncharged species, or charge recombination with respective counterions. Collisions in the gas phase of neutral analytes with protonated or deprotonated matrix ions leads to proton transfer reactions, yielding either protonated or deprotonated analytes, Equation 2, where A is the analyte.



Equation 2

It has more recently been suggested that both the lucky survivor theory and gas phase protonation are parts of one overall mechanism¹⁹³. It has been shown that both models occur in parallel, although one model may occur to a greater extent than the other depending on sample preparation conditions, laser fluence, matrix and analyte proton affinities, whether protonated or neutral analyte species are embedded in the matrix clusters, as well as analyte composition and size. It was found that with neutral analytes, protonation could only be explained by the gas phase protonation model. For medium basic analytes, which exist as either neutral or protonated species depending on the pH of the sample preparation solution, it was found, for an analyte such as nicotinamide, that both neutral and protonated species were incorporated into the matrix crystals. The neutral species followed the gas phase protonation model, whereas the protonated species agreed with the lucky survivor model. On changing the pH from basic to acidic, the lucky survivor pathway was promoted and the gas phase protonation suppressed. However, on using those acidic conditions and increasing the laser fluence not only promoted the gas protonation model but also led to an increase in contribution from the lucky survivor model. When looking at positive precharged analytes such as peptides, both models were shown to occur in parallel, and the ratio between the occurrences of the two models was greatly affected by the matrix proton affinity. The lower the proton affinity of the matrix, the more the gas phase protonation model dominates. Proton affinities of 2,5-DHB and CHCA, two common MALDI matrices, are 854 KJ mol^{-1} and 766 KJ mol^{-1} , respectively.

MALDI generally produces singly charged ions, $[M + H]^+$, from peptides, proteins and glycoproteins. This is because high charge states cannot survive in the matrix/analyte clusters as the highly charged species absorb electrons that are formed during matrix photoionization processes, causing charge reduction to either a single charge or complete

neutralisation. When analysing oligosaccharides, sodiated molecules ($[M + Na]^+$) are usually observed. This is because there are few sites for protonation and so they are cationised instead. Cationisation is important where the proton affinities of the analyte are lower than those of the matrix (typical proton affinities of carbohydrates are $> 160 \text{ KJ mol}^{-1}$, compared to $\sim 850 \text{ KJ mol}^{-1}$ for matrices¹⁹⁴) and therefore they cannot compete with the matrix for available protons in the secondary reactions in the matrix-analyte plume. Sodium ions are present in the matrix, solvents, and glassware and possibly from residual Na^+ from permethylation and so is readily available.

2.6. Electrospray ionisation

Electrospray ionisation (ESI), like MALDI, is a soft ionisation technique. ESI was developed by Dole¹⁹⁵ during his work on ionisation of polystyrene molecules. Although at the time this was not for mass spectrometry, it had solved a problem that would help the application to mass spectrometry. Until then it was an impossible task to analyse large polyatomic molecules by mass spectrometry, without extensive fragmentation. The electrospray principles developed by Dole was applied by Fenn et al^{196,197} to use in mass spectrometry of large biomolecules, proteins.

In ESI, analytes are introduced to the ion source in solution and pass through a small capillary in an electric field gradient. As the solution reaches the tip of the capillary, charge builds up on the solvent surface and deforms the initially spherical profile of the emerging droplet, by coulombic forces, into what is known as a Taylor cone¹⁹⁸, Figure 22. A counterflow of nebuliser gas (usually N_2) aids evaporation of the solvent and as this happens, the charge density on the emerging liquid increases until the Rayleigh limit¹⁹⁹ is exceeded and the repulsive pressure exceeds the surface tension, causing a coulombic explosion, creating small droplets in a fine 'spray'. The solvent continues to evaporate, again increasing the charge density until it once again exceeds the surface tension causing coulombic explosion; the cycles continue until only desolvated gas-phase analyte ions remain. This process of analyte ion formation is known as the charge residue model, proposed by Dole¹⁹⁵, Figure 23.

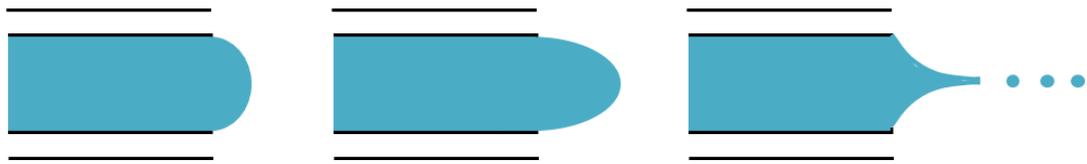


Figure 22: Formation of a Taylor cone¹⁹⁸

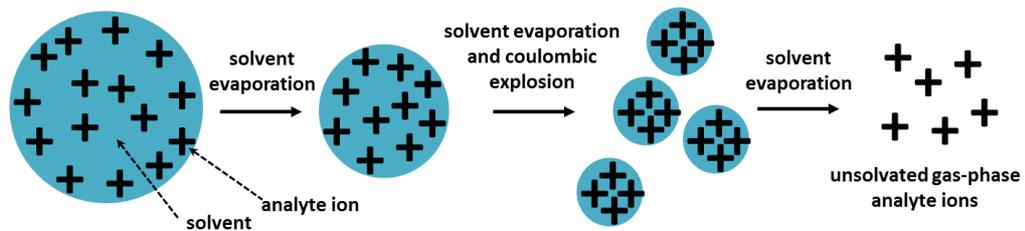


Figure 23: Charge residue model¹⁹⁵ for gas-phase ion production in ESI.

A second model proposed by Iribarne and Thompson²⁰⁰ is the ion evaporation model. Here the solvent evaporation and subsequent increase in charge density occurs as in the charge residue model but instead of repeated coulombic explosion, once the charge density is too high, an analyte ion is desorbed from the liquid surface. This decreases the charge density for a moment. The droplet continues to evaporate, again increasing the charge density and then further analyte ions desorb from the droplet and the cycle continues, Figure 24. Once the gas-phase analyte ions are formed they are sent to the mass analyser, which is under vacuum, to be analysed, Figure 25.

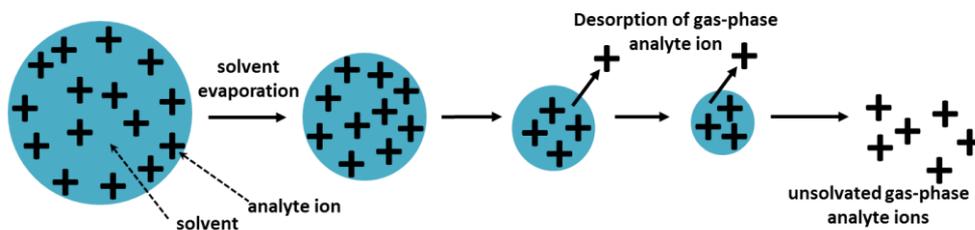


Figure 24: Ion evaporation model²⁰⁰ for gas-phase ion formation in ESI

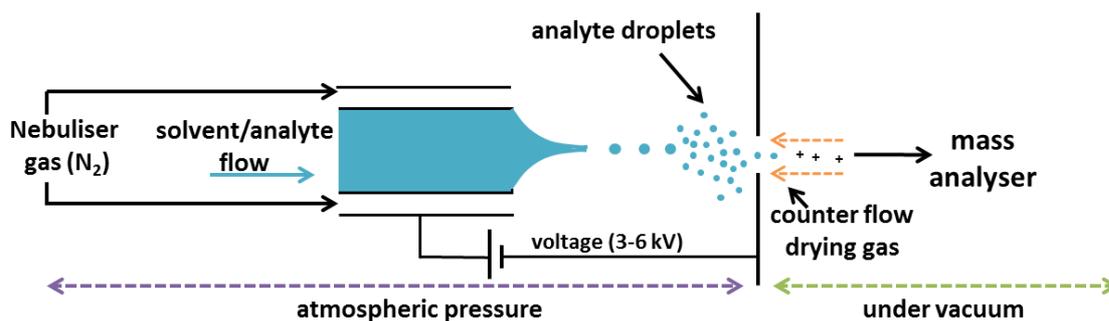


Figure 25: Schematic of ESI source.

ESI, when used in positive ion mode produces either protonated or cationised analyte molecules, and for large biological molecules such as proteins and peptides where there are multiple sites for charging, ESI can produce multiply charged ions, making detection of larger molecules possible on regular m/z range mass spectrometers due to the resulting lower m/z values compared to those ionisation techniques where only singly charged species are formed resulting in much higher m/z values which may be difficult to detect.

2.7. Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer

FT-ICR works on the principle that the trajectory of an accelerated ion can be controlled by a magnetic field^{201,202}. There are four main components to ICR instruments. Firstly, a superconducting magnet is used in modern instruments. Performance of the instrument improves on increasing the magnetic field strength. Secondly, there is an analyser cell where ions are trapped and stored, mass analysed and detected. The most common analyser cell is the cubic cell. This is composed of six plates consisting of two trapping plates, two excitation plates and two detection plates, Figure 26. As well as this, an ultra-high vacuum system is required to achieve high resolution by avoiding collisions with background gas molecules. FT instruments also have a complex data system to be able to acquire, process and analyse the data.

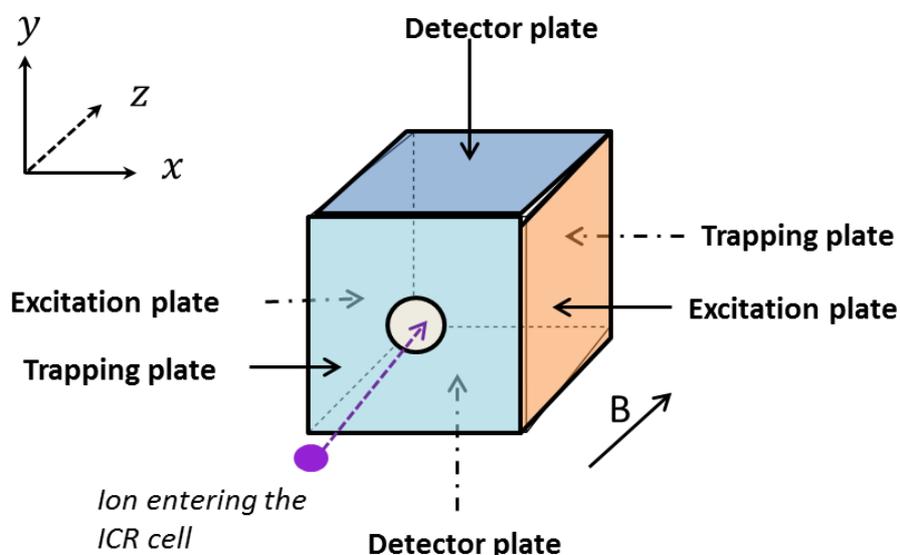


Figure 26: A cubic FT-ICR analyser cell. The cell consists of six plates made up of two trapping plates, which sit perpendicular to the magnetic field, two excitation plates and two detection plates, which sit parallel to the magnetic field.

2.7.1. Ion motion

The motion of an ion in the analyser cell is a result of three combined oscillations (cyclotron, trapping and magnetron motion), governed by electric and magnetic fields. The cyclotron motion arises from the interaction of an ion with the magnetic field. An ion moving in the presence of a uniform magnetic field experiences a force, the Lorentz force F , Equation 3, (where q and v are the charge and velocity of the ion, B is the magnetic field, and x shows the direction of the magnetic component is perpendicular to the plane determined by v and the magnetic field, (B) causing the ion to travel in a circular orbit, Figure 27. The magnetic field is constant and the m/z ratio of an ion is determined by measuring the cyclotron frequency (f_c), the frequency with which an ion repeats an orbit, Equation 4, where m is the mass of the ion and q is the charge. The cyclotron frequency is independent of the velocity and kinetic energy of the ion, which is why FT instruments can achieve such high resolution; in some mass spectrometers the resolution is limited by the spread of an ion's kinetic energy distribution, but this does not affect measurement of f_c and thus m/z in an ICR instrument. The radius of the cyclotron orbit is affected by the kinetic energy and velocity; on increasing the kinetic energy of the ions or their velocity, the radius of the orbit increases.

$$F = qv \times B$$

Equation 3

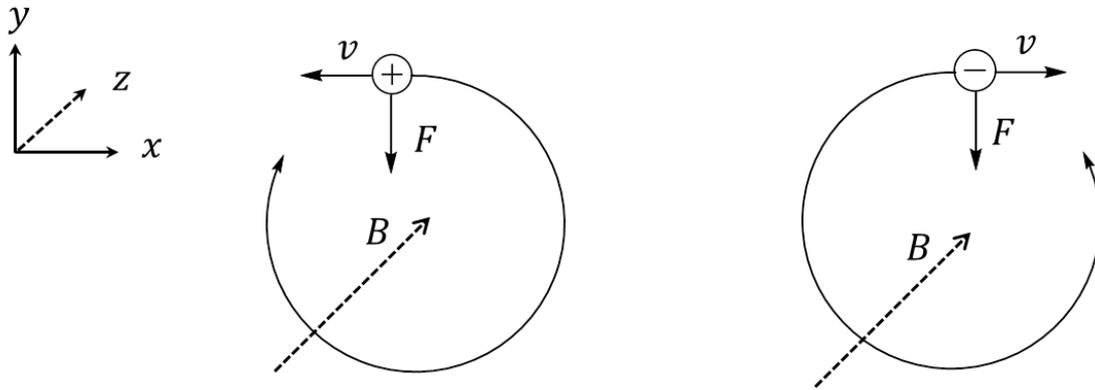


Figure 27: Ion cyclotron motion. The direction of travel is the opposite for positively or negatively charged ions²⁰¹.

$$f_c = \frac{qB}{2\pi m}$$

Equation 4

An ion moving parallel to the magnetic field experiences no force from that field. In the cubic analyser cell, two trapping plates are perpendicular to the magnetic field, creating a potential well that traps ions inside. A positive or negative voltage is applied to the plates to trap positive or negative ions respectively. Ions undergo harmonic oscillation between the trapping plates along the axis of the magnetic field, Figure 28.

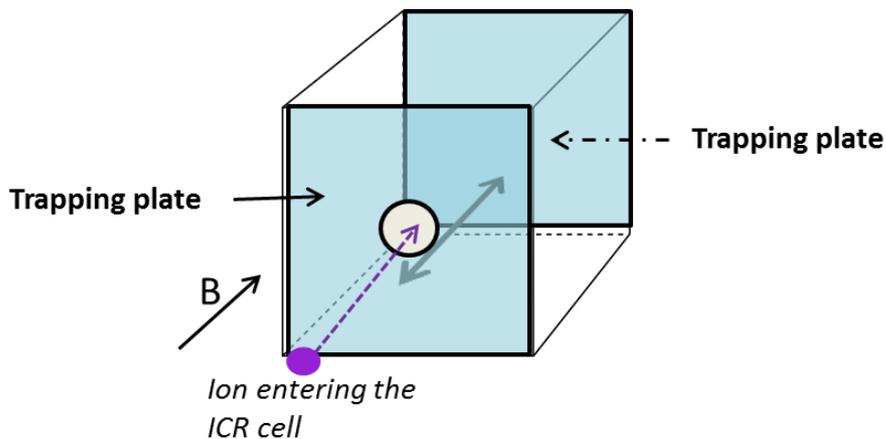


Figure 28: Ion motion as a result of trapping potential application in ICR cell. Ions oscillate back and forth, parallel to the magnetic field (B), between the trapping plates.

Magnetron motion arises from a combination of the magnetic and electric fields. The magnetic field keeps ion motion (cyclotron motion) in the xy -plane perpendicular to the z -axis. The electric trapping potential constrains the ion motion along the z -axis. The electric potential is at a maximum at the trapping plates and at a minimum in the centre of the cell, trapping the ions to the centre of the cell. The trapping potential provides a repulsive force that keeps the ions along the z -axis and repels them in the xy -plane and so the electric field attempts to move ions away from the centre of the cell. The magnetic field stops ions hitting the cell walls, resulting in the magnetron motion, Figure 29.

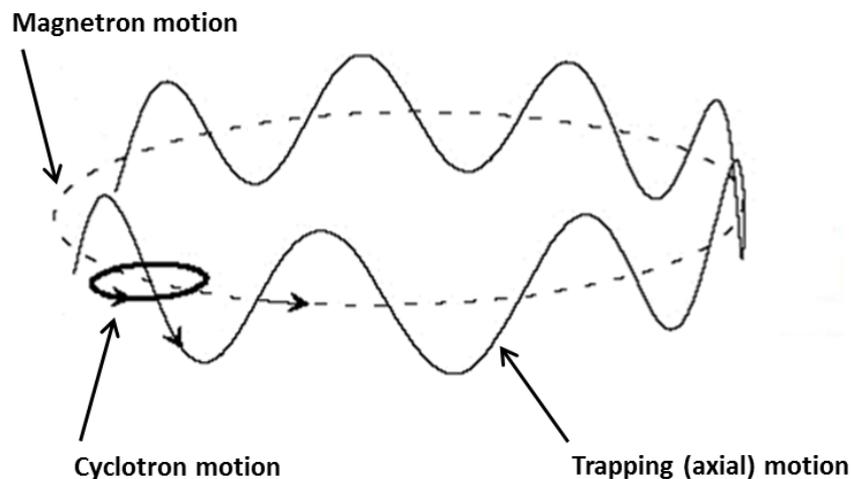


Figure 29: The motion of an ion in the analyser cell has three oscillation components: cyclotron motion, trapping (axial) motion and magnetron motion, which results from a combination of the magnetic and electric fields.

2.7.2. Excitation and detection

When ions enter the cell, no signal is detectable at the detection plates as the ions oscillate in small circular orbits. To detect ions, they are excited by applying a sinusoidal voltage to the two opposing excitation plates, which sit parallel with the magnetic field. When the frequency of the applied electric field is in resonance with the cyclotron frequency of the ions, energy is transferred to the ions, which increase their kinetic energy and so the radius of the orbit increases (Figure 30). All ions of the same m/z ratio are excited coherently, in tight packets, to the same energy, and rotate with the same frequency in the same orbit. As the ions pass close to one detector plate an “image-current” is induced in that plate by attraction of electrons to the ion packet in the positive ion detection mode, or by repulsion

of electrons in the negative ion detection mode. As ions move away from one detector plate and towards the opposite plate, the electrons in that plate are attracted towards the ion packet. The cyclotron motion of the ions thus produces a sinusoidal image signal, the frequency of which is equal to the difference between the cyclotron and the much smaller magnetron frequencies. Image-current detection is non-destructive as the ions remain in the cell during and after detection.

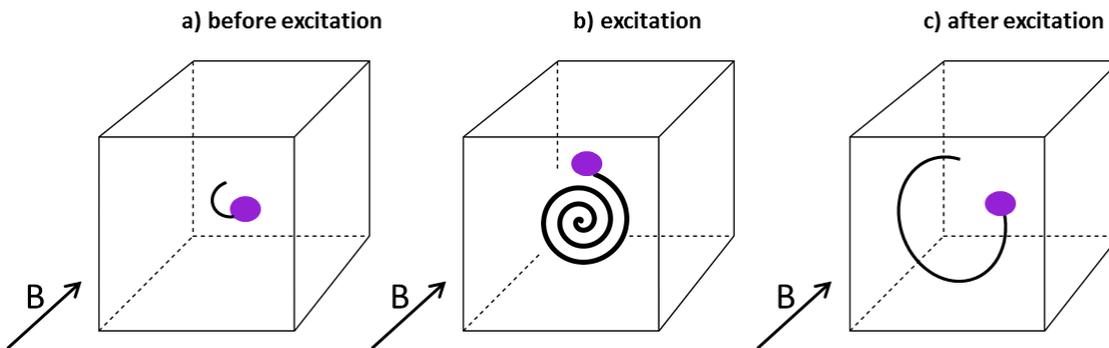


Figure 30: Excitation and detection of an ion in the cell of an ICR-MS: a) ions before excitation have a small cyclotron radius and are therefore not detected. b) During excitation, ions whose cyclotron frequency is in resonance with the applied electric field absorb energy and their cyclotron radius increases. c) When the applied electric field is turned off, the ions continue to orbit in the larger radius until a different electric field is applied, when they relax back.

Ions with a large range of different m/z ratios can be detected simultaneously by applying a sweep of a range of frequencies, causing all ions with cyclotron frequencies in the chosen range to be excited and detected as a collection of sinusoidal image currents of different frequencies and amplitudes. A complex wave is detected as a time-dependent function which, through Fourier transformation, is converted to a frequency-dependent function and hence into an intensity to m/z relationship (the mass spectrum), Figure 31.

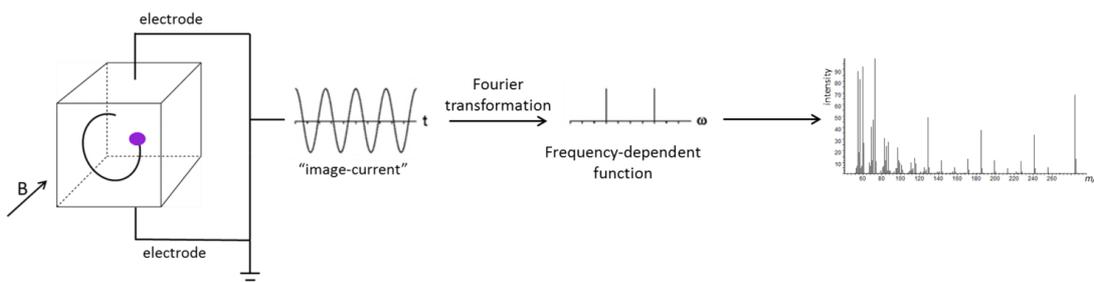


Figure 31: Fourier transform-ion cyclotron resonance (FT-ICR)

2.8. Time of flight mass spectrometry

Time of flight (ToF) mass analysers were first described in the 1940's^{203,204}, with the first commercialised instrument developed in 1955 by Wiley and McLaren²⁰⁵. ToFs are now amongst the most widely used mass analysers for the analysis of biological molecules. The first application to amino acids, arginine and cysteine, was described in 1974 by McFarlane et al, using plasma desorption/ionisation. Later application were seen for ToFs using LDI²⁰⁶, followed by MALDI¹⁷⁹. Since ToFs distinguish between different m/z values based on their flight times, through a field-free flight tube, due to different velocities, ToFs are well suited to being combined with pulsed ionisation techniques such as MALDI, as ions produced are expelled in packets. Continuous ion sources such as ESI are also used with ToF²⁰⁷ however, since the ions are produced continuously there needs to be a pulsed extraction applied to produce packets of ions in order to maintain good mass resolution.

Ions produced in the source are accelerated by an electric field, applied between an electrode and the extraction grid, towards a field-free flight tube, where they are separated according to their velocities, dependent on their m/z values. m/z values are determined by measuring the time it takes for ions to travel through the field-free flight tube between the source and the detector.

For an ion of mass (m) and charge ($q = ze$) leaving the source, accelerated by a potential difference (V_s), its electric potential energy (E_{el}) is converted to kinetic energy (E_k), given by Equation 5, and so on rearrangement, the ion's velocity can be given by Equation 6. After acceleration, an ion travels at constant velocity through the flight tube towards the detector. The time taken for an ion to travel the length of the flight tube (L) is given by Equation 7. Combining Equation 6 and Equation 7 gives Equation 8, which shows that the m/z ratio is related to t^2 . The length of the flight tube and the accelerating potential are constant for a given instrument and measurement.

$$E_k = \frac{mv^2}{2} = qV_s = zeV_s = E_{el}$$

Equation 5

$$v = \left(\frac{2zeV_s}{m} \right)^{\frac{1}{2}}$$

Equation 6

$$t = \frac{L}{v}$$

Equation 7

$$t^2 = \frac{m}{z} \left(\frac{L^2}{2eV_s} \right) \quad \text{or} \quad m/z = \left(\frac{2eV_s}{L^2} \right) t^2$$

Equation 8

Linear ToF mass spectrometers have poor mass resolution, affected by a distribution in flight times amongst ions with the same m/z ratio (Figure 32a). The difference in flight times is caused by a number of factors²⁰⁸. One is that there is a slight time difference in ion formation in the MALDI plume and so ions of the same mass, with the same initial kinetic energy, reach the detector at slightly different times. Secondly, the ions can form in different locations in the electric field and so experience slightly different accelerations, which affects the time it takes for ions of the same mass to reach the detector. Additionally, ions can move in different directions in the MALDI plume and so there is a time lag for the ions moving away from the flight tube as they take time to slow down, stop and then be accelerated in the correct direction to reach the detector. Furthermore, differences in the initial kinetic energy transmitted to the ions from the acceleration pulse, as well as energy randomisation through collisions in the plume also result in variation of ions' flight times. The reflectron and delayed pulse extraction are two techniques that have been developed to improve mass resolution by each mitigating some of these effects.

A reflectron, developed by Mamyrin²⁰⁹, is an electrostatic reflector, comprised of ring electrodes, that can correct the distribution of kinetic energy of ions with the same m/z ratio. The reflectron is positioned at the far end of the flight tube at an angle so that ions entering do not follow the same path on exiting the reflectron. The reflectron creates a decelerating electric field causing the ions to slow down upon entering and then reverses the direction of the ions, sending them back through the flight tube towards the detector, which is placed on the source side. For ions of the same m/z ratio but different kinetic

energies, those ions with more kinetic energy travel further into the reflectron than those with lower kinetic energy, meaning they stay in the reflectron for longer before being turned around to reach the detector. This improves mass resolution by allowing ions with the same m/z ratio to reach the detector simultaneously, Figure 32b. However, being able to achieve this is dependent on the values for the electrical field, the voltage applied, and the flight path length. A reflectron cannot correct for turnaround time, or for the spatial and time of formation distributions. For this, delayed pulse extraction is required.

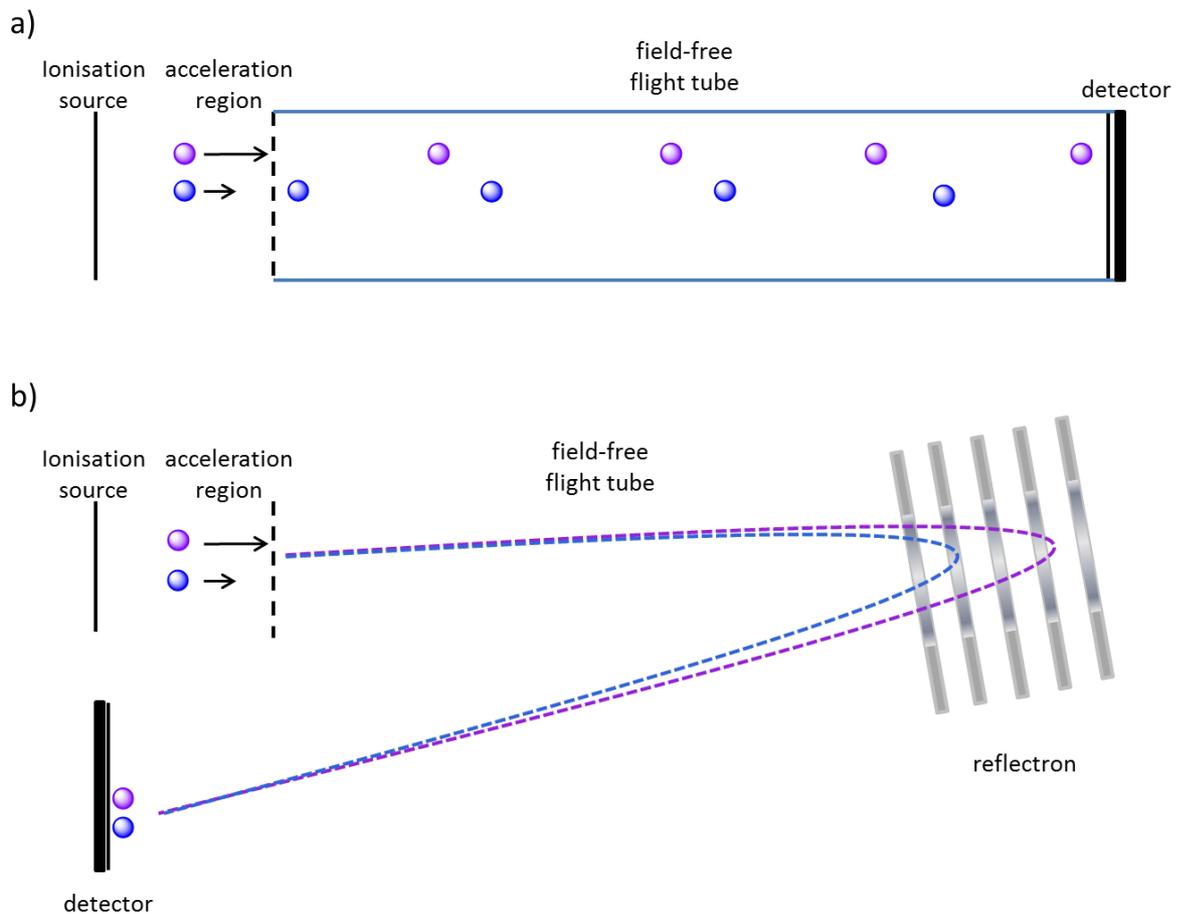
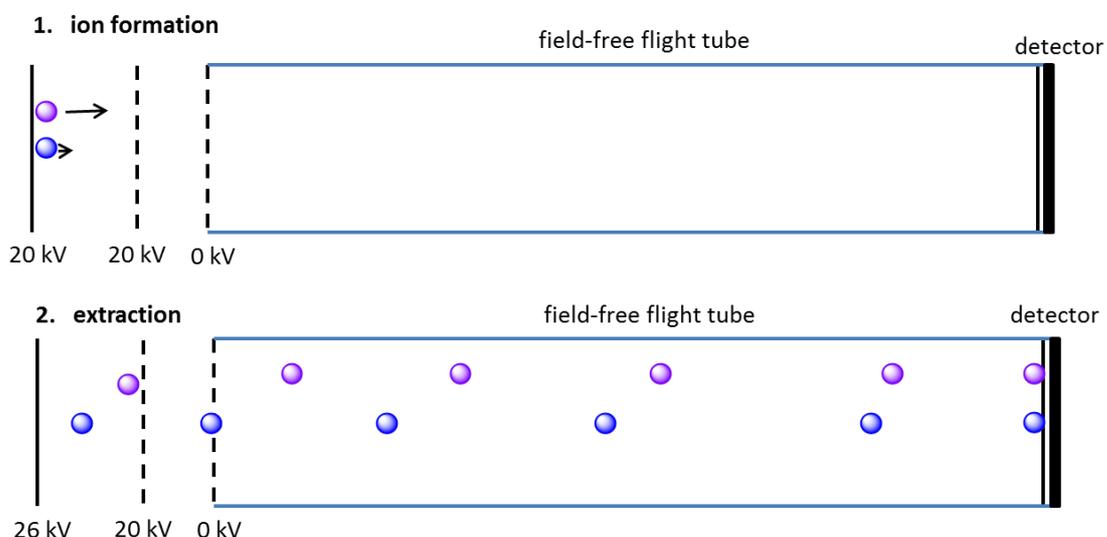


Figure 32: Time of flight mass spectrometer. a) Linear. Poor mass resolution as ions with the same m/z ratio but different kinetic energies reach the detector at different times. b) Reflectron. Improved mass resolution as ions with the same m/z ratio but different kinetic energies reach the detector simultaneously.

Delayed pulse extraction, first described by Wiley and McLaren²⁰⁵ as time-lag focussing, helps to correct for the spatial, directional and time of formation distribution by introducing a time delay between ion formation and extraction. Their ToF contained a dual-stage source

by using two extraction regions. Introducing a delayed extraction pulse gives the ions time to separate in the source and acceleration region according to their initial velocities. For those ions with the same m/z ratio but different initial velocities, the ones with higher velocity will travel further (but does not necessarily mean those ions will travel closer to the detector as they may travel on different trajectories). After a short time delay, allowing the ions to rearrange themselves, the extraction pulse is applied. The pulse transmits more energy to those ions which had less energy to begin with as they will have moved a shorter distance towards the detector and so feel a bigger pull when the pulse is applied, in addition the delayed pulse can also work to correct the trajectories of those ions not travelling directly towards the detector. These ions can therefore join the ions that had more energy to begin with and reach the detector at the same time (Figure 33a). This offers an improvement to mass resolution when compared with using a continuous extraction, where ions are immediately extracted upon formation by a continuous extraction voltage. This results in ions of the same m/z ratio but having had slight differences in time of formation, direction and speed of travel and reaching the detector at different times (Figure 33b).

a) Delayed pulse extraction



b) Continuous extraction

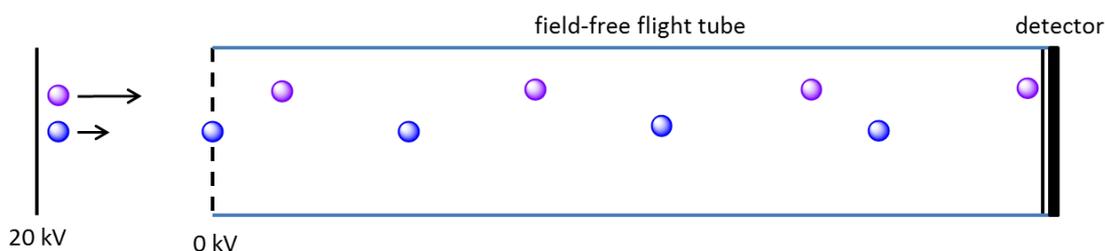


Figure 33: a) delayed pulse extraction, using dual stage extraction. Ions are formed and allowed time to separate according to their initial velocities. For ions with the same m/z ratio but different velocities, ones with a higher velocity travel further towards the detector. On application of the delayed pulse more energy is supplied to those ions which had less energy meaning they join the ions that had more energy and reach the detector simultaneously with appropriate choice of voltage and delay time. b) In continuous extraction, ions with the same m/z ratio but different velocities reach the detector at different times as ions, upon their formation, are immediately extracted.

2.9. Orbitrap

The Orbitrap mass analyser, invented by Alexander Makarov^{210,211,212} can be considered a modification of the the Kingdon trap, and the Paul trap (quadrupole ion trap.) Essentially the Orbitrap works by trapping ions orbitally and m/z values are measured from the

frequency with which ions oscillate along an electric field. The frequency is measured by acquiring an image current in the time domain (similarly to FT-ICR) and a mass spectrum is obtained after subsequent Fourier transformation.

Analyte ions are transported from the ion source via multipole ion guides and they are injected into a C-trap, a curved RF quadrupole ion trap, where ions are collisionally cooled as they experience a low level of collisions with gas molecules causing them to lose energy, slow down and collect into a packet of ions. When enough ions are accumulated, a strong electric field is applied to the C-trap rapidly ejecting the ions out of storage and injecting the ions, at a highly precise trajectory, into the Orbitrap mass analyser, Figure 34. Ions oscillate across the trap at frequencies (ω) related to their m/z ratios, Equation 9:

$$\omega = \left(\frac{Z}{m}\right)^{\frac{1}{2}}$$

Equation 9

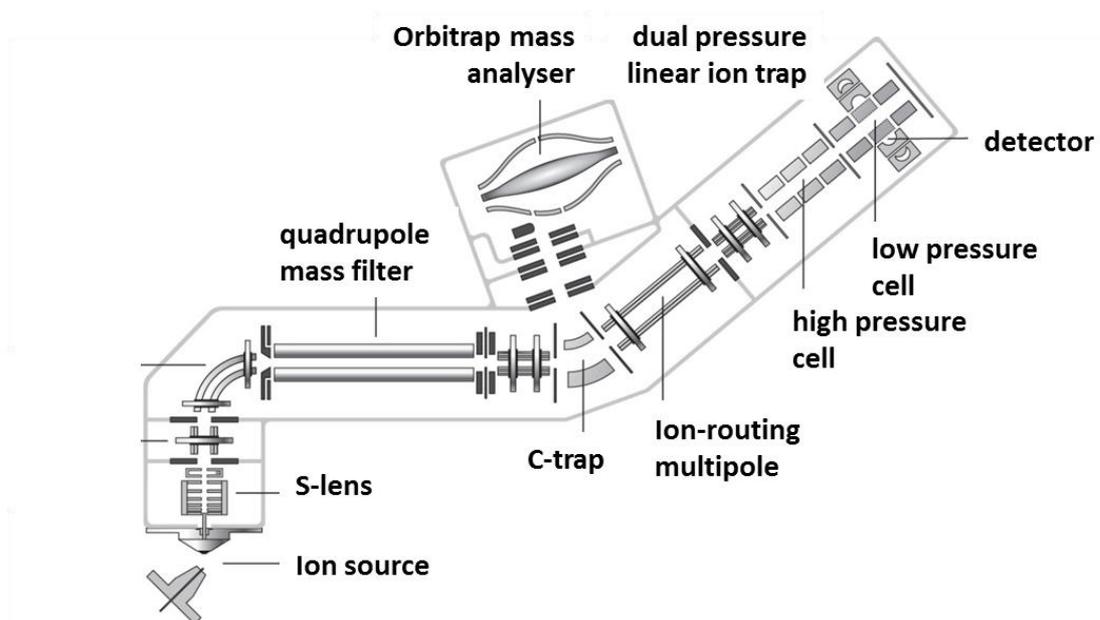


Figure 34: Orbitrap Fusion hybrid Orbitrap instrument schematic. Reproduced from²¹³

The orbitrap consists of a central spindle-like electrode and an outer barrel electrode. The outer electrode is split in the middle, $z=0$, like the 'knight' Kingdon trap, to allow collection of an image current, Figure 35. The image current is amplified from each half of the outer electrode before an analog to digital conversion for processing. Ions are injected into the

Orbitrap off set from $z=0$. Ions rotate around the central electrode and oscillate along the z -axis of the trap. Ions of the same m/z move along the z -axis coherently (with the same frequency).

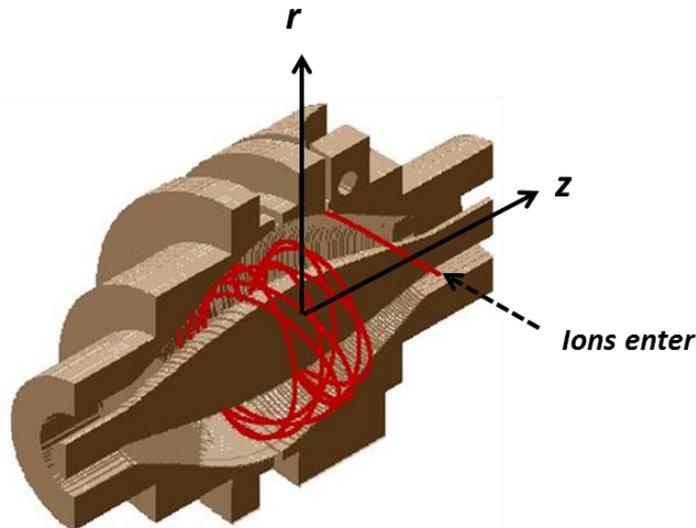


Figure 35: Orbitrap, to show how the outer electrodes are split in half along the z axis ($z=0$). Taken from Hu et al²¹¹.

As the ions enter the Orbitrap in their respective packets, of each m/z , there is a short spread of m/z -dependent flight times (around a few hundred nano seconds) although, since the pulse into the Orbitrap is fast, this time spread is much shorter than the time it takes to complete an oscillation half the length of the trap, meaning coherent axial oscillation occurs without needing excitation. As ions approach the opposite electrode, the increasing electric field contracts the radius of the ion packet and decreases the trajectory close to the z -axis of the trap, giving time for further ion packets to arrive, **Figure 36a,b**. Once all ions have entered the trap, the voltage of the central electrode stabilises, stabilising the ion trajectories. Frequencies of radial or rotational motion vary for ions with different initial ion kinetic energy, velocity and radius of rotation, as the frequency is dependent on initial ion energy, velocity and radius of rotation when the ions enter the trap and so the ions in these directions dephase much more quickly than the axial motion. The axial motion is independent of all initial ion parameters therefore ions with the same m/z oscillate along the z -axis coherently for hundreds of oscillations. The dephasing causes the ion packet to form a ring made up of uniformly distributed ions of each m/z , **Figure 36c**. On detecting image currents, ions opposite each other on the ring give opposite image currents and therefore cancel out the signal meaning radial rotational motion is not detected. Image currents that are detected therefore come from the frequency of the z -axial motion. The

image current is acquired in the time domain before it is converted to the frequency domain and into a mass spectrum to measure m/z .

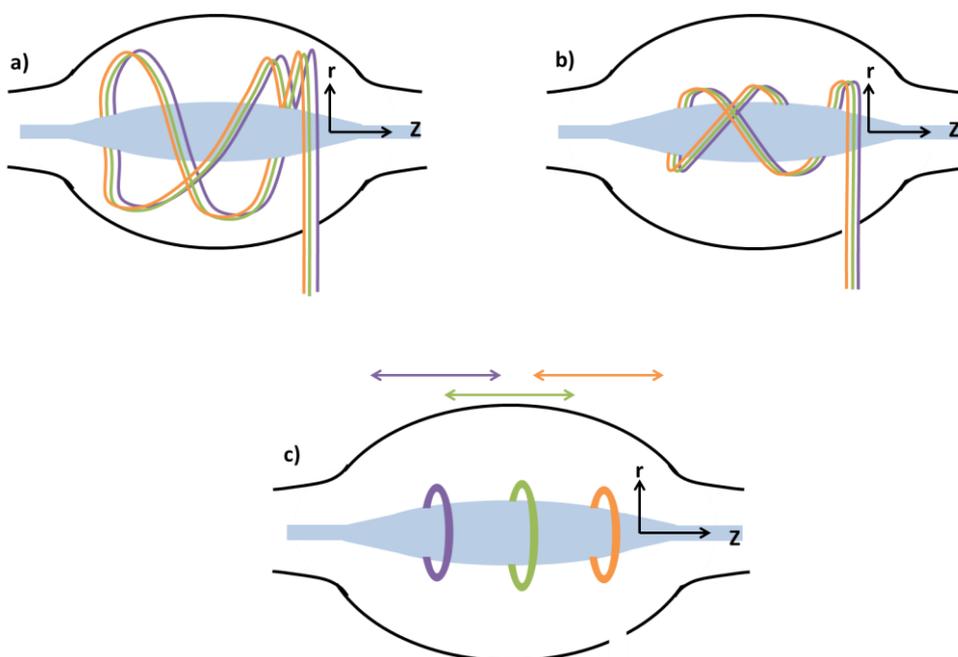


Figure 36: Movement of analyte ions in an Orbitrap mass analyser. a) and b) as ions approach the opposite electrode, the electric field contracts the radius of the ion packet and decreases the trajectory close to the z-axis of the trap, giving time for further ion packets to arrive. c) Once all ions have entered the trap, the voltage of the central electrode stabilises, stabilising the ion trajectories. Frequencies of radial or rotational motion vary, as the frequency is dependent on ion velocity and radius of rotation when the ions enter the trap and so the ions dephase much more quickly than the axial motion. The dephasing causes the ion packet to form a ring made up of uniformly distributed ions of each m/z .

2.10. Tandem mass spectrometry

To gain structural information on analyte molecules tandem mass spectrometry experiments are important, because the ionisation methods used to analyse large biological molecules are soft ionisation techniques that leaves them intact. This is helpful for measuring intact molecular mass, but the absence of fragment ions means that structural detail is not provided. To induce fragmentation, additional energy has to be applied to the analyte ions. This can be done by collision-induced dissociation, CID²¹⁴. In CID, a gas such as N₂ is

introduced between the ion source and the mass analyser. The analyte ions then collide with the gas molecules converting some of the kinetic energy of the ions into internal energy, breaking analyte molecule bonds to generate fragment ions; if these occur in a structurally diagnostic manner, they can provide structural information.

2.10.1. Fragmentation of carbohydrates

The accepted nomenclature for the mass spectrometric fragmentation of carbohydrates follows that described by Domon and Costello²¹⁵. The lowest energy fragmentations usually occur at glycosidic bonds, and so these fragmentations are most commonly observed. Fragment ions generated in this way are designated by B and Y ions. In positive ion mode, B type fragmentation occurs from single protonation of the glycosidic bond, which then breaks to give an oxonium ion from the non-reducing end and a smaller neutral glycoconjugate from the reducing terminal portion of the molecule. The Y type ion is formed from the cleavage of the glycosidic bond accompanied by proton transfer from the non-reducing part of the oligosaccharide to the reducing part, Figure 37. There may also be C and Z type fragments, which result from cleavages on the reducing side of the glycosidic oxygen. There can also be fragments across a ring, which require more energy and so are generally less abundant, denoted by A and X, Figure 38. In addition to being given labels A,B,C, X, Y or Z, the fragments generated are also numbered depending on the number of monosaccharide residues in the fragment e.g. B₂, B₃ etc, Figure 38. Furthermore, if a glycan is branched then the branches that lead off the core structure are denoted as α or β and are still then labelled in the same way as the main chain (e.g. B or Y ion and the number of monosaccharides in the fragment). If there is any further branching then labels such as α' and α'' are added, see Figure 39 for an example.

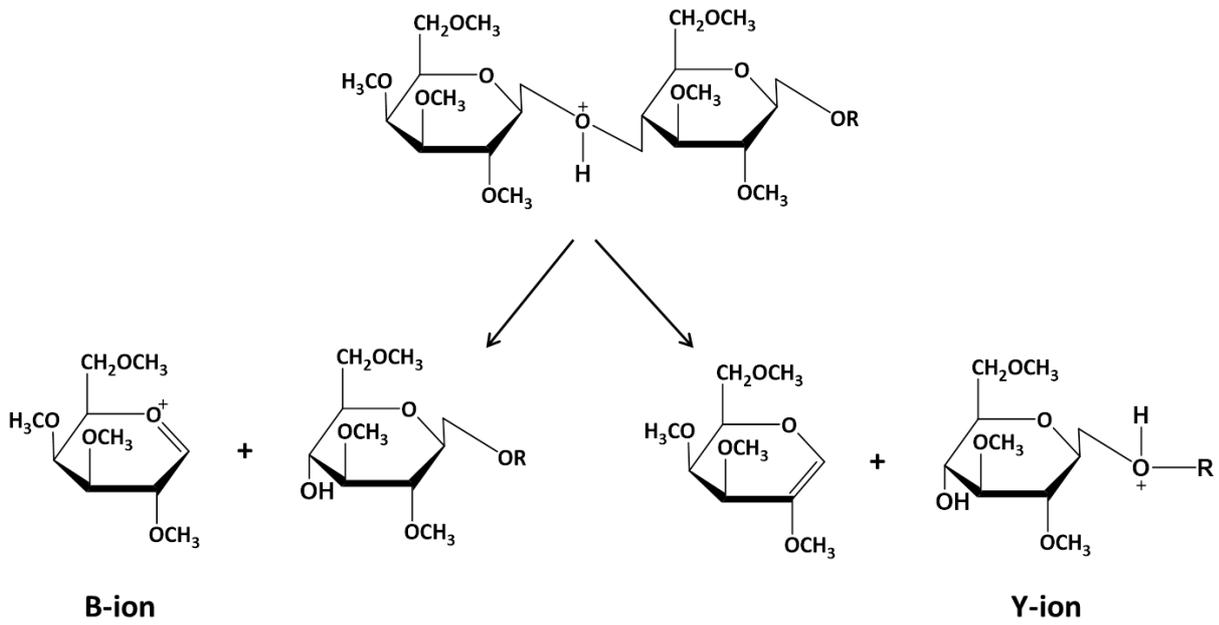


Figure 37: Fragmentation of glycans to produce B and Y ions. B ions are produced from cleavage of the protonated glycosidic bond, and Y ions are produced from cleavage of the protonated glycosidic bond followed by H-transfer²¹⁵.

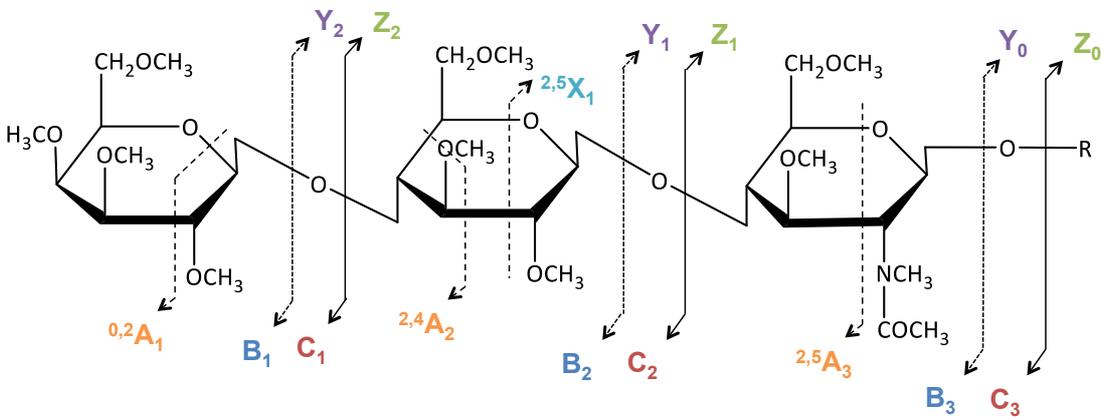


Figure 38: Fragmentation nomenclature of carbohydrates, adapted from Domon and Costello²¹⁵.

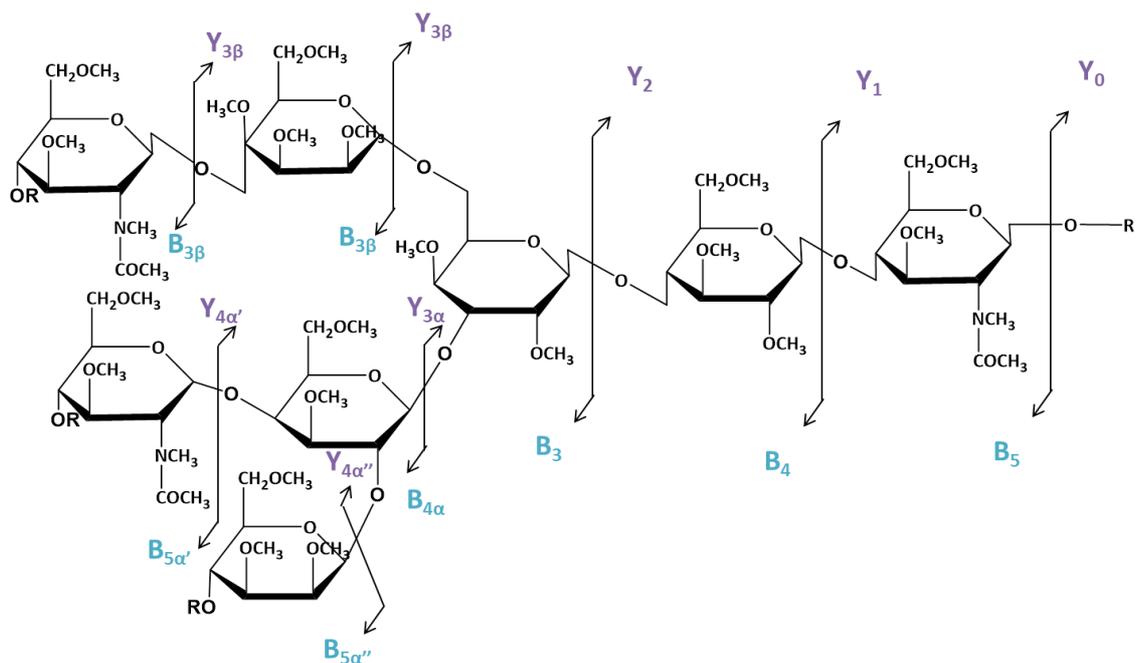


Figure 39: fragmentation and nomenclature of branched glycans by Domon and Costello²¹⁵. The branches that lead off the core structure are denoted as α or β and are then labelled in the same way as the main chain (e.g. B or Y ion and the number of monosaccharides in the fragment). If there is any further branching then labels such as α' and α'' are added.

2.10.2. Fragmentation of peptides

Peptides, like carbohydrates, fragment in a predictable manner²¹⁶. The bonds along the peptide backbone, in particular the amide bond, are most susceptible to fragmentation. Currently used nomenclature for peptide fragmentation was proposed by Roepstorff in 1984²¹⁷, refining that originally introduced by Biemann²¹⁸. When the bonds break, if the charge is on the N-terminal portion of the molecule then the fragments are denoted a-,b- or c-. If the charge is on the carboxy terminal portion then the fragments are given x-,y- or z- labels, with b- and y- ions being the most commonly observed fragments. Similarly to fragmentation of carbohydrates, the fragments produced are also numbered, according to the number of amino acid residues in the fragment, Figure 40.

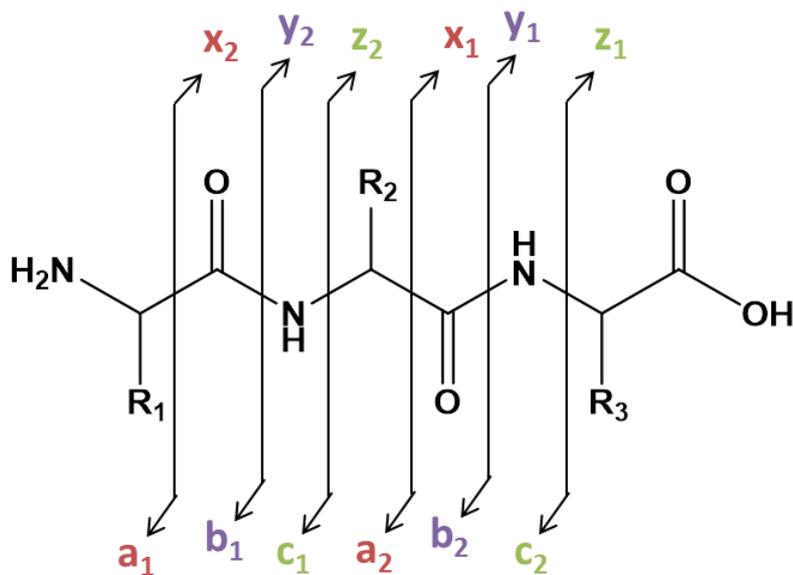


Figure 40: Peptide fragmentation and nomenclature

2.11. Coupling high performance liquid chromatography (HPLC) to mass spectrometry (LC-MS)

Coupling chromatography techniques to mass spectrometry is particularly useful when studying complex mixtures such as glycans or peptides. Coupling liquid chromatography to mass spectrometry (LC-MS) involves injecting a sample onto an LC column and separating the mixture, the components of which elute from the column and are transferred to the MS instrument, usually through an electrospray source.

In liquid chromatography, the analyte mixture is carried by a mobile phase through a column containing a stationary phase with which the analyte components interact. The stationary phase is usually made up of porous particles, generally of silica, the surface of which can be chemically modified depending on the analysis required. This is because the components in the analyte mixture need to have varying affinities for the mobile and stationary phases to allow separation, with some components being retained by the stationary phase more easily than others. Reversed phase HPLC is the most common HPLC experiment. It uses a non-polar stationary phase, such as silica derivatised with long alkyl chains, C₁₈, and a polar mobile phase (usually a mixture of water with acetonitrile or methanol). In this case, the more hydrophobic analytes are retained longer than more polar ones and therefore take longer to elute from the column. HPLC methods can use isocratic mobile phase flow in which the

composition of the mobile phase remains constant throughout the run, or a gradient can be used, in which the composition of the mobile phase is varied over the course of the run, which can give significant improvements in separation as well as speeding up analyses.

For efficient chromatographic separation, a low plate height is needed. Plate height is a theoretical concept based on the length of column needed for one equilibration event of analyte partitioning between the mobile and stationary phase. The lower the plate height, the more efficient the separation is because there are more plates and thus separation events in a column of fixed length. Plate height is affected by mobile phase flow rate through the column and the size of the particles in the stationary phase, shown by the Van Deemter equation, Equation 10:

$$H = A + \frac{B}{u} + Cu$$

Equation 10

where H is plate height, A is eddy diffusion (due to multiple flow paths), B is longitudinal diffusion, C is for mass transfer and u is the flow rate of the mobile phase.

Analytes can take multiple paths (referred to as eddy diffusion, A term in the Van Deemter equation) through a packed column, where some pathways are longer than others causing broadening of peaks and compromising separation efficiency. Eddy diffusion can be minimised by using columns packed with small particles to minimise the spread of lengths of flow paths around them. Longitudinal diffusion (B term) is affected by flow rate and is caused by analytes diffusing along the column. By increasing the flow rate, the analyte passes through the column quickly, minimising the opportunity for diffusion. However, if the flow rate is too high then broadening of peaks can also result since the mass transfer term (C), which reflects the time taken for equilibration of analyte between the mobile and stationary phases, is affected by flow rate; increasing the flow rate compromises efficient equilibration and thus impacts separation efficiency.

All the analytical techniques described in this section are amongst the fundamental analytical tools used widely in the analysis of biological molecules such as carbohydrates and peptides. The work presented in this thesis, makes use of the described mass spectrometric

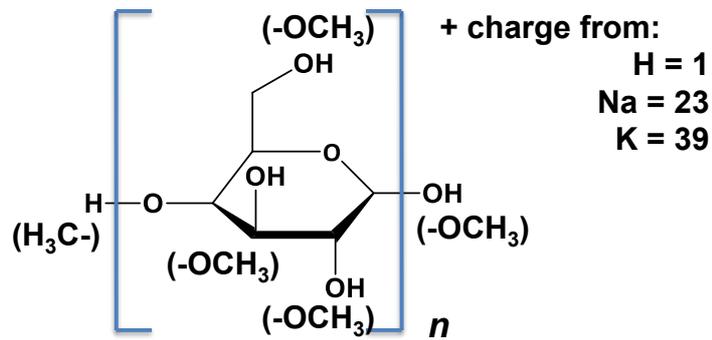
ion sources (MALDI) and mass analysers (FT-ICR and ToF) for glycan profiling and LC-MS, using ESI and Orbitrap, for the analysis of (glyco)peptides

2.12. Mass spectrometric characterisation of glycans

The nomenclature used for the representation of glycans in this thesis is presented in Table 2. The names of each of the monosaccharide residues are usually shortened. For example, *N*-acetylgalactosamine becomes GalNAc (or HexNAc for a generic composition). In mass spectra, shapes represent glycans. For example, GalNAc is represented by a yellow square. To calculate the mass of the glycans, the mass of each of the monosaccharide moieties are summed (taking into account whether the glycan is in its native state or has been permethylated), followed by addition of either H or CH₃ (native or permethylated respectively) followed by addition of 16, for oxygen, a further H or CH₃ and finally account for the charge, (e.g. protonated (+1) or sodiated (+23)), Figure 41. For example the mass of the permethylated glycan of composition NeuAc₁Gal₁GalNAc₁ that is sodiated would be 879 Da (15 + (361 + 204 + 245) + 16 +15 +23).

Table 2: Glycan nomenclature for characterisation by mass spectrometry, including the symbols and masses for native and permethylated structures.

Monosaccharide		Symbol	Native mass (Da)	Permethylated mass (Da)
<i>N</i> -acetylglucosamine (<i>N</i> -acetylhexosamine)	GlcNAc (HexNAc)		203	245
<i>N</i> -acetylgalactosamine (<i>N</i> -acetylhexosamine)	GalNAc (HexNAc)		203	245
Galactose (hexose)	Gal (Hex)		162	204
Mannose (hexose)	Man (Hex)		162	204
Glucose (hexose)	Glc (Hex)		162	204
Fucose	Fuc		146	174
<i>N</i> -acetylneuraminic acid	NeuAc		291	361
<i>N</i> -glycolylneuraminic acid	NeuGc		307	375



$$\text{native} = 1 + [\text{monosaccharide units}]_n + 16 + 1 + 1$$

$$\text{Permethylated} = 15 + [\text{monosaccharide units}]_n + 16 + 15 + 23$$

Figure 41: Calculation of the mass of a glycan in both native and permethylated forms.

Chapter 3: Experimental Procedures

3. Experimental procedures

3.1. Cell culture: Mesenchymal stromal cells

3.1.1. Basal media

Cells were cultured in basal media composed of Dulbecco's Modified Eagle Media (high glucose, pyruvate, no glutamine) supplemented with 10 % fetal bovine serum, 1% Penicillin/streptomycin and 1 % Gluta-Max-I.

3.1.2. htert MSC culture

htert MSCs are immortalized cells created from primary human mesenchymal stromal cells. Cell lines Y201 and Y101 were used for this research. Cells were grown in a humidified incubator at 37 °C and 5 % CO₂. Cells were grown in tissue culture flasks or 10 cm petri dishes. Before cells reached full confluence, cells were split 1/6 or as required, by incubation with trypsin/EDTA for 10 min or until detached.

3.1.3. Adipogenesis

In order to induce adipogenesis cells were cultured in basal media described above that also contained 1 µM dexamethasone, 500 µM 3-isobutyl-1-methylxanthine, 1 µg/ml insulin and 100 µM indomethacin. During differentiation (21 days), the adipogenic media was changed every 3-4 days.

3.2. Histochemistry: Oil red O staining

Culture media was removed and cells were washed with 1x PBS. The cells were then fixed with 4 % paraformaldehyde for 10 min and then washed with deionised water. Cells were then incubated for 5 min in 60 % isopropanol, before removing and incubating for 10 minutes in 0.3% Oil red O solution. The 0.3% Oil red O solution was made by diluting a stock solution of 0.5% Oil red O (in isopropanol) with deionised water (3:2, v:v), this was left overnight for precipitate to settle as well as being filtered with a 0.45 µm syringe filter. Cells were then washed once with 60% isopropanol and three times with deionised water. 100 µl of deionised water was added to keep the cells moist and were imaged using bright field microscopy.

3.3. Extraction and collection of glycoproteins from CHO and MSC cells.

A 10 cm dish of confluent CHO or MSC cells was washed gently with warm phosphate buffered saline (PBS) six times. Cells were then scraped with a cell scraper and with 1 mL PBS into a plastic microfuge tube. The tube was centrifuged at 16000 g, 5 min, 4 °C. The supernatant was discarded and lysis buffer (4% SDS, 100 mM Tris pH 7.6, 100 mM DTT) was added at approximately 10x volume of the cell pellet. The solution was pipetted to mix and to break up the cell pellet. The sample was heated at 95 °C for 5 min and then centrifuged for 5 min; the pellet was removed and the supernatant was stored at -80°C if not needed immediately, otherwise, the steps in section 3.5 were followed.

3.4. Extraction, collection of glycoproteins from porcine bladder urothelial tissue

Three bladders were obtained from an abattoir near York and processed the morning they were collected. The bladders were dissected using forceps and washed with Hank's balanced salt solution (HBSS) , containing calcium and magnesium, supplemented with HEPES buffer solution 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid(5 mL in 500 mL HBSS), 20 KIU aprotinin (1 mL in 500 mL), Fungizone (2 mL in 500 mL to give a final concentration of 1 µg/mL) and Pen-Strep (5 mL in 500 mL to give a final concentration of 100 µg/mL).

Urothelium was removed from dissected bladders by surface scraping using a scalpel. The urothelial scrapings were then washed in HBSS medium and centrifuged. The medium was removed and the cell pellet was re-suspended in 750 µL HBSS medium. Aliquots of ~50 µL of cell suspensions were transferred into microfuge tubes (each aliquot around 50 mg), centrifuged and the supernatant was removed and samples were stored at -80 °C until needed.

3.5. Solubilisation of standard glycoproteins in SDS (using FASP).

20 pmol of porcine stomach mucin (Sigma Aldrich, UK) was solubilised in 10 µL lysis buffer (4% SDS, 100 mM Tris pH 7.6, 100 mM DTT). The sample was heated at 95 °C for 5 min. The solubilised glycoprotein sample was diluted with exchange buffer (8 M urea in 100 mM Tris/HCl, pH 8.5) in a ratio of 10:1 exchange buffer to sample by volume. A 450 µL aliquot of the diluted solution was transferred to a centrifugal filter unit (Amicon® Ultra-0.5 centrifugal filter unit, nominal mass cut off 30 kDa). The filter unit was centrifuged for 5 min and the process repeated until the entire sample had passed through the filter unit. The sample retained above the filter membrane was rinsed three times with 300 µL of exchange buffer

(centrifuging after each rinse for 5 min), and the filtrate was discarded. The filter was then rinsed three times with 300 μ L 50 mM ammonium bicarbonate (centrifuging after each rinse for 5 min), and the filtrate was discarded. The filter unit was transferred to a clean collection tube ready for O-glycan release.

3.6. Solubilisation of standard glycoproteins and urothelial cell samples, and preparation for O-glycan release, using an adapted enhanced filter aided sample preparation (eFASP)¹⁷⁷

20 pmol of porcine stomach mucin (Sigma Aldrich, UK) was solubilised in 10 μ L eFASP lysis buffer¹⁷⁷ (4 % SDS, 0.2 % deoxycholic acid (DCA), 50 mM Tris(2-carboxyethyl) phosphine hydrochloride (TCEP), 100 mM ammonium bicarbonate, pH 8). The sample was heated at 90 °C for 10 min. The sample was vortexed to mix and cooled to 37 °C before adding the alkylation stock solution (500 mM 4-vinylpyridine in ethanol) to give a final concentration of 25 mM 4-vinylpyridine. The sample was incubated at 37 °C, 15 min. Quench buffer (1 M dithiothreitol, 100 mM ammonium bicarbonate, pH 8) was added to give a final concentration of 40 mM DTT. The glycoprotein sample was diluted with exchange buffer (8 M urea, 100 mM ammonium bicarbonate, with or without 0.2 % DCA) solution in either 10:1 or 30:1 ratio of exchange buffer to sample solution by volume. An aliquot of up to 450 μ L of the resulting solution was transferred to a filter unit (Amicon® Ultra-0.5 centrifugal filter unit, nominal mass cut off 30 kDa), centrifuged for 5 min, and the filtrate discarded. This was repeated until all the sample solution had passed through the filter membrane. The sample retained above the filter membrane was rinsed three times with 300 μ L of exchange buffer (centrifuging after each rinse for 5 min), and the filtrate was discarded. The filter was then rinsed three times with 50 mM ammonium bicarbonate (centrifuging after each rinse for 5 min), and the filtrate was discarded. The filter unit was transferred to a clean collection tube ready for N- or O-glycan release.

3.7. Normal human urothelial cells

Cell culture and Trans-epithelial electrical resistance (TER) measurements carried out by Ros Duke. Cell harvests carried out by Kirsty Skeene.

Three cell lines, Y1419, Y1860 and Y1862 were grown to full confluency on a 6 cm dish with keratinocyte serum-free medium. At confluence, undifferentiated cells were harvested by

removing the medium, washing 3 x with 1 mL PBS before scraping into a microfuge tube with 0.5 mL PBS. For the differentiated cell samples, each of the cell lines were grown to full confluency on a 6 cm dish in keratinocyte serum-free medium supplemented with 5 % bovine serum. At confluence, cells were seeded onto Griener membranes (12 well thincert 0.4um cat no#665641), and maintained in the same medium. Cells were seeded at 5×10^5 per membrane in 500 μ l volume and 1 mL culture medium placed into outer well of the Griener membrane. Differentiation was induced by increasing the exogenous calcium concentration (CaCl_2) to 2 mM.

TER measurements carried out on cells in the Greiner membranes for 8 days to monitor formation of a tight barrier/differentiation. Electrodes were rinsed three times with 0.15 M KCl. The prodes were placed above and below the membrane and the TER reading taken.

After the 8 days, differentiated cells were harvested by removing the medium, rinsing 3 times with 0.5 mL PBS and then scraping the cells into a microfuge tube with 0.5 mL PBS. Cells were lysed and prepared for glycan release by following the procedure outline in section 2.6.

3.8. Non-reductive β -elimination – using overnight incubation

300 μ L NH_4OH solution, (28-30%), (Sigma-Aldrich, UK) was added to the filter device containing the glycoprotein sample remaining in the filter unit after the solubilisation and washing steps. The glycoprotein sample was incubated in NH_4OH , at 45°C for 16 h in a heating block. Following incubation, the filter device was centrifuged for 5 min at 14000 x *g*. 150 μ L water was added to the filter device and then centrifuged for a further 5 min. This was repeated once. The filtrates containing the released O-glycans were all collected in the same tube and were transferred to a glass tube and dried using a vacuum centrifuge.

3.8.1. Non-reductive β -elimination – using sonication

300 μ L NH_4OH solution, (28-30%), (Sigma-Aldrich, UK) was added to the filter device containing the glycoprotein sample remaining in the filter unit after the solubilisation and washing steps. The sample was sonicated at 45 °C for 5 min then left to incubate at 45 °C in the water bath for 10 min, followed by another 5 min of sonication. The sonication/incubation was repeated giving a total sonication of 20 min and total incubation

time of 40 min. The filter device was centrifuged for 5 min at 14000 x *g*. 150 µL water was added to the filter device and then centrifuged for a further 5 min. This was repeated once. The filtrate was then transferred to a glass tube and dried using a vacuum centrifuge.

3.8.2. PNGase F treatment – *N*-glycan release

100 µL of 50 mM ammonium bicarbonate solution pH 8 was added to the filter unit, followed by 8 U (4 µL of a 2 U/µL solution in 5 mM potassium phosphate, pH 7.5) of PNGase F. The sample was incubated at 37 °C for 16 h in a heating block. Following incubation, the filter device was centrifuged for 5 min at 14000 x *g*. 150 µL water was added to the filter device and then centrifuged for a further 5 min. This was repeated once. The filtrate containing the released *N*-glycans were all collected in the same tube and was transferred to a glass tube and dried using a vacuum centrifuge.

3.9. Per-O-methylation

The dried released O-glycans were re-dissolved in approx. 1 mL DMSO (Sigma Aldrich, UK). Approximately 20 mg of NaOH powder was added, immediately followed by 150 µL, dropwise, of iodomethane (Sigma Aldrich, UK) and the solution was left to stand for 10 min. A further 150 µL, dropwise, of iodomethane was added and the solution was left to stand for 10 min. A final 300 µL, dropwise, of iodomethane were added to the sample solution and left to stand for 20 min. The reaction was quenched with 1 mL of 100 mg/mL sodium thiosulfate solution, followed by 1 mL dichloromethane (DCM) and the sample solution was mixed using a vortex mixer. The solution was left to stand until the organic/aqueous layers had separated and then the top aqueous layer was removed. The organic layer was washed with about 1 mL water a total of five times, finally removing the top aqueous layer and drying the organic layer in a vacuum centrifuge.

The internal standard, maltotetraose, was deuterio-permethylated using iodomethane- d_3 following the same procedure outlined for permethylation above. A stock solution of 400 pmol was prepared and the deuterio-permethylated maltotetraose was spiked into the permethylated glycan samples to 4 pmol prior to spotting onto the MALDI target plate.

3.10. MALDI-FT-ICR MS

The dried permethylated O-glycans were re-dissolved in 10 μ L acetonitrile. The matrix was prepared by dissolving 2,5-dihydrobenzoic acid in 50 % acetonitrile, giving a 20 mg/mL solution. The sample spot was prepared by mixing 4 μ L of the matrix solution with 2 μ L sample solution. 2 μ L of the mixture was transferred to a stainless steel MALDI target plate and dried under vacuum.

Mass spectra were acquired on a Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometer. A 9.4 T solariX FT mass spectrometer (Bruker Daltonics), with a smart beam nitrogen UV laser (337 nm), was operated in positive ion mode MALDI. The mass spectrometer was calibrated externally using a permethylated maltoheptaose standard. Mass spectra were recorded over an m/z range 400-4000, acquiring 12 scans, with each scan being derived from 500 laser shots. The laser power was set at 40 %. Spectra were acquired using solariXcontrol software and processed with DataAnalysis 4.0.

3.11. Glycan quantitation and statistical analysis

The relative abundance of glycans was calculated by comparing the peak intensities of each of the O-glycans with peak intensity of the internal standard, and the relative abundance was expressed as a percentage. Statistical analysis was carried out using a T test with Holm Sidak post-hoc test (* for $P = < 0.05$, ** for $P = < 0.01$ and *** for $P = < 0.0001$)

3.12. Tandem mass spectrometry (MS/MS)

MS/MS spectra were acquired using Nanospray Flex ion source with off-line borosilicate emitters on an Orbitrap Fusion hybrid mass spectrometer (Thermo Scientific). Positive ESI-MS and MS² spectra were acquired in the Orbitrap using Xcalibur software (version 4.0, Thermo Scientific). Fragmentation was induced using either collision induced dissociation (CID) or higher energy collision dissociation (HCD). For CID, collision energy was varied between 25-35 %, and for HCD the collision energy was varied between 32-43 %, depending on the m/z of the glycans, and the quadrupole isolation window set to 2 m/z .

3.13. Tryptic digestion

The protein remaining in the filter unit following *N*- and *O*-glycan release was treated with trypsin at 37 °C for 16 h. Following incubation the filter was centrifuged to collect the tryptic peptides. Samples were acidified with 0.1 % TFA before LC/MS/MS analysis.

3.14. Peptide analysis

Samples were loaded onto an UltiMate 3000 RSLCnano HPLC system (Thermo) equipped with a PepMap 100 Å C₁₈, 5 µm trap column (300 µm x 5 mm Thermo) and an Acclaim PepMap RSLC, 2 µm, 100 Å, C₁₈ RSLC nanocapillary column (75 µm x 150 mm, Thermo). The trap wash solvent was 0.05% (v/v) aqueous trifluoroacetic acid and the trapping flow rate was 15 µL/min. The trap was washed for 3 min before switching flow to the capillary column. The separation used a gradient elution of two solvents (solvent A: aqueous 2 % (v/v) acetonitrile containing 1% (v/v) formic acid; solvent B: acetonitrile containing 1 % (v/v) formic acid). The flow rate for the capillary column was 300 nL/min. Column temperature was 50°C and the gradient profile was: linear 3-10% B over 7 mins, linear 10-35 % B over 30 mins, linear 35-99 % B over 5 mins then proceeded to wash with 99 % solvent B for 4 min. The column was returned to initial conditions and re-equilibrated for 15 min before subsequent injections.

The nanoLC system was interfaced with an Orbitrap Fusion hybrid mass spectrometer (Thermo) with a Nanospray Flex ionisation source (Thermo). Positive ESI-MS and MS² spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were: ion spray voltage, 2,200 V; sweep gas, 2 Arb; ion transfer tube temperature; 275 °C. MS¹ spectra were acquired in the Orbitrap with: 120,000 resolution, scan range: *m/z* 375-1,500; AGC target, 4e⁵; max fill time, 100 ms; data type, profile. MS² spectra were acquired in the linear ion trap specifying: quadrupole isolation, isolation window, *m/z* 1.6; activation type, HCD; collision energy, 32 %; scan range, normal; scan rate, rapid; first mass, *m/z* 110; AGC target, 5e³; max injection time, 100 ms; data type, centroid. Data dependant acquisition was performed in top speed mode using a 3 s cycle. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at 5e³.

3.15. Protein identification

Peak lists were generated in MGF format using Mascot Distiller (version 5, Matrix science), stipulating a minimum signal to noise ratio of 2 and correlation (Rho) of 0.6. MGF files were searched against the porcine subset of the UniProt database (34,316 sequences; 14,316,901 residues), with a decoy database, using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.5.1). Search criteria specified: enzyme, trypsin; fixed modifications, pyridylethyl (C); variable modifications, deamidated (NQ), oxidation (M), ser->diamino-propanoate (S), thr->Dab (T), Thr-> diamino-butyrates (T); peptide tolerance, +/- 5 ppm; MS/MS tolerance, 0.5 Da; Instrument, ESI-TRAP. Results were passed through Mascot percolator to achieve a false discovery rate of <1% and further filtered to accept only peptides with an expect score of 0.05 or lower.

3.16. Preparation of protein, glycopeptides and glycans for NMR analysis

3.16.1. "intact" urea-denatured glycoprotein (monoclonal antibody), IgG

150 µL IgG was diluted to 1 mL with 7 M urea/tris solution. 100 mM DTT was added and the glycoprotein sample was incubated at 60 °C for 30 min and then transferred to a centrifugal filter unit (molecular weight cut off 10 kDa) and centrifuged for 8 min at 10 000 rcf. The sample was washed two times with 200 µL urea/tris buffer, to reduce the DTT concentration, before adding 40 mM iodoacetamide to the sample and incubating for 15 min, at 37 °C. The sample was centrifuged to remove residual iodoacetamide and the sample was washed twice with 200 µL urea/tris buffer. The denatured protein sample was concentrated to 100 µL ready for NMR. Luciano Mueller performed sequential dissolving and lyophilising to ensure minimal losses from the sample sticking to the sides of the tube.

3.16.2. Preparation of "large" N-glycan sample from IgG

2 mL of IgG was used. 150 µL aliquots were transferred to 12 microcentrifuge tubes and made up to 1 mL with 7 M urea/tris buffer. 100 mM DTT was added to each of the samples and incubated at 60 °C for 30 min. Each reduced sample was transferred to a centrifugal filter unit (10 kDa) and centrifuged for 8 min at 10 000 rcf. The samples were washed twice with 7 M urea/tris, centrifuging each time at 10 000 rcf for 8 min. 40 mM iodoacetamide was added to each sample and incubated at 37 °C for 15 min. The samples were centrifuged to remove the iodoacetamide and washed, twice, with 7 M urea/tris. The urea was reduced to 2 M with 50 mM ammonium bicarbonate. 6 U of PNGase F was added to each sample, to

release the *N*-glycans, and incubated at 37 °C for 16 h. Following incubation the samples were centrifuged (using a clean/washed collection tube) at 10 000 rcf for 8 min to collect the released *N*-glycans. The filter was washed with 50 µL water, twice, each time centrifuging at 10 000 rcf for 8 min and the filtrate collected.

The released glycans were cleaned up by PGC SPE. PGC glycosep SPE cartridge was pre-treated with 2 mL of 1 M NaOH, 2 mL pure water, 2 mL 30 % acetic acid and 2 mL pure water. SPE cartridge was primed with 2 mL 50 % acetonitrile, 0.1 % TFA, followed by 2 mL 5 % acetonitrile, 0.1 % TFA. The glycan sample was loaded onto the column. Glycan sample washed with 2 mL pure water, followed by 2 mL 5 % acetonitrile, 0.1 % TFA. Glycans were eluted with ~ 2 mL 50 % acetonitrile, 0.1 % TFA into a labelled tube. The cleaned glycan sample collected from the first SPE was dried under nitrogen. The glycan sample filtrate from the second SPE was collected into the tube containing the first dried glycan sample and this was also dried down under nitrogen. The process was continued to give a large glycan sample from the combined 12 glycan samples. The glycan sample was lyophilized ready for NMR analysis.

3.16.3. Preparation of a “large” (glyco)peptide sample from GITR

2 mL GITR sample was diluted into 10 mL, 7M, urea/tris buffer. 100 mM DTT was added and the sample incubated at 60 °C for 30 min. The sample was transferred in a large centrifugal filter unit (10 kDa) and centrifuged for 25 min. The sample was washed with 4 mL, 7 M, urea/tris buffer, centrifuging for 25 min at 3800 rcf. The sample was treated with 40 mM iodoacetamide at 37 °C for 15 min, followed by centrifuging for 25 min to remove iodoacetamide. The sample was washed with 4 mL 7 M urea/tris buffer, centrifuging at 3800 rcf for 25 min. The 7 M urea was reduced to 2 M urea with 50 mM ammonium bicarbonate. 200 µL of 3 mg/mL trypsin was added and the sample was incubated at 37 °C for 16 h. Following incubation, the tryptic digest sample was centrifuged (in a clean/washed collection tube) for 30 min at 3800 rcf to collect the tryptic peptides. The sample was washed with 150 µL water and the sample was centrifuged for a further 30 min.

3.17. Nuclear magnetic resonance spectroscopy

The dried glycosylated antibody (IgG) samples were re-dissolved in 99.9 % D₂O, 8 M urea (perdeuterated), 50 mM tris-d11, 30 % CD₃CN, pH 2.5. The samples were transferred to 4

mm NMR tubes. NMR measurements were acquired on a Bruker AvanceIII HD 600 MHz NMR spectrometer equipped with a TCI cryoprobe. A suite of 2D NMR experiments (set up by Luciano Mueller) were carried out; [$^1\text{H}^{13}\text{C}$]-HSQC, [$^1\text{H}^{13}\text{C}$]-dept-HSQC, [$^1\text{H}^1\text{H}$]-TOCSY, [$^1\text{H}^{13}\text{C}$]-HMBC, [$^1\text{H}^1\text{H}$]-NOESY, [$^1\text{H}^{13}\text{C}$]-HMQC-COSY, and [$^1\text{H}^1\text{H}$]-dqf-COSY. All NMR experiments were carried out at 308 K. Spectra were recorded with 4-64 scans using an excitation sculpting pulse sequence for solvent suppression (spectral widths of 14 and 160 ppm in the proton and carbon dimension respectively, relaxation delay 1 s, acquisition time 0.2 s). Spectra were acquired and processed using Bruker TopSpin 3.5 pl5.

Chapter 4: Method Development

4. Method development

A method has been developed, designed to be an extension of FANGS and to take advantage of the convenience, throughput and ability to work with realistic numbers of cells that it offers, to release and isolate O-glycans from cellular glycoproteins for subsequent mass spectrometric analysis. The method makes use of a centrifugal filter as a simple reaction vessel in which glycoproteins can be treated with ammonium hydroxide solution to release the O-glycans by non-reductive β -elimination. Following centrifugation, glycans can be retrieved in the filtrate as in FANGS, as the released glycans pass through the filter membrane while proteins are large enough to remain above the membrane and so are retained. The isolated O-glycans can then be permethylated for mass spectrometric analysis. The method can be conveniently combined with the established FANGS protocol, to release, isolate and analyse both *N*- and O-glycans using one sample, in the same filter unit, enabling analysis of the whole glycome from the same sample, using only around one million cells.

4.1. Method development for O-glycan release

Porcine stomach mucin was chosen as a glycoprotein standard that is well characterised and commercially available with which to carry out method development. O-glycans were released from 20 pmol mucin, by first dissolving it in a solution of SDS (as for FASP¹⁴³ or FANGS¹⁴¹) and exchanging the SDS for urea and then ammonium bicarbonate solution, before carrying out O-glycan release with NH_4OH . After collecting and permethylating the released glycans, a MALDI-FT-ICR mass spectrum was obtained, revealing the expected glycan signals²¹⁹ between m/z 500-2214.

A recent publication describes enhanced filter-aided sample preparation¹⁷⁷ (eFASP), which uses alternative reagents to those first described by Manza et al¹⁴² and later in FASP by Wisniewski et al¹⁴³, and that reports increased sensitivity and sample recovery for a proteomics study. To see whether eFASP sample preparation could also yield improved results for a glycomics study, samples of porcine stomach mucin were prepared for O-glycan release, using either FASP or eFASP protocols for SDS solubilisation and removal/exchange. The filter was then used as a reaction vessel to release the O-glycans by treatment with NH_4OH , to compare the two glycoprotein preparation methods side by side. MALDI-FT-ICR mass spectra of permethylated O-glycans, ionised as $[\text{M} + \text{Na}]^+$, were acquired, revealing expected O-glycan signals for mucin glycans²¹⁹ between m/z 500-2214. The data from the

two sample preparation methods show the same O-glycans were released, but better signal to noise was observed for the sample prepared using eFASP, Figure 42.

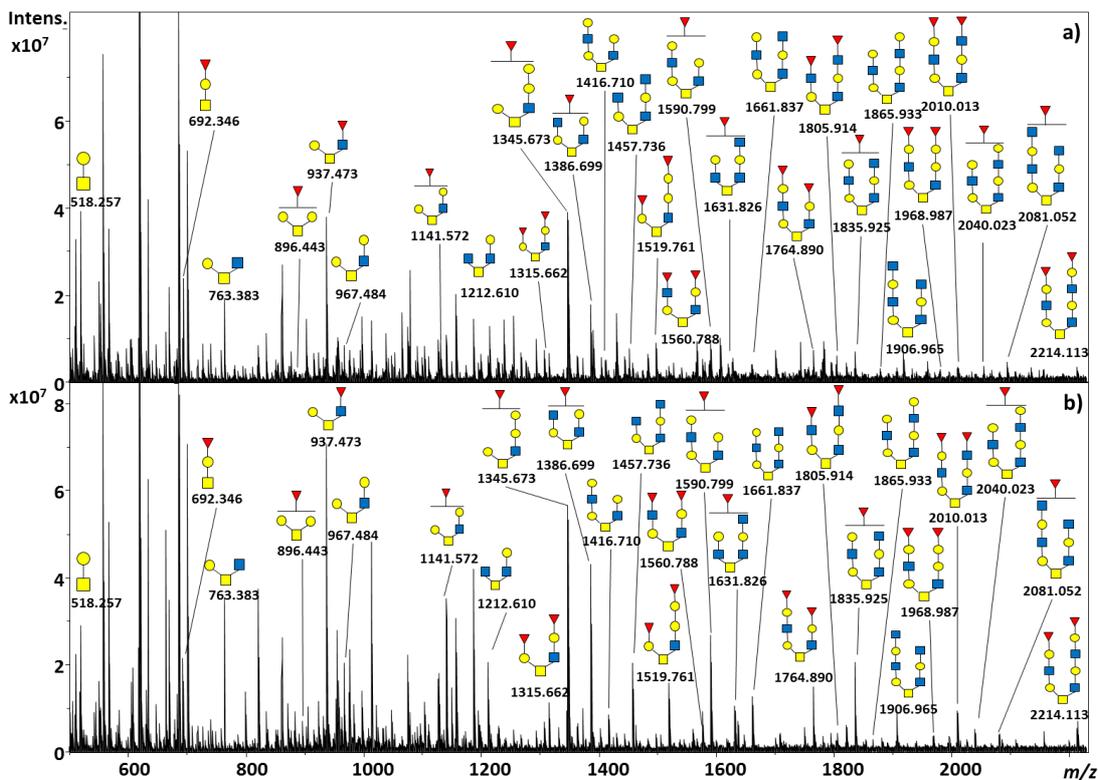


Figure 42: MALDI-FT-ICR spectrum of permethylated O-glycans, ionised as $[M + Na]^+$, released from 20 pmol of porcine stomach mucin a) following FASP for SDS removal/exchange, followed by O-glycan release, in the filter, with NH_4OH . b) following eFASP for SDS removal/exchange, followed by O-glycan release, in the filter, with NH_4OH

The alkylation step in FASP/FANGS is carried out using 50 mM iodoacetamide in the filter unit, for 15 min. In comparison, in eFASP the recommended alkylation step takes 1 h at 37 °C, with 50 mM iodoacetamide (in the authors' standard protocol) or with 4-vinylpyridine (in their express protocol, where the alkylating reagent is added before the sample is transferred to the filter, eliminating a washing step). The alkylation step was investigated further, to determine whether it is possible to keep the reaction time at 15 min, as in FASP/FANGS. Samples of porcine stomach mucin (20 pmol) were thus prepared for O-glycan release (using eFASP reagents). For the alkylation steps, samples were alkylated for 15 min or 1 h, using iodoacetamide or 4-vinylpyridine. Permethylated O-glycans were identified following alkylation using each of the two, alkylation reagents. Overall, 4-vinylpyridine gave

spectra with better signal to noise than those from the samples prepared using iodoacetamide, and there appeared to be no discernible difference in the results following alkylation for 15 min or 1 h, Figure 43a-d, suggesting 15 min with 4- vinylpyridine is suitable.

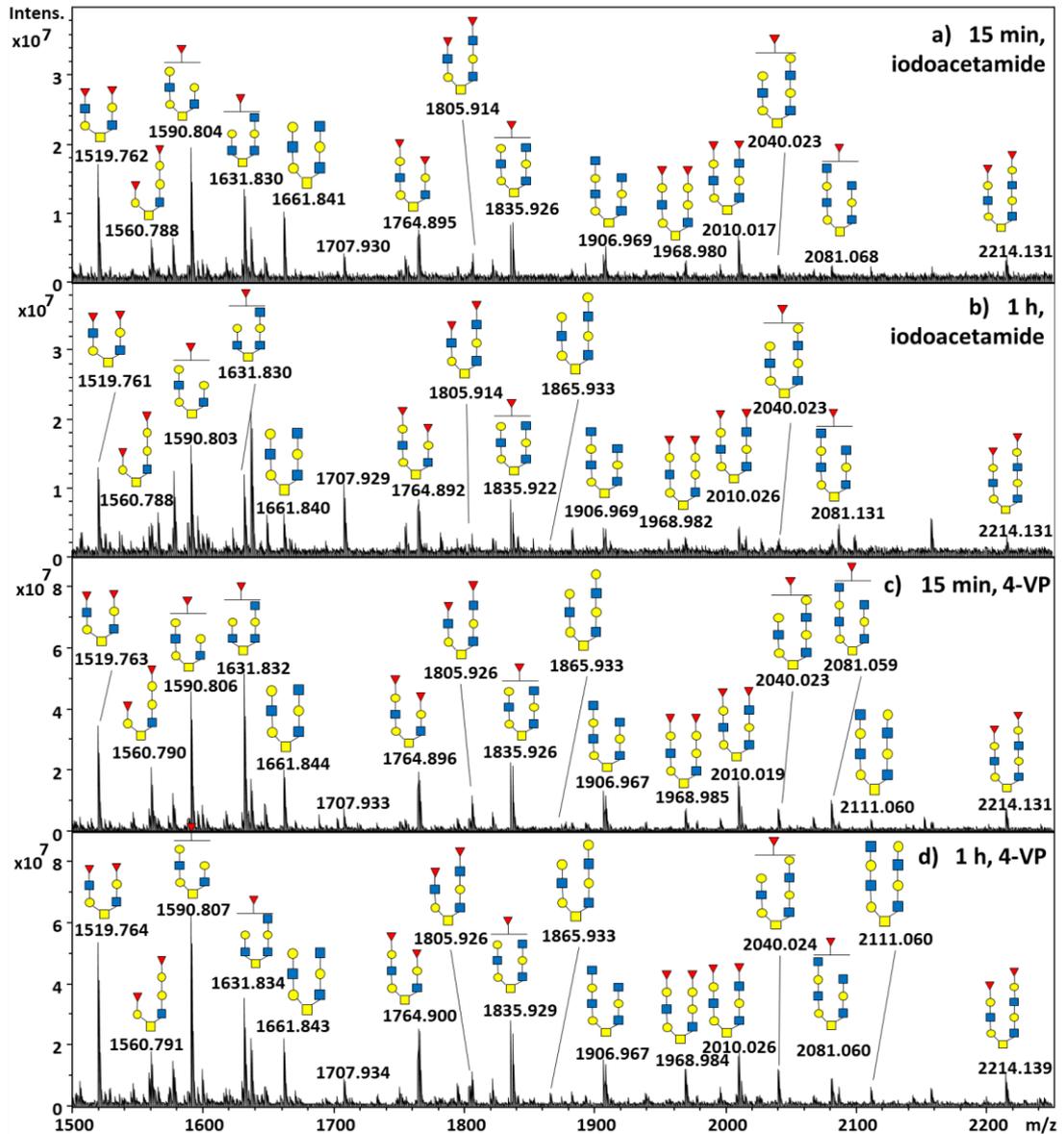


Figure 43: MALDI-FT-ICR spectrum (m/z 1200-2220) of permethylated O-glycans, ionised as $[M + Na]^+$, released from 2 pmol mucin. a) alkylation using 4-vinylpyridine for 1 h, b) alkylation using iodoacetamide for 1 h, c) alkylation using 4-vinylpyridine for 15 min, d) alkylation using iodoacetamide for 15 min

The user manual for the centrifugal filters lists chemical compatibilities. For SDS, the suggestion is to use a concentration below the critical micelle concentration (CMC), less than or equal to 0.1 %, since micelles are too large to pass through the filter membrane. The lysis buffer used in this work is a 4 % SDS solution. In the FANGS protocol, the lysis buffer is diluted by a factor of 1 in 10 with exchange buffer, giving 0.4 % SDS, which is above the CMC and higher than the recommended concentration. Samples of mucin were prepared in 100 μ L lysis buffer – a typical volume of lysis buffer needed for a whole cell lysate from a 10 cm culture dish. This was diluted with exchange buffer either 1 in 10 (standard FANGS protocol) or 1 in 30 to reduce the SDS concentration to 0.13 %. A dilution of 1 in 30 increases the total volume of solution to be washed through the filter to 3 mL. This increases the time needed for sample preparation, with a maximum of 450 μ L at a time being loadable into the reservoir above the filter using the 30 kDa filters supplied by millipore. In FANGS, each centrifugation step was carried out for 10 min. The centrifugation time was investigated and it was found that, using the 1 in 30 dilution, only 5 min centrifugation was needed for all the liquid to pass through; this made the sample preparation time comparable with that needed for the 1 in 10 dilution that used 10 min centrifugation.

In addition, it was considered whether adding deoxycholic acid (DCA) to the exchange buffer, as recommended in eFASP, was necessary in a glycomics workflow. The use of DCA may be irrelevant when dealing with glycans because DCA was used in eFASP to enhance trypsin digestion. Therefore, mucin sample preparation using the 1 in 10 or 1 in 30 dilution with exchange buffer was carried out either with or without addition of DCA to the exchange buffer. Following O-glycan release in the filter with NH_4OH and permethylation of the isolated O-glycans, MALDI-FT-ICR mass spectra were acquired. The full range of O-glycans was observed as in the preceding experiments, giving signals over the range m/z 518-2214. Each sample gave the full range of O-glycans expected, with all spectra showing similar signal to noise. What is clear from the data is that there appear to be more peaks that do not correspond to glycan signals in the samples that were prepared using DCA than in those prepared without including DCA, highlighted by the red box in Figure 44, so subsequent protocols omitted the DCA.

In a parallel experiment, a masters student, Matthew Walker (in Daniel Ungar's group) had also investigated further diluting SDS, from 1 in 10 to 1 in 40, in an attempt to optimise PNGase F digestion for *N*-glycan release. The reasoning behind this was to test whether

more efficient SDS removal would improve PNGase F activity and therefore lower amounts of PNGase F could be used. In his experiment, when analysing *N*-glycans released from RNase B using PNGase F, no *N*-glycans were observed for the sample prepared using 1 in 10 SDS dilution but, on diluting 1 in 40, all expected *N*-glycans were observed.

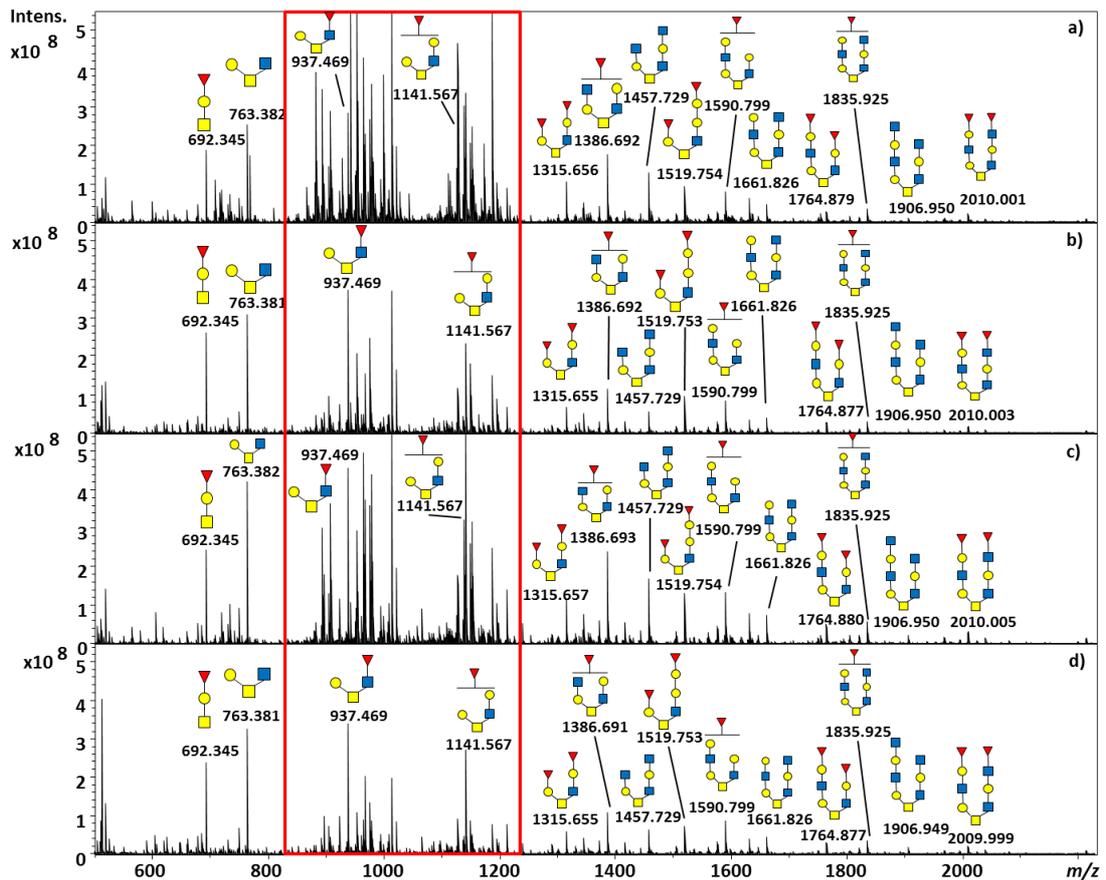


Figure 44 MALDI-FT-ICR mass spectrum of permethylated O-glycans released from porcine stomach mucin using overnight incubation, prepared using SDS solution following 1 in 10 dilution with exchange buffer a) containing DCA b) without DCA and 1 in 30 dilution with exchange buffer, c) containing DCA d) without DCA. The red box highlights that increasing the dilution and removing DCA gives cleaner glycan mass spectra. All glycans ionised as $[M + Na]^+$.

4.2. Reducing incubation time for O-glycan release

Classic methods for releasing O-glycans generally involve overnight incubation in base (16 – 18 h). In an attempt to reduce the time for glycan release, sonication in a sonic bath, and microwave irradiation using a laboratory microwave reactor were investigated.

Sonication times of 5, 10, 15 and 20 min were tested to aid O-glycan release. Glycan signals could be observed after only 5 min of sonication, but it was difficult to avoid the temperature in the sonic bath rising above 45 °C on sonication for longer periods. The temperature of 45 °C was chosen for direct comparison with the overnight incubation method, and with the intention of avoiding hydrolysis of labile glycans. Sonicating for just 5 min, followed by holding the sample in the sonication bath for 10 min without sonicating and then repeating rounds of sonication/rest ensured the temperature went no higher than 45 °C. Using a total of 15-20 min sonication interspersed with 30-40 min rest (a total O-glycan release time of 45 min – 1 h) gave glycan profiles that were comparable with those obtained on overnight incubation, Figure 45. The sonication method was thus chosen as the method to pursue because it is compatible with glycan release in the filter unit. When using the microwave reactor, samples had to be transferred from the filter unit to a microwave tube and back, which made it a lot less convenient, reduced throughput, and would incur inevitable sample loss. In addition, because the sonication was carried out in a sonic bath, multiple samples can be handled side-by-side, and so the approach is amenable to the handling of multiple samples in parallel, which is in line with the aims of the method we have developed.

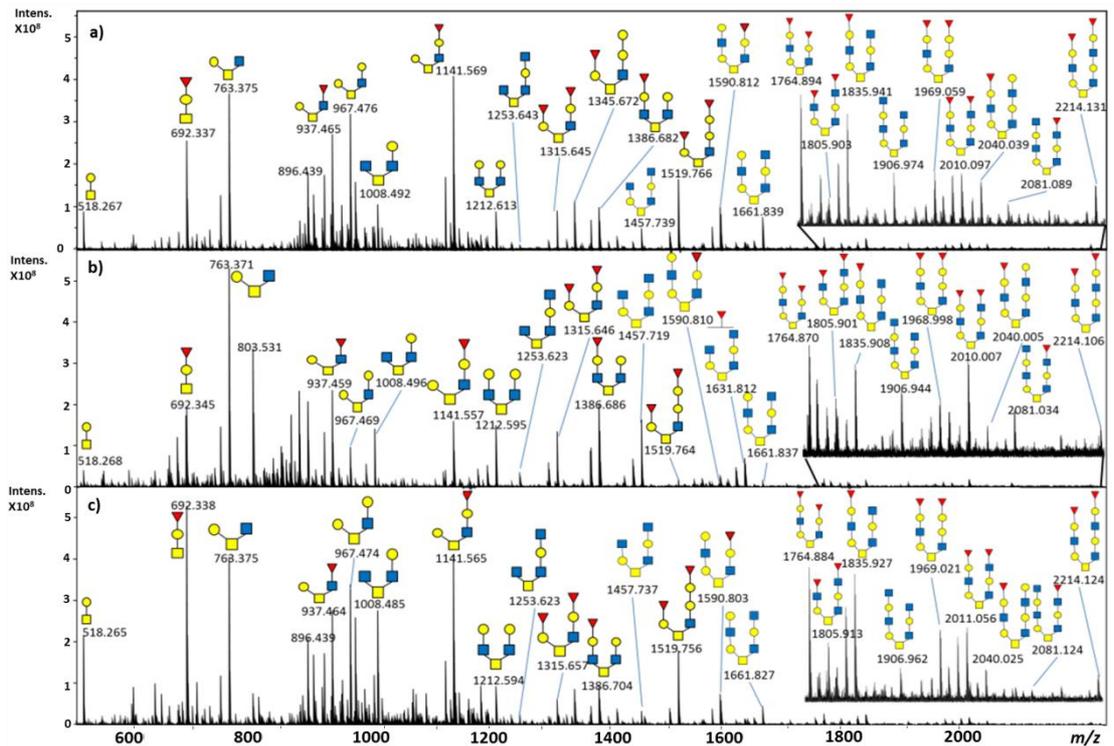


Figure 45 MALDI-FT-ICR mass spectrum of permethylated O-glycans, ionised as $[M + Na]^+$, released from porcine stomach mucin using 16 h incubation (top panel), 5 min sonication/10 min rest (x4) (middle panel) and 10 min microwave irradiation (bottom panel).

After optimising the O-glycan release method using the glycoprotein standard porcine stomach mucin, the method was applied to porcine bladder urothelium cells to demonstrate its application to a primary cell sample; urothelial cells were chosen as they are known to contain mucins and are thus likely to bear O-glycans. Samples of porcine bladder urothelium were prepared for O-glycan release from a 50 mg aliquot of bladder urothelial scrapings, Figure 46, and following our optimised eFASP/FANGS protocol.

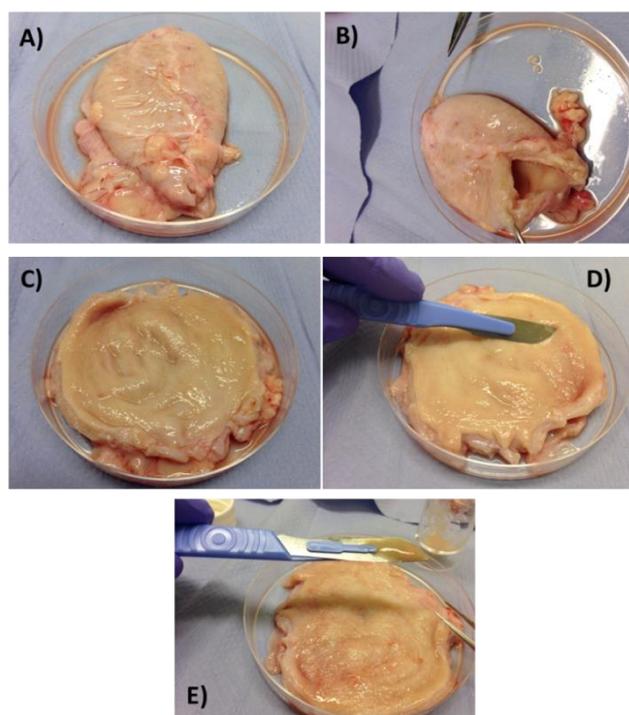


Figure 46: Collection of porcine bladder urothelium. a) Porcine bladder b) bladder cut open through the urethra c) inside the porcine bladder. d) surface of the bladder scraped with a scalpel e) scrapings collected for glycomic analyses. This method of harvesting porcine bladder urothelium has been developed and validated by Chung-Yao Wang in the group, in collaboration with Prof Jenny Southgate.

O-glycans were released by treating with NH_4OH in the filter unit using overnight incubation (16 h) or 20 min sonication (interspersing each of the four 5 min sonications with 10 minute rests to avoid local overheating). Glycans were permethylated and on obtaining a MALDI-FT-ICR mass spectrum, O-glycans were identified in both the overnight incubation and sonication-aided treated samples, between m/z 518- 1777. Glycans at m/z 518 and 722, $\text{Hex}_1\text{HexNAc}_1$ and $\text{Hex}_2\text{HexNAc}_1$, are the only two glycans which are not sialylated or fucosylated. Fucosylated glycans were identified at m/z 692, 896, 937, 1141, 1315, and 1328 and sialylated glycans were identified at m/z 879, 1083, 1124, 1154, 1328, 1358, 1485, 1573, 1777, figure 5. Even those glycans of low abundance could be readily identified in the sonicated sample (m/z 692, 1083, 1358, 1485, 1573, 1777). Importantly, the sonication method has left the labile sialic acid moieties intact, and shows comparable signal to noise for each of the sialylated glycan structures with those in samples prepared using the overnight method, Figure 47.

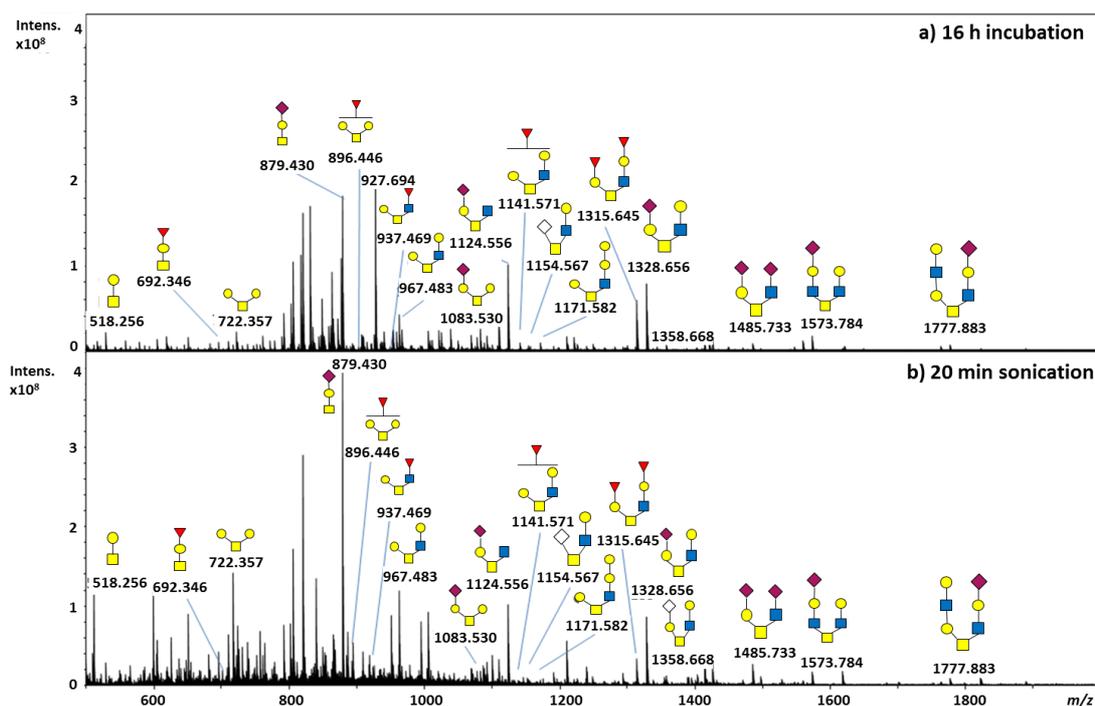


Figure 47 MALDI-FT-ICR mass spectrum of permethylated glycans released from porcine bladder urothelium, ionised as $[M + Na]^+$. a) O-glycans released using 16 h incubation b) O-glycans released using 20 min sonication at 45 °C

4.3. One-sample, one-pot; the whole glycome

Both *N*- and *O*-glycans are important when studying disease, and so access to a protocol which would allow analysis of the whole glycome from one sample handled in the same reaction pot is desirable, especially one based around the use of centrifugal filters that enable use of small cell numbers as well as low volumes of reagents. The workflows for releasing *N*-glycans and *O*-glycans have thus been combined to test whether *N*-glycans and *O*-glycans can be released from the same sample in the same reaction vessel (with the potential of also being able to analyse the residual protein afterwards if required). A sample of porcine bladder urothelium was treated using the optimised sample preparation workflow based on eFASP, and then treated with PNGase F for 16 h at 37 °C. Following incubation, the filter was centrifuged to collect the released *N*-glycans. The filter was transferred to a clean collection tube and 300 μ L NH_4OH was added to release the *O*-glycans, aided by 20 min sonication at 45 °C, and the released *O*-glycans retrieved on centrifugation of the filter unit. *N*- and *O*-glycans were permethylated and MALDI-FT-ICR mass spectra were acquired. *N*-glycans were observed from m/z 967-2966, corresponding mainly to oligomannose species, with some complex type *N*-glycans, some of which were sialylated

and others fucosylated. The O-glycan mass spectrum revealed O-glycans between m/z 518 – 1777 – the same as for the urothelial O-glycan sample from which *N*-glycans had not previously been released. In addition, focussing on m/z 1500 – 3000 in the O-glycan spectrum, it is clear that there are no signals corresponding to *N*-glycans, suggesting that all the *N*-glycans have been efficiently washed out of the filter before O-glycan release was carried out. However, there are signals observed in both the *N*- and the O-glycan spectra at m/z 967, m/z 1141 and m/z 1171. Although these signals could formally be due to carry over of *N*-glycans into the O-glycan sample, these three signals are observed in a spectrum of O-glycans that had been prepared without previously releasing *N*-glycans. This, together with the fact that none of the other more abundant *N*-glycans are observed contaminating the O-glycan sample, suggests that it is unlikely that the m/z 967, 1141 and 1171 signals derive from *N*-glycan carryover, Figure 48.

To further investigate whether these common signals could indeed be due to carry over, HCD product ion spectra were acquired for both the *N*- and O-glycans (released following *N*-glycan release) for the glycans detected at m/z 967, 1141 and 1171. This was to obtain structural information to provide evidence that the glycans in the O-glycan sample were O-glycans and not carry over from the *N*-glycan samples. Firstly, the HCD spectrum for m/z 967, for Hex₂HexNAc₂ in the *N*-glycan sample shows diagnostic B and Y fragment ions that indicate that the *N*-glycan structure is linear (as expected from the known *N*-glycan structures^{220,30}) and contains two HexNAc units at the reducing terminal followed by two Hex units joined together. This is shown by the Y₂ (m/z 545.27, HexNAc₂) and B₂ (m/z 445.24, Hex₂) ions, Figure 49. In contrast, in the HCD spectrum for the O-glycan at m/z 967, the fragments show that the structure is branched, and does not contain adjacent pairs of Hex or HexNAc residues, and is consistent with the structure expected for the commonly occurring mucin type O-glycans^{43,219,123}. This is shown by the diagnostic Y_{1α} (m/z 504.24) and B_{2α} (m/z 486.23) ions for Hex₁HexNAc₁ substructures, Figure 49. This evidence demonstrates that the m/z 967 glycan is an O-glycan and not due to carry over of an *N*-glycan.

Similarly, the HCD spectrum for m/z 1141 (Hex₂HexNAc₂Fuc) in the *N*-glycan sample shows fragment ions that indicate a linear structure, as for the m/z 967 glycan but with a fucose on the core HexNAc unit. Y_{2α} (m/z 719.34 for HexNAc₂Fuc) and B_{2α} (m/z 445.21 for Hex₂) demonstrate the expected core fucosylated *N*-glycan truncated core structure.

For the O-glycan HCD spectrum of m/z 1141, there is a $Y_{1\alpha}$ fragment ion for $\text{Hex}_1\text{HexNAc}_1$ at m/z 504, in addition to a $C_{3\alpha}$ ion at m/z 678 for $\text{Fuc}_1\text{HexNAc}_1\text{Hex}_1$. These ions were not detected in the spectrum from the equivalent species in the *N*-glycan sample, although they would not be expected anyway in an *N*-glycan. The spectrum gives enough evidence to show that this structure is an O-glycan, and clearly a different structure than the *N*-glycan, Figure 50.

Finally, the HCD spectrum for the *N*-glycan at m/z 1171 shows diagnostic fragment ions for three hexose units glycosidically linked together, as expected (formally three mannose units in an *N*-glycan), shown by Y_2 (m/z 545.27) and B_1 (m/z 649.30). In the O-glycan HCD spectrum this is not the case and there are only fragment ions to show the loss of one or two hexose units, $Y_{3\alpha}/Y_{1\beta}$, (m/z 953.46), $Y_{2\alpha}$ (m/z 749.37) and $B_{1\alpha}/B_{1\beta}$ (m/z 241.24) The loss of two hexose units in the *N*-glycan sample would only be possible from a double cleavage, and this is not observed, Figure 51.

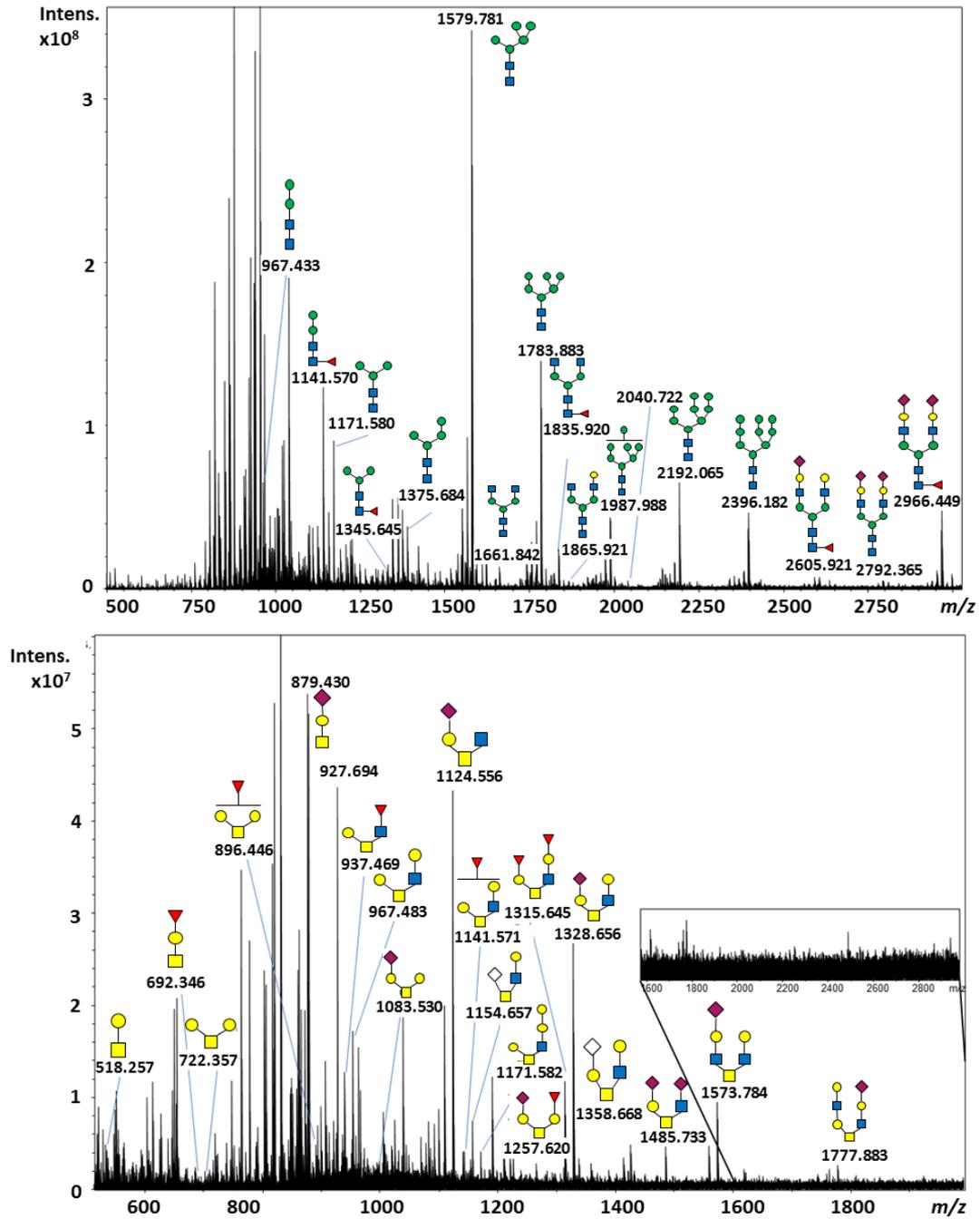


Figure 48 MALDI-FT-ICR mass spectrum of permethylated glycans released from porcine bladder urothelium, ionised as $[M + Na]^+$. a) *N*-glycans released before *O*-glycans b) *O*-glycans released after *N*-glycans in the same filter unit from the same sample

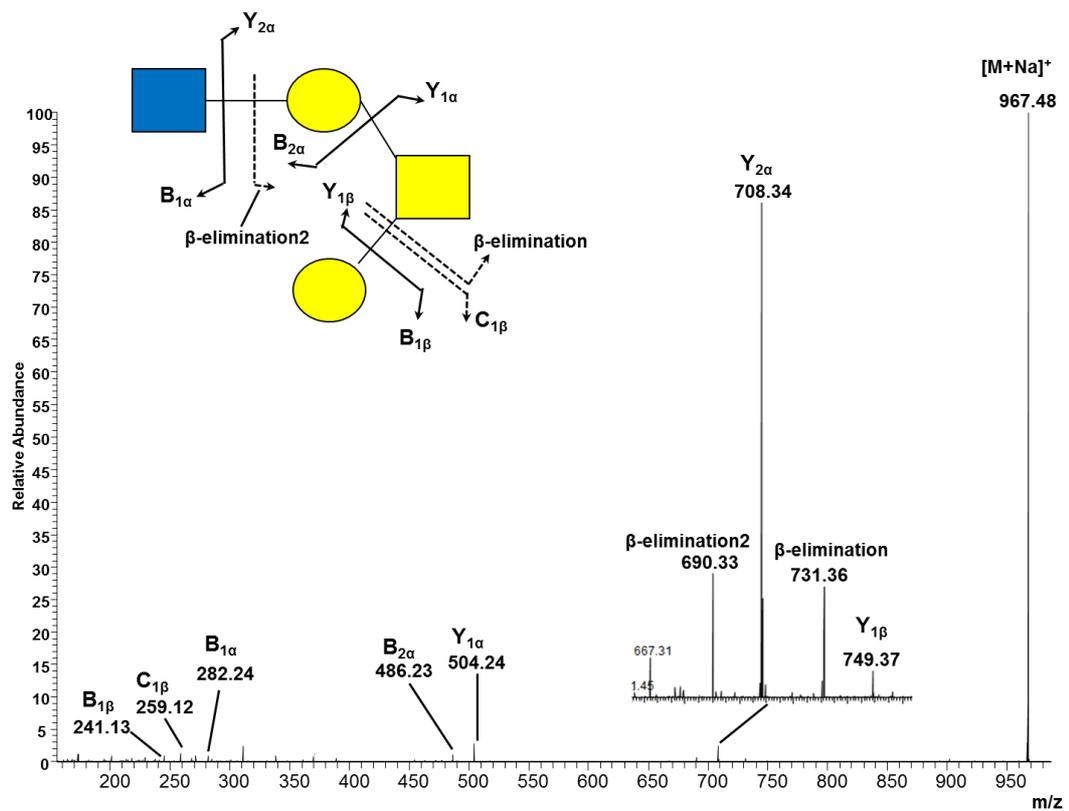
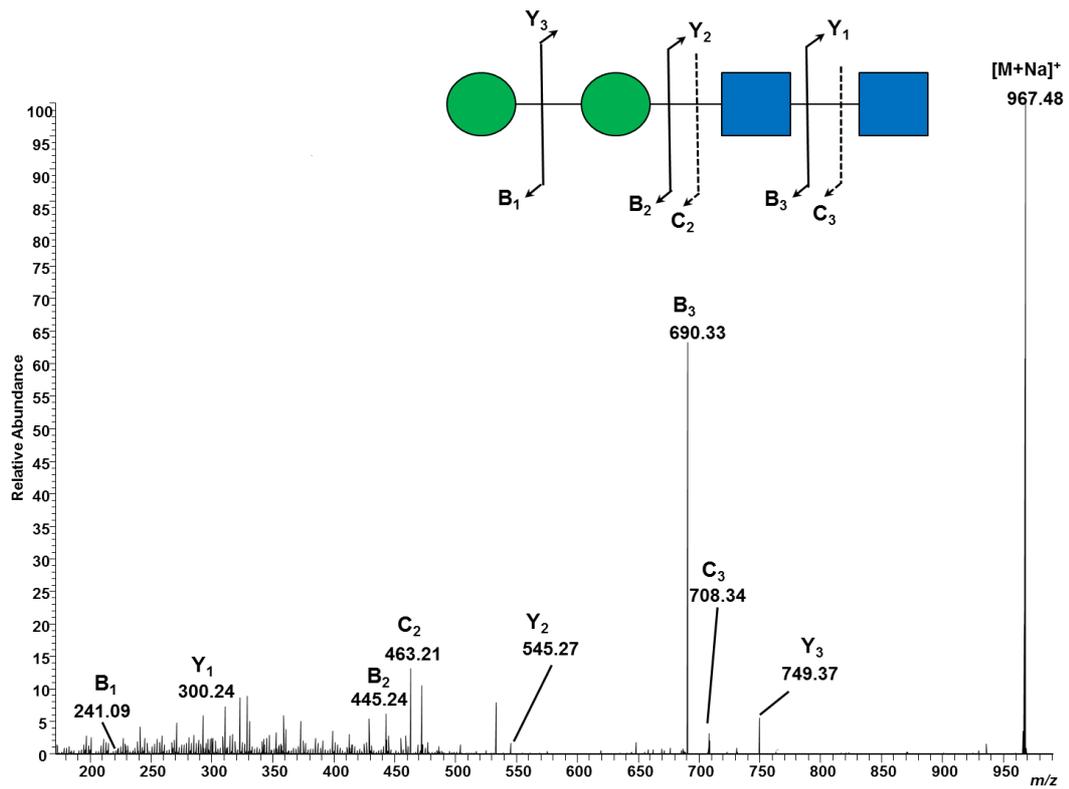


Figure 49: HCD spectra of precursor m/z 967, $[M + Na]^+$ ($Hex_2HexNAc_2$) from porcine urothelial *N*-glycans (top) and *O*-glycans (bottom). Glycans are permethylated.

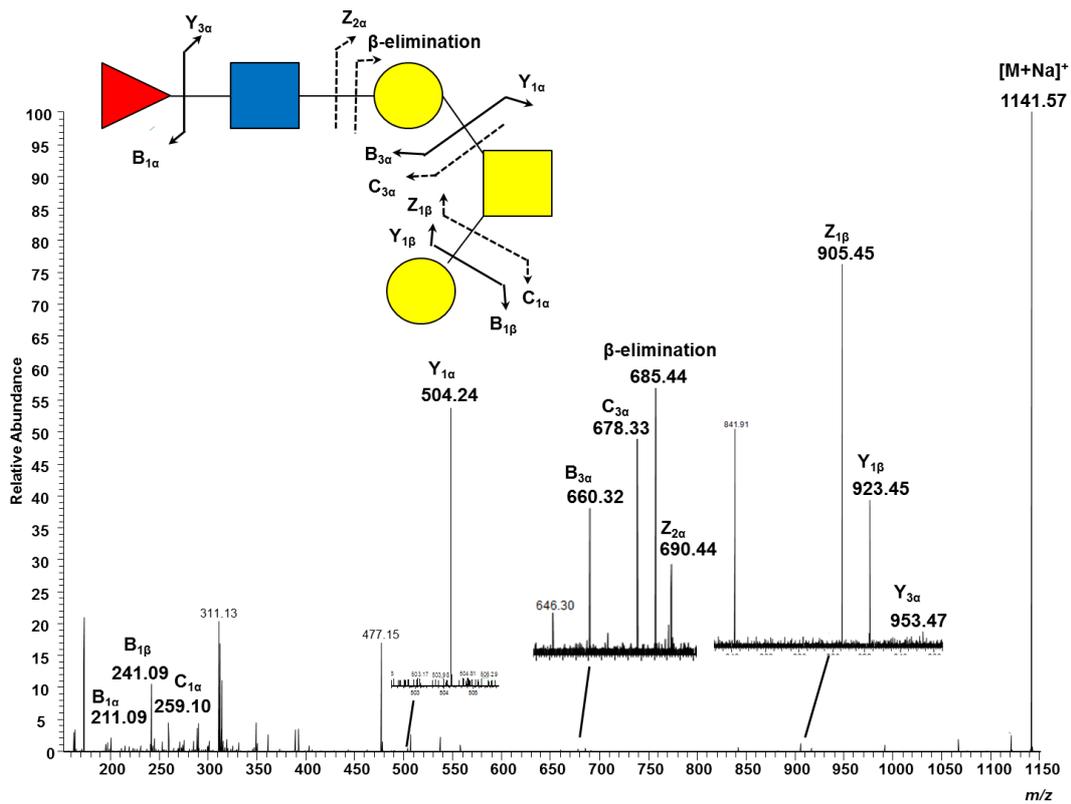
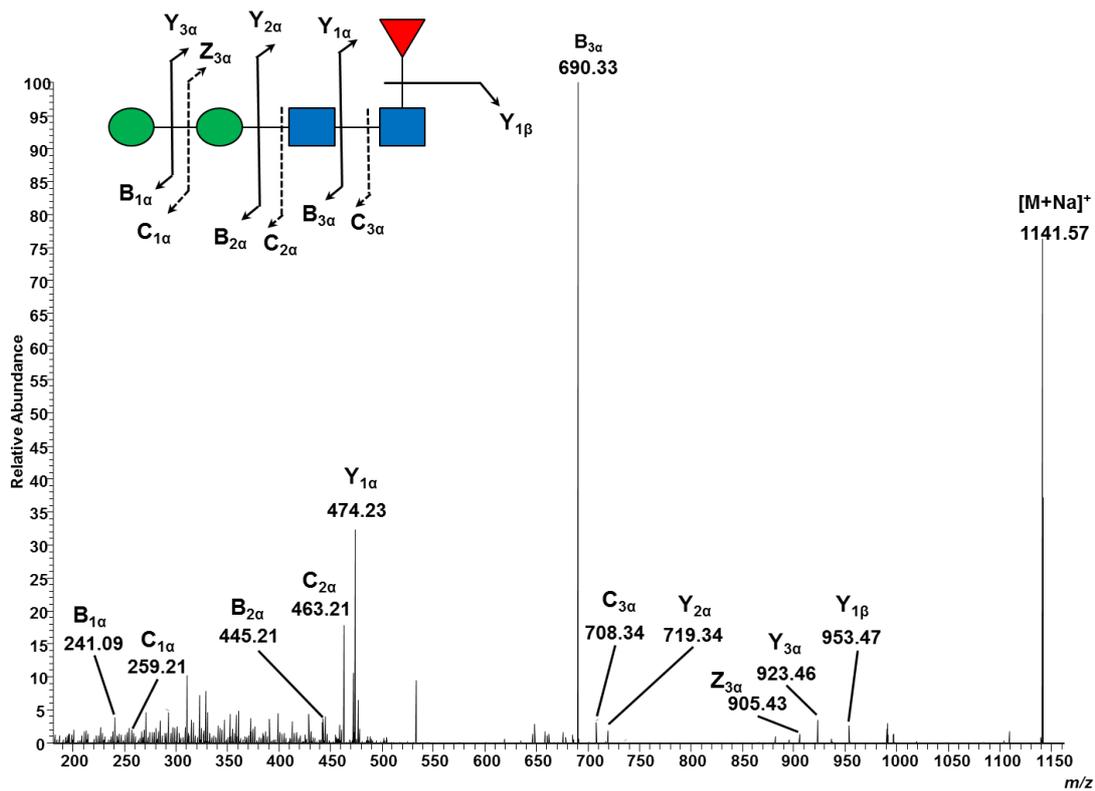


Figure 50: HCD spectra of precursor m/z 1141, $[M + Na]^+$ (Fuc₁Hex₂HexNAc₂) from porcine urothelial *N*-glycans (top) and *O*-glycans (bottom). Glycans are permethylated.

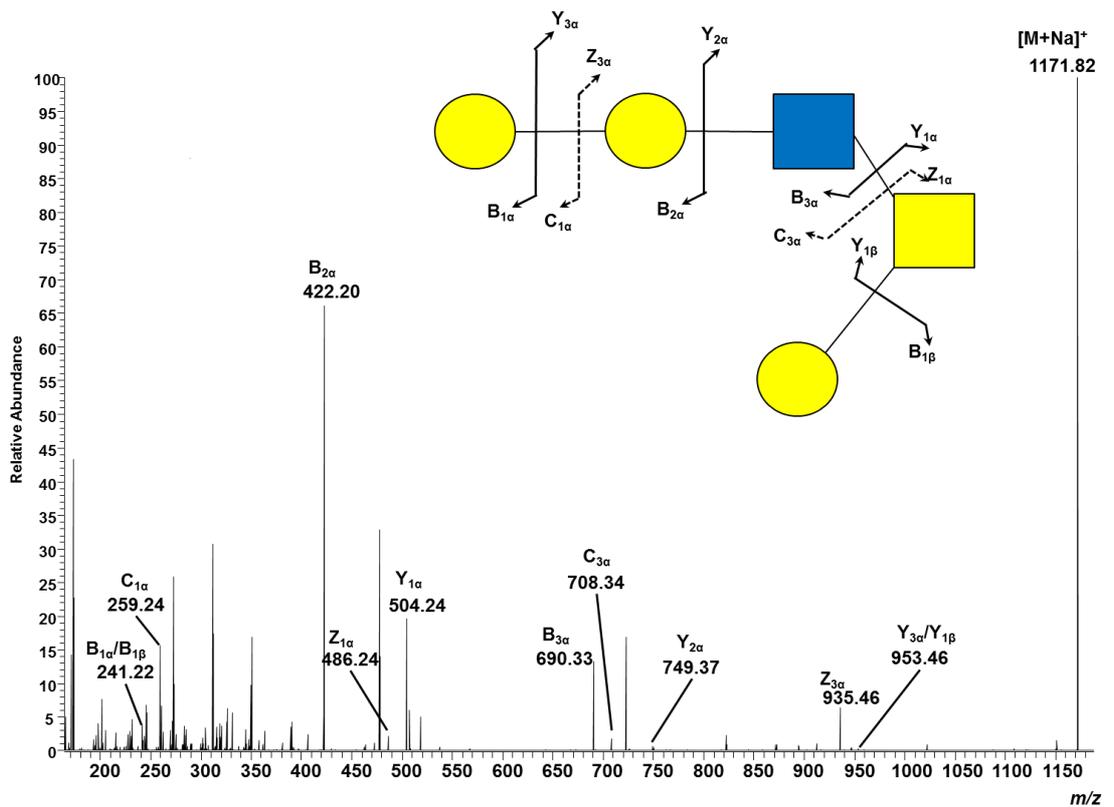
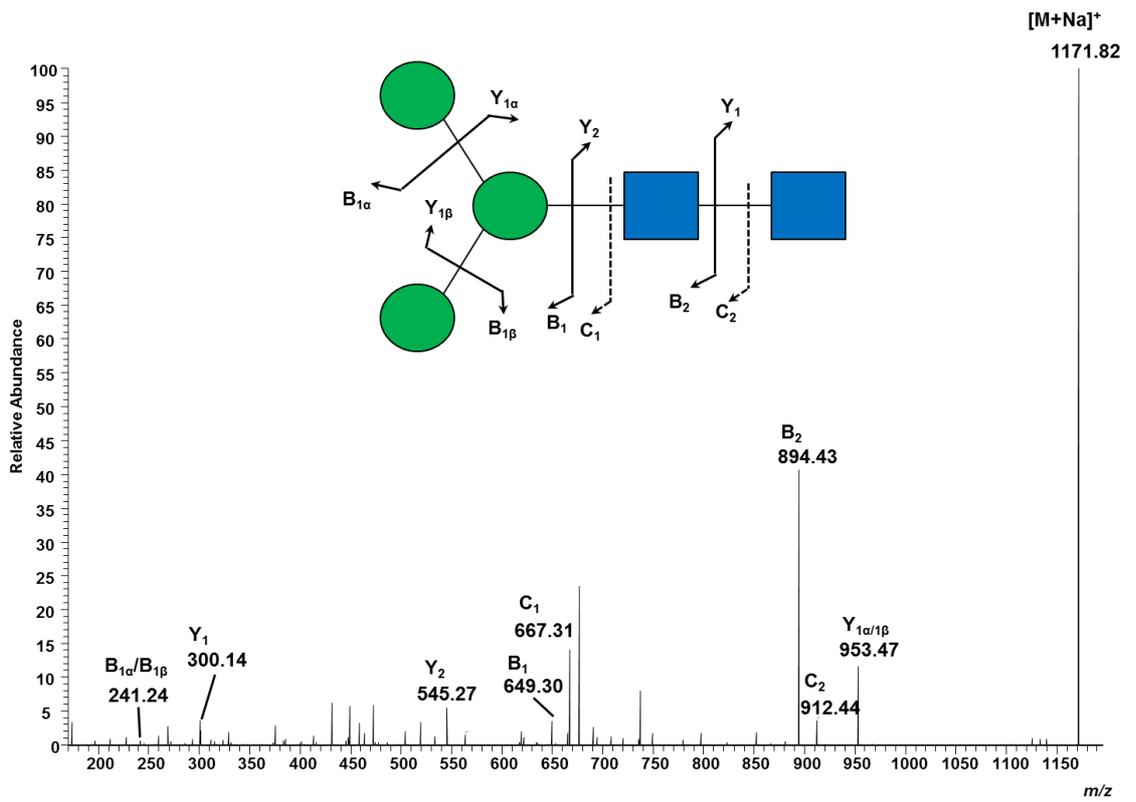


Figure 51: HCD spectra of precursor m/z 1171, $[M + Na]^+$ ($\text{Hex}_3\text{HexNAc}_2$) from porcine urothelial *N*-glycans (top) and *O*-glycans (bottom). Glycans are permethylated.

To show that the developed method can be used as a truly one pot method to study glycans but also protein, after both the *N*- and *O*-glycans had been removed, the remaining protein in the filter unit was treated with trypsin. The tryptic peptides were collected in the filtrate after centrifugation and analysed by LC-MS/MS. Mascot searching matched 5005 peptides, and identified 1301 proteins at an FDR of 1 %. These results were compared with those from porcine urothelium cells that were treated with trypsin without prior removal of the glycans. Mascot searching of those data matched 1348 proteins and 5323 peptides at an FDR of 1 %. Proteins identified were similar across the two sets of data, with 848 proteins the same in the two sets of data. Although this experiment was only carried out once (no replicates) as a proof of concept, the results suggest that the proteomic analysis is not substantially compromised by prior glycan release.

4.4. Optimum time for O-glycan release?

To study glycan structures, especially if we want to determine the relative abundance of glycans in a sample, it would be useful to know that glycan release methods are efficient and consistent in releasing all the different glycan structures from the protein, whether using incubation or sonication.

Methods in the literature for *O*-glycan release use incubation; incubation times typically range from 16 -24 h. Rademaker et al¹⁵³ looked at optimising the time needed for *O*-glycan site analysis. A time study was carried out treating glycoprotein samples with NH_4OH in a non-reductive β -elimination reaction for 0-36 h. The authors concluded that as the incubation time increased, signals for de-glycosylated peptides increased. Interestingly there was a variation in optimal release time; 8 h for *O*-glycans attached to a serine residue and 18 h for those attached to a threonine residue. It was suggested that this could be because the threonine-bound *O*-glycans could be less accessible to the NH_4OH reagent, compared to the serine bound, because of the steric hindrance from the presence of a methyl group on threonine's side chain, that is absent in serine. An incubation time of 15-16 h was recommended to yield sufficient information from an unknown glycoprotein structure.

To determine whether incubation time for *O*-glycan release is also linked to what glycans are present in a glycoprotein sample and whether some glycan moieties are released more

quickly than others, a time study was carried out by MChem student Sam Black, under the supervision of the author. The study was to determine whether the relative abundance of glycans detected changes as a function of the duration of NH₄OH treatment.

Porcine stomach mucin was treated with 300 µL of NH₄OH, samples were incubated at 45 °C, for 1 h, 2 h, 4 h, 6 h, 8 h, 16 h, 18 h, 20 h, 22 h and 24 h, before removing from the heating block and collecting the released O-glycans by centrifugation. O-glycans were permethylated before analysis by MALDI-FT-ICR MS. An internal standard (deuteromethylated maltotetrose) was spiked at a fixed level into the O-glycan samples after permethylation, and so the relative abundance of each glycan signal was calculated as a percentage of the internal standard peak.

In general, the abundance of each glycan signal increased with increasing time of incubation. A closer inspection of the changes in glycan abundance, for example of Fuc₁Hex₁HexNAC₁, shows that the abundance increases between 2-8 hours and then there is little change from 8-24 h, suggesting all of this O-glycan had been released within 8 h. This is similarly the case for Hex₁HexNAC₁, Fuc₁Hex₁HexNAC₂, Fuc₁Hex₂HexNAC₃ and Fuc₂Hex₂HexNAC₃. However, for Fuc₁Hex₂HexNAC₁, Fuc₁Hex₃HexNAC₃ and Fuc₂Hex₃HexNAC₂ the glycan signals increase in relative intensity between 2-8 h but then for these structures there is a decrease in abundance after 8 h, possibly suggesting that these glycans begin to degrade if incubated for longer than 8 h, Figure 52.

Overall, results showed that 16 h incubation resulted in the largest number of different glycan compositions released (a total of 44 glycans) while the fewest different compositions detected, 33, was after just 2 h. Comparing the relative abundance of each individual glycan across the entire time study showed that it was in the 8 h sample that the majority of glycans reached their highest relative abundance (Table 3). On 8 hours of incubation, 17 glycans reached their peak relative intensities; accounting for 38.6 % of the total of 44 different glycan compositions detected overall, Table 3.

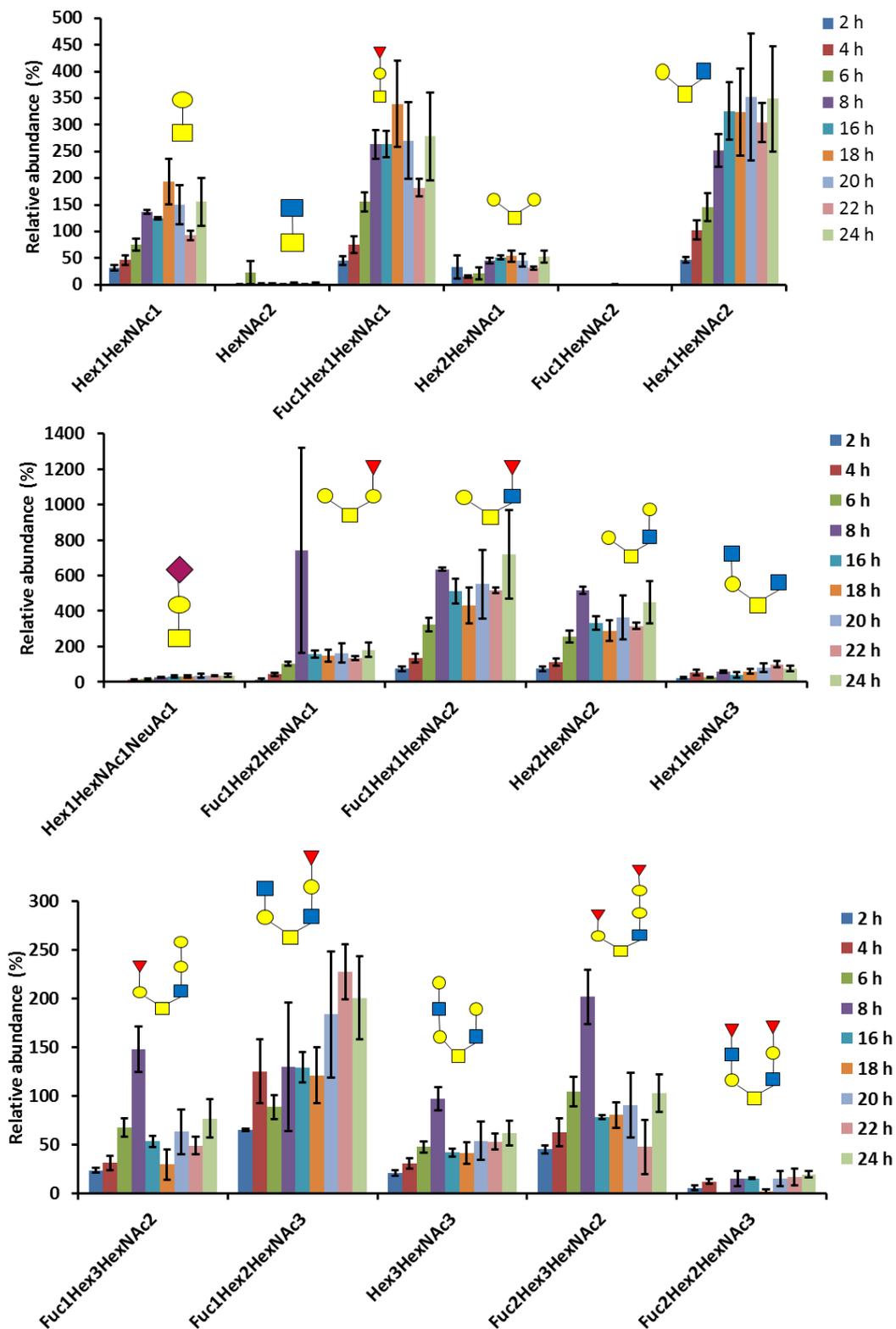


Figure 52: Relative abundance of permethylated O-glycans released from porcine stomach mucin over 2 - 24 h incubation with NH₄OH. Glycan relative abundance calculated as a percentage of the signal from a spiked internal standard.

Table 3: Number of different glycan compositions detected at range of times treated with NH₄OH

Incubation time (h)	2	4	6	8	16	18	20	22	24	Total
Number of glycans reaching their peak relative intensity	0	1	1	17	3	5	4	6	7	44
Total glycans reaching their peak relative intensity of the total (%)	0	2.3	2.3	38.6	6.8	11.4	9.1	13.6	15.9	100
Number of different glycan compositions detected	33	40	35	36	44	35	37	35	39	-

To see how the relative abundance of glycans changed between 8 and 16 h, additional mucin samples were prepared for O-glycan release at 8, 10, 14 and 16 h. After permethylating the released glycans and spiking with an internal standard, a mass spectrum was obtained and the relative abundance of glycans was calculated as a percentage of the spiked internal standard. It was found that 14 h of incubation resulted in the largest number of different glycan compositions released (a total of 34 glycans) while the fewest different compositions detected, 28, was after 16 h. Comparing the relative abundance of each individual glycan across the 8-16 h time study showed that it was in the 10 h sample that the majority of glycans reached their highest relative abundance. On 10 hours of incubation, 14 glycans reached their peak relative intensities; accounting for 38.9 % of the total of 36 different glycan compositions detected overall, Table 4. This study suggested that perhaps an incubation time of 8-10 h is optimal, compared with the 15-16 h as recommended by Rademaker et al²²¹.

Table 4: Number of different glycan compositions detected at range of times treated with NH₄OH

Incubation time (h)	8	10	14	16	Total
Number of glycans reaching their peak relative intensity	7	14	7	8	36
Total glycans reaching their peak relative intensity of the total (%)	19.4	38.9	19.4	22.2	100
Glycans detected	31	29	34	28	-

In a similar experiment, also by Sam Black supervised by the author, porcine stomach mucin was also used to study O-glycan release with the aid of sonication, to determine how sonication time affects the abundance of O-glycans released. O-glycans were released after 5 min sonication, 10 min sonication, 5 min sonication followed by 10 min rest in the sonic bath), two 5 min sonications each followed by 10 min rest periods in the sonic bath) and four 5 min sonications each followed by with 10 min rest in the sonic bath. O-glycans were permethylated and spiked with an internal standard before recording a mass spectrum. In total, 39 different glycan compositions were detected. Results for the four x 5 min sonication regime (a total of 20 min sonication to give a total of 60 min glycan release time) detected 19 glycan compositions at their peak abundance, accounting for 48.7 % of the 39 different compositions released, Table 5.

Table 5: number of glycans detected when treated with NH₄OH and a combination of sonication and rest periods at the incubation temperature

Glycan release time (min)	5	15	30	45	60	Total
Number of most dominant glycans	1	16	0	3	19	39
% dominant glycans of the total	2.6	41.0	0	7.7	48.7	100
Glycans detected	27	33	35	29	32	-

In general, the abundance of each glycan signal increased with increasing time of sonication/rest periods, (a few examples of the identified O-glycans are displayed in Figure 53). A closer inspection of the changes in glycan abundance, for example of Hex₁HexNAC₃, shows that after only 15 min the abundance of that glycan is at its highest compared to the

other time points. In addition, for Hex₁HexNAc₂ and Hex₃HexNAc₃ there is little change in the abundance of these glycans between 15 – 60 min suggesting the majority of those glycans are released by 15 min of sonication/rest.

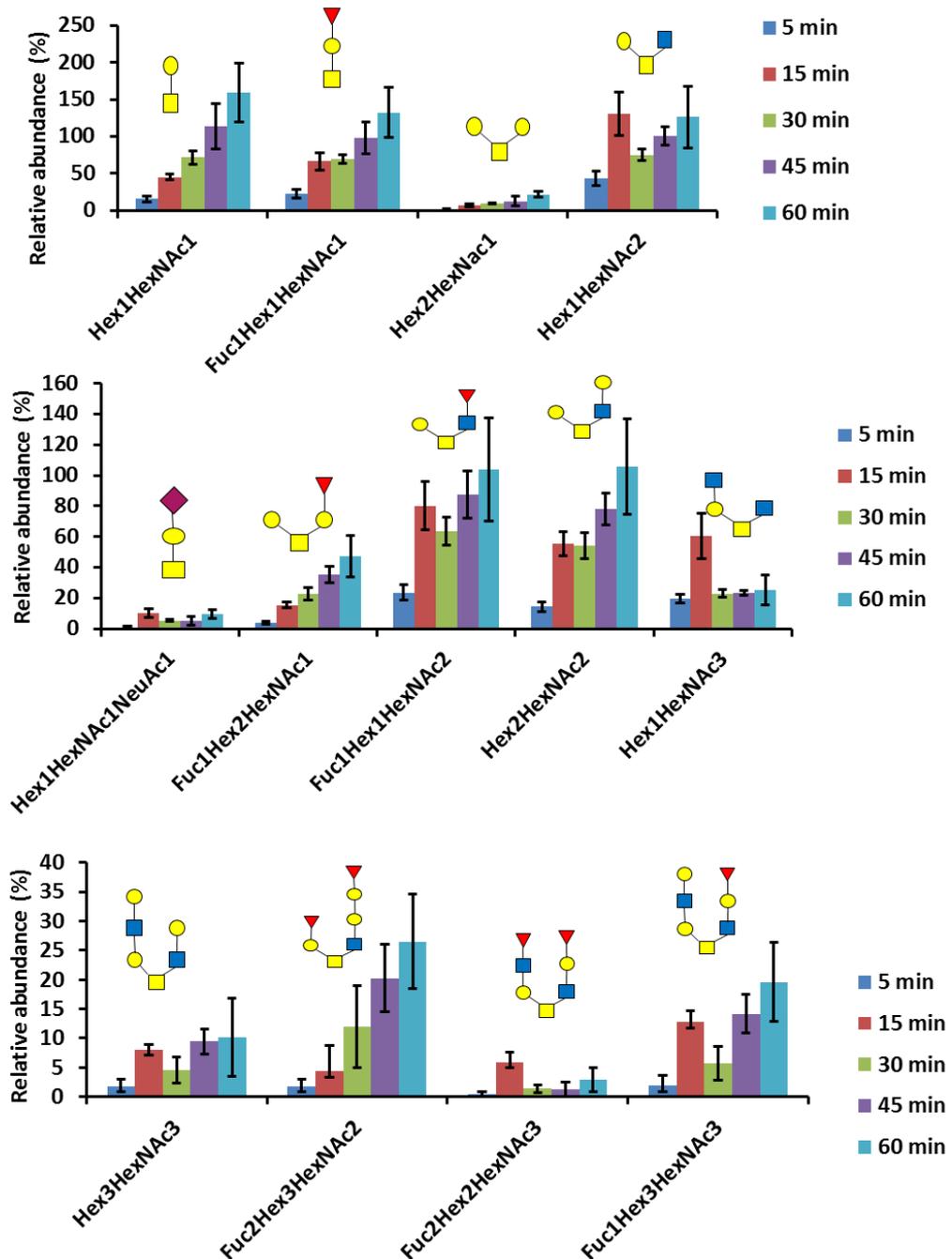


Figure 53: Relative abundance of permethylated O-glycans released from porcine stomach mucin from 5 min sonication – 60 min (alternating 5 min sonication followed by 10 min rest period) with NH₄OH. Glycan relative abundance calculated as a percentage of the signal from a spiked internal standard.

4.5. O-glycan release from Chinese hamster ovary cells (CHO)

CHO cells are easily grown in culture and are readily manipulated, and so these are a useful cell line to use to generate models of congenital disorders of glycosylation. These cells were therefore chosen to test the developed O-glycan release method for its applicability to cultured cells; the aim was to test its robustness for glycomic analysis using a small number of cultured cells (around 1 million).

Wild type CHO cells and a mutant cell line, IdIC, bearing a genetic mutation in Cog2 of the COG complex (named IdI because of defects in the uptake of low-density lipoprotein into the cell), were prepared for O-glycan release. The cell cultures were each grown on 10 cm culture dishes (giving approximately 2×10^6 cells). Cells were harvested and treated with lysis buffer containing SDS, before exchanging the SDS for urea and carrying out non-reductive β -elimination with NH_4OH in a heating block at 45°C , for 16 h, to release the O-glycans. O-glycans were permethylated before MALDI-MS analysis. A MALDI-FT-ICR mass spectrum was obtained in which O-glycans were identified from their characteristic signals¹¹⁷ at m/z 879, 1083 and 1240, from both WT and IdIC CHO cells, Figure 54.

At first glance it would appear that the mass spectra look very similar for the WT and the mutant cell lines, although the signal:noise levels are better in the glycans released from the WT cells. An internal standard, deuteromethylated maltotetraose, was spiked into each sample and the abundance of O-glycans calculated with respect to that. This revealed that the overall signal levels for O-glycans were of lower abundance from the IdIC sample than the WT CHO sample, shown in Figure 54. Kingsley et al⁸⁶ reported IdIC mutants (along with IdIB and IdID) to display aberrant glycosylation, with incomplete processing of high mannose *N*-glycans to the complex type. In addition lowered sialic acid content in both *N*- and O-glycan types was found through SDS-PAGE experiments and sialidase treatment. IdI mutants were found to have a rapid turnover of the IdI receptor at the cell surface. As the evidence suggests in Kingsley's paper along with the data shown here, the aberrant glycosylation and decrease in abundance of glycosylation could perhaps be linked to the reduced ability of the IdI receptor to take up IdI into the cell.

While the O-glycan profiles of CHO cells are not particularly interesting since there are only three O-glycans¹¹⁷, this analysis serves to show that it is possible to use the filter-based O-

glycan release approach to reveal small changes in O-glycan profiles of cultured cells when genetic modifications have been introduced, Figure 55.

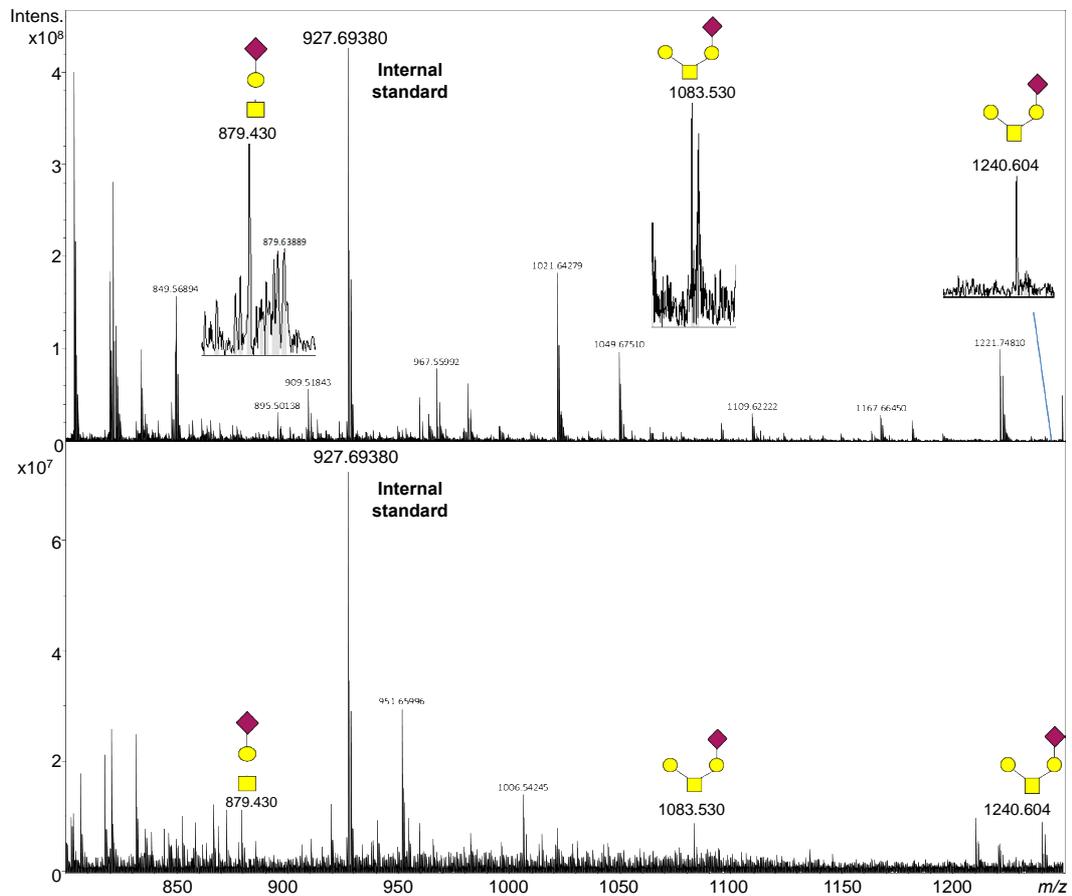


Figure 54 MALDI-FT-ICR mass spectrum of permethylated O-glycans released and isolated from CHO cells. Glycans ionised as $[M + Na]^+$. A) CHO IdC b) WT CHO.

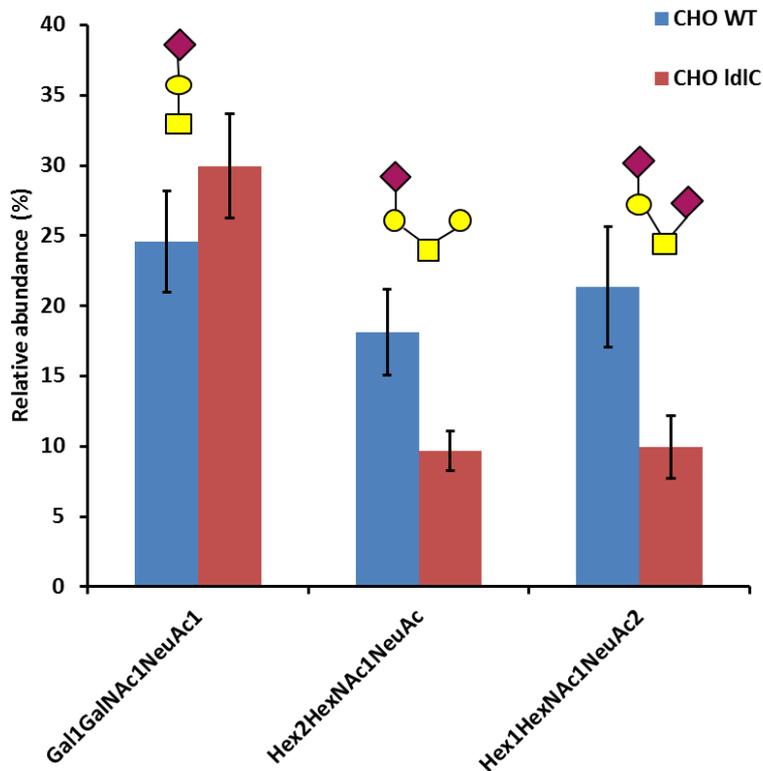


Figure 55: Comparison of the relative abundance (%) of O-glycans in WT CHO cells and IdIC CHO cells.

4.6. Conclusions and future work

A method has been developed, based on FANGS, to release and isolate O-glycans from glycoproteins for subsequent mass spectrometric analysis. The method makes use of a centrifugal filter as a simple reaction vessel in which glycoproteins can be treated with ammonium hydroxide solution to release the O-glycans by non-reductive β -elimination. Following centrifugation, glycans can be retrieved in the filtrate, as the released glycans pass through the filter membrane while proteins are large enough to remain above the membrane and so are retained. The isolated O-glycans can then be permethylated for mass spectrometric analysis. The method can be conveniently combined with the established FANGS protocol, to release, isolate and analyse both *N*- and O-glycans using one sample, in the same filter unit, enabling analysis of the whole glycome from the same sample.

This work makes use of an internal standard, showing that it is suitable for enabling relative quantification of glycans by MALDI mass spectrometry. The internal standard used in this work was permethylated maltotetraose. This was chosen as it is a chain of hexose units and

deuteropermethylation meant that it gave a unique mass not expected to be in a glycan sample. However, since an internal standard should be as closely related to the target analyte as possible it would be better to have a glycoprotein glycan as an internal standard. For studying mammalian samples, using a glycan that contains a xylose constituent would also give a unique structure, as mammalian glycans do not contain xylose. This glycan could be synthesised to be linear which would also be useful since most glycans are branched structures. In addition, a true internal standard would follow the sample analyte the whole way through sample preparation. In this work the internal standard was introduced into the glycan samples just before spotting onto the MALDI target plate. To be able to follow the sample throughout sample preparation, as a true internal standard should, the internal standard would need to be introduced on solubilising the cell lysate/glycoprotein sample in SDS. This would mean that the internal standard would need to be a glycoprotein so that it stays above the filter membrane before glycan release. The standard would then need to be released on either treatment with PNGase F for *N*-glycan release or NH_4OH for *O*-glycan release. This could be a very complex internal standard to generate since the glycoprotein standard would need to contain a single *N*-glycan and a single *O*-glycan or two separate glycoprotein standards would be needed.

A very novel mass spectrometric technique, 2D-FT-ICR-MS, could be a useful tool for glycomic analysis. 2D-FT-ICR-MS is being developed in Peter O'Connor's group at the university of Warwick. O'Connor's group have already demonstrated the use of 2D FT-ICR MS with both small molecules and larger molecules such as collagen digests. Since glycan samples often contain a large mixture of different structures and isomers, 2D FT-ICR MS could be a convenient method of simultaneously collecting MS/MS data on each glycan in one experiment rather than having to carry out many separate MS/MS experiments, and may also have the potential to be used as a quantitative tool. Initial collaborations with the O'Connor group in this area have begun, following discussions between the author and Prof O'Connor after one of the author's presentations.

Further work on the one-pot method development is demonstrated in chapters to follow, using mesenchymal stem cells and human urothelial cells, to demonstrate how the method can be used to monitor changes in glycosylation in wild-type/healthy cells compared with those that have differentiated or been genetically manipulated, using a small number of cells to make it possible to study replicates.

Chapter 5: Characterisation of antibody *N*-glycans, released by FANGS, by a combination of two-dimensional nuclear magnetic resonance spectroscopy experiments.

5. Characterisation of antibody *N*-glycans, released by FANGS, by a combination of two-dimensional nuclear magnetic resonance spectroscopy experiments.

The work in this section was carried out over a seven week work placement at Bristol-Myers Squibb, Lawrenceville, New Jersey.

5.1. Antibody research and glycome analysis

Antibody research is now a major area of research for the pharmaceutical industry^{14,222,223}, as part of the development of protein biotherapeutics. Research and development is underway to create monoclonal antibodies for treatment of many serious diseases. Glycosylation of antibodies increases product heterogeneity, affecting both structure and function²²⁴. Due the non-templated nature of glycan biosynthesis, glycosylation of antibodies can cause variation between batches of biotherapeutics and therefore it is important to know/understand the glycan profiles, since they can affect both the safety and effectiveness of a drug product.

Antibodies belong to the globular protein family, immunoglobulins²²⁵. Antibodies bind to antigens, causing pathogens to be destroyed by the immune system. The antibody structure is a “Y” shape consisting of four polypeptide chains; two “light” chains (shorter chains of amino acids) and two “heavy” chains (longer chains of amino acids). The four chains are linked together by disulfide bonds²²⁶. The heavy chain contains a hinge region that offers the antibody flexibility to help binding to antigens on the surface of pathogens. There are two main parts to most antibodies: the Fab (fragment antigen binding) region and the Fc (fragment crystallisable) region. The Fab region contains the antigen binding site²²⁷ and the Fc region contains the glycosylation sites, and has the function of being able to bind to cell-surface receptors and activate or enhance the immune response. The amino acid sequence in the Fab region is variable and so this is also called the variable region; the amino acid chain closest to the antigen-binding site varies the most between different antibodies. The amino acid sequence further down the polypeptide chain and leading into the Fc region is known as the constant region; where the amino acid sequence is similar amongst different antibodies, Figure 56. The differences in amino acid sequences amongst the antibodies gives

each antibody the ability to bind to antigens with specificity and also influences the biological activity of an antibody.

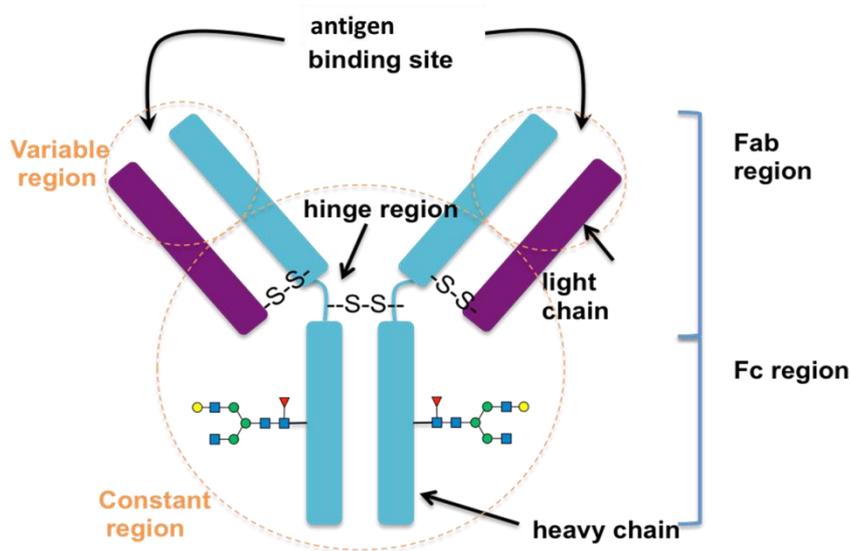


Figure 56: Generic structure of an antibody. An antibody consists of four polypeptide chains (two light chains and two heavy chains) each held together by disulfide bonds. The Fab region contains the antigen binding site. The Fc region contains glycans and can bind to cell surface receptors and activate the immune response.

When an antibody binds to an antigen to form an antibody-antigen complex this activates 'complement'²²⁸, a system of plasma proteins, as the fc region of the antibody binds to the first protein in the complement cascade, C1q²²⁹ (note: C refers to complement protein, the numbers indicates an additional protein in the complement cascade and the letters q,r and s are subsets of a particular protein). This then activates C1r and C1s which cleave C4 and C2 to form a protease, C3 convertase, that cleaves C3 into C3a and C3b. C3b marks the pathogen for destruction by phagocytes (by acting as phagocyte receptors) and so the phagocytes can engulf the pathogen. C3b also binds to C3 convertase to form C5 convertase that then cleaves C5 into C5a and C5b. C5b is the first protein of the membrane attack complex, which consists of C5b bound to C6, C7, C8 and C9. The membrane attack complex binds to the pathogen cell membrane (only pathogens that do not have a cell wall), creating pores in the surface of pathogens, which leads to cell lysis. C3a and C5a are mediators of inflammation and recruit phagocytes to the pathogen. This type of immune response is humoral immunity. There is also cell-mediated immunity involving effector cells²³⁰, that

contain fc receptors where the fc region of an antibody can bind, such as on macrophages, lymphocytes²³¹ and natural killer cells. One mechanism is antibody-dependent cell-mediated cytotoxicity (ADCC)^{232,228}. Here the antibody is bound to the antigen on the invading pathogen and the fc region of the antibody will bind to natural killer cells. Natural killer cells are then activated and release cytotoxic chemicals to destroy the pathogens²³³.

There are five classes of immunoglobulins^{225,234,235}, based on their heavy chain structures; IgG, IgA, IgM, IgD and IgE. IgG is the most common antibody found in blood plasma²³⁴ and is the most widely used antibody in biotherapeutics²³⁵. There are four subclasses of IgG; IgG1, IgG2, IgG3 and IgG4, named in order of relative abundance in plasma, with IgG1 being the most abundant. The subclasses differ according to their amino acid sequence in the constant region of the heavy chains, including differences to the number and location of disulphide bonds²³⁶ and the length of the hinge region²³⁵. IgA is the main antibody found in bodily secretions (e.g. tears, saliva and mucus). IgM is the largest antibody and remains in the blood to destroy bacteria. IgD is membrane bound on B cells, and IgE, although found only in trace amounts in blood, triggers allergic responses.

In developing therapeutic proteins such as monoclonal antibodies, it is essential that full protein characterisation is carried out to establish critical quality attributes. These are the aspects of the therapeutic that could affect the stability, efficacy, shelf-life, bioavailability, etc. For example, the number and location of disulfide bonds affect the folding, shape and stability of the protein²³⁷ which, if changed, may mean that the protein is unable to perform its intended function. The secondary structure of a protein, made up of α -helices and β -sheets, folds and holds together by non-covalent interactions to give the tertiary protein structure. Changes to conformation and folding, essential for protein function, can be monitored by methods such as circular dichroism²³⁸. Unpaired cysteines can cause intermolecular binding and aggregation²³⁹, which can affect solubility and immunogenicity (the ability to induce an immune response).

In addition, peptide mapping/fingerprinting analysis^{240,241}, using LC-MS, is carried out for quality control between batches. In addition peptide mapping is used for monitoring protein degradation, which can happen over time in storage, through handling in sample preparation, and due to changes in pH and temperature. In addition, in the early stages of drug development degradation profiles can be established through forced/accelerated

degradation studies²⁴² to help select suitable drug candidates and in formulation development. Common protein degradation/modifications include oxidation of methionine^{243,244} and deamidation of asparagine²⁴³ to succinamide, which can then undergo hydrolysis to form aspartic and isoaspartic acid^{245,246}. These changes can also be monitored by isoelectric focussing²⁴⁷ and SDS-PAGE.

Glycosylation profiling is important, as the structures present influence efficacy of the drug, stability^{19,20}, solubility, the half-life of the drug (oligomannose glycans are associated with faster clearance from serum than other glycans)^{21,22} and therefore affect dosing. Glycans can also affect immunogenicity^{248,249,250} and even though glycans are usually found in the Fc region of an IgG they can sometimes be found in the Fab region, as reported for IgA²⁵¹, and can therefore influence antigen binding. Sialylated glycans are known to have an anti-inflammatory effect²⁵², and it has also been shown that high levels of sialylation correlate with lowered affinity for Fc receptors and reduction in antibody-dependent cell-mediated cytotoxicity²⁵³.

Analysis of glycoproteins, glycopeptides and glycans is typically carried out using a combination of methods including mass spectrometry or high performance liquid chromatography combined with mass spectrometry, and can require sample preparation such as labelling and derivatisation. Mass spectrometry is good for characterising glycan or peptide compositions, and can be used to determine structure using tandem mass spectrometric methods. It is particularly appropriate for glycosylation site analysis²⁵⁴, glycan sequencing, and assignment of the location of branching. What mass spectrometry cannot readily do is distinguish between the isomeric hexose units glucose, galactose and mannose, or the *N*-acetylhexosamines (*N*-acetylglucosamine and *N*-acetylgalactosamine), nor is it the method of choice for determining anomericity or routinely assigning linkage position in glycan chains.

NMR is useful for identifying glycan linkages, stereochemistry and anomericity, and can be very powerful in combination with mass spectrometric methods to provide complementary information²⁵⁵.

Tools are needed to assess the reproducibility of glycosylation in batches of biologics to help determine the glycans that tend to vary in population between batches. Schubert et al²⁵⁶

recently published an approach for intact glycoprotein analysis of the monoclonal antibody Cetuximab, using a suite of 2D-NMR experiments. Combining this set of experiments allowed establishment of monosaccharide connectivities and types of glycosidic linkages, based on chemical shifts and magnitudes of J-couplings. In addition, Wiegandt et al²⁵⁷ described a method for analysis of PNGase F released *N*-glycans using LC-MS/MS combined with 1D ¹H-NMR. They used only 60 µg of protein, not enough for 2D NMR but this amount is enough for 1D NMR, which is more sensitive. Analysis was carried out using three dimensional cross coupling correlation^{258,255}, enabling them to distinguish glycans which differed only in the branching position of one monosaccharide.

The aim of the work described in this chapter was to apply the approach of Schubert²⁵⁶ for analysis of an intact glycoprotein to an IgG at BMS, and extend this to a tryptic digest, as well as to the *N*-glycans released and isolated from the glycoprotein using FANGS¹⁴¹, for analysis by a suite of 2D-NMR techniques. The proposal was that performing 2D NMR would eliminate the need for prior chromatographic separation of the glycans as carried out by Wiegandt et al²⁵⁷. An additional aim was to investigate the use LC-MS/MS combined with NMR as described by Wiegandt et al, but to use Schubert's 2D NMR instead of 1D NMR approach for more detailed structural information. Matlab software was to be investigated to separate spectra for similar but distinct glycan structures.

5.2. Nuclear magnetic resonance spectroscopy (NMR)

Protons, electrons and neutrons all carry a fundamental property, spin. Each unpaired electron, proton or neutron in an atom has a spin, denoted by the spin quantum number l , of $l = \frac{1}{2}$. An NMR spectrum can be generated for nuclei with non-zero spin, most commonly the proton, ¹H (an unpaired proton therefore with a spin of $l = \frac{1}{2}$) and the stable carbon isotope, ¹³C (an unpaired neutron therefore with a spin of $l = \frac{1}{2}$). If the number of neutrons and protons is even, then the nuclei have no overall 'spin' and an NMR spectrum cannot be obtained. Helium has an even number of both protons and neutrons and therefore their spins cancel, giving no overall spin, and a NMR spectrum cannot be obtained.

When an external magnetic field (B_0) is applied, the spin of the proton, which has a magnetic moment μ , aligns with B_0 either in the lower energy, spin up, state or the higher energy, spin down, state. The protons can absorb a photon of energy, where the energy is related to its

frequency, ν , by planck's constant, h , Equation 11, to make a transition from the lower to the upper energy state when the energy of the photon is in resonance with the energy difference between the two energy states²⁵⁹.

Equation 11

$$E = h\nu$$

The magnetic moment of a nucleus can be described as:

$$\mu = \gamma J$$

where J is the spin angular momentum and γ is the gyromagnetic ratio. J is quantized and can only take discrete energy levels:

$$J = m \frac{h}{2\pi}$$

where m is the magnetic quantum number and can take the values of $-l$ to l , so for a nucleus with $l = 1/2$, m can be $+1/2$ or $-1/2$, shown in energy level diagram, Figure 57.

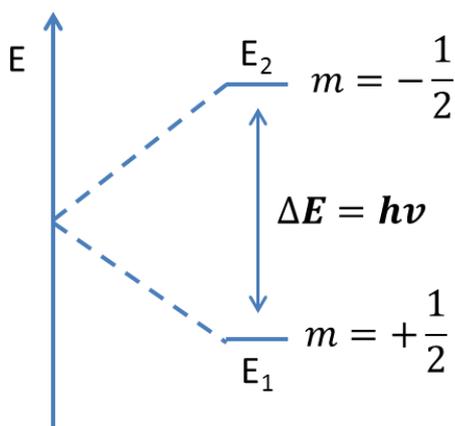


Figure 57: Energy level diagram

The signal observed in an NMR spectrum results from the difference between the energy absorbed, moving from a lower to a higher energy spin state, and the energy emitted when moving from a higher to a lower energy spin state. The energy difference can be shown by Equation 12, where γ is the gyromagnetic ratio, a constant for a particular nucleus. The NMR signal is proportional to the energy difference between the spin states, and because the

energy difference is very small, this can produce a weak signal. The signal can be enhanced by increasing the magnetic field strength and by increasing the sample concentration.

Equation 12:

$$\Delta E = \frac{\gamma h B_0}{2\pi}$$

An NMR experiment does not just measure one spin from one nucleus; it instead measures many spins, from a number of nuclei. The sum of all the magnetic moments in the nuclei of a molecule is its magnetization, and produces a net magnetization vector²⁶⁰ that points towards the externally applied magnetic field, B_0 . To record an NMR spectrum, a radiofrequency (RF) field pulse is applied perpendicular to the magnetic field²⁶⁰. Without the RF field, all the spin packets rotate about the z-axis in the general direction of B_0 , but non-coherently. When the pulse ends, the nuclei relax and return to their equilibrium positions and the signal decays.

During the RF pulse, a magnetic field, B_1 is induced and all the nuclear spins with a frequency matching this induced magnetic field feel the effect of that field (i.e. they are in resonance with the RF field). The energy absorbed by the nuclear spins causes transitions from lower to higher energy and higher to lower energy. The effect of the RF pulse is to bring about coherence of the spin packets by moving them all to the direction of the pulse, rotating about the axis of the RF field²⁶⁰, Figure 58a. In doing so this creates sine waves in the detector, and when the pulse ends the nuclei relax, and the signal decays in a spiral trajectory, known as free induction decay (FID). The FID is converted to a frequency domain spectrum by Fourier transformation²⁶¹, Figure 58b.

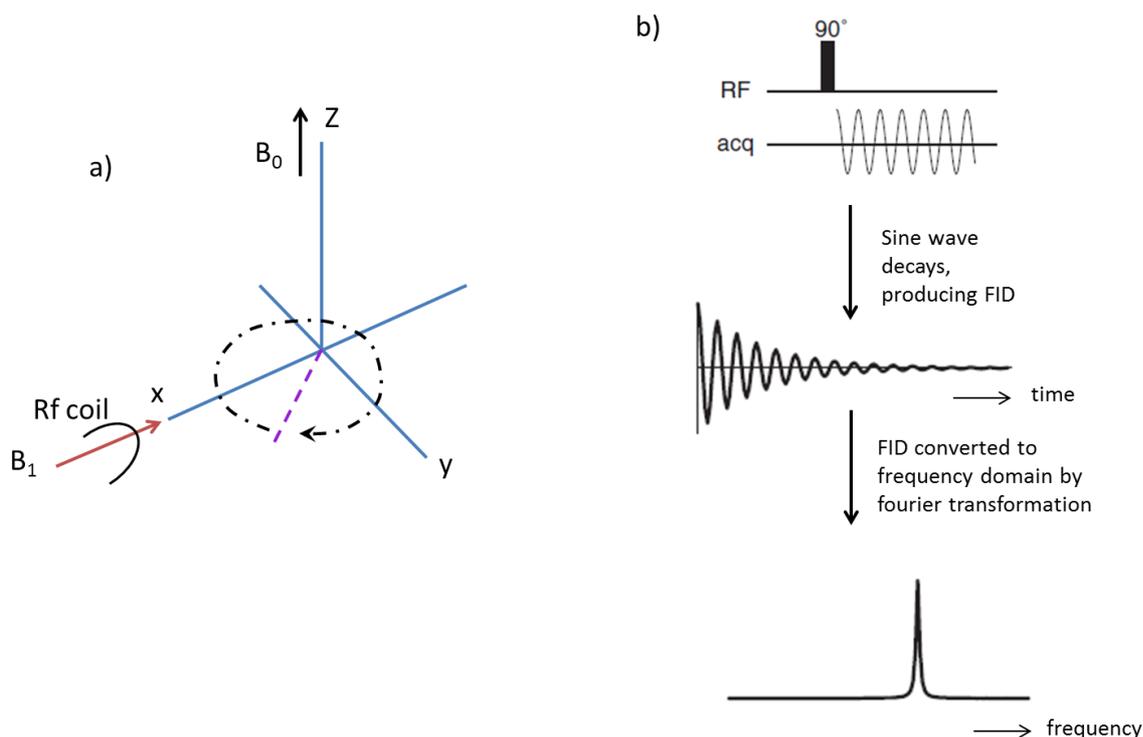


Figure 58: NMR pulse to signal. a) RF field is applied, inducing B_1 magnetic field, perpendicular to B_0 . b) the induced magnetic field from the RF field creates sine waves at the detector as B_1 oscillates about the x,y axis. When the pulse ends the signal decays, known as free induction decay (FID) and the FID is converted to a frequency domain by Fourier transformation²⁶¹.

5.2.1. Instrumentation and experimental set-up

An NMR instrument comprises a magnet, a probe, an RF source, amplifiers, analogue/digital converters, lock and shim systems and a computer system. In the bore of the magnet is a set of shim coils, and within the shim coils sits the probe. The probe contains the RF coil, sample spinner and temperature control. The probe accepts the sample, sends the RF pulse to the samples and detects resulting signals.

Samples for NMR analysis are prepared in a deuterated solvent and should not contain solid particles as this can affect the homogeneity of the magnetic field. Once the sample is inserted into the probe, the probe needs tuning to get the most effective detection from the sample. Tuning is achieved by adjusting two capacitors on the RF probe²⁶², one being the matching capacitor and the other the tuning capacitor. The latter changes the resonance

frequency of the RF coil and the former matches the effective resistance of the loaded probe to the cable from the spectrometer.

The system then needs to be “locked” to keep the magnetic field stable, to produce a high-resolution spectrum. A deuterium lock measures the resonance frequency of the deuterium signal from the solvent the sample is dissolved in, and corrects the magnetic field, B_0 , so that the resonance of the frequency of the deuterium signal remains constant. Also, the deuterium lock provides the $\delta = 0$ reference, so trimethylsilane (TMS) is not required.

Following this, the magnet is “shimmed”. The shim system²⁶² is a series of small coils, each of which has a current running through it, to produce a small magnetic field, to correct for minor spatial differences in the magnetic field. Differences could arise from variations in the thickness of the sample tube or sample permeability. A homogenous magnetic field means it is possible to measure small differences in frequencies and also prevents line broadening.

5.2.2. Chemical shift and spin-spin coupling

When an atom is placed into a static magnetic field B_0 , the electrons of the atom rotate about the axis of the magnetic field. This creates a current, which then induces a magnetic field at the nucleus, which opposes the static magnetic field, known as the effective field. The electron density surrounding each nucleus in an atom of a molecule varies, according to the types of functional groups/bonds in a molecule. The effective field induced at the nucleus of different atoms will cause shifts in frequency for the NMR signal, known as chemical shift. Chemical shift allows magnetically inequivalent nuclei in a molecule to be distinguished. Nuclei are more exposed to/deshielded from the applied magnetic field by electronegative neighbouring atoms and will have higher chemical shifts as a result, measured in ppm. Nuclei are more shielded from the applied magnetic field when the neighbouring atoms have lower electronegativity and will therefore have a lower chemical shift.

Nuclei in the same chemical environment are called ‘equivalent’. In a molecule, not all nuclei are equivalent and so those which are close to each other in space, or through bonding, will influence the other’s effective magnetic field, known as spin-spin coupling (or J coupling), resulting in characteristic splitting of the peaks in an NMR spectrum. For example, a proton that is close to one other non-equivalent proton, through bonding, appears as a doublet

peak and one that is close to two non-equivalent protons appears as a triplet. The distance between the split peaks can be measured to give a coupling constant, J . Coupling constants can be used to provide spatial information about two protons coupling constants are larger for bonds that are closer together and the value decreases as the bonds become further away. For example, for vicinal coupling, a coupling through three bonds, if the protons are axial to each other, they produce a larger coupling constant than two protons that are equatorial to each other, while geminal coupling (through two bonds) produces a larger coupling constant than vicinal couplings).

5.2.3. Two-dimensional NMR spectroscopy

Two-dimensional NMR introduces a second frequency axis that allows correlation of the two frequencies. Firstly a pulse is applied and the resulting magnetisation is left to evolve for a set time, then a second pulse is applied (the mixing time). After the mixing time, the signal is recorded by FID followed by two fourier transformations to give a two dimensional spectrum as a function of two frequencies²⁶³. Homonuclear two-dimensional spectra correlate frequencies of the same nuclei in the two dimensions, e.g. $^1\text{H}^1\text{H}$, and contain a series of peaks along the diagonal, that represents the one dimensional spectrum, in addition to peaks that sit away from the diagonal, which are known as cross peaks, from spins that correlate. Heteronuclear two-dimensional spectra contain frequencies of different nuclei in each of the two dimensions, e.g. $^1\text{H}^{13}\text{C}$.

5.2.3.1. Total correlation spectroscopy

Total correlation spectroscopy, TOCSY, is a through bond correlation technique and plots correlations between all protons within a spin system, providing there is coupling between every intervening proton. This means that correlations can be observed between distant protons and not just between geminal (through 2-bond) or vicinal (through 3-bond) protons. This is useful for characterizing the monosaccharide units of glycans. This correlation cannot be transferred across to an adjacent monosaccharide or neighbouring amino acid in a peptide because the oxygen atom in the glycosidic bond disrupts the TOCSY transfer. Transfer steps around the monosaccharide ring can be controlled by altering the spin-lock time, a pulse sequence that keeps the magnetization in the transverse phase for a specified period of time. A short spin-lock time gives one to three step transfers. A longer spin-lock

time can be used to give up to six transfer steps. For example, in galactose resonances are observed between the anomeric H1 with H3 and with H4, Figure 59a and b, respectively.

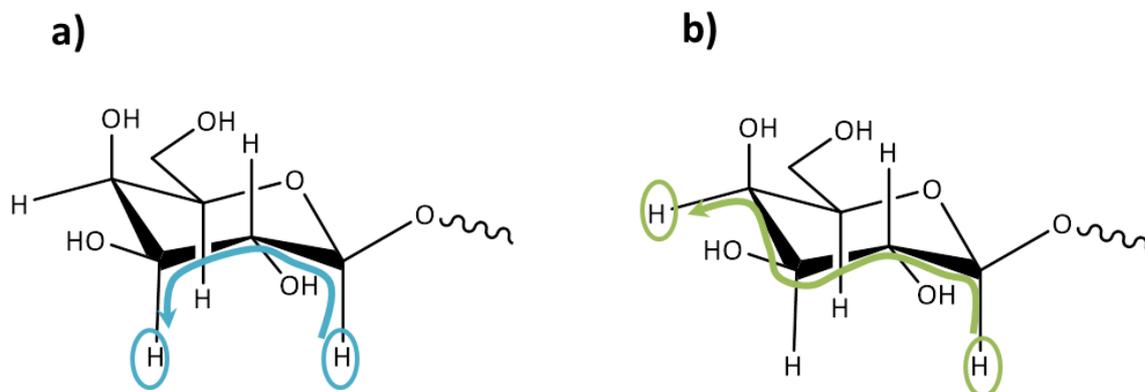


Figure 59: [$^1\text{H}^1\text{H}$]-TOCSY. Example of the type of $^1\text{H}^1\text{H}$ correlations that could be observed in a [$^1\text{H}^1\text{H}$]-TOCSY experiment for galactose, e.g. a) H1 with H3 and b) H1 with H4.

5.2.3.2. Heteronuclear single quantum correlation

Heteronuclear single quantum correlation, HSQC, determines the single bond correlations between two different nuclei, e.g. ^1H and ^{13}C . This can be useful for assigning the correlations between the anomeric carbon and anomeric proton in monosaccharide units to determine what monosaccharide units are present, Figure 60.

In addition, dept-HSQC can be carried out (where dept is distortionless enhancement polarisation transfer). The ^1H is excited and magnetisation is transferred to carbon. This transfer alters the amplitude and size of the ^{13}C signal in relation to the number of ^1H s directly attached. For example, a CH_2 resonance becomes a negative signal and a signal for a carbon with no directly attached ^1H s disappears. This is particularly useful for assigning C6 on a hexose ring, as C6 is a CH_2 .

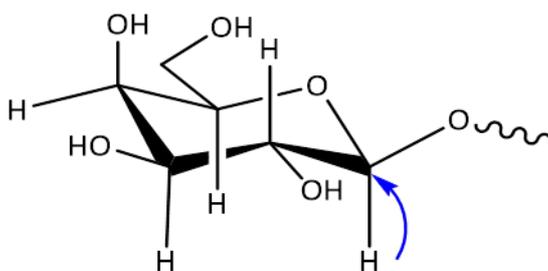


Figure 60: $^1\text{H}^{13}\text{C}$ -HSQC. Example of a single bond correlation observed in a [$^1\text{H}^{13}\text{C}$]-HSQC experiment of galactose, for the anomeric carbon with the anomeric proton.

5.2.3.3. Heteronuclear multiple bond correlation

Heteronuclear multiple bond correlation, HMBC, determines multiple bond correlations between two different nuclei e.g. ^1H and ^{13}C . This can be useful for assigning the correlations between the anomeric proton, and a carbon multiple bonds away. This method is useful to assign carbons around a ring system and to assign linkages across a glycosidic bond in a glycan, Figure 61.

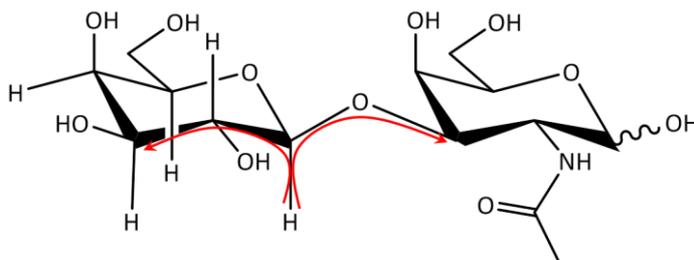


Figure 61: $^1\text{H}^{13}\text{C}$ -HMBC. Example of multiple bond correlation in a $^1\text{H}^{13}\text{C}$ -HMBC experiment of glycans. HMBC can provide linkage information from the anomeric proton to the carbon at the other side of the glycoside bond on the neighbouring monosaccharide unit. In addition, HMBC can provide multiple bond correlations of ^1H and ^{13}C around a single ring system.

5.2.3.4. Heteronuclear multiple quantum correlation – correlation spectroscopy

$^1\text{H}^{13}\text{C}$ -HMQC-COSY determines bond correlations two bonds apart between two different nuclei (^1H and ^{13}C). This is useful for assigning correlations between a carbon and the ^1H on the adjacent carbon, which can offer more confidence when assigning a ring system in glycans, Figure 62.

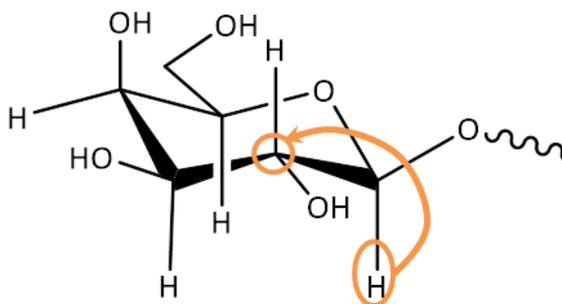


Figure 62: $^1\text{H}^{13}\text{C}$ -HMQC-COSY. Example of the type of correlation that could be observed in galactose, showing the correlation of the anomeric proton with C2 in the ring.

5.2.3.5. Nuclear Overhauser effect spectroscopy

Nuclear Overhauser effect spectroscopy, ^1H -NOESY, determines through space correlations of nuclei that are close in space. This can be used to show the presence of a HexNAc by NOE of the CH_3 in the amide acetyl group, with the anomeric proton, Figure 63a. In addition NOESY can also be used to determine which monosaccharide units are close in space, for example through interaction of the methyl group on a fucose with a ^1H on an adjacent monosaccharide, Figure 63b.

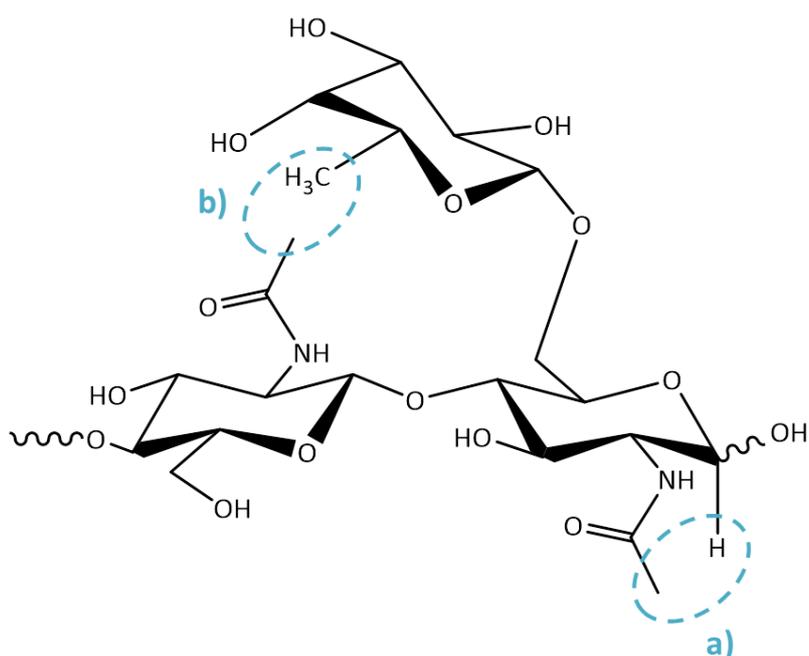


Figure 63: ^1H -NOESY. Example of the type of nOes that could be observed on a glycan. a) the nOe between the CH_3 in the amide acetyl group of a HexNAc, with the anomeric proton. b) through space interaction of the methyl group on a fucose with a ^1H on an adjacent monosaccharide.

5.3. Results and discussion

A method has been developed to prepare, from antibodies: the denatured 'intact' protein, a tryptic digest, or enzymatically released *N*-glycans, for subsequent analysis of the antibody glycome by NMR. The method makes use of a centrifugal filter as both a simple sample clean-up device¹⁴³ and reaction vessel, where the antibody can either be denatured and then concentrated for NMR analysis of the protein, denatured then treated with trypsin¹⁴³ followed by NMR analysis of the tryptic digest, or denatured and treated with PNGase F to

release the *N*-glycans, as for FANGS¹⁴¹. Centrifuging the filter can retrieve the glycans, as the released glycans pass through the filter membrane and are retrieved in the filtrate, while proteins are large enough to remain above the membrane and so are retained.

The antibody used for this work at BMS, is an IgG and it is well known that IgGs have one site of glycosylation²³⁴. Following tryptic digestion, the peptide expected to contain the site of glycan attachment²³⁴ has the sequence EEQYNSTYR, since trypsin cleaves on the carboxyl side of K or R except when either is followed by P, Figure 64. The four most abundant/ most common glycans expressed on an IgG consist of oligomannose and complex structures bearing zero, one or two terminal galactose units. The oligomannose structure does not contain a core fucose^{234,264,265,266}, Figure 65, and there are no O-glycans.



Figure 64: Primary amino acid sequence of the Fc region of the IgG around the glycosylation site. Red lines depict tryptic cleavage sites. The N, in orange, is the glycosylation site²³⁴.

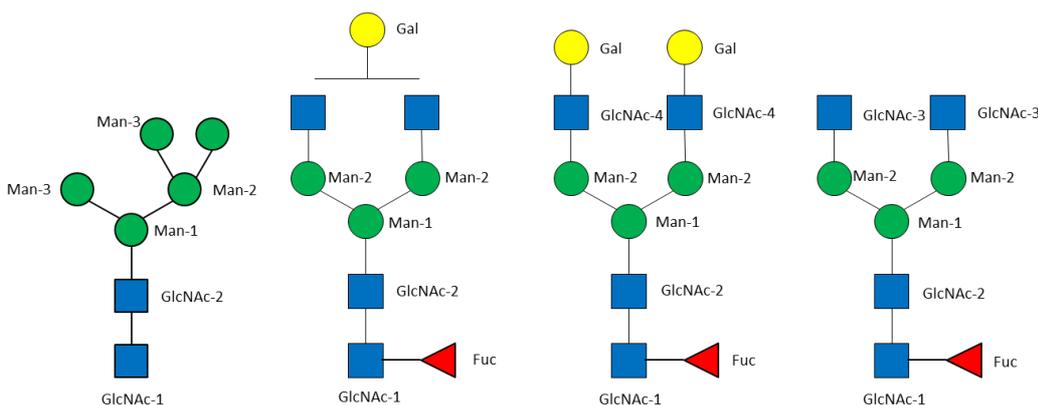


Figure 65: *N*-Glycans found in the IgG^{264,265,266}.

Samples of IgG were prepared for ‘intact’ protein analysis, tryptic digestion and *N*-glycan release. Each glycoprotein sample was denatured in 7 M urea/Tris buffer, before reducing and alkylating. The 7 M urea was carefully diluted to 2 M urea, with 50 mM ammonium bicarbonate. Careful dilution was needed to maintain the protein in solution while reducing

the concentration of urea sufficiently to avoid denaturing trypsin or PNGase F for carrying out tryptic digestion or *N*-glycan release. The sample for intact protein analysis was concentrated in the filter unit and then was ready for NMR analysis. The tryptic digest and the released *N*-glycans were purified using SPE prior to NMR analysis.

The released *N*-glycans were subjected to a suite of 2D NMR experiments to characterise them. Experiments carried out were [$^1\text{H}^{13}\text{C}$]-HSQC, [$^1\text{H}^{13}\text{C}$]-dept-HSQC, [$^1\text{H}^1\text{H}$]-TOCSY, [$^1\text{H}^{13}\text{C}$]-HMBC, [$^1\text{H}^1\text{H}$]-NOESY, [$^1\text{H}^{13}\text{C}$]-HMQC-COSY, and [$^1\text{H}^1\text{H}$]-dqf-COSY. Results depicted in tables 5-12 show the assigned chemical shifts from the released *N*-glycan spectra.

Firstly, the [$^1\text{H}^{13}\text{C}$]-HSQC data for the denatured intact protein, tryptic digest and released glycan samples were compared. It is clear that the released glycan data show the simplest spectra as there is no longer interference from peaks that correspond to the amino acids/peptides. The section of the spectra that offers the easiest characterisation of the glycans is the anomeric region, with ^1H chemical shifts between $\delta = \sim 4.4\text{-}5.5$ ppm and ^{13}C chemical shifts between $\delta = \sim 80\text{-}110$ ppm. The anomeric region reveals the resonances indicating the connection between the anomeric carbon and the ^1H on the anomeric carbon, and gives well resolved peaks that are not masked by peptide/protein signals, Figure 66.

A closer inspection of the anomeric region of the [$^1\text{H}^{13}\text{C}$]-HSQC spectra reveals well resolved peaks in all three spectra: intact protein, tryptic digest and released *N*-glycans. Assignments for each of the anomeric ^1H and ^{13}C peaks on fucose, galactose, *N*-acetylglucosamine and mannose were established, Figure 67, with the help of a 'glycosciences' chemical shift predictor²⁶⁷, the data from which are presented, with the assignments, in tables 1-8. In most instances the chemical shifts obtained were well matched with the predicted chemical shift, with the exception being for GlcNAc-1. This gave a higher than expected anomeric ^1H chemical shift and a much lower than predicted ^{13}C chemical shift. From the HSQC data, the methyl from fucose and the methyl of the acetyl group in GlcNAc were identified easily from the glycan spectra, Figure 68, and although they could also be identified for the protein and peptide HSQC data, the peaks were not as well resolved.

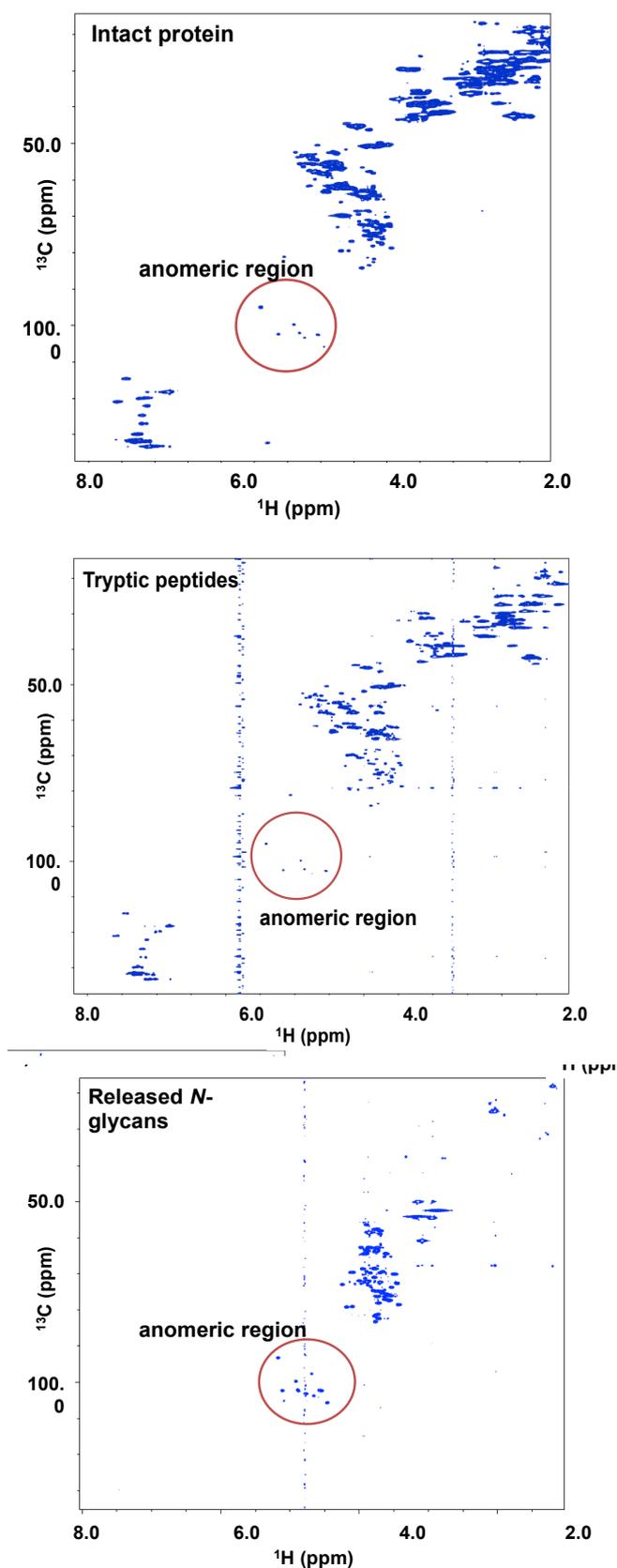


Figure 66: ^1H - ^{13}C -HSQC NMR spectra of the denatured intact protein, tryptic digest, and released *N*-glycans. The anomeric region, depicting the ^1H - ^{13}C correlations between the anomeric carbon and the anomeric proton, is circled.

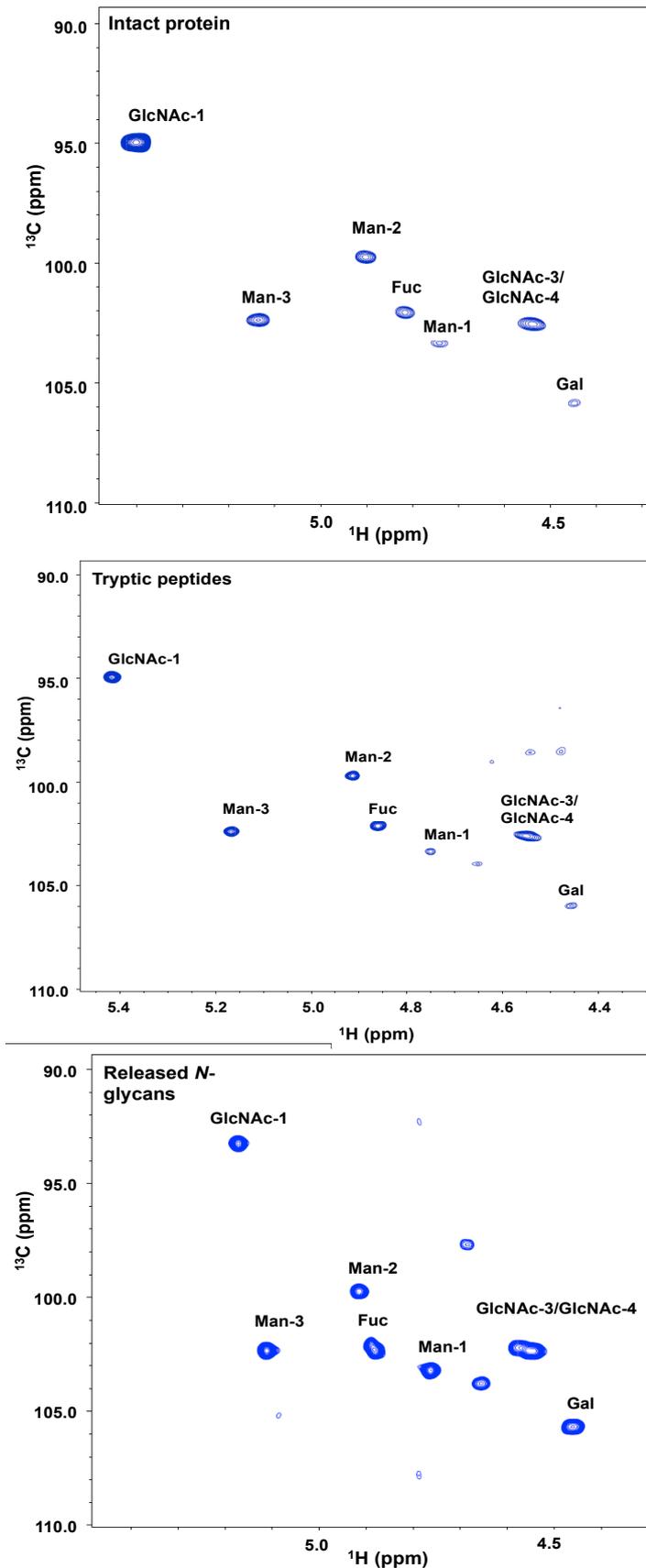


Figure 67: ^1H - ^{13}C -HSQC NMR spectra of the denatured intact protein, tryptic digest, and released *N*-glycans, expanded in the anomeric region.

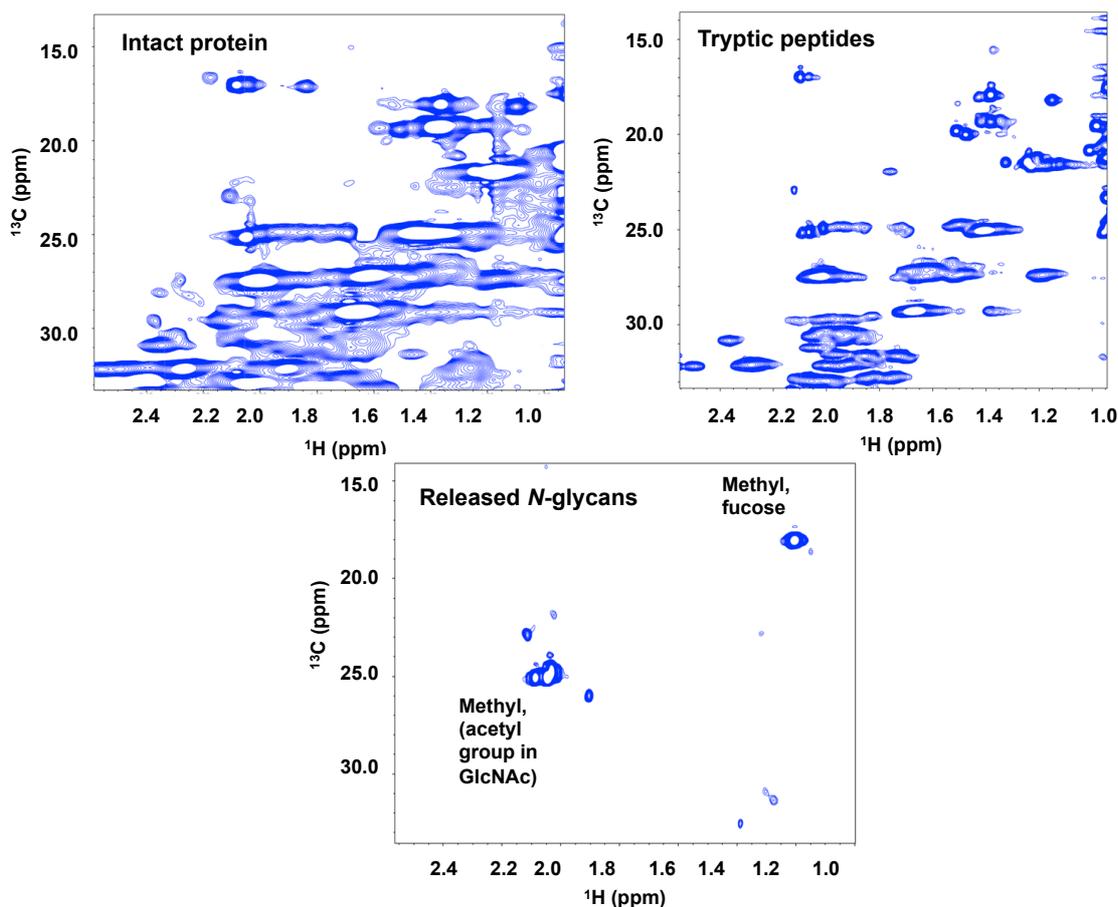


Figure 68: $^1\text{H}^{13}\text{C}$ -HSQC NMR spectra, showing the methyl/acetyl group region, of the denatured intact protein, tryptic digest, and released *N*-glycans.

TOCSY experiments were carried out to help determine the ^1H - ^1H correlations around each ring system. TOCSY with a short mixing time (13 ms) gives correlations of atoms 2-3 bonds apart and TOCSY with a long mixing time (120 ms) can reveal correlations between atoms up to six bonds apart. TOCSY spectra are much more complex than a $[\text{}^1\text{H}^{13}\text{C}]$ -HSQC spectrum and so to begin with it was easier to analyse the glycan spectra (rather than the peptide or protein spectra) as this offers the simplest data, to make assignments easier. Focusing on the anomeric region of the TOCSY spectra (short and long mixing times), a series of $^1\text{H}^1\text{H}$ cross peaks are observed for protons around each ring system correlating with the anomeric ^1H . Some signals are well resolved and others merge together, making full assignments of all the ring systems a difficult task without using additional experimental data.

In the short mixing time TOCSY there are up to two peaks present that align with each of the anomeric protons, Figure 69. These same signals are present in the long mixing time TOCSY

spectrum along with additional signals, Figure 70. This suggests that the peaks present in the short mixing time spectra can be assigned to H2 and H3 for each ring system, although this would need to be confirmed by additional experiments. For galactose, there are peaks at δ 3.53 ppm and δ 3.66 ppm in the 13 ms TOCSY spectrum. These also appear in the 120 ms TOCSY spectrum, in addition to a peak at δ 3.91 ppm. This suggests that, and because of the predicted chemical shifts, the cross peak connecting the signals at 4.46/3.53 ppm is that for ^1H -1 to ^1H -2 correlation on the galactose ring, and the cross peak at δ 4.46/3.66 ppm is for ^1H -1 to ^1H -3 correlation on the galactose ring. There are well resolved signals for the ^1H - ^1H correlations around GlcNAc-1, allowing assignments to be made for resonances of ^1H -1 with ^1H -2- ^1H -6, Figure 70.

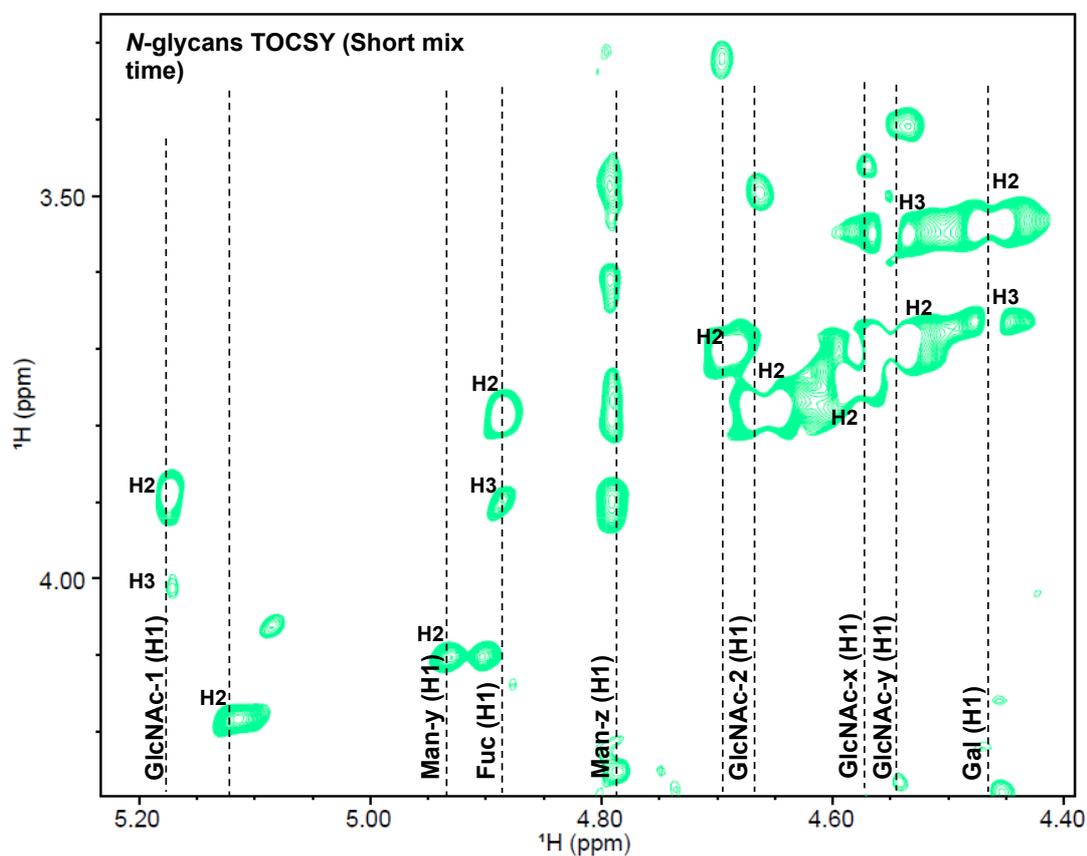


Figure 69: TOCSY short mixing time (13 ms) showing assignments of the anomeric protons of the released *N*-glycans with other protons around each monosaccharide ring system.

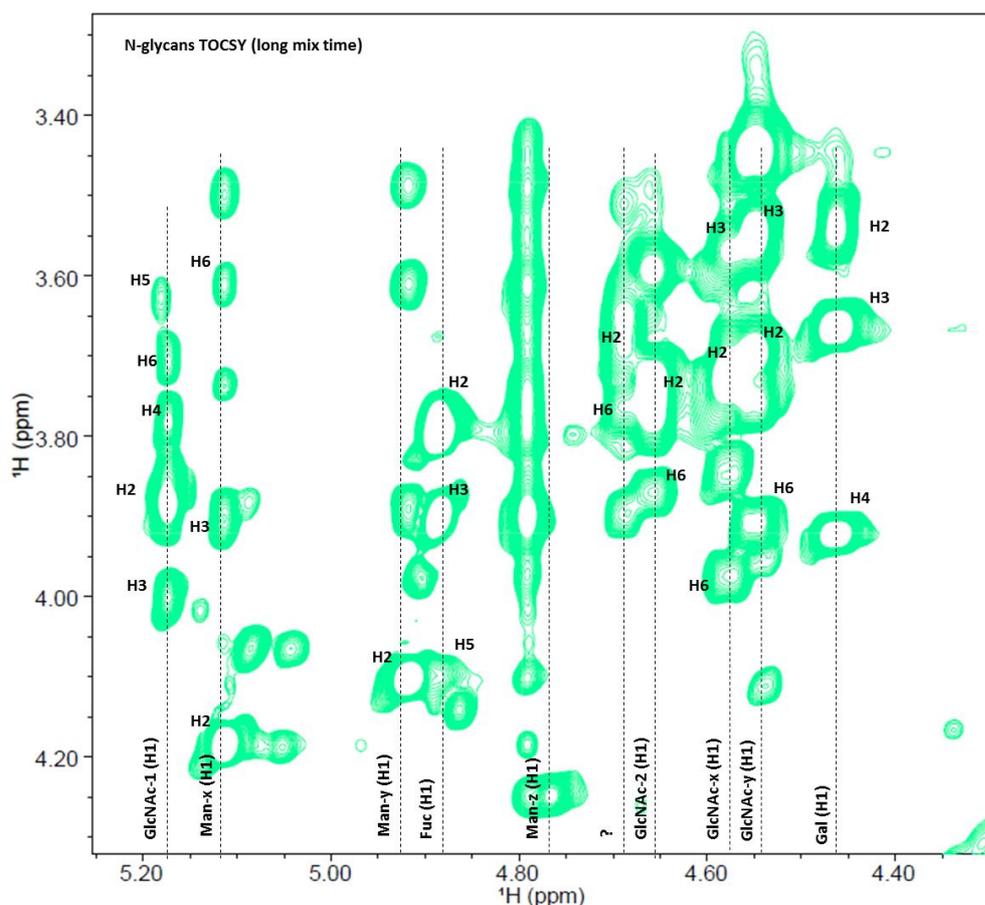


Figure 70: $^1\text{H}^1\text{H}$ -TOCSY long mixing time (120 ms) showing $^1\text{H}^1\text{H}$ correlations of anomeric protons, of released *N*-glycans, with protons around each monosaccharide ring system.

After assigning all of the anomeric protons using $^1\text{H}^{13}\text{C}$ -HSQC, peaks corresponding to C2-H2, H1-C2 and H2-C1 correlations of each monosaccharide were investigated using $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^1\text{H}$ -dqf-COSY, and $^1\text{H}^{13}\text{C}$ -dept-HSQC. The method of assignment is illustrated using fucose and GlcNAc-1. For example, a cross peak in the $^1\text{H}^{13}\text{C}$ -HMQC-COSY data at 4.88/71.01 ppm shows the interaction between the anomeric proton of fucose with C2 of fucose, in addition to showing a cross peak in the $^1\text{H}^{13}\text{C}$ -dept-HSQC spectrum at 3.78/71.01 ppm for the interaction of H2 of fucose with C2 of fucose. The $^1\text{H}^1\text{H}$ -dqf-COSY spectrum confirms correlation between anomeric fucose proton with its H2, Figure 71 and Figure 72. A second example shows evidence for H1C1 to H2C2 correlations of GlcNAc-1. A cross peak in $^1\text{H}^{13}\text{C}$ -HMQC-COSY at 5.17/56.39 ppm shows the interaction between the anomeric proton of GlcNAc-1 with C2 of GlcNAc-1, in addition to showing a cross peak in the $^1\text{H}^{13}\text{C}$ -dept-HSQC spectrum at 3.89/56.39 ppm for interaction of H2 of GlcNAc-1 with C2 of GlcNAc-1. The $^1\text{H}^1\text{H}$ -dqf-COSY spectrum confirms the H1-H2 correlation between the anomeric GlcNAc-1 proton with its H2, Figure 73 and Figure 74.

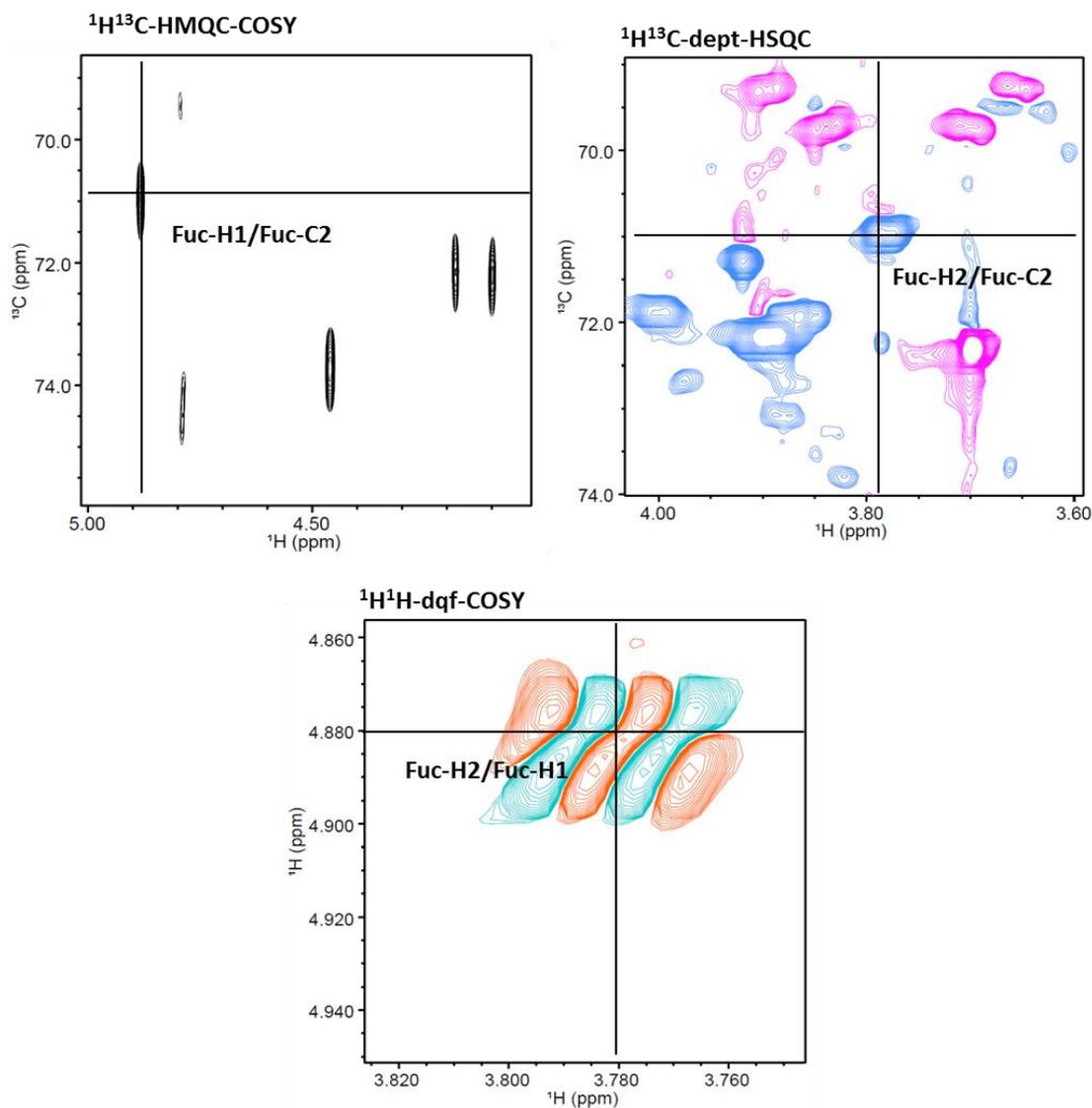


Figure 71: Evidence of connectivity between fucose H1C1 with fucose H2C2, using a combination of $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY. Cross peaks are indicated by the intersecting lines.

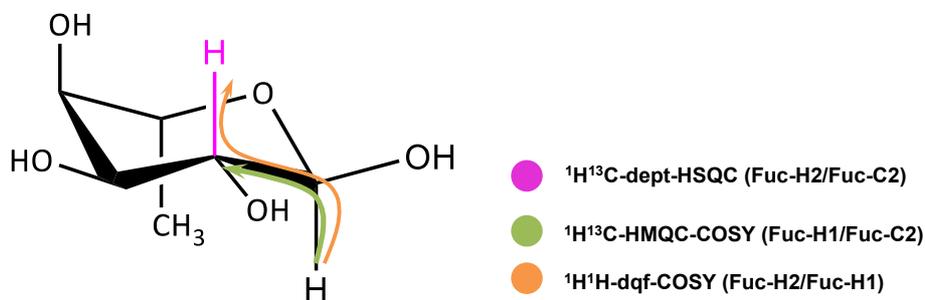


Figure 72: Evidence of connectivity of H1C1 of GlcNAc-1 to H2C2 of GlcNAc-1, using a combination of $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY.

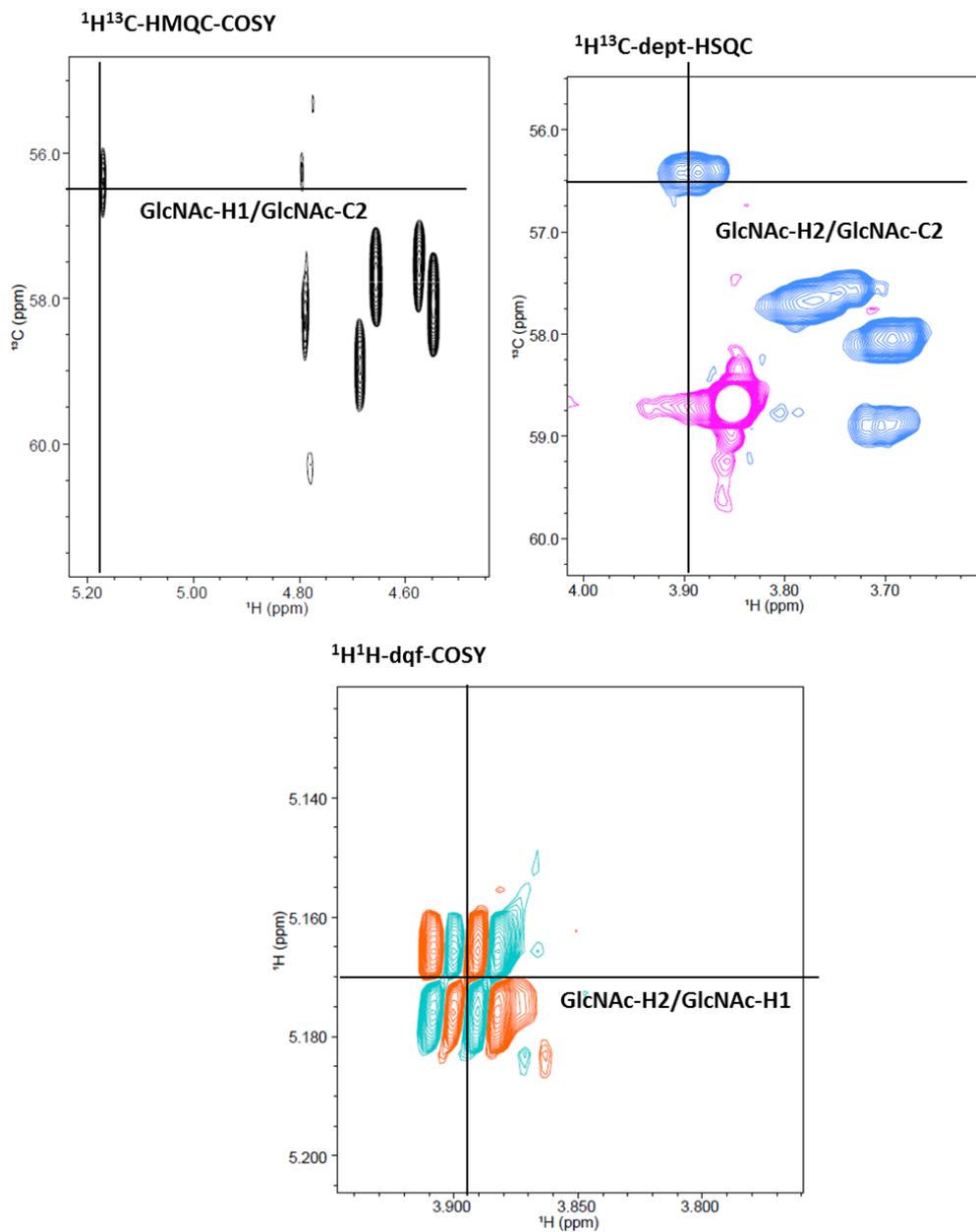


Figure 73: Evidence of connectivity of H1C1 of GlcNAc-1 to H2C2 of GlcNAc-1, using a combination of $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY. Cross peaks are indicated by the intersecting lines.

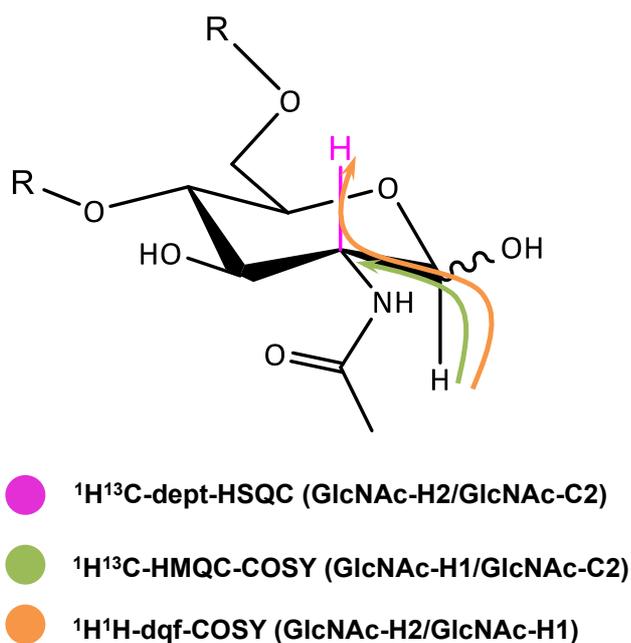


Figure 74: Evidence of connectivity of H1C1 of GlcNAc-1 to H2C2 of GlcNAc-1, by $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC and $^1\text{H}^1\text{H}$ -dqf-COSY.

To determine the linkages between the monosaccharide units, an [$^1\text{H}^{13}\text{C}$]-HMBC experiment was carried out. Three of the four glycans in the IgG show Fuc-GlcNAc in a 1-6 linkage. To determine this linkage between fucose and GlcNAc, NMR spectra from the $^1\text{H}^{13}\text{C}$ -HMBC, $^1\text{H}^{13}\text{C}$ -dept-HSQC, $^1\text{H}^1\text{H}$ -NOESY and $^1\text{H}^1\text{H}$ -TOCSY experiments were compared. The $^1\text{H}^{13}\text{C}$ -HMBC spectrum revealed a cross peak at 4.88/72.30 ppm for the correlation of the anomeric proton of fucose with what would be expected to be the C6 of GlcNAc. Looking also at the anomeric carbon of fucose, a cross peak was identified at 102.30/3.78 ppm for the H1-H6 correlation with H6 of GlcNAc. $^1\text{H}^{13}\text{C}$ -dept-HSQC revealed a cross peak at 3.78/72.30 ppm for the CH_2 group of GlcNAc (in pink). The proton resonance peak at 3.78 ppm for H6 of GlcNAc also aligns with an nOe cross peak, showing spatial connectivity with the anomeric proton of fucose, along with the cross peak in $^1\text{H}^1\text{H}$ -TOCSY which confirms the correlation with GlcNAc, Figure 75 and Figure 76.

A cross peak showing the linkages between GlcNAc-2 and GlcNAc-1 as a 1-4 linkage was identified at 103.72 ppm for the anomeric carbon of GlcNAc-2, and 3.78 ppm for the H4 of GlcNAc-1. However, there was no cross peak to show interaction of the GlcNAc-2 anomeric ^1H and GlcNAc-1 C4. In addition, there was no other cross peak in the HMBC data that could show the connectivity between any of the other glycan units.

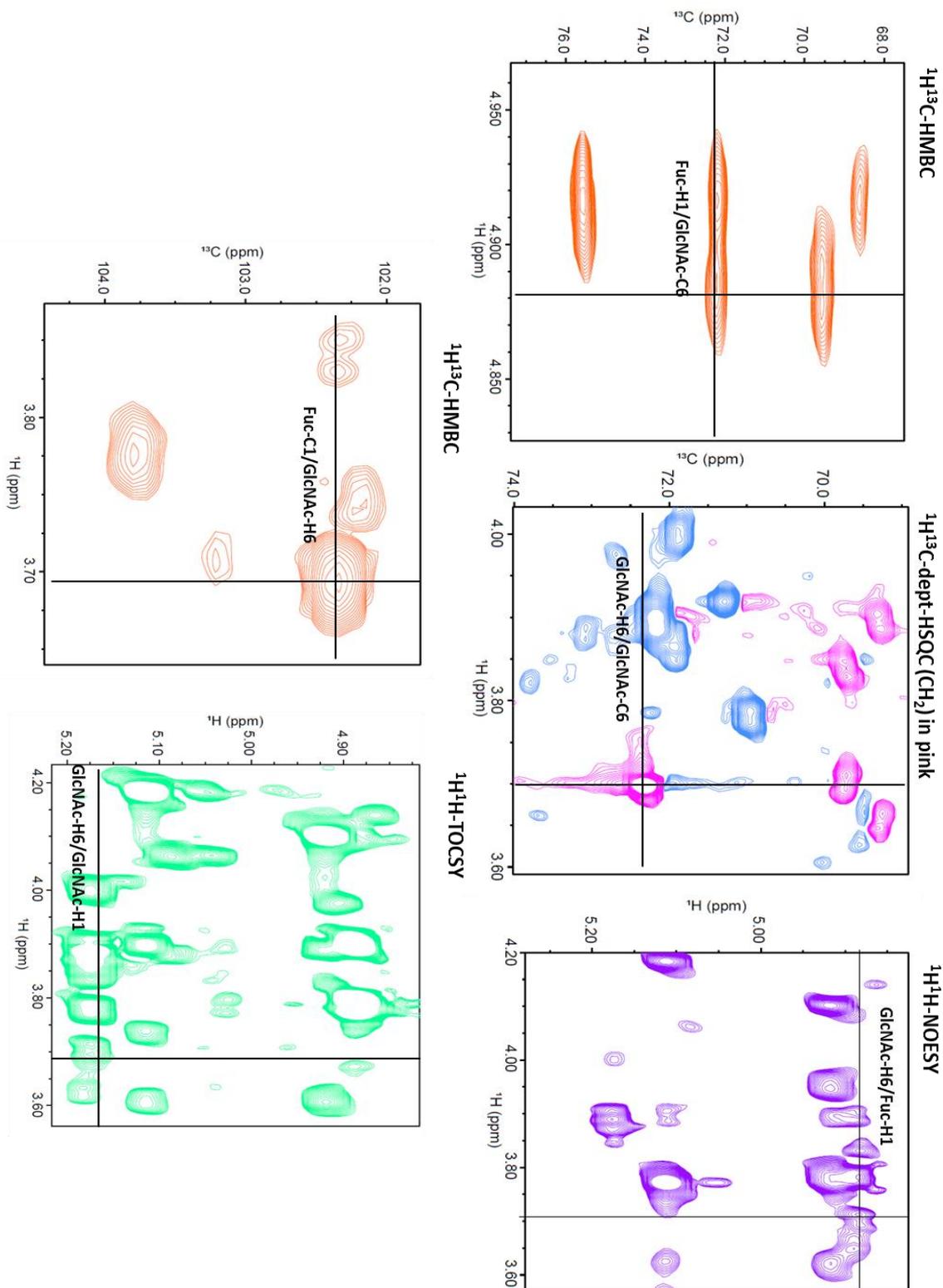


Figure 75: Comparison of $^1\text{H}/^{13}\text{C}$ -HMBC, $^1\text{H}/^{13}\text{C}$ -dept-HSQC, $^1\text{H}/^1\text{H}$ -NOESY and $^1\text{H}/^1\text{H}$ -TOCSY to show the Fuc(1-6)GlcNAc-1 linkage.

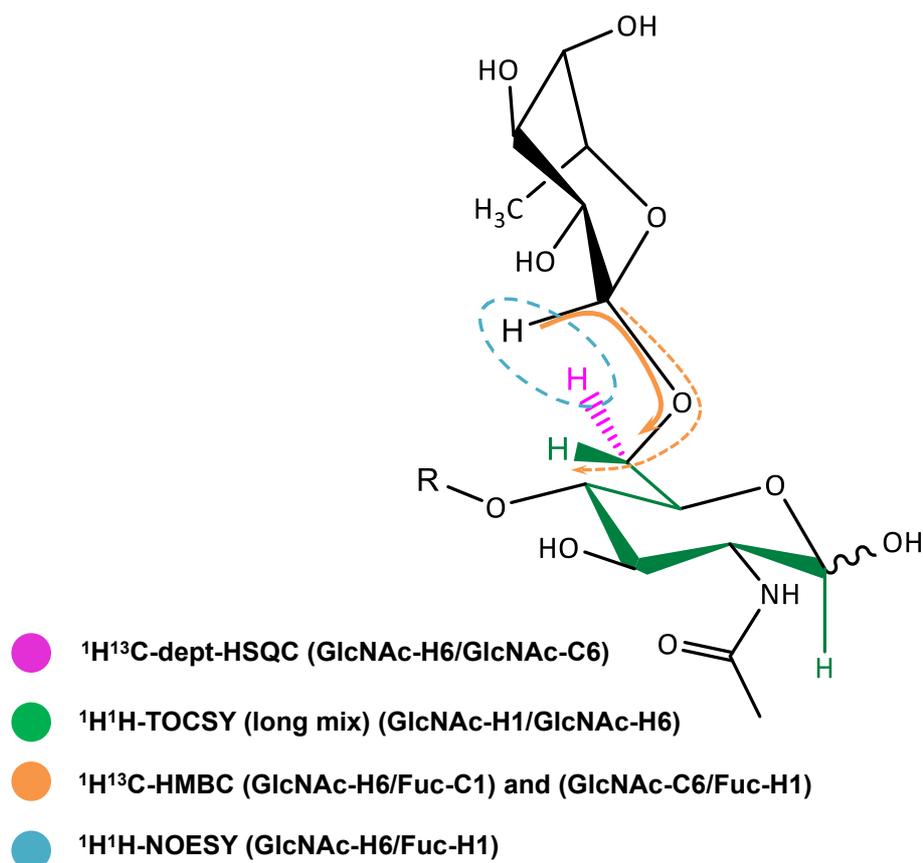


Figure 76: Comparison of $^1\text{H}^{13}\text{C}$ -HMBC, $^1\text{H}^{13}\text{C}$ -dept-HSQC, $^1\text{H}^1\text{H}$ -NOESY and $^1\text{H}^1\text{H}$ -TOCSY to show the Fuc(1-6)GlcNAc-1 linkage.

Since GlcNAc contains an acetamide methyl group, described in the earlier HSQC data, this feature can also be used to try to distinguish between different GlcNAcs using NOESY, which detects through-space connectivity. Figure 77 shows an expanded region of the NOESY spectrum connecting GlcNAc-methyl resonances with anomeric proton resonances. The NOESY spectrum shows a cross peak at 5.17/2.03 ppm, for the nOe between the anomeric proton of GlcNAc-1 with ^1H from the methyl group of GlcNAc-1. There is a cross peak from the nOe between the anomeric ^1H of fucose with the methyl of GlcNAc-2 (4.88/2.08 ppm), as well as the nOe for GlcNAc-2's anomeric ^1H with its methyl ^1H s (4.64/2.08 ppm), Figure 78. In addition, there are cross peaks for an additional GlcNAc between its anomeric proton and methyl group, as well as well as two cross peaks for nOes of mannose with GlcNAc, although it cannot be unambiguously be determined which one is for the oligomannose glycan and which one is for the complex glycan.

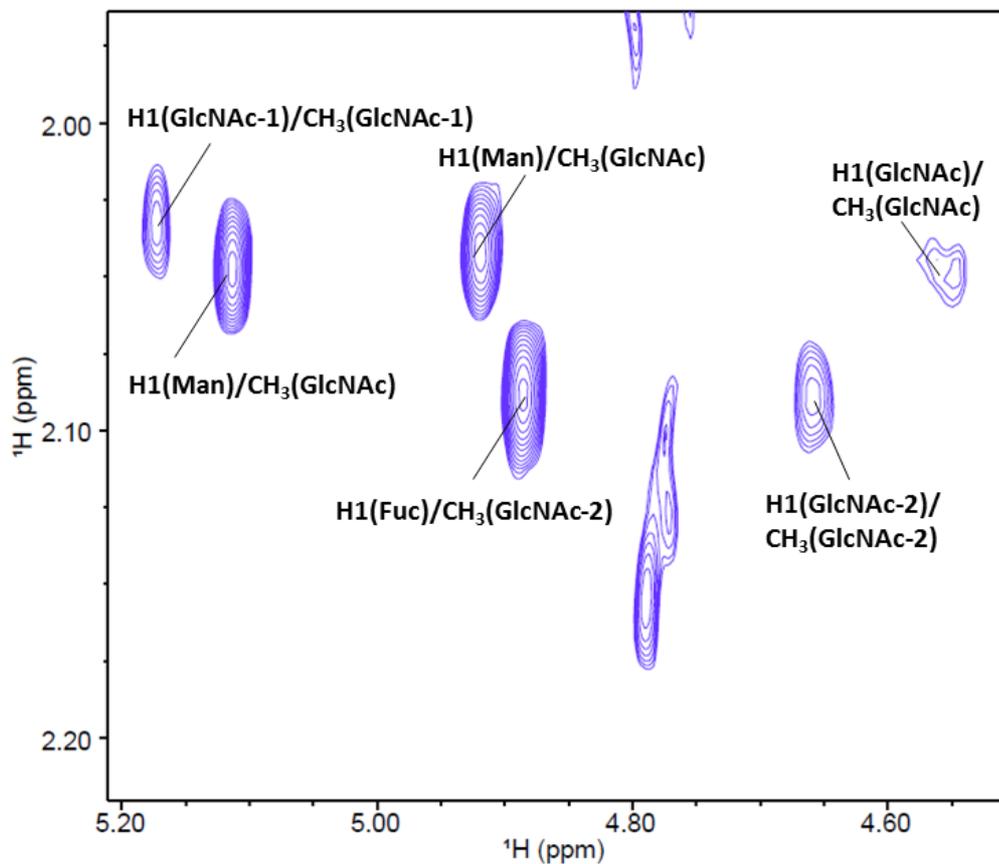


Figure 77: ^1H - ^1H -NOESY spectrum showing nOe cross peaks connecting GlcNAc-methyl resonances with anomeric proton resonances, from Fuc, Man and GlcNAc.

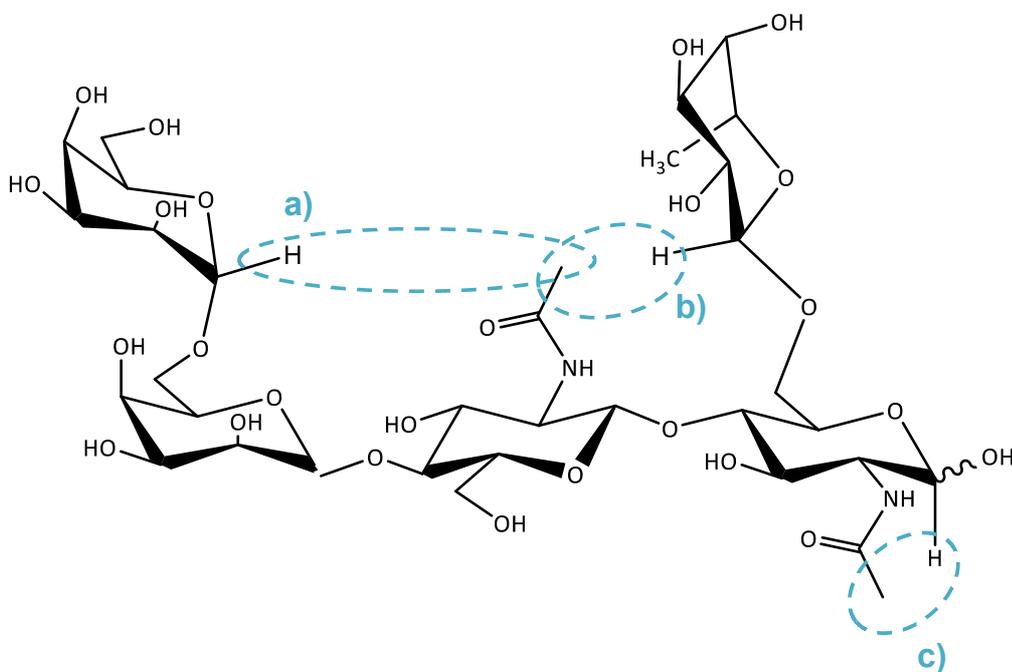


Figure 78: ^1H - ^1H -NOESY interactions connecting the GlcNAc-methyl with anomeric proton from a) Man, b) Fuc and c) GlcNAc.

After assigning all the anomeric carbons and protons C1-C3 and H1-H3 of each of the glycan monosaccharide units as well as the second and third carbons and protons around each of the monosaccharide ring systems using $^1\text{H}^1\text{H}$ -TOCSY, the $^1\text{H}^1\text{H}$ -TOCSY data were compared with those from $^1\text{H}^{13}\text{C}$ -dept-HSQC to establish connectivities for the H6-C6 correlations with the anomeric ^1H . In $^1\text{H}^{13}\text{C}$ -dept-HSQC all of the CH_2 peaks, important for the C6H6 correlations, are inverted and are shown in pink, Figure 79. Only a few $^1\text{H}^1\text{H}$ -TOCSY peaks could be assigned unambiguously. For GlcNAc-1, there is a cross peak for the H6-H1 correlation in TOCSY that lines up with the cross peak for the C6-H6 correlation in dept-HSQC, at 3.7 ppm. In addition, there is a cross peak for the H6-H1 correlation in TOCSY that lines up with the C6-H6 correlation in dept-HSQC at 3.87 ppm.

All the assignments that have been made are summarised in tables 6-13, showing clearly that it is a difficult and laborious task to assign such complex NMR data manually and completely. Fucose and GlcNAc-1 have full assignments of all carbon atoms and protons in the ring systems, whereas only partial assignment was possible for the remaining monosaccharides. Some assignments, although known to be for a GlcNAc or mannose, cannot be unambiguously identified as to which mannose or GlcNAc it is for a particular glycan structure (e.g. is it a mannose in the oligomannose, or in one of the complex glycans?). In spite of not completing the assignments, each monosaccharide type that was present was identified. For example, there were cross peaks among the NMR spectra that were due to the presence of GlcNAc, galactose, fucose and mannose. Identifications of the core fucosylated glycans were established through HMBC, and NOESY helped to determine the close in space connection between a mannose and a GlcNAc. Since there were cross peaks missing in the HMBC spectrum complete structural assignments were not possible.

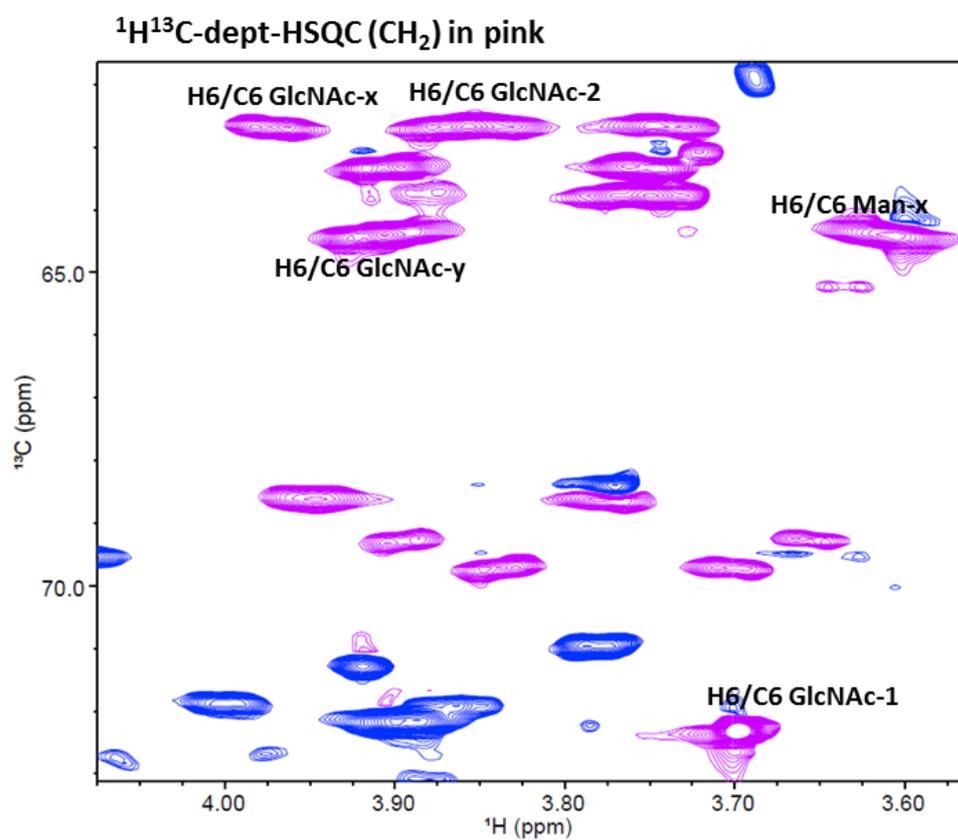
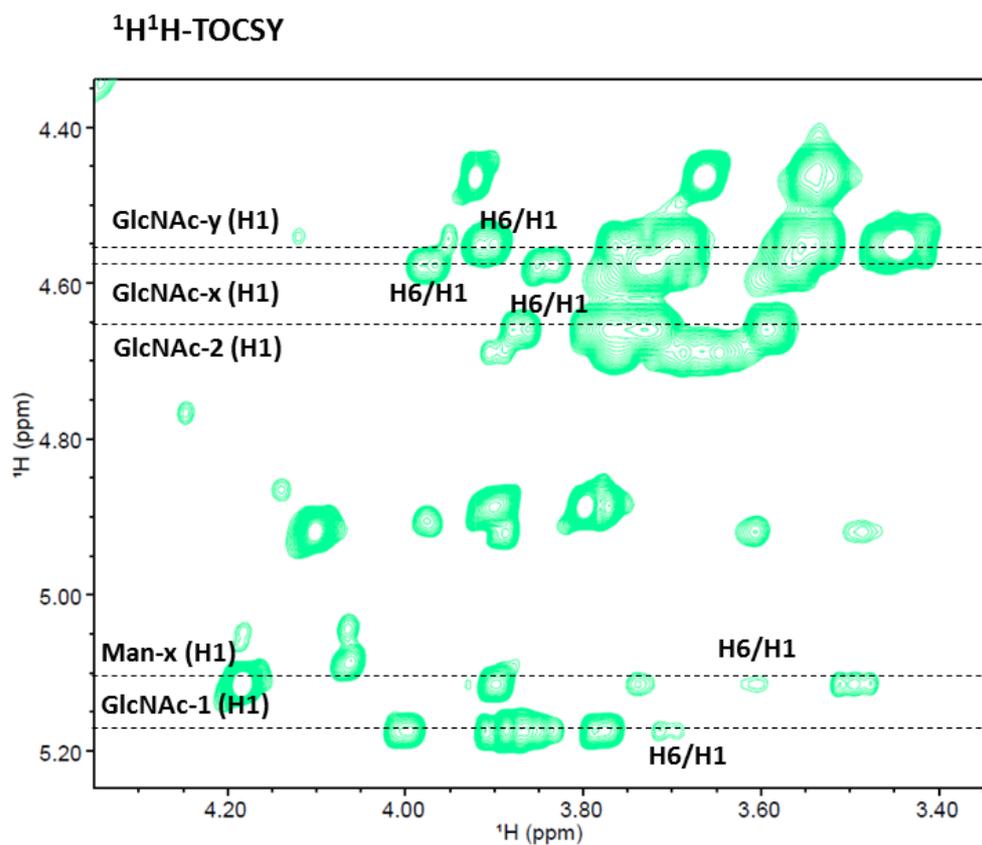


Figure 79: $^1\text{H}^1\text{H}$ -TOCSY and $^1\text{H}^{13}\text{C}$ -dept-hsqc to show H6C6 connectivities with the anomeric proton.

Table 6: Assignments of fucose ^1H and ^{13}C resonances confirmed by matching cross peak patterns in $[\text{}^1\text{H}^{13}\text{C}]$ -HSQC, $[\text{}^1\text{H}^{13}\text{C}]$ -HMBC, $[\text{}^1\text{H}^1\text{H}]$ -TOCSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC, $^1\text{H}^1\text{H}$ -NOESY, $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^1\text{H}$ -dqf-COSY.

$^1\text{H}/^{13}\text{C}$ number	Glycoscience chemical shift predictor		measured chemical shifts	
	^1H	^{13}C	^1H	^{13}C
1	4.88	100.02	4.88	102.30
2	3.77	69.05	3.78	71.01
3	3.89	70.35	3.89	72.17
4	3.81	72.83	3.80	74.63
5	4.13	67.71	4.10	69.50
6	1.22	16.18	1.20	18.35

Table 7: Assignments of GlcNAc-1 ^1H and ^{13}C resonances confirmed by matching cross peak patterns in $[\text{}^1\text{H}^{13}\text{C}]$ -HSQC, $[\text{}^1\text{H}^{13}\text{C}]$ -HMBC, $[\text{}^1\text{H}^1\text{H}]$ -TOCSY, $^1\text{H}^{13}\text{C}$ -dept-HSQC, $^1\text{H}^1\text{H}$ -NOESY, $^1\text{H}^{13}\text{C}$ -HMQC-COSY, $^1\text{H}^1\text{H}$ -dqf-COSY.

$^1\text{H}/^{13}\text{C}$ number	Glycoscience chemical shift predictor		measured chemical shifts	
	^1H	^{13}C	^1H	^{13}C
1	5.07	102.07	5.17	93.28
2	3.85	56.85	3.89	56.39
3	3.75	74.73	4.01	69.82
4	3.66	71.02	3.78	70.94
5	3.58	77.14	3.85	
6	3.98	61.84	3.70	72.30
Carbonyl ^{13}C	-	-	-	177.51
Methyl	-	-	2.04	25.05

Table 8: Assignments of GlcNAc-2 ¹H and ¹³C resonances confirmed by matching cross peak patterns in [¹H¹³C]-HSQC, [¹H¹³C]-HMBC, [¹H¹H]-TOCSY, ¹H¹³C-dept-HSQC, ¹H¹H-NOESY, ¹H¹³C-HMQC-COSY, ¹H¹H-dqf-COSY.

¹ H/ ¹³ C number	Glycoscience chemical shift predictor		measured chemical shifts	
	¹ H	¹³ C	¹ H	¹³ C
1	4.64	102.70	4.64	103.72
2	3.32	56.85	3.77	57.68
3	3.74	74.73		
4	3.60	71.02		
5	3.59	77.14		
6	3.92	61.84	3.87	62.41
Carbonyl ¹³ C	-	-	-	
Methyl	-	-	2.09	25.06

Table 9: Assignments of galactose ¹H and ¹³C resonances confirmed by matching cross peak patterns in [¹H¹³C]-HSQC, [¹H¹³C]-HMBC, [¹H¹H]-TOCSY, ¹H¹³C-dept-HSQC, ¹H¹H-NOESY, ¹H¹³C-HMQC-COSY, ¹H¹H-dqf-COSY.

¹ H/ ¹³ C number	Glycoscience chemical shift predictor		measured chemical shifts	
	¹ H chemical shift	¹³ C chemical shift	¹ H chemical shift	¹³ C chemical shift
1	4.47	103.75	4.46	105.67
2	3.53	72.17	3.53	73.74
3	3.68	73.72	3.66	73.70
4	3.92	69.63	3.91	
5	3.68	76.37	3.66	75.28
6	3.75	62.33		

Table 10: Assignments of Man-z ¹H and ¹³C resonances confirmed by cross peak patterns in [¹H¹³C]-HSQC, [¹H¹³C]-HMBC, [¹H¹H]-TOCSY, ¹H¹³C-dept-HSQC, ¹H¹H-NOESY, ¹H¹³C-HMQC-COSY, ¹H¹H-dqf-COSY.

¹ H/ ¹³ C number	Glycoscience chemical shift predictor		measured chemical shifts	
	¹ H chemical shift	¹³ C chemical shift	¹ H chemical shift	¹³ C chemical shift
1	4.77	103.80	4.76	103.25
2	4.24	71.80	4.24	72.92
3	3.77	71.90		
4	3.80	68.40		
5	3.66	74.90		
6	3.71	62.60		

Table 11: Assignments of Man-y ¹H and ¹³C resonances confirmed by cross peak patterns in [¹H¹³C]-HSQC, [¹H¹³C]-HMBC, [¹H¹H]-TOCSY, ¹H¹³C-dept-HSQC, ¹H¹H-NOESY, ¹H¹³C-HMQC-COSY, ¹H¹H-dqf-COSY.

¹ H/ ¹³ C number	Glycoscience chemical shift predictor		measured chemical shifts	
	¹ H chemical shift	¹³ C chemical shift	¹ H chemical shift	¹³ C chemical shift
1	4.92	99.10	4.92	99.66
2	4.11	77.70	4.10	79.10
3	3.89	70.95	3.89	
4	3.49	68.57	3.49	
5	3.62	74.32	3.61	
6	3.62	62.78	3.61	

Table 12: Assignments of Man-x ¹H and ¹³C resonances confirmed by cross peak patterns in [¹H¹³C]-HSQC, [¹H¹³C]-HMBC, [¹H¹H]-TOCSY, ¹H¹³C-dept-HSQC, ¹H¹H-NOESY, ¹H¹³C-HMQC-COSY, ¹H¹H-dqf-COSY.

¹ H/ ¹³ C number	Glycoscience chemical shift predictor		measured chemical shifts	
	¹ H	¹³ C	¹ H	¹³ C
1	5.11	101.50	5.11	102.32
2	4.06	77.70	4.18	79.17
3	3.90	70.95	3.90	
4	3.51	68.57	3.49	
5	3.75	77.70	3.74	
6	3.62	62.40	3.61	64.38

Table 13: Assignments of GlcNAc-x/y (terminal? on Gal?) ¹H and ¹³C resonances confirmed by cross peak patterns in [¹H¹³C]-HSQC, [¹H¹³C]-HMBC, [¹H¹H]-TOCSY, ¹H¹³C-dept-HSQC, ¹H¹H-NOESY, ¹H¹³C-HMQC-COSY, ¹H¹H-dqf-COSY.

¹ H/ ¹³ C number	Glycoscience chemical shift predictor		measured chemical shifts	
	¹ H chemical shift	¹³ C chemical shift	¹ H chemical shift	¹³ C chemical shift
1	4.59	101.2	4.55/4.58	102.34
2	3.75	56.52	3.69/3.73	58.10/57.57
3	3.69	73.58	3.70	
4	3.70	80.16	3.72	81.15
5	3.58	76.30	3.55	
6	3.83	61.54	3.91/3.97	64.39/62.70

5.4. Conclusions and future work

A FANGS-based¹⁴¹ approach was successfully applied to preparation of an IgG for *N*-glycan release, in a centrifugal filter, for 2D-NMR glycomic analysis, and to potentially use it as a fingerprinting tool to monitor glycan profiles in batches of biotherapeutic.

Assignments of the remaining glycan peaks is time consuming and difficult especially in the crowded regions of the NMR spectra. In NMRview, the software used to analyse the data, there are slidertools which allow the predicted glycan chemical shifts to be imported and enables matching of the predictions with the cross peaks in each of the NMR spectra. Luciano Mueller at BMS is aiming to build a database of the glycans in the mixture in the program XPLORE. Currently, not all of the carbohydrate rings exist in the XPLORE library and therefore these need building in. Having a database to use in line with NMRview would make interpreting spectra of complex glycan mixtures much easier, as the assignments made in this work using a sample containing only four glycans was not straightforward.

The concentrated samples of the tryptic peptide mixture are now being fractionated by HPLC at BMS so that both NMR and mass spectrometric data on the glycans can be collected. The aim is that the fraction containing the peptide with the glycans attached can be analysed by NMR, rather than as here analysing the entire peptide mixture. This should eliminate urea contamination and simplify the spectra to make glycan assignments easier. In addition, once both the NMR and LC-MS data are collected, they can be correlated using Matlab software using a recently developed three dimensional cross correlation^{258,255,257} to help separate isomeric glycoforms.

What also needs to be done now is to test the fully developed method on different batches of the IgG to see whether it is a useful tool to establish small differences in batches and whether fingerprints can be generated to help with semi-automated bath quality control. If the fingerprinting works then it may not be necessary to make complete assignments of all the glycans. In addition, testing the method on different antibodies containing different glycans such as sialylated, more complex mixture of glycans or even those that may also contain O-glycans could be useful to test the robustness of the method.

Chapter 6: Demonstration of the potential of the one-pot method to study glycosylation, and to monitor changes to glycosylation, in experimental cellular systems

6. Demonstration of the potential of the one-pot method to study glycosylation, and to monitor changes to glycosylation, in experimental cellular systems.

6.1. Mesenchymal stromal cells (MSCs)

Mesenchymal stromal cells, usually derived from bone marrow, are heterogeneous populations of cells which contain varying levels of different cells that have the ability to differentiate (change from one cell type to another) into adipocytes, chondrocytes, osteoblasts and myoblasts to make fat, cartilage, bone or muscle tissue^{268,269}. Some cells in the heterogeneous mixture are able to differentiate into all lineages, others can only differentiate into one or two and some cannot differentiate at all. Stem cells, including bone marrow-derived stem cells, can migrate through the blood to damaged tissues/organs for repair following injury. Damage can also occur to stem cells through genetic alterations with age^{270,271} that can lead to degenerative disorders (e.g. of brain²⁷², eye, bone²⁷³ and muscle), cardiovascular disease^{274,275}, and cancer. The ability of stem cells to differentiate into different cell types means there is a potential use of MSCs in regenerative medicine^{276,277,278}, cell replacement²⁷⁹ and in tissue engineering²⁸⁰. This could be achieved by transplantation of MSCs or even stimulating patient stem cells using growth factors to treat disorders such as osteoporosis and osteoarthritis. Studying what influences differentiation may help to identify novel targets that can be manipulated to influence stem cell behaviour.

In investigating what influences differentiation, studies have used gene profiling^{281,282}, proteomics²⁸³ and transcriptomics²⁸⁴, but it is also known that there are changes to glycosylation upon cellular differentiation^{285,139}. There are glycoproteins on the surface of MSCs and it is known that glycans affect protein function, suggesting that glycosylation could play an important role in MSC differentiation. These published studies²⁸¹⁻¹³⁹ all use primary cells that are by nature heterogeneous and so they do not necessarily all behave the same way on differentiation and so are not ideal to use to assess the functions of glycans in differentiation. In addition, primary MSCs are limited in the number of times they can divide in culture and therefore this limits the number and depth of studies into functionality that can be carried out. In addition, some studies using mass spectrometry to look into the function of glycosylation in differentiation do not include permethylation of the glycans before analysis by mass spectrometry^{91,139}. This is likely to undermine the accuracy of quantification of relative abundances⁹¹, due to the well-documented lability of sialic acid

groups in positive mode, and the over-representation of sialylated species in negative mode analysis, as well as the poorer sensitivity of detection without permethylation⁹¹.

MSCs used in the studies presented in this thesis have therefore come from a cell line immortalised using human telomerase reverse transcriptase to give h-TERT-MSCs²⁸⁶. Where normally cells would not proliferate/undergo cell division indefinitely and therefore cease to divide after a particular number of divisions (known as replicative senescence), immortalised cells can continue to divide for more much longer periods in vitro. The cells are derived from a single donor, removing the heterogeneity of the cellular population for a more reliable comparison of changes in glycosylation. Two cell lines have been used, Y201 h-TERT-MSCs and Y101 h-TERT-MSCs. The Y201 cell line is capable of differentiating into both adipocytes and osteoblasts, whereas the Y101 can differentiate into osteoblasts but has limited ability to differentiate into adipocytes. Since MSCs can easily be grown in culture they are also useful to use as models for congenital disorders of glycosylation and to help to gain further understanding of glycan biosynthesis by comparing wild-type MSCs with those that have genetic modifications targeted to the COG complex, to study differences in glycan profiles.

Profiling of *N*-glycans from MSCs has been investigated and reported in the literature. *N*-glycan profiling of the immortalised cell lines Y201 and Y101 has been investigated by a former PhD student from the Ungar group, Katherine Wilson²⁸⁷. However, there has to date been no analysis of the *O*-glycan profile of these cells to see if the *O*-glycans play a role in differentiation, and if the glycan profiles change after genetic modification of the cell lines.

6.2. Data analysis

All mass spectrometric data collected in this chapter of work was acquired using MALDI-FT-ICR. All glycan samples were permethylated before analysis. Glycan assignments have been confidently assigned based on their mass accuracies, which, for the majority were all below 1 ppm, and also from the knowledge of possible *O*-glycan structures produced by the *O*-glycan biosynthetic pathway. All glycan samples were spiked with an internal standard, deuteropermethylated maltotetraose, to carry out relative quantification of the glycans to test whether there were relative quantitative differences in the glycan profiles of compared glycan sample sets. The relative abundance was calculated by comparing the peak intensities of each of the *O*-glycans with peak intensity of the internal standard, and the relative

abundance was expressed as a percentage. Statistical analysis was carried out using a T test with Holm Sidak post-hoc test (* for $P = < 0.05$, ** for $P = < 0.01$ and *** for $P = < 0.0001$).

6.3. O-glycans of Y201 and Y101 MSCs

The developed O-glycan release method was applied to proliferating wild-type Y201 and Y101 MSCs, to see if O-glycans could be detected and identified. Cells were grown in a 10 cm culture dish until just under confluency, before being harvested. MS analysis of the subsequently released O-glycans revealed signals at m/z 518, 722, 879, 967, 1083, 1141, 1171, 1240, and 1328 for the Y101 cell line and at m/z 518, 722, 879, 926, 967, 1083, 1141, 1240, 1328 and 1345, for the Y201 cell line, Figure 80a and b respectively and Table 14.

On comparing the resulting glycan profiles for each of the cell lines, Y201 had three O-glycans that were significantly higher in abundance than in the Y101 cell line, Figure 81. These were $\text{Hex}_2\text{HexNAc}_1$ ($P = 0.0028$ **) $\text{Hex}_3\text{HexNAc}_1$ ($P = < 0.0001$ ***) and $\text{Hex}_2\text{HexNAc}_2\text{Fuc}_1$ ($P = < 0.0001$ ***), Table 15. These differences in the abundance of O-glycans could perhaps be attributed to the different potentials of the two cell lines to differentiate.

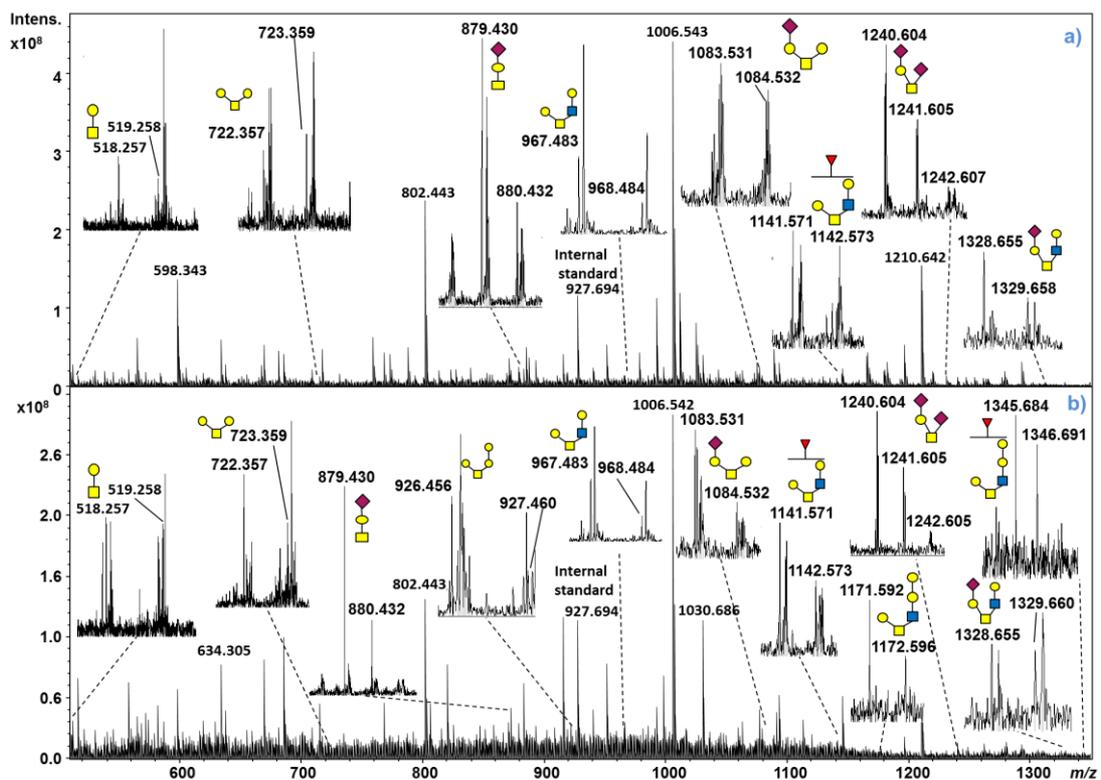


Figure 80: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from a) Y101 and b) Y201 MSC cell lines. All glycans ionised as $[M + \text{Na}]^+$.

Table 14: List of O-glycans identified from Y201 and Y101 MSC cell lines, showing the mass accuracy

glycan composition	calculated <i>m/z</i>	Y201 WT average measured <i>m/z</i> (n=5)	Y201 WT mass accuracy (ppm)	Y101 WT average measured <i>m/z</i> (n=5)	Y101 WT mass accuracy (ppm)
Hex ₁ HexNAc ₁	518.2572	518.2571	-0.12	518.2570	-0.41
Hex ₂ HexNAc ₁	722.3569	722.3568	-0.19	722.3567	-0.32
Hex ₁ HexNAc ₁ NeuAc ₁	879.4304	879.4301	-0.40	879.4304	0.05
Hex ₃ HexNAc ₁	926.4567	926.4559	-0.87	-	-
Hex ₂ HexNAc ₂	967.4833	967.4826	-0.72	967.4826	-0.75
Hex ₂ HexNAc ₁ NeuAc ₁	1083.5306	1083.5299	-0.63	1083.5307	0.12
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	1141.5714	-0.95	1141.5744	1.64
Hex ₃ HexNAc ₂	1171.5831	-	-	1171.5830	-0.12
Hex ₁ HexNAc ₁ NeuAc ₂	1240.6045	1240.6034	-0.88	1240.6032	-1.04
Hex ₂ HexNAc ₂ NeuAc ₁	1328.6569	1328.6535	-2.53	1328.6561	-0.59
Hex ₃ HexNAc ₂ Fuc ₁	1345.6723	1345.6726	-0.22	-	-

Table 15: P values for significant differences between Y201/Y101 WT O-glycan levels. Statistical analysis carried out by T test with Holm Sidak post-hoc test. ns = not significant

glycan composition	P values for significant differences between Y201 and Y101 WT O-glycans (n=5)
Hex ₁ HexNAc ₁	ns
Hex ₂ HexNAc ₁	0.0028
Hex ₁ HexNAc ₁ NeuAc ₁	ns
Hex ₃ HexNAc ₁	< 0.0001
Hex ₂ HexNAc ₂	ns
Hex ₂ HexNAc ₁ NeuAc ₁	ns
Hex ₂ HexNAc ₂ Fuc ₁	< 0.0001
Hex ₃ HexNAc ₂	ns
Hex ₁ HexNAc ₁ NeuAc ₂	ns
Hex ₂ HexNAc ₂ NeuAc ₁	ns
Hex ₃ HexNAc ₂ Fuc ₁	ns

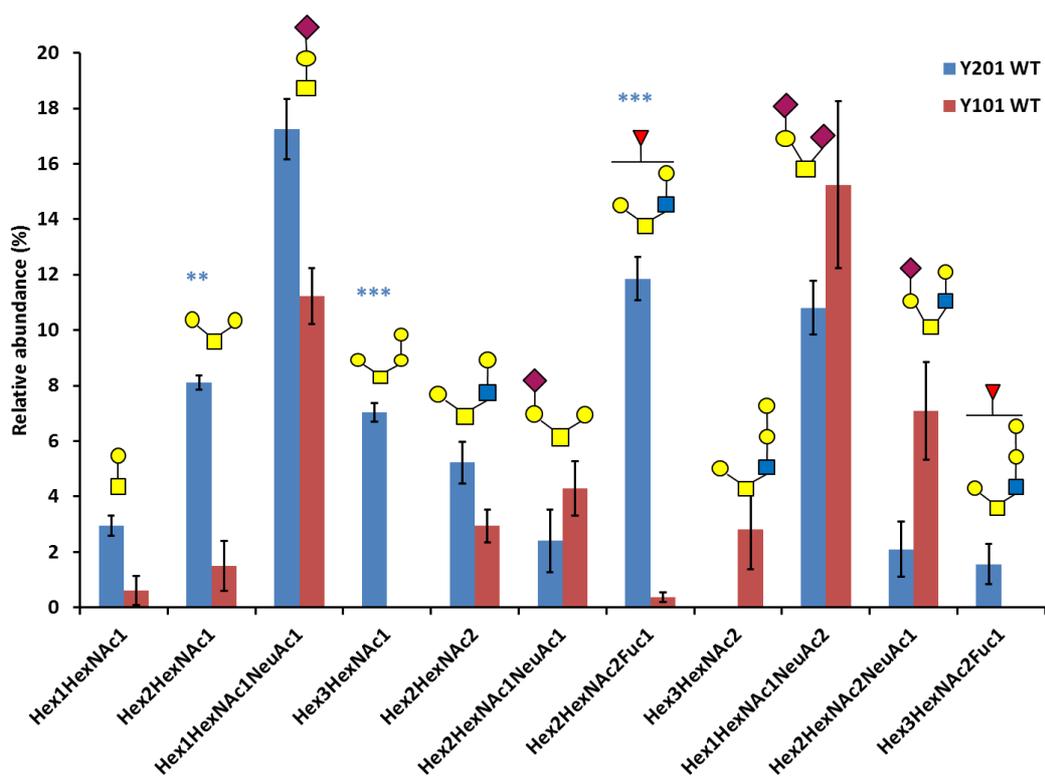


Figure 81: Comparison of the O-glycan profiles of the Y201 and Y101 wild-type (WT) MSCs. Relative abundance calculated by calculating each glycan signal as a percentage of the spiked internal standard signal. P = * < 0.05, ** < 0.01, * < 0.001.**

To confirm the presence of O-glycans, tandem mass spectrometry was carried out. Three examples of tandem mass spectrometry data are shown in Figure 82 - Figure 84. The first is a HCD product ion mass spectrum for NeuAc₁Hex₁HexNAc₁ (*m/z* 879). The spectrum shows diagnostic B and Y fragment ions of a sialic acid attached to a hexose unit, shown by Y₂ at *m/z* 504, Y₁ at *m/z* 300 and B₁ at *m/z* 398. The second is a HCD product ion mass spectrum of NeuAc₂Hex₁HexNAc₁ (*m/z* 1240). The spectrum shows diagnostic Y and B ions for the presence of two terminal sialic acids, shown by Y_{2α}/Y_{1β} at *m/z* 865, along with B_{1α}/B_{1β} at *m/z* 398. The third shows the CID product ion spectrum of NeuAc₁Hex₂HexNAc₂ (*m/z* 1328). The spectrum shows diagnostic fragment ions for a terminal sialic acid unit, glycosidically linked to a hexose unit, shown by Y_{2α} at *m/z* 953, B_{1α} at *m/z* 398 and B_{2α} at *m/z* 731. In addition, the B_{2β} at *m/z* 486 and Y_{1β} at *m/z* 865 show the second branch of the glycan that contains the terminal HexNAc and a hexose.

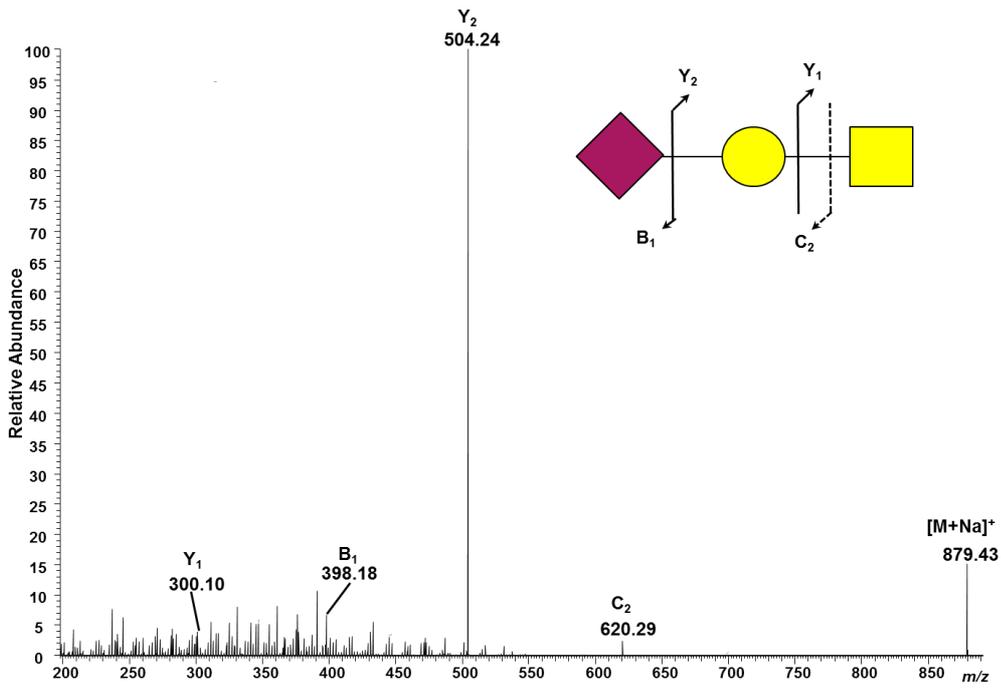


Figure 82: HCD product ion mass spectrum of permethylated O-glycan NeuAc₁Hex₁HexNAc₁ (precursor *m/z* 879).

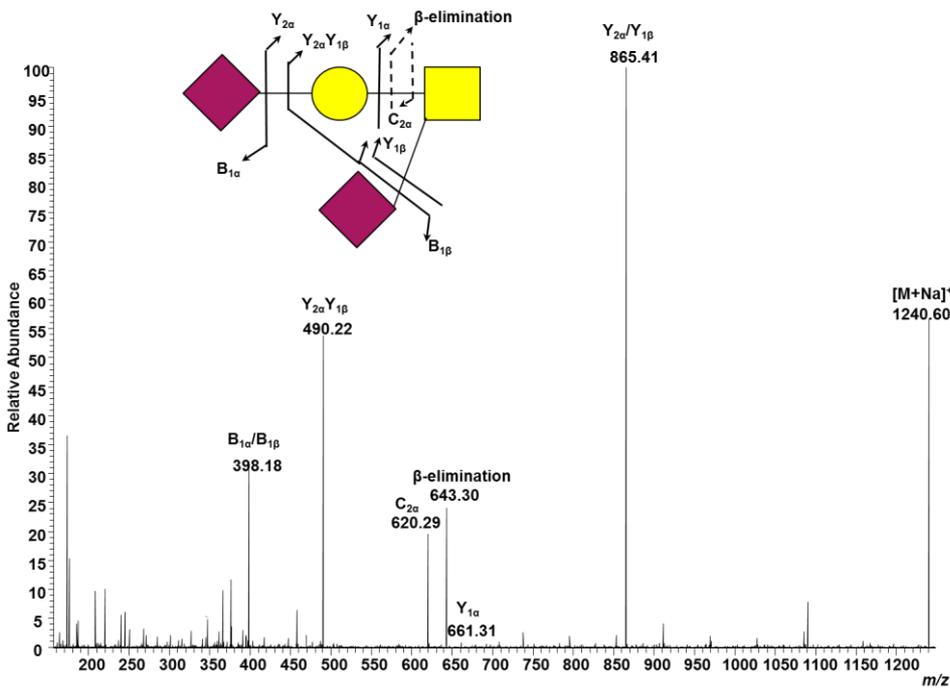


Figure 83: HCD product ion mass spectrum of permethylated O-glycan NeuAc₂Hex₁HexNAc₁ (precursor *m/z* 1240).

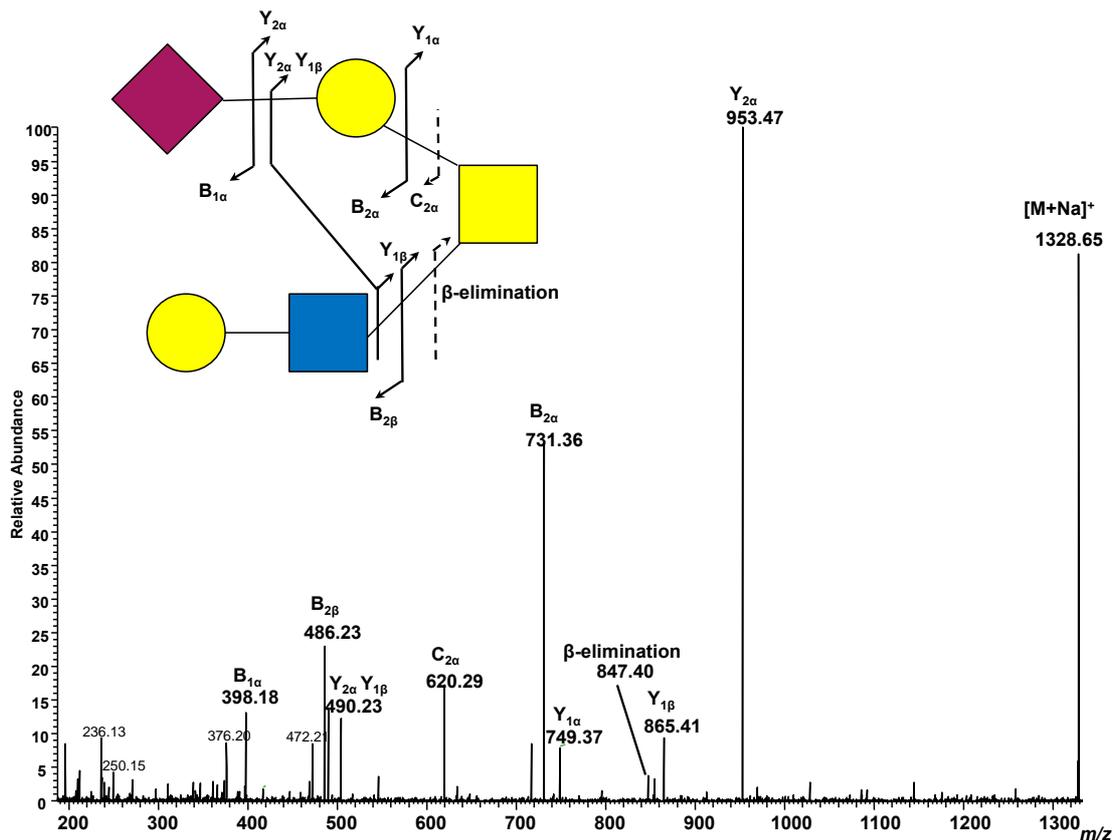


Figure 84: CID product ion mass spectrum of permethylated O-glycan, NeuAc₁Hex₂HexNAc₂ (precursor *m/z* 1328)

6.3.1. Chemical inhibition of O-glycan biosynthesis

Since the developed O-glycan release method had given such interesting data with the wild-type MSCs (from Y101 and Y201 cell lines), the cell lines were treated with an O-glycan biosynthesis inhibitor, benzyl 2-acetamido-2-deoxy- α -D-galactopyranoside (BadGal), for 48 hours. BadGal treatment was carried out to test whether disrupting O-glycan biosynthesis affected the O-glycan profile of the MSCs and to further test whether the developed O-glycan release method was able to detect differences between the O-glycan profiles of these different cell samples following inhibitor treatment.

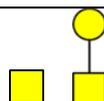
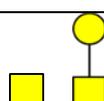
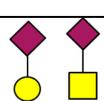
BadGal is a synthetic analogue of N-acetylgalactosamine²⁸⁸ and acts as a competitive substrate of the glycosyltransferase which normally adds Gal to GalNAc to give the T-antigen, Gal β 1-3GalNAc²⁸⁹. Studies investigating the effects of BadGal on glycan biosynthesis on mucins have found, using western blotting and ELISA, that after 24 h of BadGal treatment there was an increase in Tn-antigen (GalNAc-Ser) and an unexpected increase in Gal β 1-

3GalNAc, as well as a decrease in sialic acid content²⁹⁰. The decrease in sialylation of mucin O-glycans was found to be caused by in situ β -galactosylation of BadGal, which then acts as a competitive substrate of Gal β 1-3GalNAc α 2,3-sialyltransferase, seen by the accumulation of NeuAc α 2-3Gal β 1-3GalNAc α -O-benzyl in BadGal-treated cells²⁹⁰. Another study found that inhibition of O-glycosylation by BadGal was only partial²⁹¹.

Katherine Wilson had analysed glycans of BadGal-treated MSCs using lectin staining and flow cytometry to determine whether the treatment altered glycan biosynthesis. It was found that after treatment with BadGal there was a significant difference in the binding of the lectins. Three of the lectins used, peanut agglutinin (PNA), jacalin, and *Vicia villosa* lectin (VVL), preferentially bind to structures in O-glycans. PNA binds to Gal-GalNAc (when not sialylated). Jacalin and VVL bind to GalNAc, and Gal-GalNAc (and can still bind when sialylated). *Sambucus nigra* agglutinin (SNA) was also used and binds to NeuAc-Gal, which is a linkage found in both N- and O-glycans or NeuAc-GalNAc, and is present only in O-glycans. Details of the lectins used can be found in Table 16.

It was found that PNA binding increased. PNA binds to terminal Gal-GalNAc, and since binding of this lectin only occurs to unsialylated structures, the increase in binding for PNA could suggest a decrease in sialylated structures. This result would also suggest that BadGal treatment results in truncated glycans, as expected. What lectin staining fails to show is which glycan structures are missing with respect to those in untreated cells.

Table 16: Lectins used for lectin staining, by K. Wilson, for analysis of N- and O glycans

Lectin	Binding specificity
Peanut agglutinin (PNA)	Gal-GalNAc (when not sialylated) 
Jacalin	GalNAc, Gal-GalNAc (will also bind if sialylated) 
<i>Vicia villosa</i> lectin (VVL)	GalNAc, Gal-GalNAc (will also bind if sialylated) 
<i>Sambucus nigra</i> agglutinin (SNA)	NeuAc-Gal, NeuAc-GalNAc 

The proliferating BadGal-treated MSCs were harvested after 48 h treatment and the O-glycans were released and permethylated. MS analysis revealed O-glycans at m/z 518, 722, 879 and 967, for Y201 and at m/z 518, 879 and 967, for Y101, Figure 85 and Table 17.

On comparing the resulting glycan profiles for each of the wild-type cell lines with those treated with BadGal, the most noticeable difference was the decrease in the abundance of sialylated glycans, Figure 86. For both Y201 and Y101 cell lines, Hex₁HexNAc₁NeuAc₁ was still present in the BadGal-treated samples, but the abundance was lower than in the wild-type samples, with the decrease in the Y101 cell line being significant ($P = 0.0461$ *), Table 18. Hex₂HexNAc₁NeuAc₁ is not detected in either Y201 or Y101 BadGal-treated samples, but since this glycan was present at relatively low abundance in the wild-type samples, the change was not found to be significant. Hex₁HexNAc₁NeuAc₂ was also not detected in either Y201 or Y101 BadGal-treated cells, with the change being significant ($P = 0.0001$ *** and $P = 0.0188$ *, respectively), Table 18. For Hex₂HexNAc₂NeuAc₁, which was also undetected in the BadGal-treated samples from both cell lines, the only significant change was for the Y101 samples ($p = 0.0481$ *), Table 18, since this glycan was present at relatively low abundance in the wild-type cells for Y201. A decrease in abundance of sialylated glycans was also reported for mucins following BadGal treatment²⁹⁰. For both the Y201 and Y101 cell lines BadGal treatment resulted in an increase in abundance of T-antigen, also reported using lectin staining of mucins following BadGal treatment²⁹⁰. For the Y201 cell line, there was little difference in the abundance of Hex₂HexNAc₁ and Hex₃HexNAc₁ between the wild-type and BadGal-treated cells, suggesting addition of a second and third galactose onto GalNAc is not affected by BadGal treatment. There was also no change in the abundance of Hex₂HexNAc₁ in Y101 wild-type and BadGal-treated cells. There was a significant increase, for both Y201 and Y101, in Hex₂HexNAc₂ in the BadGal-treated cells compared with the wild-type ($P = 0.0016$ ** and $P = 0.0229$ *, respectively). The large increase in abundance of this glycan could be because the O-glycan structure cannot be extended further, either by the addition of sialic acids, or by fucosylation; there are no detected fucosylated glycans in the BadGal-treated samples, shown by a significant decrease in the abundance of Hex₂HexNAc₂Fuc₁ ($P = < 0.0001$ ***) in the Y201 BadGal-treated sample compared with the wild-type sample, and no detection of Hex₃HexNAc₂Fuc₁. Hex₂HexNAc₂Fuc₁ was also not detected in the Y101 BadGal-treated samples, but this glycan was only at low abundance in the wild-type cells and therefore this was not a significant change. The results reveal that BadGal treatment of MSCs resulted in truncated O-glycans, with a reduced number of sialylated structures and

the elimination of fucosylated structures, and also shows that the developed O-glycan release method has the potential to be applied to monitoring changes to O-glycan profiles between experimentally comparable cultured MSC cells.

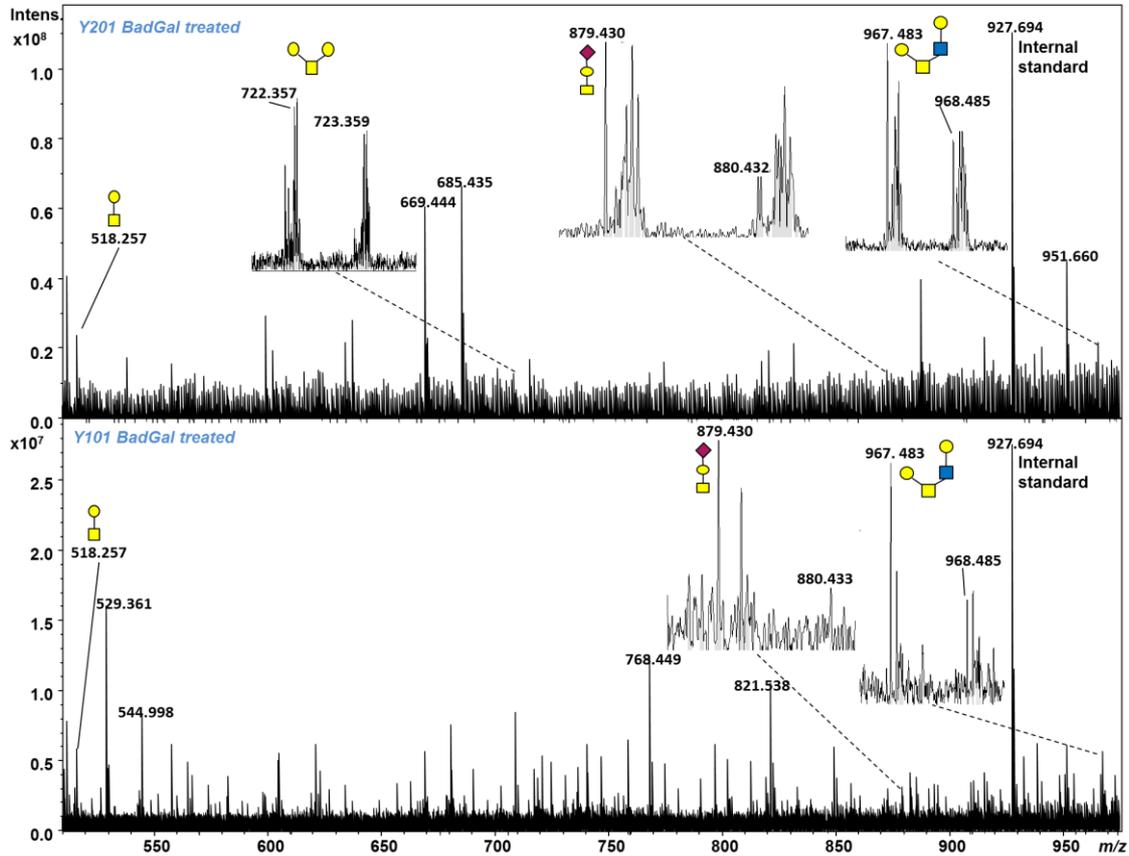


Figure 85 MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y201 and Y101 MSCs after 48 h pre-treatment with the O-glycan biosynthesis inhibitor, BadGal. All glycans ionised as $[M + Na]^+$

Table 17: List of permethylated O-glycans identified, with their mass accuracies, from Y201 and Y101 MSCs after 48 h pre-treatment with O-glycan biosynthesis inhibitor, BadGal

glycan composition	calculated <i>m/z</i>	Y201 BadGal treated (av. measured <i>m/z</i> , n=5)	Y201 BadGal treated mass accuracy (ppm)	Y101 BadGal treated (av. measured <i>m/z</i> , n=5)	Y101 BadGal treated mass accuracy (ppm)
Hex ₁ HexNAc ₁	518.2572	518.2571	-0.11	518.2572	-0.09
Hex ₂ HexNAc ₁	722.3569	722.3567	-0.25	722.3568	-0.08
Hex ₁ HexNAc ₁ NeuAc ₁	879.4304	879.4302	-0.18	879.4305	0.14
Hex ₂ HexNAc ₂	967.4833	967.4830	-0.30	967.4834	0.06

Table 18: List of P values for the significant differences between Y201 and Y101 WT O-glycans. Statistical analysis carried out using one-way ANOVA with Holm Sidak post-hoc test. ns = not significant

glycan composition	Y201 P values for significant difference between WT and BadGal treated (n=5)	Y101 P values for significant difference between WT and BadGal treated (n=5)
Hex ₁ HexNAc ₁	ns	ns
Hex ₂ HexNAc ₁	ns	ns
Hex ₁ HexNAc ₁ NeuAc ₁	ns	0.0461
Hex ₃ HexNAc ₁	ns	-
Hex ₂ HexNAc ₂	0.0016	0.0229
Hex ₂ HexNAc ₁ NeuAc ₁	ns	ns
Hex ₂ HexNAc ₂ Fuc ₁	< 0.0001	ns
Hex ₃ HexNAc ₂	-	ns
Hex ₁ HexNAc ₁ NeuAc ₂	0.0001	0.0188
Hex ₂ HexNAc ₂ NeuAc ₁	ns	0.0481
Hex ₃ HexNAc ₂ Fuc ₁	ns	ns

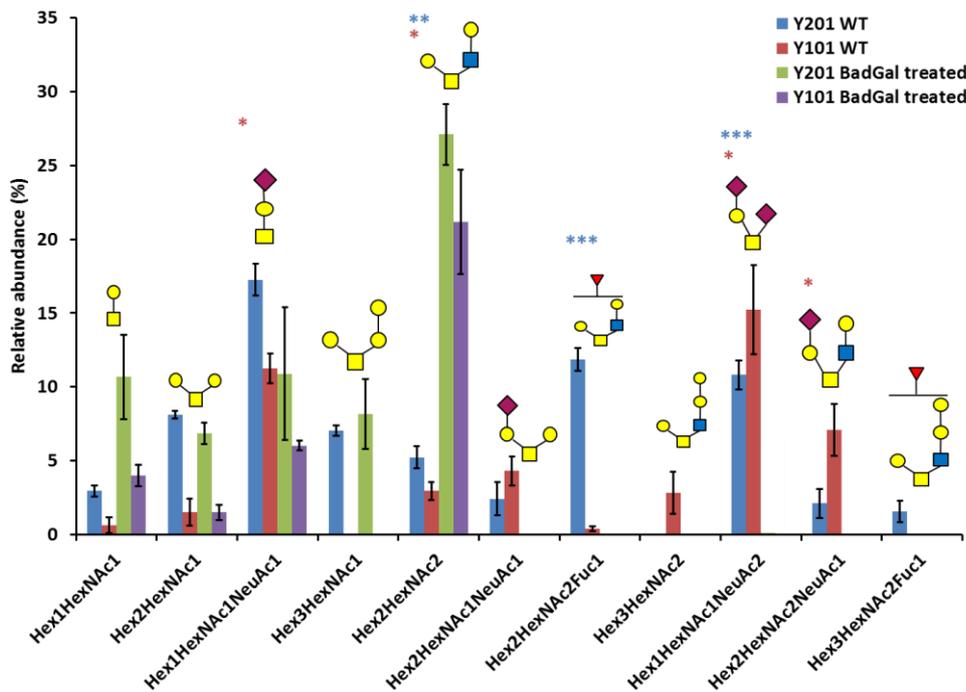


Figure 86: Comparison of relative O-glycan abundance between wild-type and BadGal treated MSCs from two cell lines (Y201 and Y101). Relative abundance calculated by calculating each glycan signal as a percentage of the spike internal standard. Blue asterisks are for statistically significant changes to the Y201 cell line and red asterisks are for the Y101 cell line, where $P < 0.05 = *$, $P < 0.01 = **$ and $P < 0.001 = *$.**

6.3.2. Genetic modification of MSCs: effect on glycan profile

Cog4 knockdown (KD) MSCs (where Cog4 refers to the Cog4 lobe of the COG complex) were prepared to test whether changes were detectable in the O-glycan profiles when compared with the wild-type. MSCs were grown in 10 cm culture dishes until just under confluency, before being harvested. MS analysis of the subsequently released O-glycans identified characteristic signals at m/z 518, 722, 879, 967, and 1328, Figure 87 and Table 19.

For glycan structures Hex₂HexNAc₁ and Hex₂HexNAc₂, both bearing a terminal galactose, there is an increase in the relative abundance in the Cog4KD cells compared with the wild-type cells, Figure 88. This could be attributed to reduced ability to transfer sialic acid to the galactose in Hex₂HexNAc₂, since in the Cog4KD sample, Hex₂HexNAc₁NeuAc₁ was not detected, a significant change when compared with wild-type cells ($P = 0.0962 *$), Table 20. However, there is a significant increase in the abundance of Hex₁HexNAc₁NeuAc₁, ($P =$

0.0069 **), Table 20. The large increase in abundance of Hex₁HexNAc₁NeuAc₁ could be due to reduced ability to transfer a second sialic acid to the growing glycan chain, shown by the absence of Hex₁HexNAc₁NeuAc₂ in the Cog4KD sample, a change compared to the wild-type sample that is significant (P = 0.0028 **), Table 20. The fucosylated O-glycan, Hex₂HexNAc₂Fuc₁, was not detected, although this glycan is only relatively low abundance in the wild-type sample.

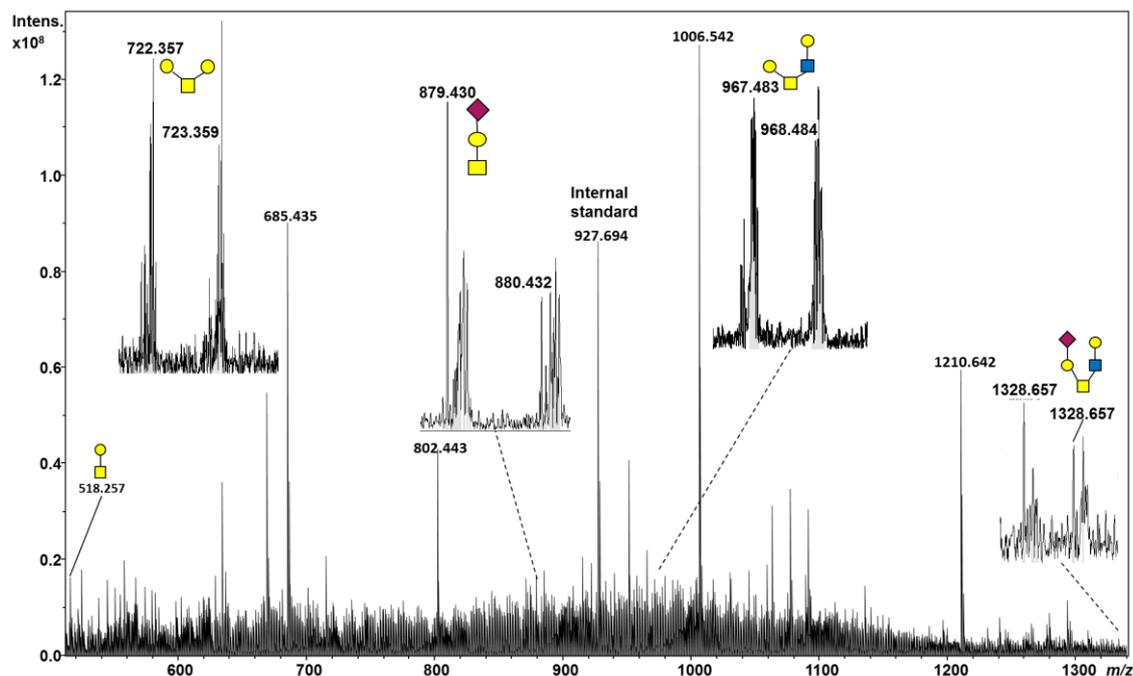


Figure 87: MALDI-FT-ICR mass spectrum of permethylated O-glycans, released from Y101 Cog4KD MSCs. All glycans ionised as $[M + Na]^+$

Table 19: List of permethylated O-glycans identified, along with their mass accuracies, in Cog4KD MSCs

glycan composition	calculated m/z	Y101 Cog4KD (av. measured m/z , n = 5)	Y101 Cog4KD mass accuracy (ppm)
Hex ₁ HexNAc ₁	518.2572	518.2571	-0.28
Hex ₂ HexNAc ₁	722.3569	722.3568	-0.15
Hex ₁ HexNAc ₁ NeuAc ₁	879.4304	879.4303	-0.11
Hex ₂ HexNAc ₂	967.4833	967.4827	-0.58
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	-	-
Hex ₃ HexNAc ₂	1171.5831	-	-
Hex ₂ HexNAc ₂ NeuAc ₁	1328.6569	1328.6559	-0.76

Table 20: P values for significant differences between Y101 wild-type and Cog4KD O-glycans. Statistical analysis used a one-way ANOVA with Holm Sidak post-hoc test.

glycan composition	Y101 P values for significant differences between WT/Cog4KD O-glycans (n = 5)
Hex ₁ HexNAc ₁	ns
Hex ₂ HexNAc ₁	ns
Hex ₁ HexNAc ₁ NeuAc ₁	0.0069
Hex ₃ HexNAc ₁	-
Hex ₂ HexNAc ₂	ns
Hex ₂ HexNAc ₁ NeuAc ₁	0.0962
Hex ₂ HexNAc ₂ Fuc ₁	ns
Hex ₃ HexNAc ₂	ns
Hex ₁ HexNAc ₁ NeuAc ₂	0.0028
Hex ₂ HexNAc ₂ NeuAc ₁	ns
Hex ₃ HexNAc ₂ Fuc ₁	-

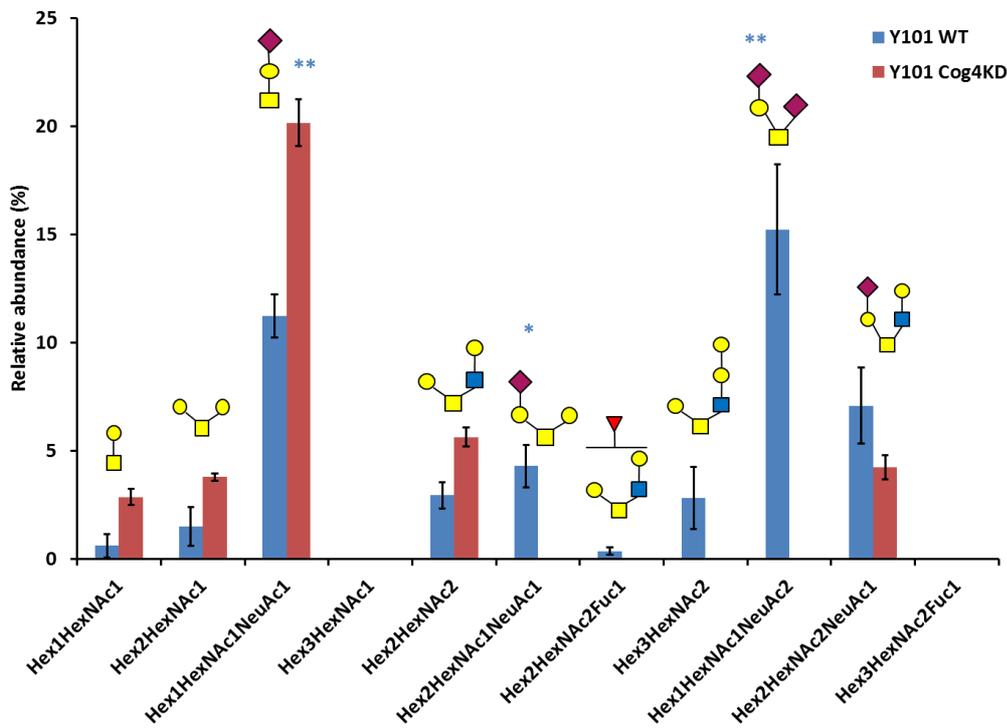


Figure 88 Comparison of relative abundance of O-glycans between Y101 WT and Cog4KD MSCs. Relative abundance of glycans were quantified by comparing peak intensities with a spiked internal standard. Blue asterisks indicate statistically different changes to the relative glycans. P < 0.05 = *, < 0.01 = **.

A study by Reynders et al analysed *N*-glycans in CDG-II patient cells with a mutation in *Cog4*, resulting in down regulation of *Cog4* expression, and found there was a decrease in sialylated *N*-glycans²⁹². It has also been reported that in disorders that affect both *N*- and *O*-glycans, sialylation of both glycan types is decreased¹². One study analysed, by mass spectrometry, *O*-glycans from serum and plasma samples of patients with mutations in *Cog4* and *Cog7* and found decreased sialylation of both glycan types. In addition, the first described example of an in-born error of metabolism affecting both *N*- and *O*-glycosylation reported a decrease in sialylated *N*- and *O*-glycans, where the *O*-glycosylation status was determined by analysing apolipoprotein C-III (a protein which is only *O*-glycosylated), using isoelectric focussing²⁹³. K. Wilson carried out *N*-glycan profiling of MSC *Cog4*KD mutants and also observed a decrease in sialylation. With lectin staining and flow cytometry, K. Wilson found there was an increase in PNA binding. PNA binds to terminal Gal-GalNAc, and since binding of this lectin only occurs to unsialylated structures, the increase in binding for PNA could suggest a decrease in sialylated structures. All the evidence described above adds confidence to the results from this *O*-glycan profiling, that was carried out using only a small number (~ 1 million) of cultured cells. There is no literature describing changes to *O*-glycosylation profiles of *Cog4* mutants of MSCs.

6.3.3. Demonstration of the one-pot *N*- and *O*-glycan release method to study the *N*- and *O*-glycome of MSCs and to monitor the change to glycosylation between the wild-type and mutant cell lines.

Since the developed one-pot method was successful on porcine bladder urothelium (chapter 4), the same MSC wild-type and *Cog4*KD samples that had already had their *O*-glycans released, were subsequently treated with PNGase F to release the *N*-glycans. MS analysis of the released *N*-glycans identified glycan signals between m/z 967 – m/z 3054, from the wild-type MSCs, showing the presence of both oligomannose and complex-type *N*-glycans. For the *Cog4*KD MSCs, *N*-glycans were identified between m/z 967 – 2396, also showing the presence of both oligomannose and complex-type *N*-glycans, Figure 89 and Table 21.

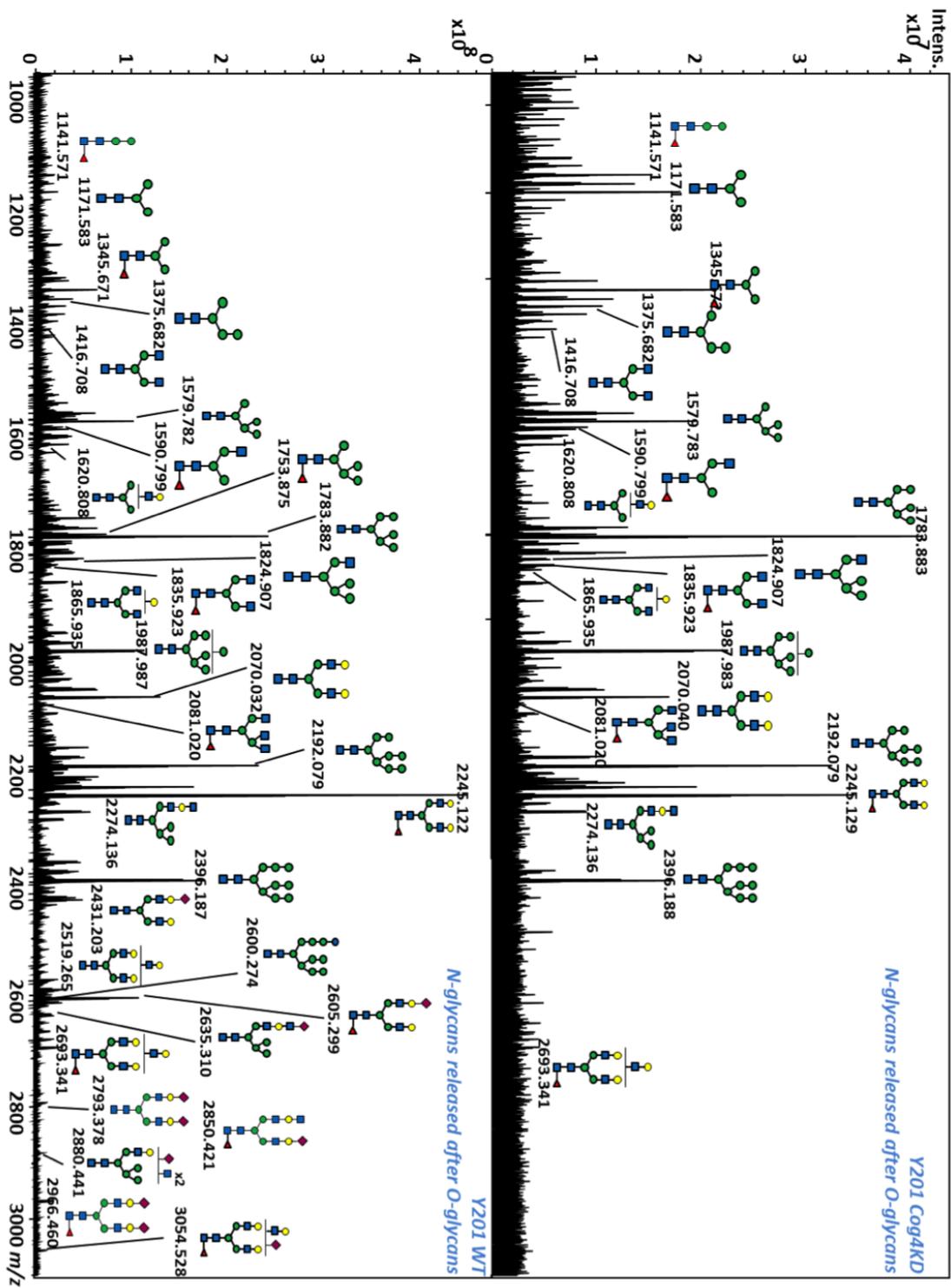


Figure 89: MALDI-FT-ICR mass spectrum of permethylated N-glycans released from Y201 Cog4KD MSCs (top panel) and WT MSCs (bottom panel). All glycans ionised as $[M + Na]^+$.

Figure 90 shows a comparison of the relative abundance of *N*-glycans from the wild-type and Cog4KD MSC samples. For both samples, oligomannose type *N*-glycans appear to be the most abundant species, but the abundance of these glycans is lower in the Cog4KD sample than in the WT sample. In addition, there is an increase in abundance of small glycans (HexNAc₂Hex₂, HexNAc₂Hex₂Fuc₁, and HexNAc₂Hex₃) for the Cog4KD samples, and sialylated glycans, which are detected in the wild-type MSCs, are not detected in the Cog4KD samples. There is also a decrease in abundance of complex-type glycans that contain a terminal galactose in the Cog4KD sample. These results are in agreement with data obtained previously by K. Wilson, who carried out *N*-glycan profiling of MSCs, and with literature (discussed above) on O-glycan profiling which showed a decrease in sialylated *N*-glycans in Cog4 CDG-II patients.

Table 21: List of permethylated *N*-glycans identified, with their mass accuracies, in Cog4KD MSCs

glycan composition	calculated <i>m/z</i>	Av. measured <i>m/z</i> (n = 5) Cog4KD	mass accuracy Cog4KD (ppm)	Av. measured <i>m/z</i> (n = 5) mass WT	mass accuracy WT (ppm)
Hex ₂ HexNAc ₂	967.4833	967.4833	0.03	967.4826	-0.72
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	1141.5726	0.05	1141.5717	-0.73
Hex ₃ HexNAc ₂	1171.5831	1171.5828	-0.28	1171.5826	-0.40
Hex ₃ HexNAc ₂ Fuc ₁	1345.6723	1345.6724	0.04	1345.6719	-0.33
Hex ₄ HexNAc ₂	1375.6828	1375.6831	0.18	1375.6822	-0.43
Hex ₃ HexNAc ₃	1416.7094	1416.7100	0.45	1416.7086	-0.58
Hex ₄ HexNAc ₂ Fuc ₁	1549.7720	1549.7361	-23.18	1549.7352	-23.75
Hex ₅ HexNAc ₂	1579.7826	1579.7829	0.16	1579.7825	-0.06
Hex ₃ HexNAc ₃ Fuc ₁	1590.7986	1590.7994	0.50	1590.7994	0.48
Hex ₄ HexNAc ₃	1620.8092	1620.8101	0.54	1620.8093	0.06
Hex ₅ HexNAc ₂ Fuc ₁	1753.8718	-	-	1753.8620	-5.61
Hex ₆ HexNAc ₂	1783.8824	1783.8837	0.72	1783.8823	-0.08
Hex ₅ HexNAc ₃	1824.9089	1824.9092	0.14	1824.9089	0.01
Hex ₃ HexNAc ₄ Fuc ₁	1835.9249	1835.9269	1.08	1835.9263	0.74
Hex ₄ HexNAc ₄	1865.9355	1865.9351	-0.21	1865.9344	-0.58
Hex ₇ HexNAc ₂	1987.9822	1987.9834	0.61	1987.9831	0.44
Hex ₆ HexNAc ₃	2029.0087	-	-	2029.0086	-0.07

Hex ₅ HexNAc ₄	2070.0353	2070.0365	0.59	2070.0340	-0.61
Hex ₈ HexNAc ₂	2192.0819	2192.0835	0.75	2192.0783	-1.65
Hex ₅ HexNAc ₄ Fuc ₁	2244.1245	2244.1268	1.03	2244.1204	-1.83
Hex ₆ HexNAc ₄	2274.1350	2274.1397	2.05	2274.1379	1.29
Hex ₉ HexNAc ₂	2396.1817	2396.1833	0.65	2396.1841	1.00
Hex ₅ HexNAc ₄ NeuAc ₁	2431.2089	-	-	2431.2031	-2.39
Hex ₆ HexNAc ₅	2519.2614	-	-	2519.2586	-1.11
Hex ₆ HexNAc ₃ Fuc ₁ NeuAc ₁	2564.272	-	-		
Hex ₁₀ HexNAc ₂	2600.2815	-	-	2600.2570	-9.42
Hex ₅ HexNAc ₄ Fuc ₁ NeuAc ₁	2605.2981	-	-	2605.2952	-1.10
Hex ₆ HexNAc ₄ NeuAc ₁	2635.3087	-	-	2635.3169	3.12
Hex ₆ HexNAc ₅ Fuc ₁	2693.3506	2693.3583	2.86	2693.3412	-3.50
Hex ₅ HexNAc ₄ NeuAc ₂	2792.3826	-	-	2792.3775	-1.84
Hex _{5H} HexNAc ₄ Fuc ₁ NeuAc ₁	2966.4718	-	-	2966.4574	-4.86
Hex ₆ HexNAc ₅ Fuc ₁ NeuAc ₁	3054.5242	-	-	3054.5358	3.81

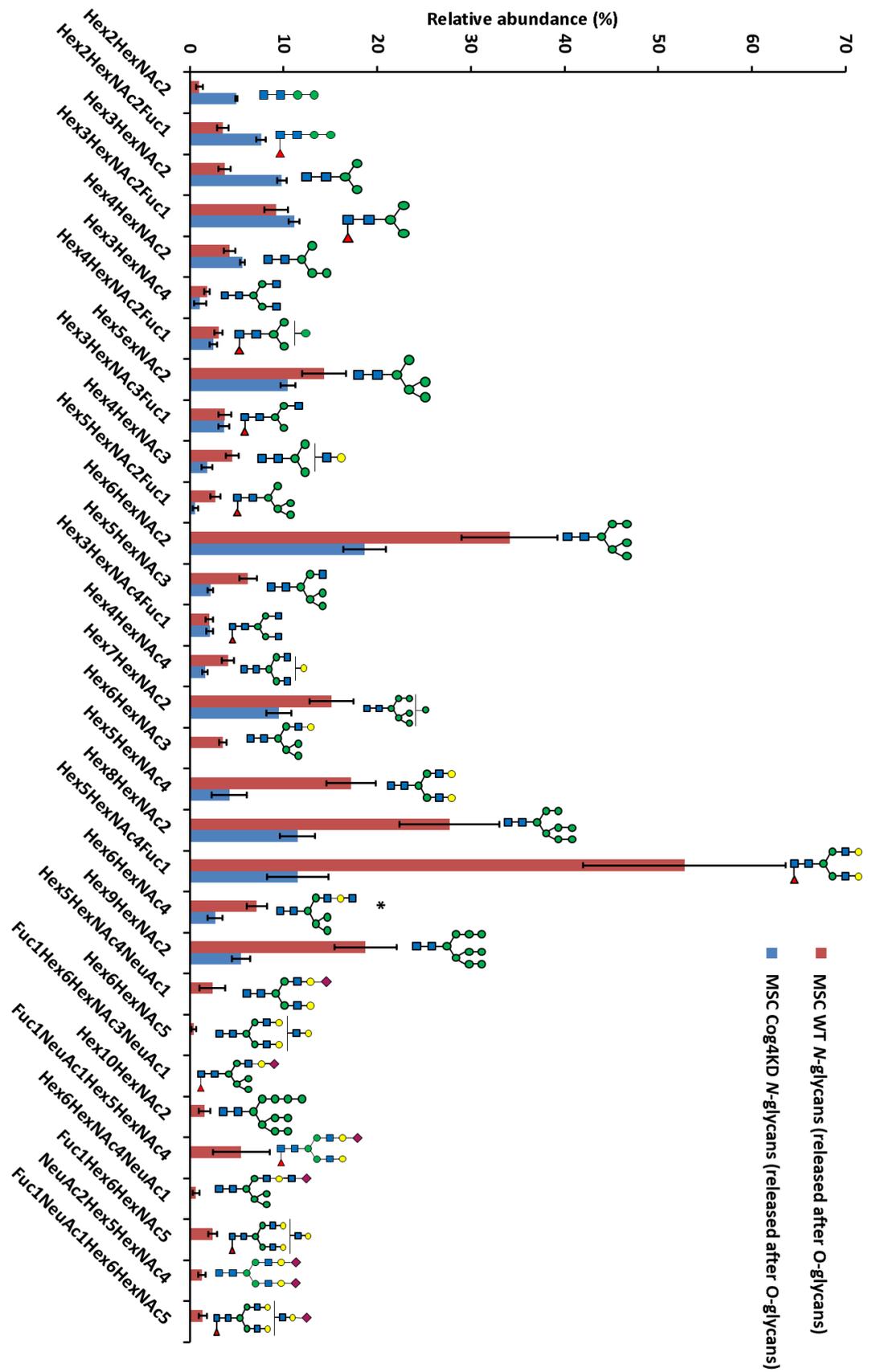


Figure 90: Comparison of the relative abundance of permethylated *N*-glycans in Y101 WT and Cog4KD MSCs.

6.3.4. Adipogenesis

To see if changes in O-glycan profiles could be observed between wild-type MSCs and those that have differentiated into adipocytes, Y201 and Y101 hTERT-MSCs were cultured in either basal or adipogenic medium for 21 days. The samples in basal medium were prepared for the purpose of identifying changes to glycan profiles associated with being cultured at full confluency (proliferating cells were cultured to close to confluency), and to distinguish them from changes occurring on differentiation into adipocytes. The Y101 cells were treated similarly to identify changes due purely to being in adipogenic medium for 21 days, as these have limited ability to differentiate into adipocytes. Some of the Y101 cells do differentiate; this is because the cell lines were produced from a heterogeneous mix of cells.

On day 21, cells were stained with Oilred O to assess the extent of formation of lipid droplets and therefore differentiation into adipocytes. Cells cultured in basal medium for 21 days showed no staining as no lipid droplets were formed. Both Y201 and Y101 cells cultured in adipogenic medium for 21 days showed staining for oil droplet formation, but staining was less prominent for the Y101 cell line than for the Y201 cell line, Figure 91.

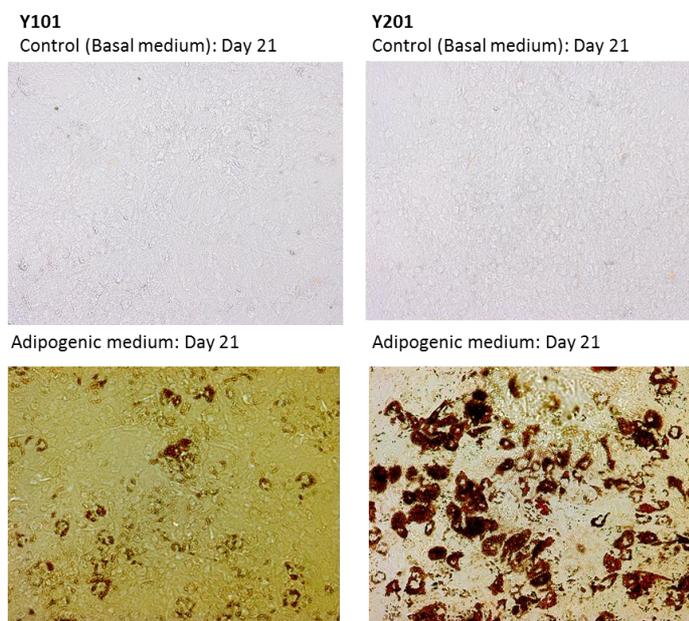


Figure 91: Oilred O staining of Y201 and Y101 cell lines after being cultured in either basal or adipogenic medium for 21 days. Oilred O did not stain the cells in the basal medium. As cells differentiate into adipocytes lipid droplets form and the Oilred O stains the lipid droplets red. Y201 differentiate well and form many lipid droplets, Y101 although some staining is seen differentiation is limited.

After 21 days, the cells were harvested and the O-glycans were released and permethylated before analysis by mass spectrometry. For both Y201 and Y101 MSCs cultured in basal media for 21 days, O-glycans were identified at m/z 518, 722, 879, 967, 1083, 1141, and 1171. For the adipocytes derived from the Y201 and Y101 MSCs, O-glycans were identified at m/z 518, 722, 879, 967, 1083, 1141, and 1171, plus three additional O-glycan signals at m/z 1328, 1345 and 1689, Figure 92, Figure 93, Table 22 and Table 23.

On comparing the difference between the abundance of each glycan in the Y201 MSCs and the adipocytes derived from them, there is an increase in abundance of Hex₁HexNAc₁, Hex₂HexNAc₁NeuAc₁ and Hex₂HexNAc₂Fuc₁ (P values = < 0.0001 ^{***}, 0.0120 ^{*} and < 0.0001 ^{***} respectively). In addition, in the adipogenic sample, additional glycans (Hex₃HexNAc₂, Hex₂HexNAc₂NeuAc₁, Hex₃HexNAc₂Fuc₁, Hex₂HexNAc₂NeuAc₂) appear, which are not present in the samples cultured in basal medium for 21 days. Statistical analysis of each of these glycan signals revealed a significant change (P = <0.0001 ^{***}, 0.0053 ^{**}, 0.0053 ^{**} and 0.0081 ^{**}), Table 24, Figure 94.

For the Y101 cell line, in contrast to Y201, there is no change in the relative abundance in Hex₁HexNAc₁ between the cells cultured in basal or adipogenic medium. Also unlike Y201, there is a decrease in the relative abundance of Hex₂HexNAc₁NeuAc₁ (P = 0.0303 ^{*}) in Y101 samples. Similarly to Y201 there is a significant increase in abundance of Hex₂HexNAc₂Fuc₁, Hex₃HexNAc₂ and Hex₃HexNAc₂Fuc₁ (P = 0.0391 ^{*}, < 0.0001 ^{***}, and 0.0391 ^{*}, respectively), Figure 94.

For the additional glycans that appear in both the Y201 and Y101 cell lines cultured in adipogenic medium there is a similar relative abundance in the two cell lines for Hex₃HexNAc₂ and Hex₃HexNAc₂Fuc₁ but a much higher abundance in the Y201 cell line for Hex₂HexNAc₂Fuc₁, Hex₂HexNAc₂NeuAc₁ and Hex₂HexNAc₂NeuAc₂. It is likely therefore that those changes that are similar in the two cell lines are due to the cells being cultured in adipogenic medium, where the change has been induced by the medium, whereas those changes in the Y201 cell line that are more prominent than in the Y101 cell lines are changes to the glycans either as a result of differentiation into adipocytes or the glycan changes enable the cells to differentiate.

Focusing on the Hex₂HexNAc₂ O-glycan, the addition of fucose to give the increased abundance of Hex₂HexNAc₂Fuc₁ could be due to the up regulation of a specific fucosyltransferase enzyme that transfers fucose to either the galactose or *N*-acetylglucosamine during adipogenesis. Experiments could be carried out to see if this biosynthetic enzyme is upregulated during adipogenesis. Fucosyltransferase gene knock-out could be carried out, followed by adipogenesis to see the affect on the glycan profile, to test whether Hex₂HexNAc₂Fuc₁ is a consequence of, or is required for, adipogenesis .

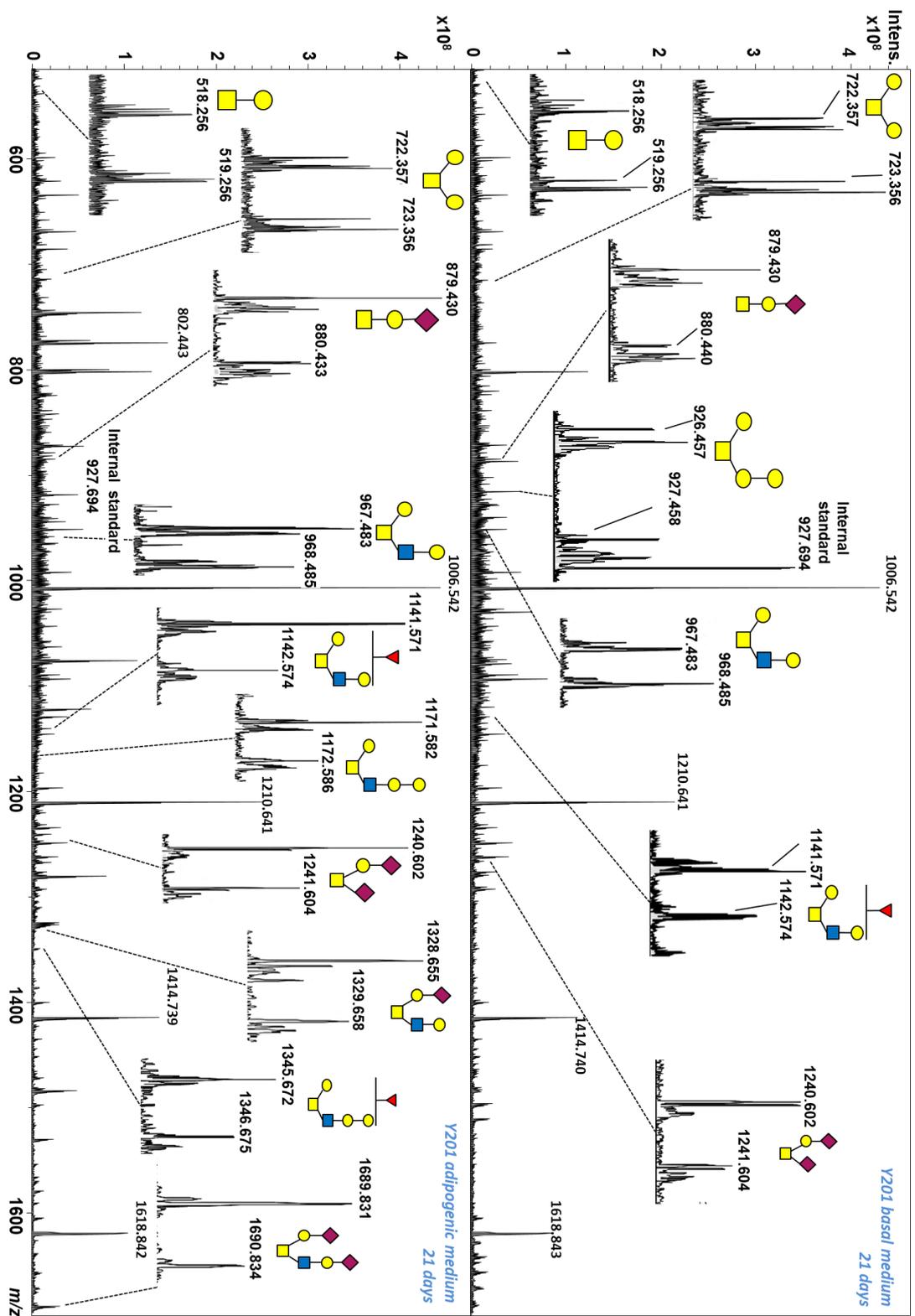


Figure 92: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y201 MSCs after 21 days in either basal medium or adipogenic medium. All glycans ionised as $[M + Na]^+$.

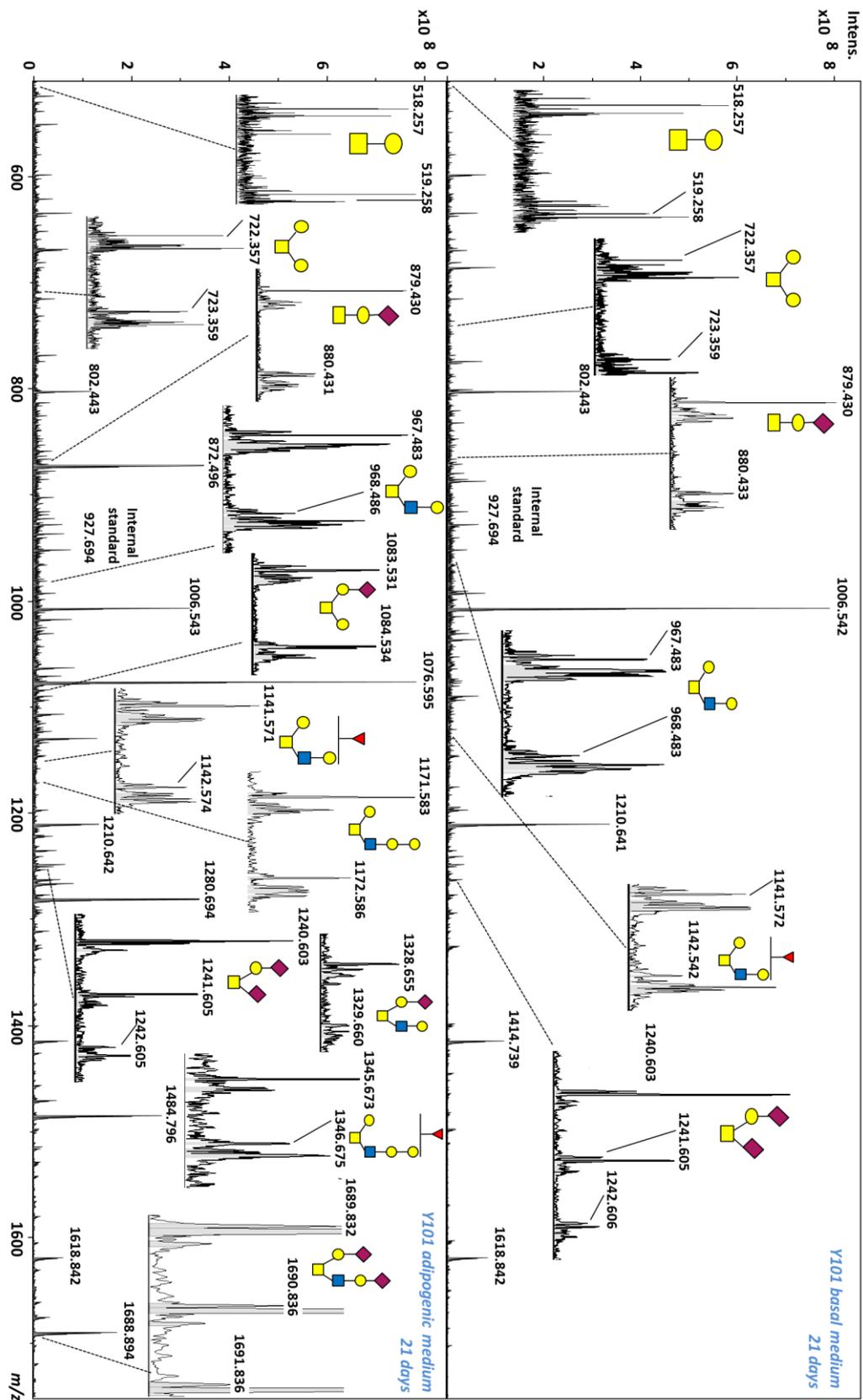


Figure 93: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y101 MSCs after 21 days cultured in either basal or adipogenic medium. All glycans ionised as $[M + Na]^+$.

Table 22: List of identified MSC permethylated O-glycans (Y201 cell line) and mass accuracies

glycan composition	calculated <i>m/z</i>	Y201 (basal) av. measured <i>m/z</i> (n = 8)	Y201 control mass accuracy (ppm)	Y201 adipogenesis av. measured <i>m/z</i> (n = 12)	Y201 adipogenesis mass accuracy (ppm)
Hex ₁ HexNAc ₁	518.2572	518.2571	-0.12	518.2572	0.01
Hex ₂ HexNAc ₁	722.3569	722.3567	-0.29	722.3569	-0.01
Hex ₁ HexNAc ₁ NeuAc ₁	879.4304	879.4300	-0.48	879.4301	-0.29
Hex ₂ HexNAc ₂	967.4833	967.4826	-0.69	967.4826	-0.69
Hex ₂ HexNAc ₁ NeuAc ₁	1083.5306	-	-	1083.5294	1.07
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	1141.5712	-0.71	1141.5710	-1.35
Hex ₃ HexNAc ₂	1171.5831	-	-	1171.5822	-0.73
Hex ₁ HexNAc ₁ NeuAc ₂	1240.6045	1240.6024	-1.67	1240.6025	-1.63
Hex ₂ HexNAc ₂ NeuAc ₁	1328.6569	-	-	1328.6555	-1.03
Hex ₃ HexNAc ₂ Fuc ₁	1345.6723	-	-	1345.6795	5.34
NeuAc ₂ Hex ₂ HexNAc ₂	1689.8306	-	-	1689.8301	-0.30

Table 23: List of identified MSC permethylated O-glycans (Y101 cell line) and mass accuracies

glycan composition	calculated mass	Y101 control (basal) av. measured <i>m/z</i> (n = 8)	Y101 control mass accuracy (ppm)	Y101 adipogenesis av. measured <i>m/z</i> (n = 12)	Y101 adipogenesis mass accuracy (ppm)
Hex ₁ HexNAc ₁	518.2572	518.25723	-0.17	518.2573	0.12
Hex ₂ HexNAc ₁	722.3569	722.3567	-0.29	722.3569	0.04
Hex ₁ HexNAc ₁ NeuAc ₁	879.4304	879.43009	-0.27	879.4302	-0.23
Hex ₂ HexNAc ₂	967.4833	1083.52893	-0.37	967.4829	-0.40
Hex ₂ HexNAc ₁ NeuAc ₁	1083.5306	967.48283	-1.11	1083.5298	-0.76
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	1141.57093	-1.13	1141.5712	-1.12
Hex ₃ HexNAc ₂	1171.5831	1171.58312	-0.21	1171.5821	-0.88
Hex ₁ HexNAc ₁ NeuAc ₂	1240.6045	1240.60294	-1.07	1240.6030	-1.23
Hex ₂ HexNAc ₂ NeuAc ₁	1328.6569	-	-	1328.6555	-1.08
Hex ₃ HexNAc ₂ Fuc ₁	1345.6723	-	-	1345.6726	0.25

NeuAc ₂ Hex ₂ HexNAC ₂	1689.8306	-	-	1689.8319	0.79
---	-----------	---	---	-----------	------

Table 24: List of P values for the significant differences in O-glycan abundance between Y201 and Y101 MSCs cultured for 21 days in either basal medium or adipogenic medium. Statistical analysis carried out using one-way ANOVA with Holm Sidak post-hoc test.

glycan composition	Y201 p values for significant difference between glycan in cells cultured (21 days) in basal or adipogenic media (n = 12)	Y101 p values for significant difference between glycan in cells cultured (21 days) in basal or adipogenic media (n = 12)
Hex ₁ HexNAC ₁	0.0010	ns
Hex ₂ HexNAC ₁	ns	ns
Hex ₁ HexNAC ₁ NeuAc ₁	ns	0.0013
Hex ₂ HexNAC ₂	ns	ns
Hex ₂ HexNAC ₁ NeuAc ₁	0.0120	0.0303
Hex ₂ HexNAC ₂ Fuc ₁	ns	ns
Hex ₃ HexNAC ₂	< 0.0001	0.0391
Hex ₁ HexNAC ₁ NeuAc ₂	< 0.0001	< 0.0001
Hex ₂ HexNAC ₂ NeuAc ₁	0.0053	ns
Hex ₃ HexNAC ₂ Fuc ₁	0.0053	0.0391
NeuAc ₂ Hex ₂ HexNAC ₂	0.0081	ns

The one-pot method for glycan release was applied by subsequently releasing *N*-glycans after O-glycans were removed from Y201 MSCs and adipocytes derived from them. *N*-glycans were identified by their accurate masses, Table 25, from knowledge of possible glycan structures from the *N*-glycan biosynthetic pathway and expected *N*-glycan structures reported in the literature^{287,285,139}.

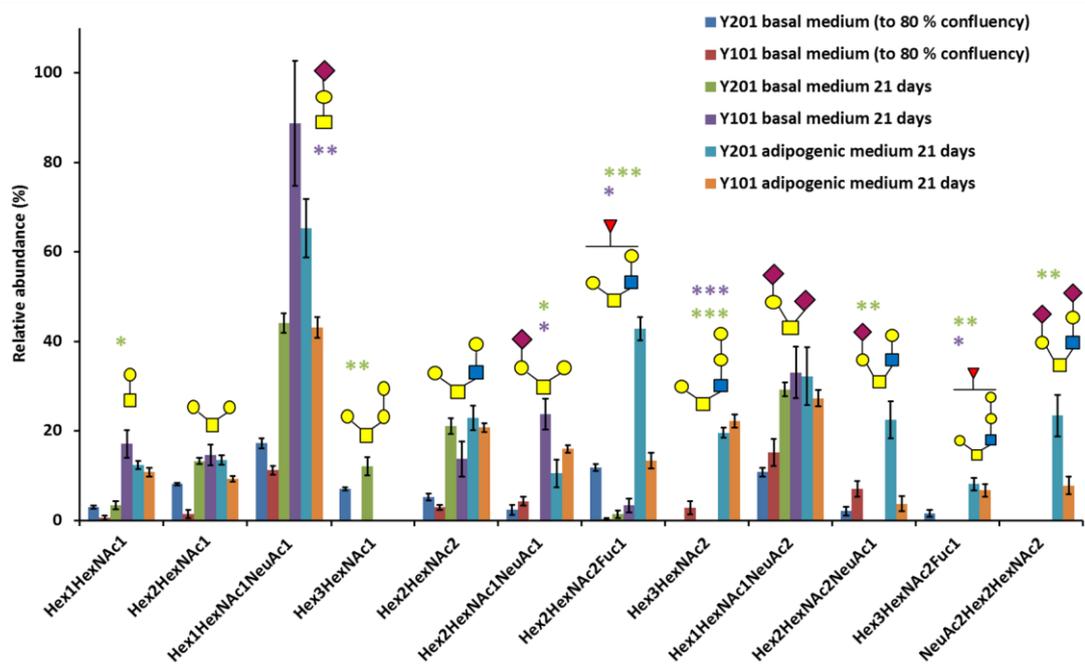


Figure 94: Comparison of the relative abundance of O-glycans in Y201 and Y101 MSC cell lines after being cultured in either basal or adipogenic medium for 21 days, alongside values for cells in basal medium for only approx 7 days (80% rather than complete confluence). Quantitation carried out by comparing the glycan peak intensities to the peak intensity of a spiked internal standard. Asterisks indicate a statistically significant change in the relative glycan abundance between the basal samples (21 days) and the adipogenic samples (green for Y201 and purple for Y101. P = * < 0.05, ** < 0.01, * < 0.001).**

For the *N*-glycans obtained from the Y201 cell line cultured in basal medium for 21 days, *N*-glycans were identified between *m/z* 967 – 3416, Figure 95 and for the Y201 cell line cultured in adipogenic medium for 21 days *N*-glycans were identified between *m/z* 967 – 3416, Figure 96. Comparing the resulting *N*-glycan profiles, oligomannose type structures were the most abundant across both cell lines. For the *N*-glycans derived from the adipogenic Y201 cell line cultured in adipogenic medium for 21 days, there was an increase in complex-type *N*-glycans and a reduction in the abundance of oligomannose-type *N*-glycans, compared with the basal medium cultured sample, Figure 97. However, on carrying out statistical analysis, none of the changes in the relative glycan abundances between the Y201 controls and the adipogenic samples was significant.

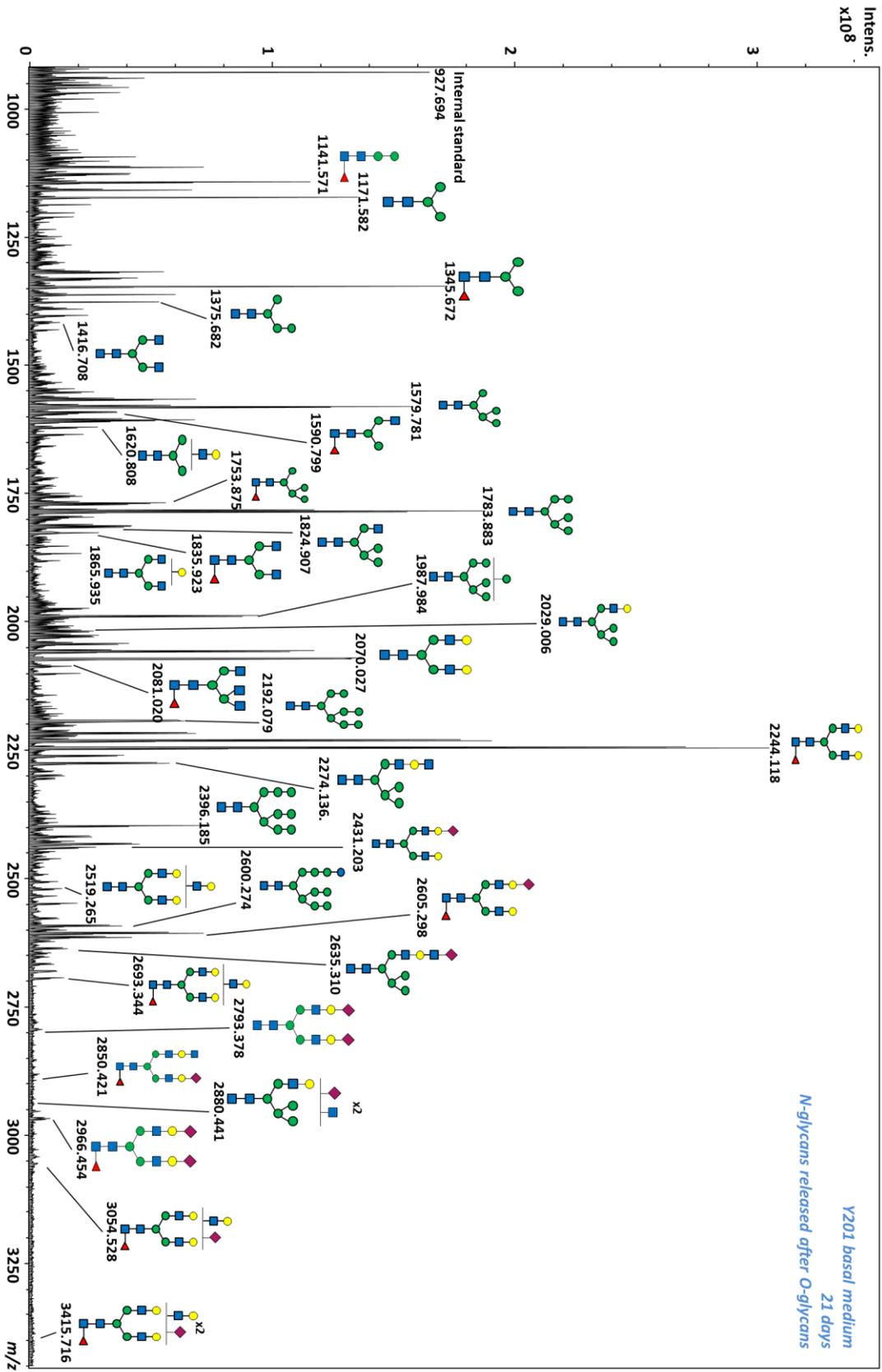


Figure 95: MALDI-FT-ICR mass spectrum of permethylated *N*-glycans released, after *O*-glycans, from Y201 MSCs cultured for 21 days in basal medium. All *N*-glycans ionised as $[M + Na]^+$.

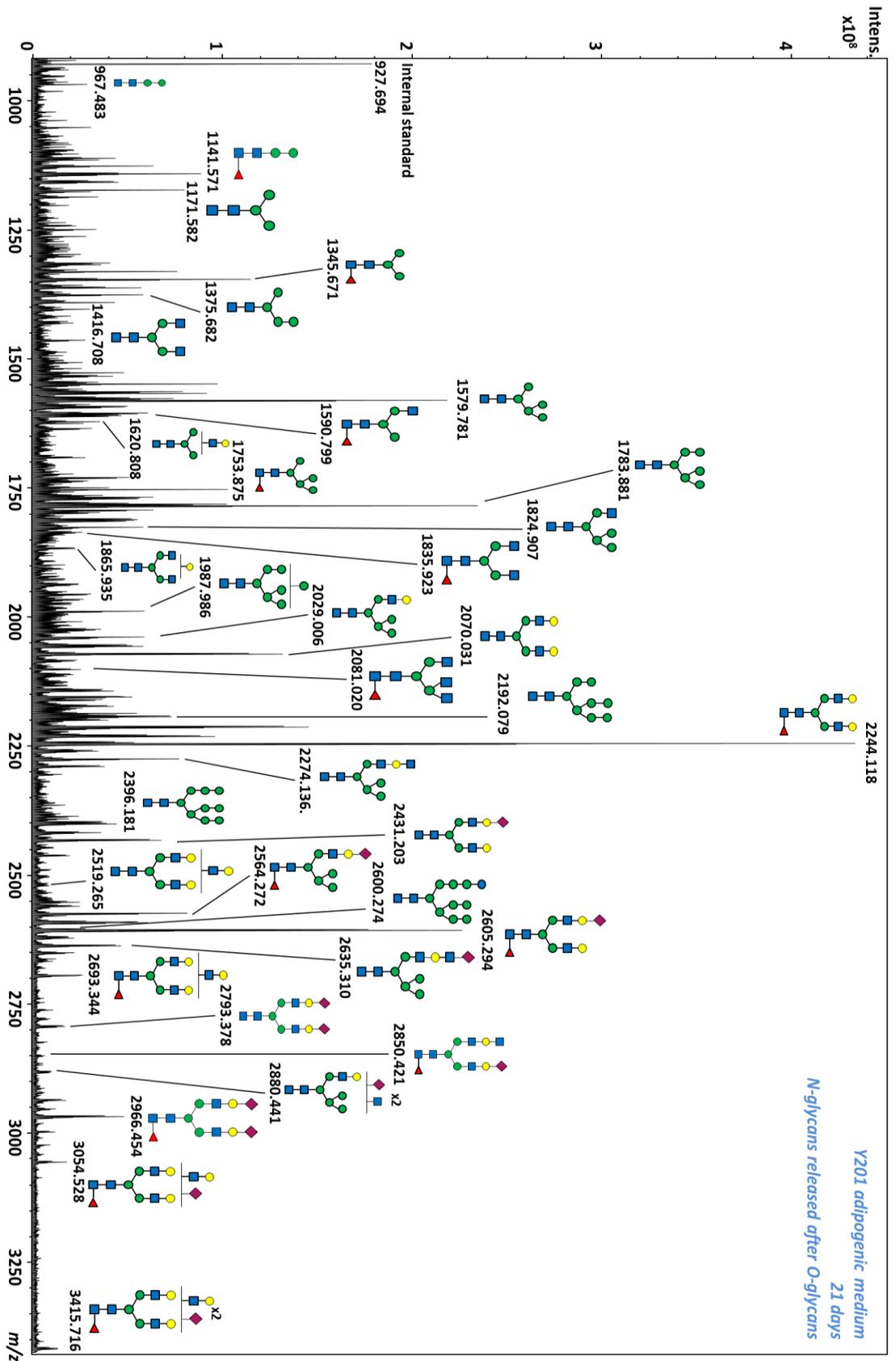


Figure 96: MALDI-FT-ICR mass spectrum of permethylated *N*-glycans (released after O-glycan release) from Y201 MSCs, following adipogenesis. All *N*-glycans ionised as $[M + Na]^+$.

Table 25: List of permethylated *N*-glycans identified in MSCs, along with their mass accuracies, after being cultured in either basal medium or adipogenic medium for 21 days.

glycan composition	calculated Mass	Av. measured <i>m/z</i> (n = 3) control	mass accuracy control (basal) (ppm)	adipogenesis Av. measured <i>m/z</i> (n = 3)	mass accuracy adipogenesis (ppm)
Hex ₂ HexNAc ₂	967.4833	967.4825	-0.78	967.4829	-0.38
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	1141.5713	-1.06	1141.572	-0.83
Hex ₃ HexNAc ₂	1171.5831	1171.5815	-1.34	1171.582	-0.97
Hex ₃ HexNAc ₂ Fuc ₁	1345.6723	1345.6714	-0.65	1345.672	-0.20
Hex ₄ HexNAc ₂	1375.6828	1375.6816	-0.86	1375.678	-3.49
Hex ₃ HexNAc ₃	1416.7094	1416.6716	-0.85	1416.709	-0.31
Hex ₅ HexNAc ₂	1579.7826	1579.7812	-0.87	1579.782	-0.50
Hex ₃ HexNAc ₃ Fuc ₁	1590.7986	1590.7622	-0.63	1590.799	0.09
Hex ₄ HexNAc ₃	1620.8092	1620.8076	-1.00	1620.808	-0.83
Hex ₅ HexNAc ₂ Fuc ₁	1753.8718	1753.8496	2.12	1753.859	-7.46
Hex ₆ HexNAc ₂	1783.8824	1783.8813	-0.63	1783.882	-0.41
Hex ₅ HexNAc ₃	1824.9089	1824.9075	-0.77	1824.908	-0.54
Hex ₃ HexNAc ₄ Fuc ₁	1835.9249	1835.9200	-2.67	1835.923	-0.92
Hex ₄ HexNAc ₄	1865.9355	1865.9311	-2.34	1865.932	-1.76
Hex ₇ HexNAc ₂	1987.9822	1987.9824	0.09	1987.983	0.43
Hex ₆ HexNAc ₃	2029.0087	2029.0089	0.12	2029.007	-0.95
Hex ₅ HexNAc ₄	2070.0353	2070.0294	-2.85	2070.03	-2.44
Hex ₃ HexNAc ₅ Fuc ₁	2081.0512	2081.020	-15.03	-	-
Hex ₈ HexNAc ₂	2192.0819	2192.0745	-3.39	2192.077	-2.35
Hex ₅ HexNAc ₄ Fuc ₁	2244.1245	2244.1167	-3.48	2244.118	-2.98
Hex ₆ HexNAc ₄	2274.1350	2274.1359	0.41	2274.136	0.57
Hex ₉ HexNAc ₂	2396.1817	2396.1820	0.14	2396.182	0.02
Hex ₅ HexNAc ₄ NeuAc ₁	2431.2089	2431.2009	-3.28	2431.203	-2.51
Hex ₆ HexNAc ₅	2519.2614	2519.2630	0.63	2519.262	0.23
Hex ₆ HexNAc ₃ Fuc ₁ NeuAc ₁	2564.2716	2564.2446	-10.55	2564.256	-5.96
Hex ₁₀ HexNAc ₂	2600.2815	2600.2568	-9.51	2600.279	-0.94
Hex ₅ HexNAc ₄ Fuc ₁ NeuAc ₁	2605.2981	2605.2929	-2.00	2605.293	-1.80
Hex ₆ HexNAc ₄ NeuAc ₁	2635.3087	2635.3160	2.78	2635.314	2.04

Hex ₆ HexNAc ₅ Fuc ₁	2693.3506	2693.3372	-4.98	2693.34	-4.11
Hex ₅ HexNAc ₄ NeuAc ₂	2792.3826	2792.3743	-2.96	2792.377	-2.05
Hex ₅ HexNAc ₅ Fuc ₁ NeuAc ₁	2850.4245	2850.416	-2.81	2850.421	-1.25
Hex ₆ HexNAc ₅ NeuAc ₁	2880.4350	2880.4452	3.56	2880.443	2.62
Hex _{5H} exNAc ₄ Fuc ₁ NeuAc ₁	2966.4718	2966.4556	-5.47	2966.457	-4.87
Hex ₆ HexNAc ₅ Fuc ₁ NeuAc ₁	3054.5242	3054.5284	1.37	3054.528	1.22
Hex ₆ HexNAc ₅ Fuc ₁ NeuAc ₂	3415.6979	3415.7182	5.93	3415.716	5.33

Next, MSCs were pre-treated with BadGal for 48 hours and then cultured in basal medium or adipogenic medium for 21 days, to test whether BadGal pre-treatment had an effect on the ability of MSCs to differentiate into adipocytes and whether there was an effect on the glycan profile.

On day 21, samples were stained with Oilred O to assess the extent of formation of lipid droplets and therefore differentiation into adipocytes. Cells pre-treated with BadGal and then cultured in basal medium for 21 days showed no staining, as no lipid droplets were formed. Both Y201 and Y101 cells pre-treated with BadGal and then cultured in adipogenic medium for 21 days showed staining, but staining was much less prominent for the Y101 cell line than the Y201 cell line, Figure 98.

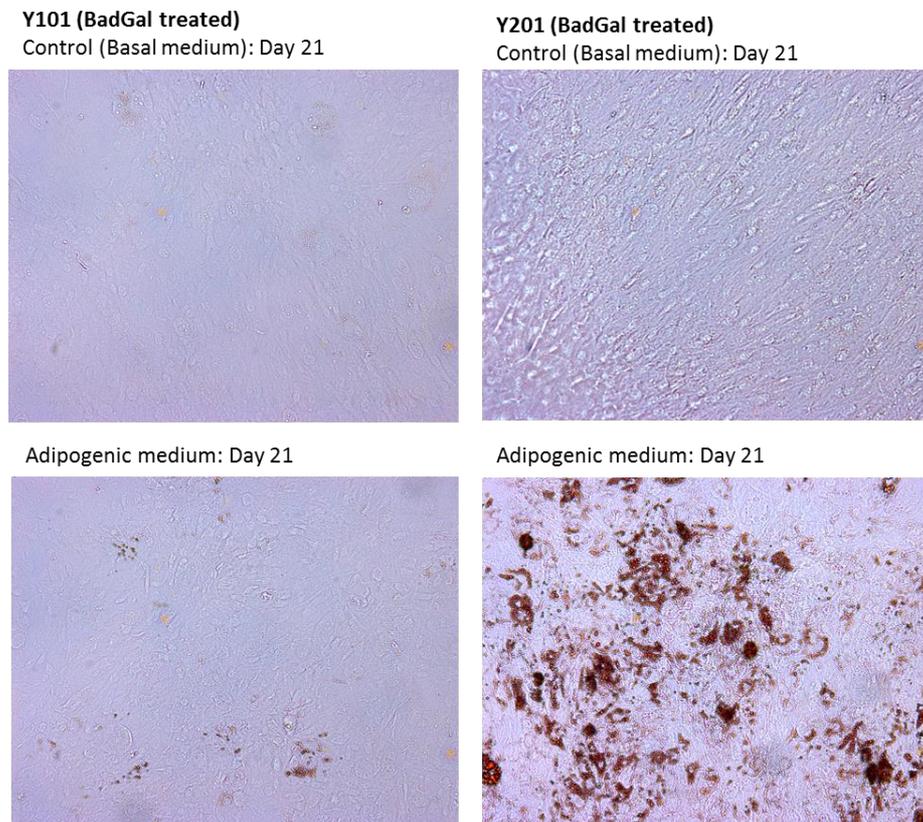


Figure 98: Oilred O staining of Y201 and Y101 MSC cell lines after pre-treatment with BadGal for 48 hours, followed by culture in basal or adipogenic medium for 21 days. Oilred O did not stain the cells in the basal medium. As cells differentiate into adipocytes lipid droplets form and the Oilred O stains the lipid droplets red. Y201 differentiate well and form many lipid droplets, Y101 although some staining is seen differentiation is limited.

After 21 days, cells were harvested and the O-glycans were released and permethylated before analysis by mass spectrometry. For Y201 MSCs pre-treated with BadGal and then cultured in basal medium for 21 days, O-glycans were identified by their characteristic signals at m/z 518, 722, 733, 879, 1141, 1240 and 1689. For the adipocytes derived from the Y201 MSCs, O-glycans were identified at m/z 518, 722, 879, 967, 1141, 1240 and 1328, Table 26 and Figure 99). For Y101 MSCs pre-treated with BadGal and then cultured in basal medium for 21 days, O-glycans were identified by their characteristic signals at m/z 518, 722, 733, 879, 967, 1240, 1328 and 1689. For the adipocytes derived from the Y101 MSCs, O-glycans were identified at m/z 518, 722, 879, 967, 1141, 1171, 1240, 1328 and 1689, Table 27 and Figure 100.

For both the BadGal pre-treated Y201 and Y101 cell lines there was no change in the relative abundance of Hex₁HexNAC₁, HexNAC₂Fuc₁, Hex₂HexNAC₂ and Hex₂HexNAC₁NeuAc₁ between basal medium and adipogenic cultures. Hex₂HexNAC₂Fuc₁ was identified in Y201 cells after growth in BadGal-treated basal medium, but was not detected from BadGal-treated adipogenic medium grown Y201 cells, nor was it detected from BadGal treated basal or adipogenic medium-grown Y101 cells, Figure 101.

This BadGal treatment study does not really add to our understanding of the functional significance of O-glycans in adipogenesis, since it does not appear to have caused much change in the glycan profiles. This could be because the cells may have recovered over the 21 days, having only been pre-treated with BadGal for 48 hours. The next step could thus be to treat with BadGal continually over the 21 days, to test whether adipogenesis and O-glycan biosynthesis are affected.

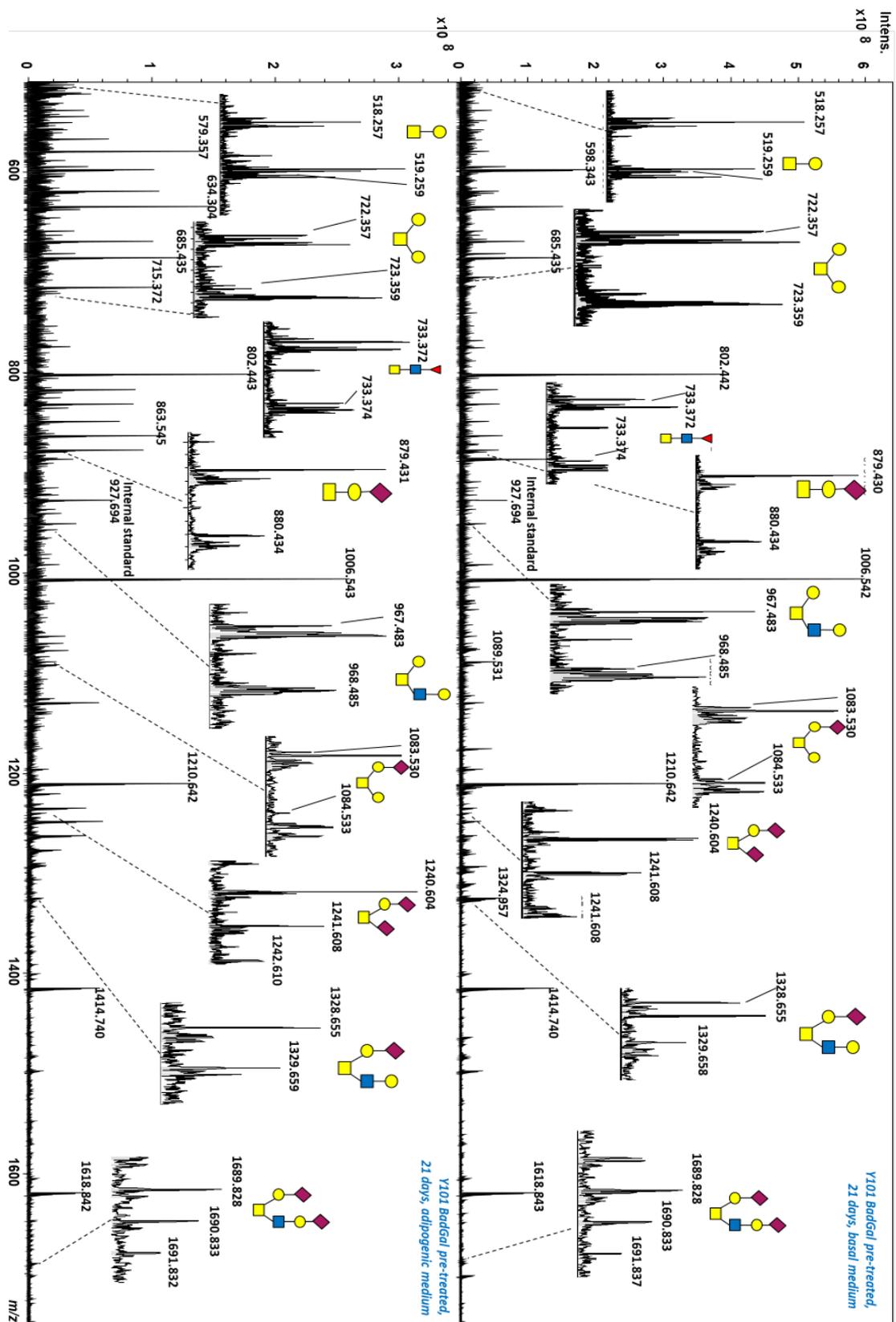


Figure 100: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from Y101 MSCs after 48 h pre-treatment with BadGal, followed by 21 days in basal or adipogenic medium. All glycans ionised as $[M + Na]^+$.

Table 26: A list of permethylated O-glycans identified from Y201 MSCs after a 48 h pre-treatment with O-glycan biosynthesis inhibitor, BadGal, followed by 21 days in either basal or adipogenic medium

glycan composition	calculated mass	Y201 BadGal treated (basal) Av. measured <i>m/z</i> (n = 8)	Y201 BadGal treated control mass accuracy (ppm)	Y201 BadGal treated adipogenesis Av. measured <i>m/z</i> (n = 8)	Y201 BadGal treated adipogenesis mass accuracy (ppm)
Hex ₁ HexNAc ₁	518.2572	518.2571	-0.25	518.2570	-0.29
Hex ₂ HexNAc ₁	722.3569	722.3568	-0.17	722.3568	-0.17
HexNAc ₂ Fuc ₁	879.4304	733.3723	-0.86		
Hex ₁ HexNAc ₁ NeuAc ₁	967.4833	879.4308	0.45	879.4308	0.44
Hex ₂ HexNAc ₂	1083.5306	-	-	967.4833	0.01
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	1141.5723	-0.18	1141.5732	0.58
Hex ₃ HexNAc ₂	1171.5831	-	-	-	-
Hex ₁ HexNAc ₁ NeuAc ₂	1240.6045	1240.6044	-0.07	1240.6044	-0.08
Hex ₂ HexNAc ₂ NeuAc ₁	1328.6569	-	-	1328.6575	0.43
NeuAc ₂ Hex ₂ HexNAc ₂	1689.8306	1689.8290	-0.95	-	-

Table 27: A list of permethylated O-glycans identified in Y101 MSCs after a 48 h pre-treatment with O-glycan biosynthesis inhibitor, BadGal, followed by 21 days in either basal or adipogenic medium

glycan composition	calculated mass	Y101 BadGal treated (basal) Av. measured <i>m/z</i> (n = 8)	Y101 BadGal treated control mass accuracy (ppm)	Y101 BadGal treated adipogenesis Av. measured <i>m/z</i> (n = 8)	Y101 BadGal treated adipogenesis mass accuracy (ppm)
Hex ₁ HexNAc ₁	518.2572	518.2571	-0.27	518.257	-0.22
Hex ₂ HexNAc ₁	722.3569	722.3566	-0.40	722.357	-0.29
HexNAc ₂ Fuc ₁	879.4304	733.3718	-1.54		
Hex ₁ HexNAc ₁ NeuAc ₁	967.4833	879.4301	-0.34	879.430	-0.16
Hex ₂ HexNAc ₂	1083.5306	967.4832	-0.09	967.483	-0.24
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	-	-	1141.571	-1.03
Hex ₃ HexNAc ₂	1171.5831	-	-	1171.581	-1.77
Hex ₁ HexNAc ₁ NeuAc ₂	1240.6045	1240.603	-1.35	1240.6028	-1.31
Hex ₂ HexNAc ₂ NeuAc ₁	1328.6569	1328.654	-1.88	1328.6544	-1.74
NeuAc ₂ Hex ₂ HexNAc ₂	1689.8306	1689.829	-1.10	1689.8287	-1.65

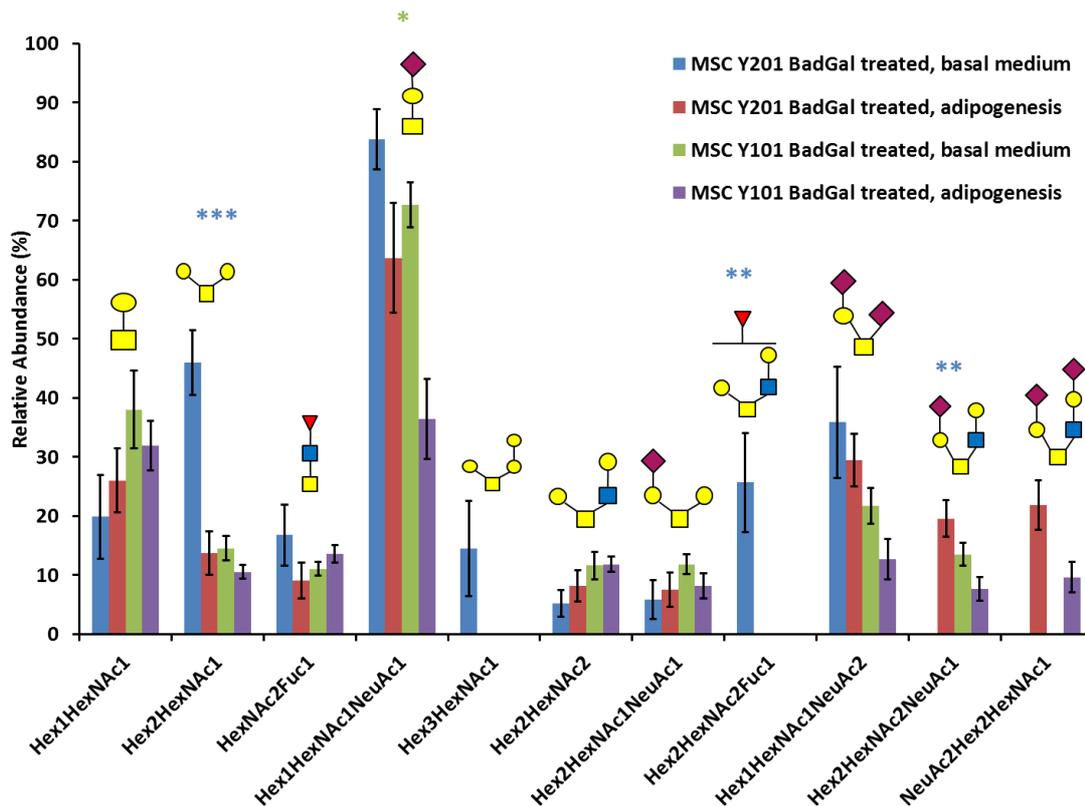


Figure 101: Comparison of the O-glycan abundance in Y201 and Y101 MSC cell lines pre-treated with BadGal for 48 h, followed by culturing in basal or adipogenic medium for 21 days. Relative abundance calculated by comparing peak intensity of spiked internal standard to each of the glycan peaks. P = * < 0.05, ** < 0.01, * < 0.001.**

6.4. Conclusions and future work

The one-sample one-pot method developed for the release and isolation of *N*- and *O*-glycans has been successfully applied to proliferating and differentiating MSCs. Mass spectrometry was used to produce quantifiable *O*-glycan profiles from two immortalised cell lines, Y201 and Y101. *O*-glycans were analysed and profiles compared from wild-type cells, those where the *O*-glycan biosynthetic pathway had been chemically disrupted using BadGal, or genetically disrupted in the Cog4KD line, and those that had been differentiated into adipocytes. Data have given an insight into which glycans are affected and what changes to the glycan profiles occur in MSCs when they are manipulated chemically or genetically. *O*-glycans are truncated on treatment with BadGal, with loss of sialylated and fucosylated glycans. In the Cog4KD samples, there was a reduction in the abundance of sialylated *O*-glycans, which was also observed with the *N*-glycans. On adipogenesis, sialylation and

fucosylation increase. These data could potentially suggest that O-glycans play a functional role of O-glycans in adipogenesis. However, even though K. Wilson et al²⁸⁷ reported changes to *N*-glycans during osteogenesis the changes were found not to be necessary for osteogenesis to occur.

Further experiments are needed to investigate whether O-glycans are important for MSC differentiation into adipocytes, or whether these changes to the O-glycan profiles are independent of adipogenesis. For example, experiments could be carried out to see if Hex₂HexNAc₂Fuc₁ is a consequence of or is required for adipogenesis. The addition of the fucose will involve a specific fucosyltransferase enzyme that transfers fucose to either the galactose or *N*-acetylglucosamine. Experiments could be carried out to test whether this biosynthetic enzyme is upregulated during adipogenesis. The gene could be knocked out and the mutant cells cultured in adipogenic medium to determine the effect of the gene knock-out on adipogenesis and the glycan profile of the cells.

In addition, rather than pre-treating the MSCs with BadGal for 48 h and then culturing in adipogenic medium, cells could instead be continuously treated with BadGal during culture in adipogenic medium, to determine the effect on glycan profiles; by only carrying out pre-treatment, the cells may have recovered over the 21 days during which they were subsequently cultured in the absence of BadGal, which could explain why very little effect was observed both on differentiation and on the glycan profile.

Since the Cog4KD MSCs showed reduction in the abundance of sialylated glycans, a Cog4 complete knock out could be analysed and compared with the wild-type and Cog4KD lines; if compromising Cog4 reduces sialylation, removing it completely would be expected to exacerbate the effect.

Chapter 7: Demonstration of the one-pot glycomics method applied to normal human urothelial cells.

7. Introduction

The urothelium is a specialised epithelial tissue, which covers the apical (luminal) surface of the urinary bladder²⁹⁴. The urothelium is arranged in layers of cells, at different states of differentiation²⁹⁵; with basal cells, intermediate cells and umbrella/superficial cells at the urothelium surface^{296,297,298} (Figure 102), which are highly specialised to perform their primary function as a protective barrier. The umbrella cells are connected via intercellular tight junctions²⁹⁹ that provide a selectively permeable barrier to prevent diffusion of urea and other waste products in the urine back into the surrounding tissue. This is important as the luminal surface is exposed to urine regularly. If the barrier were to fail, this would expose the underlying tissues to those waste products resulting in harmful effects such as urinary tract infections³⁰⁰, interstitial cystitis^{294,301} or cancer³⁰².

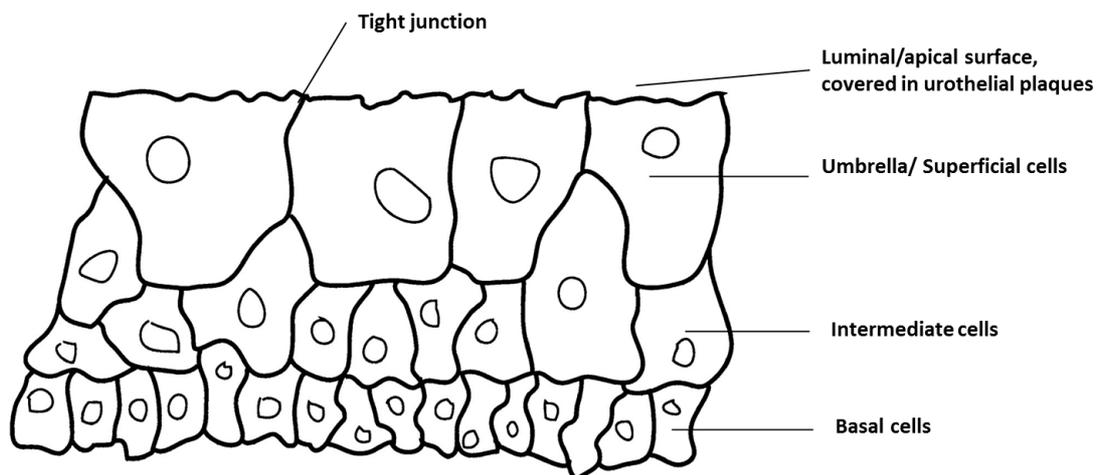


Figure 102: A schematic of urinary bladder urothelium. Cells are arranged in distinct layers; basal cells, intermediate cells and the umbrella/superficial cells which are fully differentiated to perform their function as a protective barrier; the umbrella cells are connected via intercellular tight junctions. The apical surfaces of the superficial cells are covered in urothelial plaques. Adapted from Lewis²⁹⁴.

The apical surface of the urothelium is almost entirely covered in urothelial plaques^{303,304}. Urothelial plaques are formed from integral membrane glycoproteins, uroplakins, which are arranged hexagonally^{305,306,307} in tetramers consisting of four major uroplakins^{308,309,310,311,312} UPIa, UPIb, UPII and UPIIIa. The mature form of UPII is not glycosylated but UPIa, UPIb and

UPIII are all known to contain *N*-glycans. The uroplakins are connected across cells and therefore also contributing to barrier function³¹³.

The apical surface of the urothelium is thus covered in glycoproteins and glycosaminoglycans, the presence of glycans reported as performing functions in intercellular signalling, cell adhesion, selective permeability of molecules across the blood-urine barrier³¹⁴ as well as inhibiting or promoting bacterial adhesion to the urothelium³¹⁵. In addition, the glycans protect cell membranes from proteolytic enzyme digestion by hydrolases present in urine³¹⁶.

Changes to glycans and glycosaminoglycans on the urothelium surface³⁰¹ have been reported to correlate with conditions of the urothelium such as urinary tract infections and bladder cancer³⁰². NHU cells which are known to express MUC1 and MUC4 show decreased expression of MUC4³¹⁷ and increased expression of MUC1³¹⁷ and MUC2³¹⁸ in bladder cancer and with it, aberrant glycosylation³¹⁷. Analysis of the whole glycome of urothelial tissue could have an important role in understanding more about glycosylation and glycan function in proliferating and differentiated cells that form the urothelial barrier. Differences in glycosylation between healthy cells and those in patients with urinary tract infections and bladder cancer could be investigated, which could help to work towards achieving earlier diagnosis and development of therapies and treatments for such infections and disease.

In the literature there appears to be very little reported on the study of the glycome of bladder urothelial tissue by mass spectrometry or any other direct structural approach, and instead research in this area is usually by lectin staining and histology, with many of these studies analysing glycoproteins from urine³¹⁹ and not from bladder tissue. One study that does use mass spectrometry isolated uroplakins UP1a and UPIb from murine and bovine bladders as well as human urothelial cells. *N*-glycans were released and analysed by mass spectrometry³²⁰. Differences in the class of glycans were found between the three sets of samples. For the human urothelial cells UP1a was found to contain oligomannose *N*-glycans only, whereas UP1b only had complex-type *N*-glycans.

Uropathogenic *E. coli* are responsible for almost all urinary tract infections^{321,322}. The lectin-type 1-fimbriae (fimH) of *E. coli* have been shown to bind specifically to uroplakin UPIa which contains oligomannose *N*-glycans^{323,320 320}. All binding was removed on pre-treating

the *E. coli* with mannose or treating the uroplakins with endoglycosidase-H, which cleaves the chitobiose core (GlcNAc₂) of oligomannose glycans; there was no binding observed to UPII or UPIII which are known to contain complex-type glycans. This investigation offers evidence that the urothelial surface contains mannosylated glycans and that they are involved in uropathogenic *E. coli* adhering to the urothelium.

Normal human urothelial cells can be grown in culture as monolayers, however once these cells are in culture in an artificial environment, they quickly lose the ability to differentiate and instead take on a highly proliferative, nonspecialised phenotype^{324,325}. In this form, they cannot offer much insight into urothelial tissue biology and the effect of glycosylation. Organ culture can be useful, as intact pieces of tissue can be studied *in vitro* as they can retain their *in situ* characteristics for longer periods than cultured cells³²⁶. Organ culture is of course limited by supply of human tissue, although animal tissue is more readily available.

Biomimetic *in vitro* epithelial models have been developed to allow cell culture in a less artificial environment, where cells are cultured on permeable membrane supports to allow supply of nutrients to the basal layer of cells³²⁷, replicating delivery *in situ*, and also allowing formation of cell layers, polarity and ion channels. These models allow investigation of cell differentiation, as well as investigating effects of xenobiotic drugs by adding drugs to the basal compartment (to mimic exposure to the urothelium via serum) or to the apical chamber (to mimic exposure to the luminal surface via the presence of drug in the urine).

The differentiation of NHU cells on biomimetic membranes to form a urothelial barrier can be assessed by measuring trans-epithelial electrical resistance (TER), reported by Cross et al³²⁸. TER is the measurement of electrical resistance across a cellular barrier, measured in ohms. TER provides a quantitative measure for demonstrating functional tight junctions and impermeability of the urothelial cell barrier³²⁹. The technique is non-invasive and can be used to monitor live cells during proliferation and differentiation. To make this determination, NHU cells are grown in medium containing bovine serum to full confluency. Full confluency is maintained for several days before transferring the cells on to permeable membranes. The calcium concentration above the membrane is raised to near physiological concentration (2 mM), creating a tight epithelial barrier replicating that *in situ*, and typically exhibiting a high TER reading, > 2000 Ω cm², (where a value > 500 Ω cm² is said to be a tight junction³³⁰), and low permeability to urea³³¹.

To carry out the electrical measurements, one electrode is placed in the upper (apical) chamber of the membrane well, and the other in the lower basolateral chamber of the membrane well³³², Figure 103. The resistance of the semi-permeable membrane is derived from Ohm's law. A voltage is placed across the membrane, without cells, and the current is measured and the value obtained is subtracted from the value obtained on measuring the resistance across the cell layer on the semipermeable membrane, to give the cell specific resistance³³², R_{tissue} Equation 13:

$$R_{tissue} = R_{total} - R_{blank}$$

Equation 13

Where resistance is inversely proportional to the effective area of the membrane, Equation 14:

$$R_{tissue} \propto \frac{1}{M_{area}}$$

Equation 14

TER values are reported as $\Omega \text{ cm}^2$ and calculated as in Equation 15:

$$TER = R_{tissue} \times M_{area}$$

Equation 15

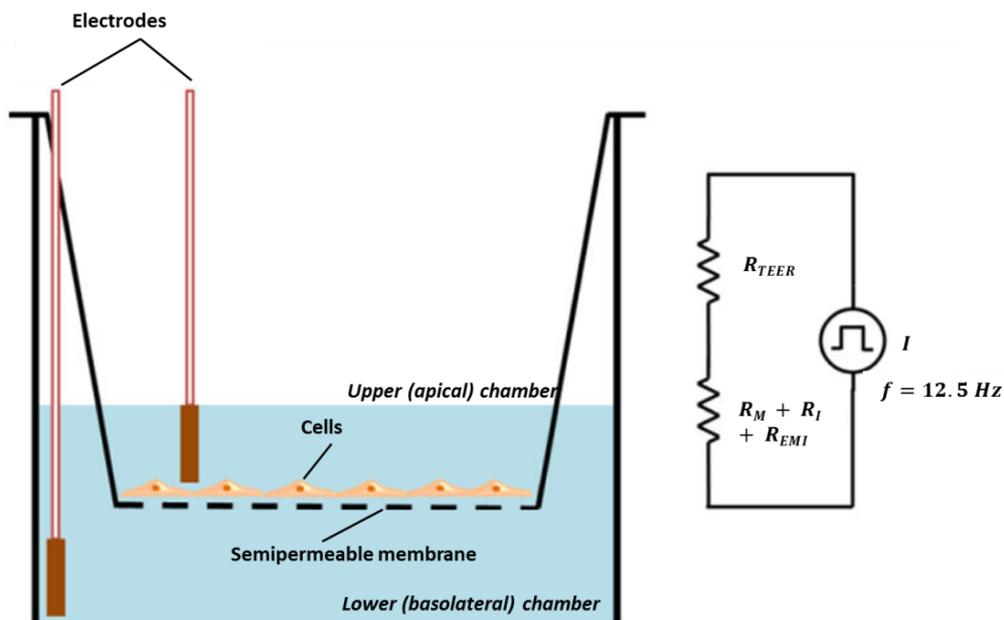


Figure 103: TER set-up. adapted from³³². To carry out the electrical measurements, one electrode is placed in the upper (apical) chamber of the membrane well, and the other in the lower basolateral chamber of the membrane well.

The aim of the work presented in this chapter was to apply the developed one-pot glycan method to analyse both *N*- and *O*-glycans from undifferentiated primary human urothelial cells and those differentiated on a semipermeable biomimetic membrane. This was to attempt to test the robustness and transferability of the approach by applying it to an independent biological system. Although *N*-glycans from NHU cells have been studied using a mass spectrometric approach, these analyses were of isolated specific uroplakins, whereas this study uses whole cell lysates to look at the whole glycome. In addition, there has to date been no analysis reported of the *O*-glycan profile of NHU cells, nor any studies of the potential role of *O*-glycan structures in differentiation of urothelial cells.

7.3. Data analysis

All mass spectrometric data collected in this chapter of work was acquired using MALDI-FT-ICR. All *O*-glycan samples were released from the protein by β -elimination, and *N*-glycans were released using PNGase F. All glycans were permethylated before analysis. Glycan assignments have been confidently assigned based on their mass accuracies, which, for the majority were all below 1 ppm, and also from the knowledge of possible *O*-glycan and *N*-glycan structures produced by the *O*-glycan and *N*-glycan biosynthetic pathway. All glycan samples were spiked with an internal standard, deuteropermethylated maltotetraose, to carry out relative quantification of the glycans to test whether there were relative quantitative differences in the glycan profiles of compared glycan sample sets. The relative abundance was calculated by comparing the peak intensities of each of the *O*-glycans with peak intensity of the internal standard, and the relative abundance was expressed as a percentage.

7.4. Application to normal human urothelial (NHU) cells

Single (unreplicated) cultures of NHU cells from three independent donor cell lines (Y1419, Y1860 and Y1862, representing three biological replicates) were kindly provided by Mrs Ros Duke from the Jack Birch Unit, Department of Biology. Non-differentiated cell cultures were grown on 6 cm diameter culture dishes and harvested at full confluence, Figure 104.

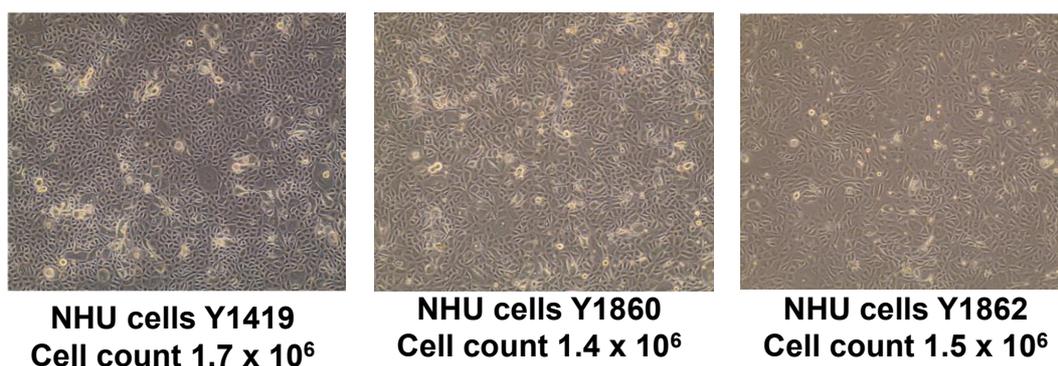


Figure 104: NHU cultures of the three cell lines Y1419, Y1860 and Y1862.

The developed one-pot method was applied. O-glycans were identified at m/z 518, 692, 722, 879, 896, 967, 1083, 1141, 1240, 1328, and 1689, for Y1419, at m/z 518, 692, 879, 896, 967, 1083, 1240, 1328, and 1689 for Y1860, for Y1860, and at m/z 518, 692, 722, 879, 896, 967, 1083, and 1141, for Y1862, Figure 105 and Table 28.

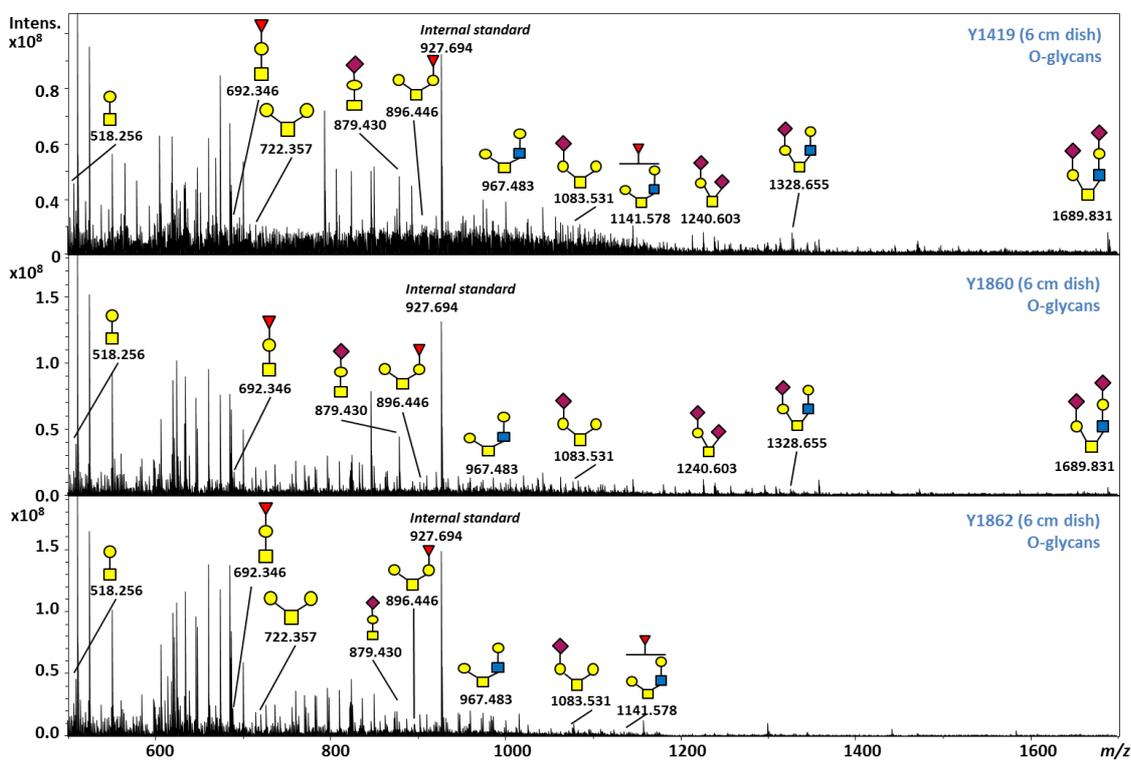


Figure 105: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from NHU cells (from three cell lines, Y1419, Y1860 and Y1862). All glycans ionised as $[M + Na]^+$

Table 28: List of permethylated O-glycans identified from Y1419, and Y1860 NHU cell lines, showing mass accuracies.

glycan composition	calculated <i>m/z</i>	measured <i>m/z</i>			Y1419 mass accuracy (ppm)		
		Y1419	Y1860	Y1862	Y1419	Y1860	Y1862
Hex ₁ HexNAC ₁	518.2572	518.2573	518.2569	518.2573	0.19	-0.50	0.21
Hex ₁ HexNAC ₁ Fuc ₁	692.3464	692.3462	-	692.3462	-0.23	-	-0.30
Hex ₂ HexNAC ₁	722.3569	722.3565	722.3565	722.3569	-0.51	-0.62	-0.03
Hex ₁ HexNAC ₁ NeuAc ₁	879.4309	879.4303	879.4305	879.4304	-0.13	0.06	0.01
Hex ₂ HexNAC ₁ Fuc ₁	896.4462	896.4459	896.4465	896.4464	-0.36	0.31	0.18
Hex ₂ HexNAC ₂	967.4833	967.4828	967.4823	967.4829	-0.52	-1.01	-0.43
Hex ₂ HexNAC ₁ NeuAc ₁	1083.5306	1083.5298	1083.5301	1083.5295	-0.73	-0.50	-0.98
Hex ₂ HexNAC ₂ Fuc ₁	1141.5725	1141.5718	-	1141.5709	-0.64	-	-1.40
Hex ₁ HexNAC ₁ NeuAc ₂	1240.6045	1240.6044	1240.6037	1240.6037	-0.06	-0.68	-0.68
Hex ₂ HexNAC ₁ Fuc ₁ NeuAc ₁	1257.6198	-	-	-	-	-	-
Hex ₂ HexNAC ₂ NeuAc ₁	1328.6569	1328.6554	1328.6582	-	-1.17	0.97	-
NeuAc ₂ Hex ₂ HexNAC ₂	1689.8306	1689.8293	1689.8276	-	-0.76	-1.76	-

On comparing the resulting glycan profiles for the three cell lines, the Y1419 cell culture appeared to produce a greater abundance of glycans than cultures from either the Y1860 or Y1862 cell lines – more different structures were identified and the glycans were present in higher relative abundance following normalisation to the internal standard. Although the cell numbers from the three different plates of cells were very slightly different, this observation prompted the protein levels from the three different dishes of cells to be determined. The culture of Y1419 had a slightly higher protein concentration than the other two cell lines; ~6 µg/µL for Y1419 compared to ~ 5 µg/µL for Y1860 and Y1862, Table 29. The glycan profiles have thus been normalised to protein concentration, Figure 106, although Y1419 still clearly yielded more glycans in this study. Of the three cell lines tested, there were female and male donors, and mixed ages (33, 73 and 50 years). Differences in glycan profile could perhaps derive from differences in age or gender of the lines' donors. Since

only two technical replicates of each of the three biological replicates were available, it is difficult to do more than speculate at the moment why one line appeared to provide much higher levels of glycans than the other two (the glycan levels of which were much more comparable).

Table 29: Protein concentration extracted from NHU cell line, measured by Bradford assay.

Cell line	Protein concentration ($\mu\text{g}/\mu\text{L}$)
Y1419 (1)	5.82
Y1419 (2)	6.69
Y1860 (1)	4.76
Y1860 (2)	5.27
Y1862 (1)	4.90
Y1862 (2)	5.32

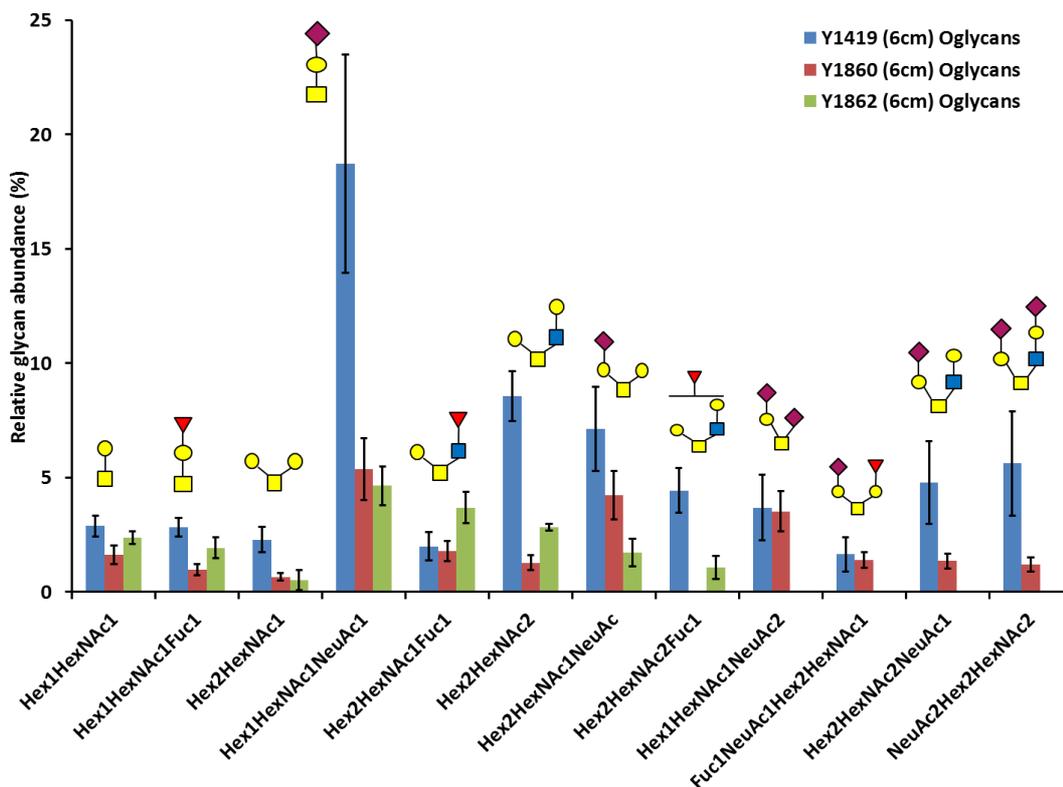


Figure 106: Comparison of the O-glycan profiles of three NHU cell lines. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. The glycan profiles have been normalised to protein concentration.

The NHU cell cultures from which the O-glycans had been released were subsequently treated with PNGase F to release the *N*-glycans. Upon MS analysis, *N*-glycans were identified m/z 967 – m/z 3054, Figure 107 and Table 30, from the NHU cell cultures showing the presence of both oligomannose and complex-type *N*-glycans. Structures are proposed on the basis of the known repertoire of *N*-linked glycans that can be biosynthesised.

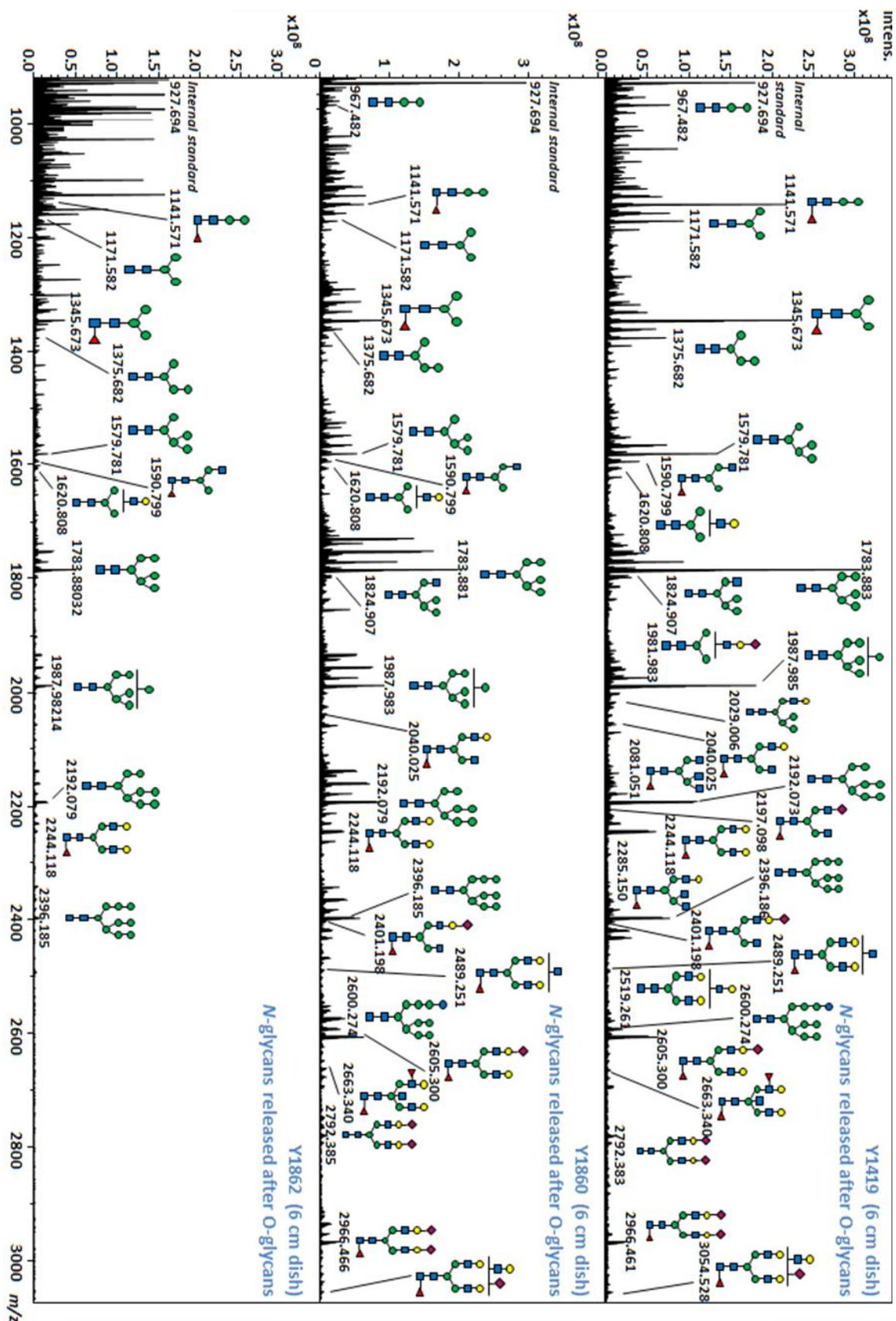


Figure 107: MALDI-FT-ICR mass spectra of permethylated N-glycans released after O-glycan release from NHU cells (from three cell lines, Y1419, Y1860 and Y1862). All glycans ionised as $[M + Na]^+$

Table 30: List of permethylated *N*-glycans identified from Y1419, and Y1860 NHU cell cultures, showing mass accuracies.

glycan composition	calculated <i>m/z</i>	measured <i>m/z</i>			mass accuracy (ppm)		
		Y1419	Y1860	Y1862	Y1419	Y1860	Y1862
Hex ₂ HexNAc ₂	967.4833	967.4830	967.4806	967.4904	-0.28	-2.84	7.31
Hex ₂ HexNAc ₂ Fuc ₁	1141.5725	1141.5711	1141.5718	1141.5713	-1.20	-0.62	-1.02
Hex ₃ HexNAc ₂	1171.5831	1171.5820	1171.5827	1171.5814	-0.92	-0.32	-1.46
Hex ₃ HexNAc ₂ Fuc ₁	1345.6723	1345.6726	1345.6727	1345.6723	0.22	0.33	0.03
Hex ₄ HexNAc ₂	1375.6828	1375.6832	1375.6838	1375.6835	0.28	0.73	0.47
Hex ₅ HexNAc ₂	1579.7826	1579.7811	1579.7811	1579.7810	-0.94	-0.92	-0.99
Hex ₃ HexNAc ₃ Fuc ₁	1590.7986	1590.7999	1590.7987	1590.7982	0.84	0.03	-0.26
Hex ₄ HexNAc ₃	1620.8092	1620.8074	1620.8062	1620.8067	-1.11	-1.84	-1.55
Hex ₆ HexNAc ₂	1783.8824	1783.8826	1783.8814	1783.8803	0.08	-0.56	-1.17
Hex ₅ HexNAc ₃	1824.9089	1824.9133	1824.9152	-	2.39	3.47	-
NeuAc ₁ Hex ₄ HexNAc ₃	1981.9828	1981.9843	1981.9821	-	0.77	-0.36	-
Hex ₇ HexNAc ₂	1987.9822	1987.9847	1987.9830	1987.9821	1.24	0.39	-0.03
Hex ₆ HexNAc ₃	2029.0087	2029.0157	-	-	3.46	-	-
Hex ₄ HexNAc ₄ Fuc ₁	2040.0247	2040.0264	2040.0227	-	0.85	-0.97	-
Hex ₃ HexNAc ₅ Fuc ₁	2081.0512	2081.0653	2081.0615	-	6.79	4.93	-
Hex ₈ HexNAc ₂	2192.0819	2192.0735	2192.0756	2192.0749	-3.84	-2.86	-3.19
NeuAc ₁ Hex ₃ HexNAc ₄ Fuc ₁	2197.0986	2197.0896	2197.0945	-	-4.09	-1.88	-
Hex ₄ HexNAc ₅ Fuc ₁	2285.1510	2285.1530	2285.1511	-	0.88	0.03	-
Hex ₅ HexNAc ₄ Fuc ₁	2244.1245	2244.1167	2244.1200	2244.1195	-3.47	-1.99	-2.22
Hex ₉ HexNAc ₂	2396.1817	2396.1838	2396.1833	2396.1835	0.89	0.65	0.74
NeuAc ₁ Hex ₄ HexNAc ₄ Fuc ₁	2401.1984	2401.1991	2401.1925	-	0.28	-2.48	-
Hex ₅ HexNAc ₅ Fuc ₁	2489.2508	2489.2540	2489.2539	-	1.30	1.25	-
Hex ₆ HexNAc ₅	2519.2614	-	-	-	-	-	-
NeuAc ₁ Hex ₇	2594.2822	2594.2885	2594.2866	-	2.41	1.70	-

HexNAc ₃							
Hex ₁₀ HexNAc ₂	2600.2815	2600.2793	2600.2787	-	-0.85	-1.07	-
NeuAc ₁ Hex ₅ HexNAc ₄ Fuc ₁	2605.2981	2605.2963	2605.2971	-	-0.68	-0.39	-
Hex ₅ HexNAc ₅ Fuc ₂	2663.3400	2663.3311	2663.3370	-	-3.34	-1.11	-
NeuAc ₂ Hex ₅ HexNAc ₄	2792.3826	2792.3784	2792.3782	-	-1.50	-1.58	-
NeuAc ₂ Hex ₅ H exNAc ₄ Fuc ₁	2966.4718	2966.4598	2966.4654	-	-4.03	-2.15	-
NeuAc ₁ Hex ₆ HexNAc ₅ Fuc ₁	3054.5242	3054.5365	3054.5384	-	4.04	4.65	-

On comparing the resulting glycan profiles for the three cell lines normalised to the internal standard, as with the O-glycan profiles, cell line Y1419 again appears to have a greater abundance of glycans than Y1860 and Y1862. The glycan profiles have been normalised to protein concentration. The most abundant *N*-glycans are the oligomannose type, with complex types containing terminal galactose, sialic acid and fucose substituents present at lower abundances. Despite the difference in the glycan abundance of the three cell lines, the glycan profiles across the three biological replicates are very similar, Figure 108.

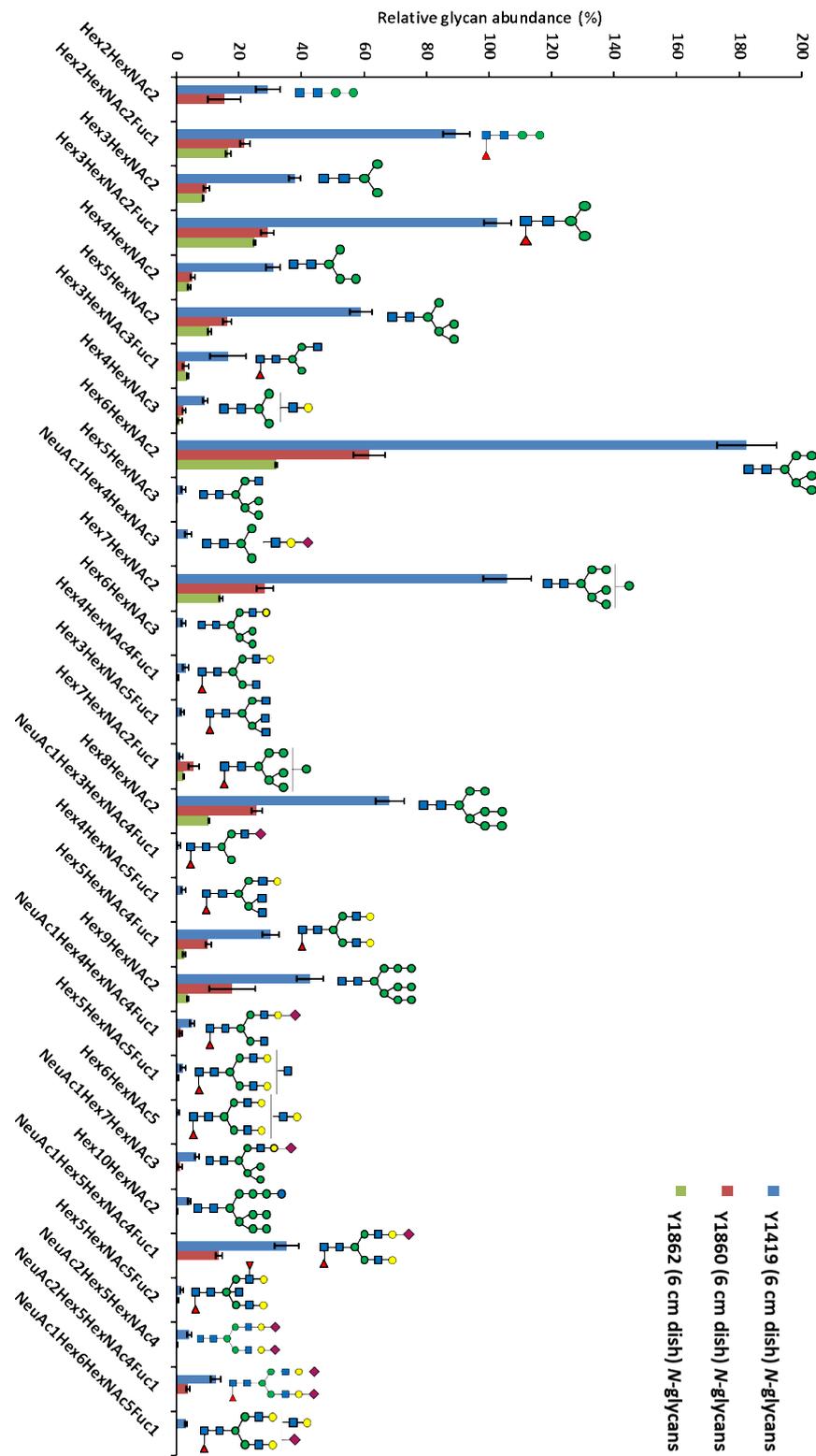


Figure 108: Comparison of the *N*-glycan profiles of cultures from Y1419, Y1860 and Y1862 NHU cell lines. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. The glycan profiles have been normalised to protein concentration.

Since the amount of internal standard spiked into the *N*-glycan samples and the O-glycan samples was the same, a direct comparison can be made between the relative abundance of *N*-glycans compared to O-glycans. Comparison of Figure 106 with Figure 108 shows that *N*-glycans are of much higher abundance than O-glycans. For example in the culture from the Y1419 cell line the three most abundant *N*-glycans are in excess of 100 % of the internal standard and in the O-glycan samples the three most abundant O-glycans are only 18- 20 % of the internal standard.

Since the one pot method was successful at isolating both *N*- and O-glycans from undifferentiated NHU cell cultures from three cell lines, NHU cell cultures were then established by Ros Duke from the same cell lines in differentiating conditions³²⁸ on semipermeable membranes, Figure 109. TER measurements for the cultures were provided by Ros Duke on days 0, 1, 2, 3, 6, 7 and 8 (for Y1860 and Y1862) and on 0, 1, 4, 5, 6, 7, and 8 days for Y1419 to monitor the formation of the tight urothelial barrier, Figure 110. The Y1862 cell line took the longest to form a tight barrier, showing a reading above 500 Ωcm^2 after 4 days, while Y1419 and Y1860 reached this on day 1, Figure 110. From day 7 there was little change in the TER readings for all lines, showing a stable barrier had formed and so the cells were then ready for harvesting.

Released and permethylated O-glycans were identified at m/z 518, 879, 896, 967, 1083, and 1240 for Y1419, at m/z 692, 879, 896, 967, 1083, 1141, 1240, 1257, and 1689 for Y1860, and at m/z 879 for Y1862, Figure 111 and Table 31.

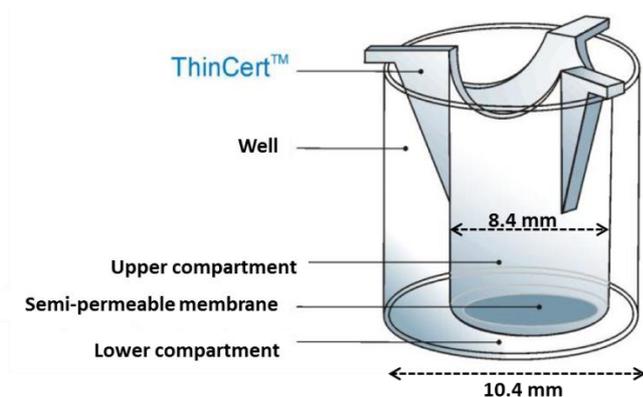


Figure 109: Schematic of the semi-permeable membrane well to which NHU cells were transferred to form a tight barrier, mimicking the barrier formed in the bladder

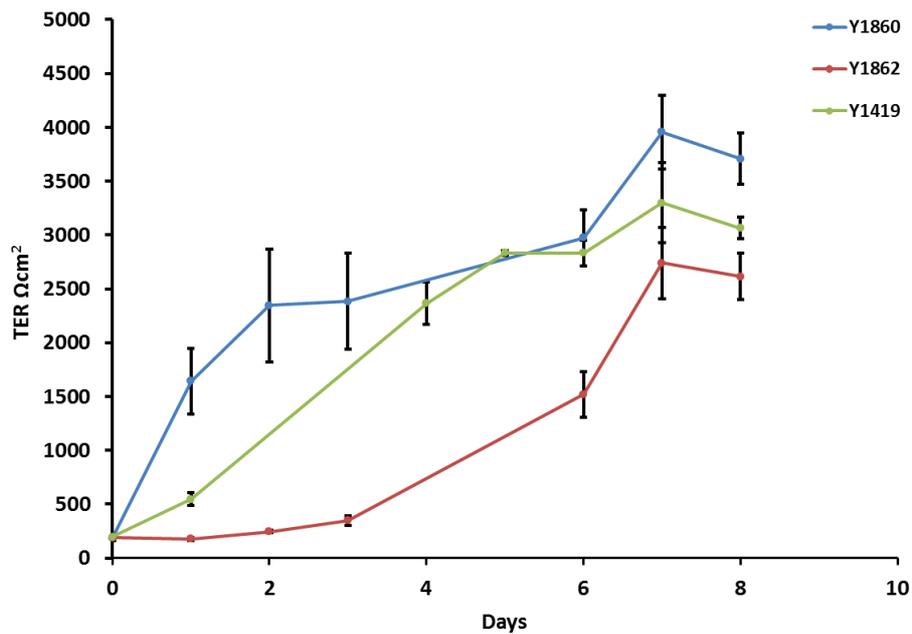


Figure 110: TER readings over 0 to 8 days for each of the NHU cell lines (Y1419, Y1860, Y1862) to show formation of a tight barrier. (data provided by Ros Duke)

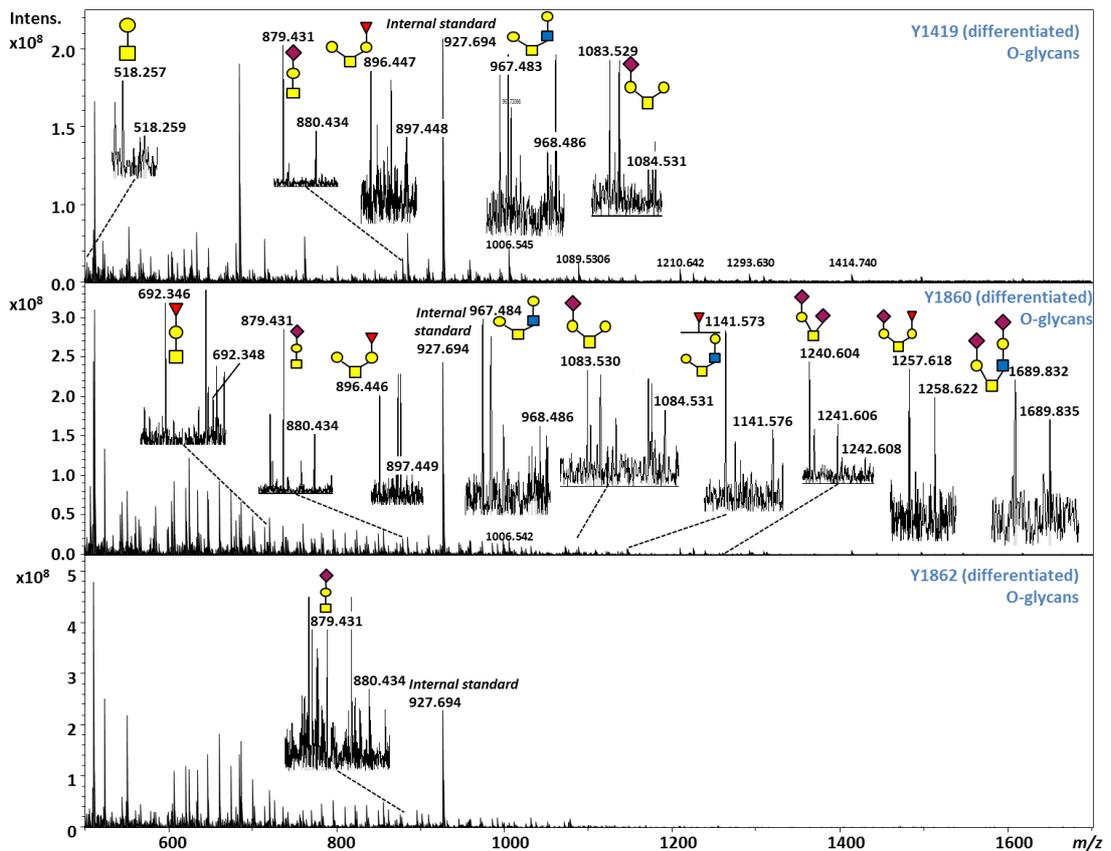


Figure 111: MALDI-FT-ICR mass spectrum of permethylated O-glycans released from NHU cells (from three cell lines, Y1419, Y1860 and Y1862). All glycans ionised as $[M + Na]^+$

Table 31: List of permethylated O-glycans identified from Y1419, and Y1860 NHU cell lines, showing mass accuracies.

glycan composition	calculated <i>m/z</i>	Y1419 measured <i>m/z</i>	Y1419 mass accuracy (ppm)	Y1860 measured <i>m/z</i>	Y1860 mass accuracy (ppm)
Hex ₁ HexNAC ₁	518.2572	518.2571	-0.14	-	-
Hex ₁ HexNAC ₁ Fuc ₁	692.3464	-	-	692.3463	-0.22
Hex ₁ HexNAC ₁ NeuAc ₁	879.4309	879.4306	0.24	879.4305	0.13
Hex ₂ HexNAC ₁ Fuc ₁	896.4462	896.4474	1.38	896.4456	-0.67
Hex ₂ HexNAC ₂	967.4833	967.4830	-0.35	967.4836	0.31
Hex ₂ HexNAC ₁ NeuAc ₁	1083.5306	1083.5294	-1.08	1083.5305	-0.08
Hex ₂ HexNAC ₂ Fuc ₁	1141.5725	-	-	1141.5731	0.50
Hex ₁ HexNAC ₁ NeuAc ₂	1240.6045	1240.6039	-0.48	1240.6040	-0.37
Hex ₂ HexNAC ₁ Fuc ₁ NeuAc ₁	1257.6198	1257.6202	-0.48	1257.6181	-1.33
NeuAc ₂ Hex ₂ HexNAC ₂	1689.8306	-	-	1689.8318	0.69

On comparing the resulting glycan profiles for each of the cell lines, the Y1419 cell line showed the highest abundance of glycans although the difference was not as marked as with the proliferating cell samples. On measuring the protein concentration by Bradford assay the concentration varied across the three cell lines, Table 32. Glycan profiles were normalised to protein concentration. As with the proliferating cells, NeuAc₁Hex₁HexNAC₁ is of higher abundance in the line Y1419 O-glycome than that from the other two cell lines, and in that from Y1862, only NeuAc₁Hex₁HexNAC₁ was identified, Figure 112.

Table 32: concentration of protein extracted from cultures from each of the differentiated NHU cell lines, measured by Bradford assay.

Cell line	Protein concentration ($\mu\text{g}/\mu\text{L}$)
Y1419 (1)	0.47
Y1419 (2)	0.30
Y1860 (1)	0.57
Y1860 (2)	0.45
Y1862 (1)	0.32
Y1862 (2)	0.22

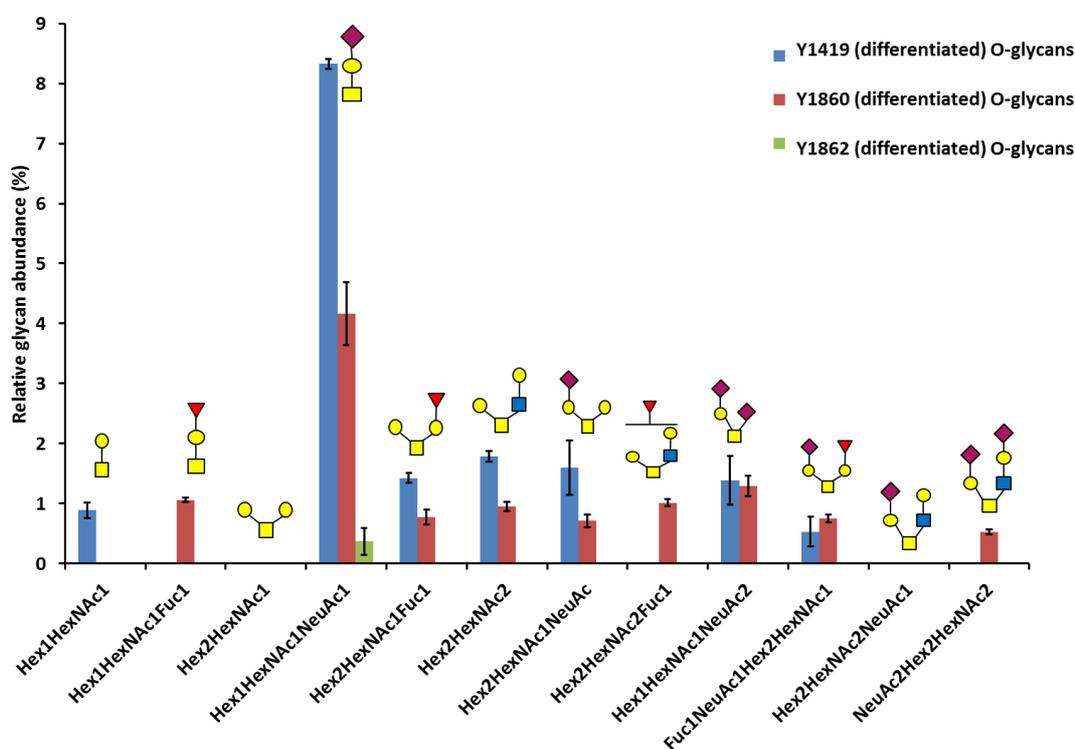


Figure 112: Comparison of the O-glycan profiles of the Y1419, Y1860 and Y1862 NHU cell cultures, differentiated on SNAP well membranes. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. The glycan profiles have been normalised to protein concentration.

For a closer inspection of how the glycan profiles differ between the undifferentiated and the differentiated cultures, the glycan profiles were normalised to protein concentration,

Figure 113. For the Y1419 cell line, there is large increase, with respect to the undifferentiated cells, in the abundance of NeuAc₁Hex₁HexNAC₁ in the differentiated sample, as well as an increase in abundance of Hex₁HexNAC₁, Fuc₁Hex₂HexNAC₁, Hex₂HexNAC₂, NeuAc₁Hex₂HexNAC₁, NeuAc₂Hex₁HexNAC₁ and NeuAc₁Fuc₁Hex₂HexNAC₁. For the Y1860 cell line, like Y1419, there is also an increase in Hex₁HexNAC₁Fuc₁, NeuAc₁Hex₁HexNAC₁, Fuc₁Hex₂Hex₁NAC₁, Hex₂HexNAC₂, NeuAc₂Hex₁HexNAC₁ and NeuAc₁Fuc₁Hex₂HexNAC₁, although there is no change in the abundance of NeuAc₁Hex₂HexNAC₁, and Hex₁HexNAC₁ was not detected in the differentiated Y1860 sample. There was also an increase in the abundance of Fuc₁Hex₂HexNAC₂ in the differentiated Y1860 sample compared to the undifferentiated Y1860 samples whereas this glycan was not detected in the Y1419 differentiated sample, Figure 113. To use these results for a more reliable quantitative comparison between glycan profiles of the undifferentiated and differentiated cells, the samples prepared would ideally be prepared from the same number of cells to begin with and a similar protein concentration as even though the data has been normalised to protein concentration, there was still a big difference in the abundance of glycans in comparing the undifferentiated cells to differentiated cells. In addition, since the membranes allow a much smaller number of cells to be used than the 6 cm dishes used for the proliferating cells, the inevitable losses due to sample handling will have a disproportionate impact on the differentiated samples' glycan yields. Consequently, while it is safe to comment on the increased abundance of certain structures in the differentiating cells with respect to their undifferentiated counterparts, it is probably not safe to comment on structures that appear to decrease in level, since cell numbers and protein are likely to be so different, e.g. in this study the protein concentration for undifferentiated cells from the Y1419 cell line was ~6 µg/µL and for the differentiated cells, ~0.38 µg/µL. Cell numbers in the biomimetic membranes cannot readily be compared without sacrificing precious differentiated donor cells since the cells form layers and so they would have to be disrupted to enable cell number estimation, unlike those proliferating in the 6 cm dishes which grow in monolayers. In addition, the differentiated donor cells will display heterogeneity as the cells on the surface of the layers maybe display different numbers/types of glycans to the cells below the surface, and so it may be more comparable to take only the cells from the apical surface of the differentiated cells to compare to the undifferentiated cells cultured on a monolayer.

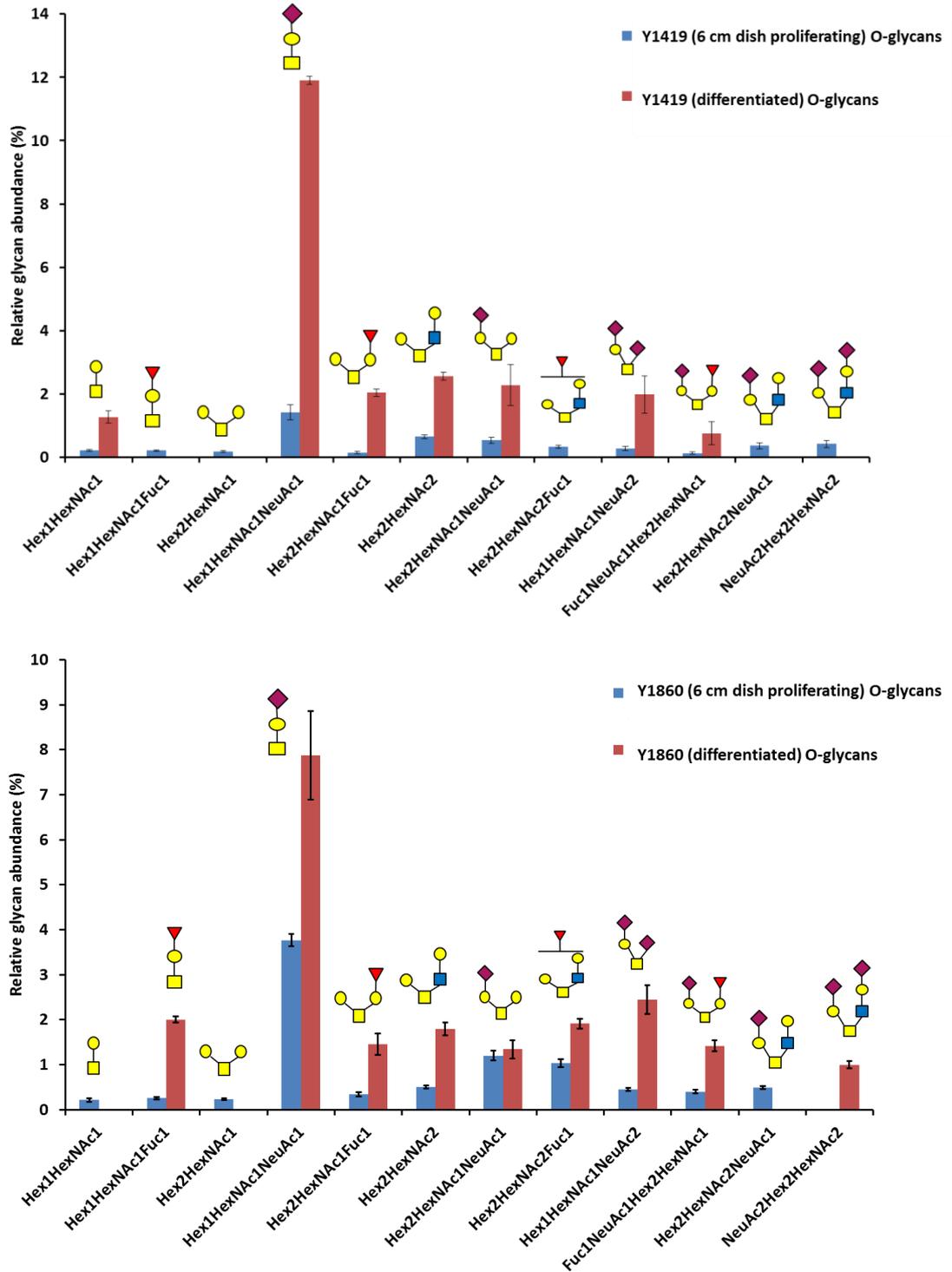


Figure 113: Comparison of the O-glycan profiles of the, proliferating cells on 6 cm dishes and differentiated cells on semi-permeable membrane wells for Y1419 NHU cell line (top graph) and Y1860 (bottom graph). Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. Glycan profiles are normalised to protein concentration.

The differentiated NHU cell samples that had already had their O-glycans released were subsequently treated with PNGase F to release the *N*-glycans. *N*-glycans were identified between m/z 967 – m/z 2396, from the NHU cell lines, showing the presence of both oligomannose and complex-type *N*-glycans, Figure 114, Table 33.

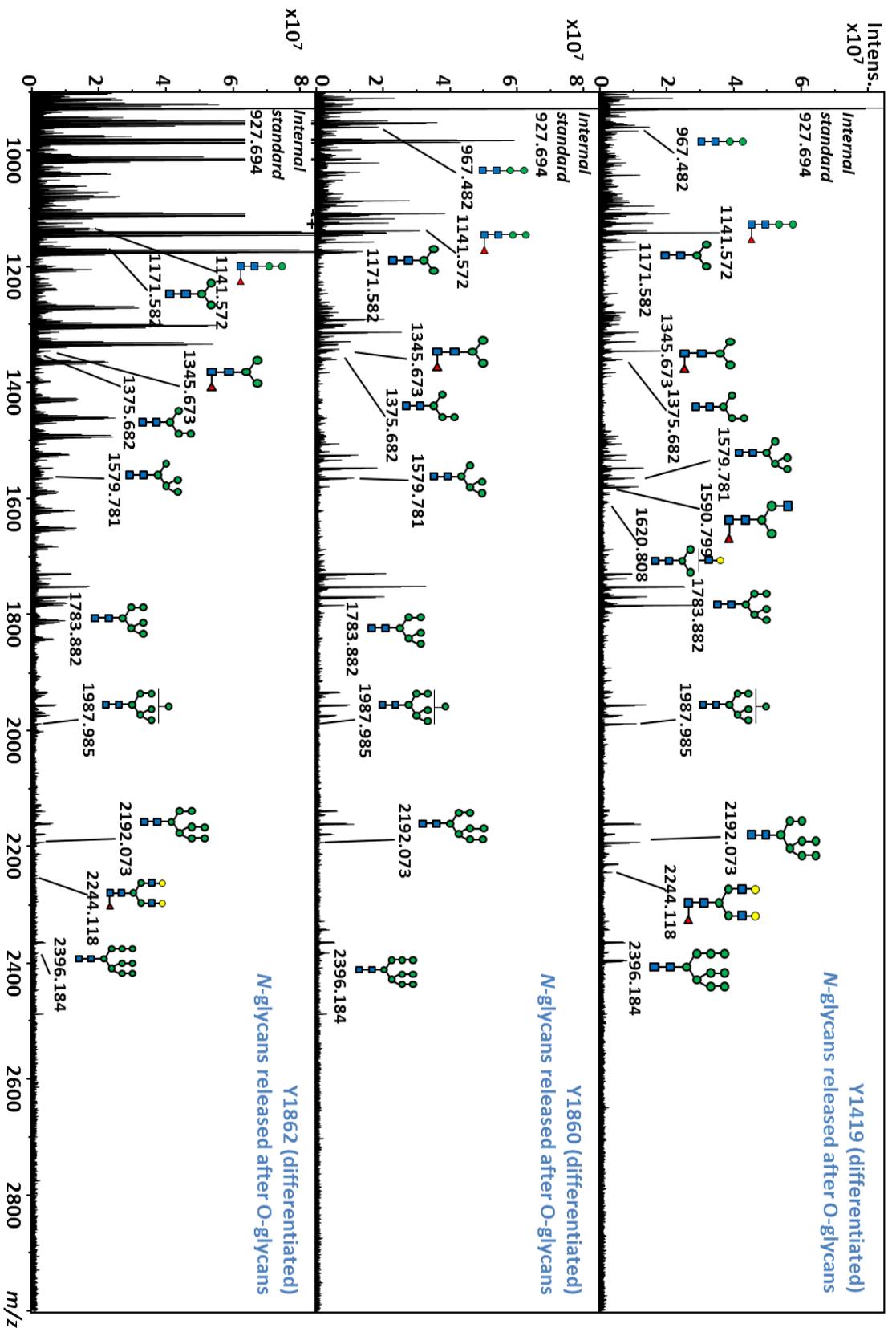


Figure 114: MALDI-FT-ICR mass spectrum of permethylated N-glycans released from NHU cells (from three cell lines, Y1419, Y1860 and Y1862) after O-glycans. All glycans ionised as $[M + Na]^+$

Table 33: List of permethylated *N*-glycans identified from differentiated Y1419 and Y1860 NHU cell lines, showing mass accuracies. All glycan compositions listed were identified in the proliferating cell lines.

glycan composition	calculated <i>m/z</i>	measured <i>m/z</i>			mass accuracy (ppm)		
		Y1419	Y1860	Y1862	Y1419	Y1860	Y1862
Hex ₂ HexNAC ₂	967.4833	967.4826	967.4807	967.4822	-0.78	-2.67	-1.12
Hex ₂ HexNAC ₂ Fuc ₁	1141.5725	1141.5722	1141.5732	1141.5724	-0.25	0.58	-0.09
Hex ₃ HexNAC ₂	1171.5831	1171.5820	1171.5829	1171.5839	-0.92	-0.19	0.70
Hex ₃ HexNAC ₂ Fuc ₁	1345.6723	1345.6726	1345.6731	1345.6734	0.23	0.60	0.82
Hex ₄ HexNAC ₂	1375.6828	1375.6848	-	-	1.45	-	-
Hex ₅ HexNAC ₂	1579.7826	1579.7818	1579.7806	-	-0.53	-1.25	-
Hex ₃ HexNAC ₃ Fuc ₁	1590.7986	1590.8004	-	-	1.14	-	-
Hex ₄ HexNAC ₃	1620.8092	1620.8006	-	-	-5.32	-	-
Hex ₆ HexNAC ₂	1783.8824	1783.8818	1783.8790	1783.8795	-0.33	-1.93	-1.61
Hex ₅ HexNAC ₃	1824.9089	-	-	-	-	-	-
NeuAc ₁ Hex ₄ HexNAC ₃	1981.9828	-	-	-	-	-	-
Hex ₇ HexNAC ₂	1987.9822	1987.9812	1987.9863	1987.9830	-0.53	2.08	0.39
Hex ₆ HexNAC ₃	2029.0087	-	-	-	-	-	-
Hex ₄ HexNAC ₄ Fuc ₁	2040.0247	-	-	-	-	-	-
Hex ₃ HexNAC ₅ Fuc ₁	2081.0512	-	-	-	-	-	-
Hex ₈ HexNAC ₂	2192.0819	2192.0817	2192.0802	2192.0783	-0.09	-0.79	-1.63
NeuAc ₁ Hex ₃ HexNAC ₄ Fuc ₁	2197.0986	-	-	-	-	-	-
Hex ₄ HexNAC ₅ Fuc ₁	2285.1510	-	-	-	-	-	-
Hex ₅ HexNAC ₄ Fuc ₁	2244.1245	2244.1297	-	-	2.31	-	-
Hex ₉ HexNAC ₂	2396.1817	2396.1882	2396.1799	2396.1798	2.73	-0.74	-0.78
NeuAc ₁ Hex ₄ HexNAC ₄ Fuc ₁	2401.1984	-	-	-	-	-	-
Hex ₅ HexNAC ₅ Fuc ₁	2489.2508	-	-	-	-	-	-
Hex ₆ HexNAC ₅	2519.2614	-	-	-	-	-	-

NeuAc ₁ Hex ₇ HexNAc ₃	2594.2822	-	-	-	-	-	-
Hex ₁₀ HexNAc ₂	2600.2815	-	-	-	-	-	-
NeuAc ₁ Hex ₅ HexNAc ₄ Fuc ₁	2605.2981	-	-	-	-	-	-
Hex ₅ HexNAc ₅ Fuc ₂	2663.3400	-	-	-	-	-	-
NeuAc ₂ Hex ₅ HexNAc ₄	2792.3826	-	-	-	-	-	-
NeuAc ₂ Hex ₅ HexNAc ₄ Fuc ₁	2966.4718	-	-	-	-	-	-
NeuAc ₁ Hex ₆ HexNAc ₅ Fuc ₁	3054.5242	-	-	-	-	-	-

On comparing the *N*-glycan profiles across the three cell lines, oligomannose-type glycans are the most abundant glycans detected. Similarly to the glycans from the undifferentiated cells, the Y1419 cell line gave a higher abundance of glycans than the other two cell lines. When comparing the *N*-glycomes of these differentiated cells with those of the undifferentiated cells, there are a lot of *N*-glycans not detected from the differentiated cells, Figure 115. This is probably an effect of using fewer cells and therefore having less protein; it would not be safe to conclude that the missing glycans are not expressed on the differentiated cells. It thus makes a direct comparison between glycan changes difficult when the number of cells used were not the same for the undifferentiated and differentiating cell sample sets.

Since the amount of internal standard spiked into differentiated O-glycan samples for the same as the differentiated *N*-glycan samples, a direct comparison could be made between the relative abundance of *N*-glycans compared to O-glycans by normalising to the internal standard spike. Comparison of **Figure 112** with **Figure 115** shows that *N*-glycans are of higher abundance than O-glycans although the difference in abundance between the two glycan types is not as exaggerated as in the proliferating samples. For example in the differentiated Y1419 cell line *N*-glycans are around 1-10 % of the internal standard and in the O-glycan samples O-glycans are around 0.5-8 % of the internal standard.

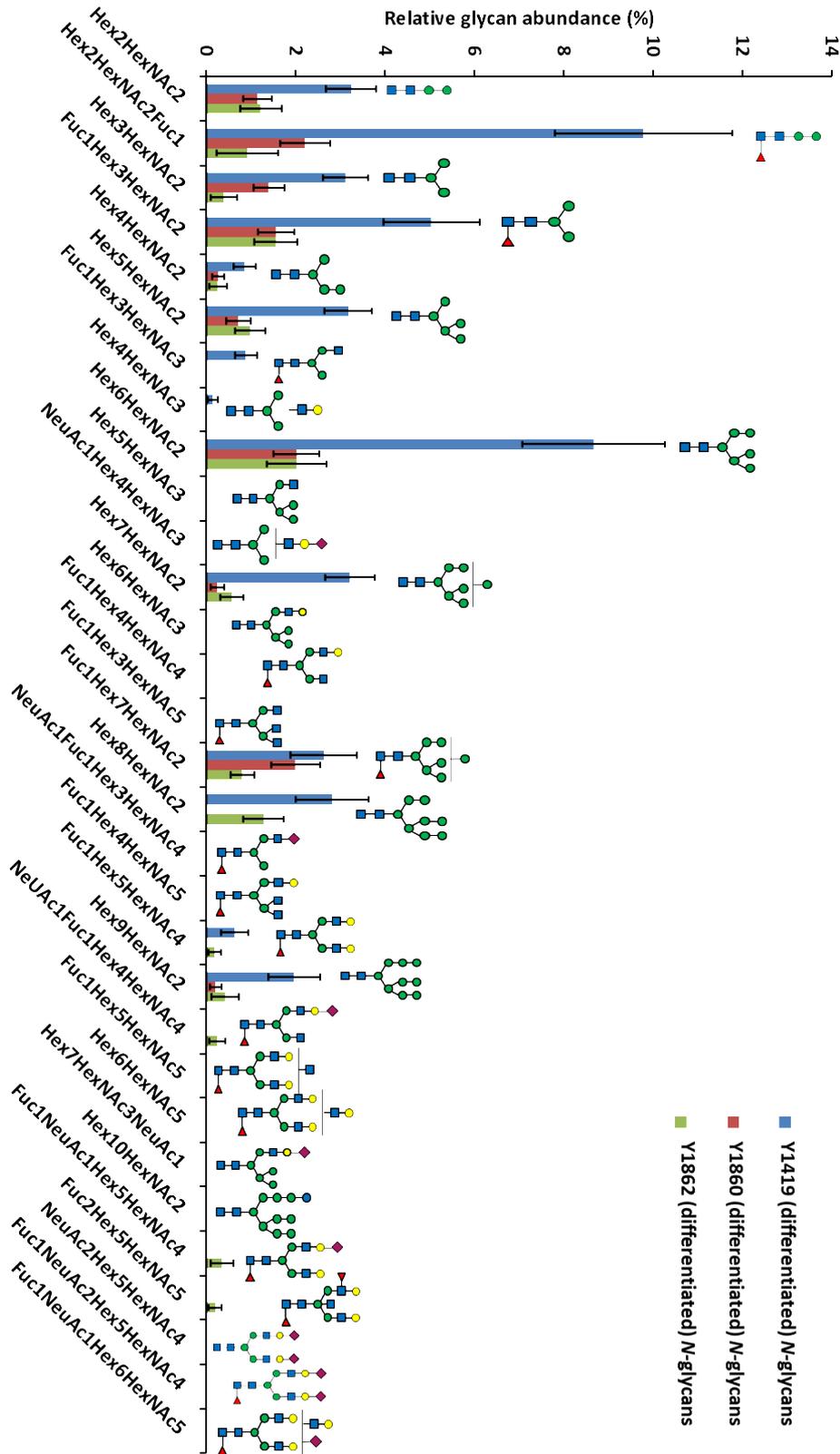


Figure 115: Comparison of the N-glycan profiles of the differentiated Y1419, Y1860 and Y1862 NHU cell lines. Relative abundance calculated by presenting each glycan signal as a percentage of the spiked internal standard signal. Glycan profiles are normalised to protein concentration.

While it is known that uroplakins contain *N*-glycans, there is nothing published in the literature about the presence or otherwise of O-glycans on these major urothelial glycoproteins. If there are O-glycans on uroplakins they may not have been detected to date if they are quantitatively very minor, or if previous studies had not attempted to isolate them for analysis. If the O-glycans described in the current work are not located on uroplakins, perhaps they come from membrane-bound or secreted mucins. Since the apical surface of the bladder is almost completely covered by uroplakins, mucins are likely to be quantitatively minor, which may explain why the urothelial O-glycans are lower in abundance than the *N*-glycans in this NHU study. Glycans in this work have been released and isolated from NHU whole cell lysates and not the isolated cell surface and so the abundance of O-glycans could appear lower if they are only present on the surface of the urothelium as they may be 'diluted' by the presence of further cell layers that may, for the majority, be *N*-glycosylated.

7.5. Conclusions and future work

The one-sample one-pot method developed for the release and isolation of *N*- and O-glycans has been successfully applied to proliferating and differentiated NHU cells. Mass spectrometry was used to produce quantifiable O-glycan and subsequently *N*-glycan profiles from cultures representing three human donor cell lines, Y1419, Y1860, Y1862. The advantage of analysing the two types of glycans from the same cells allows interesting differences in relative amounts to be very conveniently and directly compared. One cell line appeared to produce more glycans than the other two – the fact that the *N*- and O-glycans were released from the very same cells revealed that there was clearly something out of line about the general glycan levels from this cell line. If the *N*- and O-glycan analyses had been carried out on separate samples of cells, this direct comparison would not have been as readily made.

To gather more reliable data this work should be repeated with more technical and biological replicates, using comparable cell numbers for the undifferentiated and differentiated cells. By starting with a much larger amount of protein from the cells harvested from the 6 cm dishes compared with those in the semi-permeable membranes, it was difficult to make a direct comparison between changes in glycan profiles in the undifferentiated and differentiated cells. In addition, the differentiated donor cells will

display heterogeneity as the cells on the surface of the layers maybe display different numbers/types of glycans to the cells below the surface, and so it may be more comparable to take only the cells from the apical surface of the differentiated cells to compare to the undifferentiated cells cultured on a monolayer.

After repeating the study, if it is found that there are significant changes in particular glycan structures, a follow-up investigation could be carried out to determine whether those glycans play a role in differentiation and in formation of the urothelial barrier. For example, if particular genes for transferase enzymes are mutated, would it prevent formation of the tight barrier, monitored by TER, and how would it affect the glycan profile on the urothelial cell surface? Cells could also be treated with drugs under development for treating bladder diseases or infections such as UTIs or drugs and their metabolites that are thought to be toxic, to monitor the effects on glycosylation both with and without the tight barrier formation/ differentiation. For example ketamine, when used recreationally, has been shown to cause ketamine-induced cystitis^{333,334}. A study using organ cultures from human donors monitored the effect of the direct exposure of ketamine to the urothelium and found that is led to apoptosis (cell death) of urothelial cells³³⁵.

Further biological replicates could be studied to probe the effect of donor age and gender on the glycan profiles of NHU cells. It is known that females are more prone to urinary tract infections and perhaps this can be linked to differences in glycan profiles compared with males. In addition it is known that glycan structures change in age-related diseases³³⁶ so that age may be expected to impact glycan profiles and levels.

This work has shown that the abundance of O-glycans is lower than that of N-glycans in both proliferating and differentiated NHU cells. Since there is an increased expression of MUC1 genes in bladder cancer, it may be that there is an overall increase in mucin levels – this may result in an increase in total O-glycosylation of the urinary bladder surface; it would be interesting to determine whether there is an increase in O-glycan abundance of bladder cancer cells compared with healthy urothelial cells.

This work involved releasing and isolating glycans from whole cell lysates. Work needs to be done to look at the apical surface glycans, as these glycans are exposed to the luminal environment of the bladder. Analysing proliferating cells from 6 cm dishes compared with

differentiated cells cultured on SNAPwell membranes using this experimental system are probably not directly comparable, as the cell density on the 6 cm dishes is lower and cells are more likely to grow as a monolayer than those on the membrane which grow in distinct layers; the cell surfaces are likely to therefore be very different. Work to develop apical surface 'tryptic shaving' is currently being carried out by a PhD student, Chung-Yao Wang, for application to this system.

Overall, the work described in this chapter not only illustrates that the developed one-pot method works well for analysing the whole glycome using whole cell lysates from a small number of cells, but it has also generated for the first time an O-glycan profile of the urothelium. It thus represents a convenient tool for understanding more about the glycome of the urothelium that will help build knowledge that is relevant for urothelial biology, health and disease.

Chapter 8: Summary of results, overall conclusions and future work

Summary of results, overall conclusions and future work

The primary aim of this thesis was to develop a method to release and isolate O-glycans from whole cell lysates, for analysis by mass spectrometry, to be implemented in a workflow for studying disorders of protein glycosylation.

This thesis presents a FANGS-based approach¹⁴¹, making use of a centrifugal filter, to carry out O-glycan release using non-reductive β -elimination, as well as exploiting the advantages of an optimised FASP^{142,143} approach, eFASP¹⁷⁷. Method optimisation revealed how O-glycan release can be achieved with a total of 20 minutes of sonication. In addition, the thesis presents a streamlined protocol for a one-sample-one-pot approach to release *N*- and O-glycans from the same sample, in the same sample pot, with the potential to subsequently analyse the protein remaining in the filter.

Chapter 6 describes application of the O-glycan release method to mesenchymal stem cells, and the work described demonstrates how the one-pot glycan release method can be effectively used to compare changes in the glycan profiles between genetically modified MSC cells, chemically treated cells (to disrupt the O-glycan biosynthetic pathway) and those which have differentiated into adipocytes, to attempt to offer insight into the potential roles of glycans in adipogenesis. When MSCs were manipulated chemically or genetically, changes to overall glycan abundance and changes to glycan profiles (such as loss of sialylated and fucosylated glycans) could be observed.

Future experiments could investigate whether O-glycans are important for MSC differentiation into adipocytes, or whether these changes to the O-glycan profiles are independent of adipogenesis. The ability of stem cells to differentiate into different cell types means there is a potential use of MSCs in regenerative medicine^{276,277,278}, cell replacement²⁷⁹ and in tissue engineering²⁸⁰ and establishing whether glycans are either involved in or functionally relevant to differentiation could help future developments in these areas.

Chapter 7 describes application of the one-sample-one-pot method to normal urothelial cells cultured from human donor samples, enabling comparison of both *N*- and *O*- glycans between donors and between proliferating and differentiated cells. The work described in Chapter 7 not only illustrates that the developed one-pot method works well for analysing the whole glycome using whole cell lysates from a small number of cells, but it has also generated for the first time an *O*-glycan profile of the urothelium.

Future work on investigating the NHU glycome could involve looking at the apical surface glycans, as these glycans are exposed to the luminal environment of the bladder, in addition to probing the effect of donor age and gender on the glycan profiles of NHU cells.

In terms of further method development, a very novel mass spectrometric technique, 2D-FT-ICR-MS, has the potential to be a useful tool for glycomic analysis. 2D-FT-ICR-MS is being developed in Peter O'Connor's group at the University of Warwick. O'Connor's group has already demonstrated the use of 2D FT-ICR MS with both small molecules and larger molecules such as collagen digests. Since glycan samples contain often a large mixture of different structures and isomers, 2D FT-ICR MS could be a convenient method of simultaneously collecting MS/MS data on each glycan in one experiment rather than having to carry out many separate MS/MS experiments, and may also have the potential to be used as a quantitative tool. If the method can be used quantitatively it would be very useful for comparing differences in glycan profiles/abundances between *N*- and *O*-glycans in healthy and diseased cells as well as simultaneously being able to provide structural information and perhaps as a fingerprinting tool. Initial collaborations with the O'Connor group in this area have begun, following discussions between the author and Prof O'Connor after one of the author's conference presentations.

Chapter 5 describes work carried out at Bristol-Myers Squibb, Lawrenceville, New Jersey, USA. The project involved investigation of the FANGS method for larger scale *N*-glycan isolation, to enable analysis of glycans from monoclonal antibodies using a suite of 2D NMR techniques, in order to establish whether NMR could be used as a glycan fingerprinting tool to quickly and effectively monitor and analyse batches of biologics. Work on this project is now being continued at BMS, testing different batches of biotherapeutics using the FANGS method and generating fractions by HPLC of tryptic (glyco)peptides to try to isolate the glycan containing peptide to try to simplify the complex NMR spectra.

Overall, the work presented in this thesis has demonstrated a convenient way to quantitatively assess, by mass spectrometry, the whole glycome from one sample, in one reaction vessel, using only a small number of healthy or genetically manipulated/differentiated cultured cells. This work offers a method which could help in being able to further understand glycan biosynthesis, and the roles of both *N*- and *O*-glycosylation in disease. In addition FANGS can be scaled up to suit an NMR workflow.

A prepared paper on the one-sample-one-pot method development, including application to cultured cells is almost ready for submission. To follow on from this publication, there is a planned publication for the *O*-glycan profiling of the adipogenesis of the MSCs. In addition, BMS are hoping to publish recent work that is based on that carried out in chapter 5 of this thesis.

Abbreviations

μL	Microlitre
B ₀	Magnetic field
Asp	Asparagine
ATP	Adenosine triphosphate
CDG	Congenital disorders of glycosylation
CID	Collision induced dissociation
CMC	Critical micelle concentration
COG	Conserved oligomeric Golgi
COSY	Correlation spectroscopy
COPI	Coatmer protein complex 1
COPII	Coatmer protein complex 2
DCA	Deoxycholic acid
DCM	Dichloromethane
dept-HSQC	Distortionless enhancement - Heteronuclear single quantum correlation
DHB	2,5-dihydroxybenzoic acid
DMSO	dimethylsulfoxide
dqf-COSY	Double quantum filter – correlation spectroscopy
DTT	dithiothreitol
EDTA	Ethylenediaminetetracetic acid
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
FASP	Filter-aided sample preparation
eFASP	Enhanced filter-aided sample preparation
FANGS	Filter-aided <i>N</i> -glycan separation
Fuc	Fucose
FT-ICR	Fourier transform ion-cyclotron resonance
Gal	Galactose
GalNAc	N-acetylgalactosamine
GlcNAc	N-acetylglucosamine
HBSS	Hank's balanced salt solution
HCD	Higher energy collision dissociation

HMBC	Heteronuclear multiple bond correlation
HMQC	Heteronuclear multiple quantum correlation
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum correlation
LC-MS	Liquid chromatography – mass spectrometry
MALDI	Matrix-assisted laser/desorption ionisation
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSC	Mesenchymal stromal cells
NeuAc	N-acetylneuraminic acid
NHU	Normal human urothelial cells
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PBS	Phosphate buffered saline
PNGase F	Peptide N-glycosidase F
SDS	Sodium dodecyl sulphate
Ser	Serine
TEER	Trans-epithelial electrical resistance
Thr	Threonine
TOCSY	Total correlation spectroscopy

References

- 1 C. Ohyama, *Int. J. Clin. Oncol.*, 2008, **13**, 308–313.
- 2 Y. Mechref, Y. Hu, A. Garcia and A. Hussein, *Electrophoresis*, 2012, **33**, 1755–1767.
- 3 H. E. Miwa, Y. Song, R. Alvarez, R. D. Cummings and P. Stanley, *Glycoconj. J.*, 2012, **29**, 609–618.
- 4 C. Bull, T. J. Boltje, M. Wassink, A. M. A. de Graaf, F. L. van Delft, M. H. den Brok and G. J. Adema, *Mol. Cancer Ther.*, 2013, **12**, 1935–1946.
- 5 A. Palmigiano, R. Barone, L. Sturiale, C. Sanfilippo, R. O. Bua, D. A. Romeo, A. Messina, M. L. Capuana, T. Maci, F. Le Pira, M. Zappia and D. Garozzo, *J. Proteomics*, 2015, **131**, 29–37.
- 6 S. Schedin-Weiss, B. Winblad and L. O. Tjernberg, *FEBS J.*, 2014, **281**, 46–62.
- 7 O. Gornik and G. Lauc, *Dis. Markers*, 2008, **25**, 267–278.
- 8 H. F. Bunn, K. H. Gabbay and P. M. Gallop, *Science*, 1978, **200**, 21–27.
- 9 V. Ferreira, P. Briones and M. Vilaseca, *Carbohydr. Chem.*, 2012, **38**, 124–155.
- 10 H. H. Freeze, J. X. Chong, M. J. Bamshad and B. G. Ng, *Am. J. Hum. Genet.*, 2014, **94**, 161–175.
- 11 J. Jaeken, *J. Inherit. Metab. Dis.*, 2011, **34**, 853–858.
- 12 J. G. Leroy, *Pediatr. Res.*, 2006, **60**, 643–656.
- 13 J. E. Hudak and C. R. Bertozzi, *Chem. Biol.*, 2014, **21**, 16–37.
- 14 A. Beck, T. Wurch, C. Bailly and N. Corvaia, *Nat. Rev. Immunol.*, 2010, **10**, 345–352.
- 15 R. D. Astronomo and D. R. Burton, *Nat. Rev. Drug Discov.*, 2010, **9**, 308–324.
- 16 Y.-L. Huang, J.-T. Hung, S. K. C. Cheung, H.-Y. Lee, K.-C. Chu, S.-T. Li, Y.-C. Lin, C.-T. Ren, T.-J. R. Cheng, T.-L. Hsu, A. L. Yu, C.-Y. Wu and C.-H. Wong, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 2517–22.
- 17 P. H. Seeberger and D. B. Werz, *Nature*, 2007, **446**, 1046–1051.
- 18 C. Song, S. Sun, C.-X. Huo, Q. Li, X.-J. Zheng, G. Tai, Y. Zhou and X.-S. Ye, *Bioorg. Med. Chem.*, 2016, **24**, 915–920.
- 19 J. Fang, J. Richardson, Z. Du and Z. Zhang, *Biochemistry*, 2016, **55**, 860–868.
- 20 Y. Mimura, S. Church, R. Ghirlando, P. R. Ashton, S. Dong, M. Goodall, J. Lund and R. Jefferis, *Mol. Immunol.*, 2001, **37**, 697–706.
- 21 A. M. Goetze, Y. D. Liu, Z. Zhang, B. Shah, E. Lee, P. V. Bondarenko and G. C. Flynn, *Glycobiology*, 2011, **21**, 949–959.
- 22 L. Alessandri, D. Ouellette, A. Acquah, M. Rieser, D. LeBlond, M. Saltarelli, C.

- Radziejewski, T. Fujimori and I. Correia, *MAbs*, 2012, **4**, 509–520.
- 23 M. A. Tarp and H. Clausen, *Biochim. Biophys. Acta - Gen. Subj.*, 2008, **1780**, 546–563.
- 24 V. Restelli and M. Butler, *Cell engineering*, Springer, 1st edn., 2002, vol. 3.
- 25 M. Butler, *Cytotechnology*, 2006, **50**, 57–76.
- 26 A. Jamal, K. Ko, H.-S. Kim, Y.-K. Choo, H. Joung and K. Ko, *Biotechnol. Adv.*, 2009, **27**, 914–923.
- 27 H. Aghamohseni, K. Ohadi, M. Spearman, N. Krahn, M. Moo-Young, J. M. Scharer, M. Butler and H. M. Budman, *J. Biotechnol.*, 2014, **186**, 98–109.
- 28 K. Kamo, A. Takabatake, Y. Inoue and S. Izawa, *Biochem. Biophys. Res. Commun.*, 2012, **420**, 119–123.
- 29 L. Kaminski, Z. Guan, S. Yurist-Doutsch and J. Eichler, *MBio*, 2013, **4**, 1–9.
- 30 R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.*, 1985, **54**, 631–664.
- 31 J. Montreuil, J. F. G. Vliegthart and H. Schachter, *Glycoproteins*, Elsevier, Amsterdam, Vol. 30., 1995.
- 32 K. S. B. Bergstrom and L. Xia, *Glycobiology*, 2013, **23**, 1026–1037.
- 33 H. C. Hang and C. R. Bertozzi, *Bioorganic Med. Chem.*, 2005, **13**, 5021–5034.
- 34 M. Hook, *Ann. Rev. Biochem.*, 1984, **53**, 847–869.
- 35 J. Finne, T. Krusius, R. K. Margolis and R. U. Margolis, *J. Biol. Chem.*, 1979, **254**, 10295–10300.
- 36 L. Wells, *J. Biol. Chem.*, 2013, **288**, 6930–6935.
- 37 D. J. Moloney, V. M. Panin, S. H. Johnston, J. Chen, L. Shao, R. Wilson, Y. Wang, P. Stanley, K. D. Irvine, R. S. Haltiwanger and T. F. Vogt, *Nature*, 2000, **406**, 369–375.
- 38 R. Trapannone, D. Mariappa, A. T. Ferenbach and D. M. van Aalten, *Biochem. J.*, 2016, **473**, 1693–1702.
- 39 R. G. Spiro, *Glycobiology*, 2002, **12**, 43R–56R.
- 40 M. A. O'Neill and R. R. Selvendran, *Biochem J*, 1980, **187**, 53–63.
- 41 S. J. Gendler and A. P. Spicer, *Annu. Rev. Physiol.*, 1995, **57**, 607–634.
- 42 B. Garner, A. H. Merry, L. Royle, D. J. Harvey, P. M. Rudd and J. Thillet, *J. Biol. Chem.*, 2001, **276**, 22200–22208.
- 43 D. T. Tran and K. G. Ten Hagen, *J. Biol. Chem.*, 2013, **288**, 6921–6929.
- 44 T. Sørensen, T. White, H. H. Wandall, A. K. Kristensen, P. Roepstorff, H. Clausen, G. White and H. J. Biol, 1995, **270**, 24166–24173.
- 45 R. Willett, D. Ungar and V. Lupashin, *Histochem. Cell Biol.*, 2013, **140**, 271–283.
- 46 V. J. Miller and D. Ungar, *Traffic*, 2012, **13**, 891–897.

- 47 J. S. Bonifacino and B. S. Glick, *Cell*, 2004, **116**, 153–166.
- 48 B. S. Glick, T. Elston and G. Oster, *FEBS Lett.*, 1997, **414**, 177–181.
- 49 S. Emr, G. BS, A. D. Linstedt, L.-S. J, A. Luini, V. Malhotra, B. J. Marsh, A. Nakano, S. R. Pfeffer, C. Rabouille, J. E. Rothman, G. Warren and F. T. Wieland, *J. Cell Biol.*, 2009, **187**, 449–453.
- 50 R. J. Pelham HR, *Cell*, 2000, **103**, 713–719.
- 51 G. H. Patterson, K. Hirschberg, R. S. Polishchuk, D. Gerlich, R. D. Phair and J. Lippincott-Schwartz, *Cell*, 2008, **133**, 1055–1067.
- 52 N. P. Cottam and D. Ungar, *Protoplasma*, 2012, **249**, 943–955.
- 53 T. F. Roth and K. R. Porter, *J. Cell Biol.*, 1964, **20**, 313–332.
- 54 B. M. F. Pearse, *J. Mol. Biol.*, 1975, **97**, 93–98.
- 55 H. T. McMahon and I. G. Mills, *Curr. Opin. Cell Biol.*, 2004, **16**, 379–391.
- 56 B. L. Grosshans, D. Ortiz and P. Novick, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11821–7.
- 57 J. R. C. Whyte and S. Munro, *J. Cell Sci.*, 2002, **115**, 2627–2637.
- 58 E. Sztul and V. Lupashin, *Am. J. Physiol. Cell Physiol.*, 2006, **290**, C11–C26.
- 59 D. Ungar, T. Oka, M. Krieger and F. M. Hughson, *Trends Cell Biol.*, 2006, **16**, 113–120.
- 60 R. B. Jones, L. C. Ndhlovu, J. D. Barbour, P. M. Sheth, A. R. Jha, B. R. Long, J. C. Wong, M. Satkunarajah, M. Schweneker, J. M. Chapman, G. Gyenes, B. Vali, M. D. Hycza, F. Y. Yue, C. Kovacs, A. Sassi, M. Loutfy, R. Halpenny, D. Persad, G. Spotts, F. M. Hecht, T. W. Chun, J. M. McCune, R. Kaul, J. M. Rini, D. F. Nixon and M. A. Ostrowski, *J. Cell Biol.*, 2008, **183**, 607–615.
- 61 M. Fukuda, E. Kanno, K. Ishibashi and T. Itoh, *Mol. Cell. Proteomics*, 2008, **7**, 1031–1042.
- 62 V. J. Miller, P. Sharma, T. A. Kudlyk, L. Frost, A. P. Rofe, I. J. Watson, R. Duden, M. Lowe, V. V. Lupashin and D. Ungar, *J. Biol. Chem.*, 2013, **288**, 4229–4240.
- 63 T. Kudlyk, R. Willett, I. D. Pokrovskaya and V. Lupashin, *Traffic*, 2013, **14**, 194–204.
- 64 R. Willett, T. Kudlyk, I. Pokrovskaya, D. Ungar, R. Duden, V. Lupashin and L. Rock, *Nat. Commun.*, 2013, **4**, 1–28.
- 65 A. Shestakova, E. Suvorova, O. Pavliv, G. Khaidakova and V. Lupashin, *J. Cell Biol.*, 2007, **179**, 1179–1192.
- 66 M. Sohda, Y. Misumi, A. Yamamoto, N. Nakamura, S. Ogata, S. Sakisaka, S. Hirose, Y. Ikehara and K. Oda, *Traffic*, 2010, **11**, 1552–1566.
- 67 M. Sohda, Y. Misumi, S. I. Yoshimura, N. Nakamura, T. Fusano, S. Ogata, S. Sakisaka

- and Y. Ikehara, *Traffic*, 2007, **8**, 270–284.
- 68 I. B.-R. Ramirez and M. Lowe, *Semin. Cell Dev. Biol.*, 2009, **20**, 770–779.
- 69 J. E. Rothman, *Nature*, 1994, **372**, 55–63.
- 70 J. Jaeken, H. G. van Eijk, C. van der Heul, L. Corbeel, R. Eeckels and E. Eggermont, *Clin. Chim. Acta*, 1984, **144**, 245–247.
- 71 M. van Scherpenzeel, E. Willems and D. J. Lefeber, *Glycoconj. J.*, 2016, 1–14.
- 72 S. Grunewald, G. Matthijs and J. Jaeken, *Pediatr. Res.*, 2002, **52**, 618–624.
- 73 M. A. Haeuptle and T. Hennet, *Hum. Mutat.*, 2009, **30**, 1628–1641.
- 74 H. H. Freeze, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2009, **1792**, 835–840.
- 75 G. Matthijs, E. Schollen, E. Pardon, M. Veiga-Da-Cunha, J. Jaeken, J. J. Cassiman and E. Van Schaftingen, *Nat. Genet.*, 1997, **16**, 88–92.
- 76 E. Van Schaftingen and J. Jaeken, *FEBS Lett.*, 1995, **377**, 318–320.
- 77 J. Jaeken, T. Hennet, G. Matthijs and H. H. Freeze, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2009, **1792**, 825–826.
- 78 H. H. Freeze and M. Aebi, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 1999, **1455**, 167–178.
- 79 K. Panneerselvam and H. Freeze, *J. Clin. Invest.*, 1996, **97**, 1478–1487.
- 80 J. Jaeken, H. Schachter, H. Carchon, P. De Cock, B. Coddeville and G. Spik, *Arch. Dis. Child.*, 1994, **71**, 123–127.
- 81 B. Xia, W. Zhang, X. Li, R. Jiang, T. Harper, R. Liu, R. D. Cummings and M. He, *Anal. Biochem.*, 2013, **442**, 178–185.
- 82 S. Wopereis, E. Morava, S. Grunewald, M. Adamowicz, K. M. L. C. Huijben, D. J. Lefeber and R. A. Wevers, *Glycobiology*, 2005, **15**, 1312–1319.
- 83 X. Wu, R. A. Steet, O. Bohorov, J. Bakker, J. Newell, M. Krieger, L. Spaapen, S. Kornfeld and H. H. Freeze, *Nat. Med.*, 2004, **10**, 518–523.
- 84 T. Marquardt, *Nat. Med.*, 2004, **10**, 457–458.
- 85 F. Foulquier, E. Vasile, E. Schollen, N. Callewaert, T. Raemaekers, D. Quelhas, J. Jaeken, P. Mills, B. Winchester, M. Krieger, W. Annaert and G. Matthijs, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 3764–3769.
- 86 D. M. Kingsley, K. F. Kozarsky, M. Segal and M. Krieger, *J. Cell Biol.*, 1986, **102**, 1576–1585.
- 87 C. Kranz, B. G. Ng, L. Sun, V. Sharma, E. A. Eklund, Y. Miura, D. Ungar, V. Lupashin, R. D. Winkel, J. F. Cipollo, C. E. Costello, E. Loh, W. Hong and H. H. Freeze, *Hum. Mol. Genet.*, 2007, **16**, 731–741.

- 88 J. Jaeken and G. Matthijs, *Annu. Rev. Genomics Hum. Genet.*, 2007, **8**, 261–78.
- 89 Y. Wada, A. Nishikawa, N. Okamoto, K. Inui, H. Tsukamoto, S. Okada and N. Taniguchi, *Biochem. Biophys. Res. Commun.*, 1992, **189**, 832–836.
- 90 Y. Wada, J. Gu and N. Okamoto, *Biol. Mass Spectrom.*, 1994, **23**, 108–109.
- 91 Y. Wada, P. Azadi, C. E. Costello, A. Dell, R. A. Dwek, H. Geyer, R. Geyer, K. Kakehi, N. G. Karlsson, K. Kato, N. Kawasaki, K.-H. Khoo, S. Kim, A. Kondo, E. Lattova, Y. Mechref, E. Miyoshi, K. Nakamura, H. Narimatsu, M. V. Novotny, N. H. Packer, H. Perreault, J. Peter-Katalinic, G. Pohlentz, V. N. Reinhold, P. M. Rudd, A. Suzuki and N. Taniguchi, *Med. Phys.*, 2007, **17**, 411–422.
- 92 M. Guillard, J. Gloerich, H. J. C. T. Wessels, E. Morava, R. a Wevers and D. J. Lefeber, *Carbohydr. Res.*, 2009, **344**, 1550–1557.
- 93 V. Cantagrel, D. J. Lefeber, B. G. Ng, Z. Guan, J. L. Silhavy, S. L. Bielas, L. Lehle, H. Hombauer, M. Adamowicz, E. Swiezewska, A. P. De Brouwer, P. Blümel, J. Sykut-Cegielska, S. Houliston, D. Swistun, B. R. Ali, W. B. Dobyns, D. Babovic-Vuksanovic, H. van Bokhoven, R. A. Wevers, C. R. H. Raetz, H. H. Freeze, É. Morava, L. Al-Gazali and J. G. Gleeson, *Cell*, 2010, **142**, 203–217.
- 94 Y. Wada, M. Tajiri and S. Yoshida, *Anal. Chem.*, 2004, **76**, 6560–6565.
- 95 E. A. Kabat and S. Leskowitz, *J. Am. Chem. Soc.*, 1955, **77**, 5159–5164.
- 96 E. G. Gray, *Nature*, 1955, **175**, 642–643.
- 97 W. T. J. Morgan and W. M. Watkins, *Br. J. Exp. Pathol.*, 1953, **34**, 94–103.
- 98 W. Watkins, *Transfus. Med.*, 2001, **11**, 243–265.
- 99 R. D. Rosenberg and L. Lam, *Proc. Natl. Acad. Sci. U. S. A.*, 1979, **76**, 1218–22.
- 100 U. Lindahl, G. Bäckström, M. Höök, L. Thunberg, L. A. Fransson and A. Linker, *Proc. Natl. Acad. Sci. U. S. A.*, 1979, **76**, 3198–202.
- 101 M. Hricovíni, M. Guerrini, a Bisio, G. Torri, M. Petitou and B. Casu, *Biochem. J.*, 2001, **359**, 265–272.
- 102 U. Lindahl, G. Backstrom, L. Thunberg and I. G. Leder, *Proc. Natl. Acad. Sci.*, 1980, **77**, 6551–6555.
- 103 I. Capila and R. J. Linhardt, *Angew. Chem. Int. Ed. Engl.*, 2002, **41**, 391–412.
- 104 R. D. Rosenberg and P. S. Damus, *J. Biol. Chem.*, 1973, **248**, 6490–6505.
- 105 M. von Itzstein, W. Y. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. Van Phan, M. L. Smythe, H. F. White and S. W. Oliver, *Nature*, 1993, **363**, 418–23.
- 106 M. N. Matrosovich, T. Y. Matrosovich, T. Gray, N. a Roberts and H. Klenk, *Society*, 2004, **78**, 12665–12667.

- 107 Z. Shriver, R. Raman, K. Viswanathan and R. Sasisekharan, *Chem. Biol.*, 2009, **16**, 803–814.
- 108 L. Wang, A. Pulk, M. R. Wasserman, M. B. Feldman, R. B. Altman, J. H. D. Cate and S. C. Blanchard, *Nat. Struct. Mol. Biol.*, 2012, **19**, 957–63.
- 109 G. Walsh, *Nat. Biotechnol.*, 2014, **32**, 992–1000.
- 110 F. Cameron and P. L. McCormack, *Drugs*, 2014, **74**, 147–154.
- 111 J. Golay, F. Da Roit, L. Bologna, C. Ferrara, J. H. Leusen, A. Rambaldi, C. Klein and M. Introna, *Blood*, 2013, **122**, 3482–91.
- 112 Y. Shaaltiel, D. Bartfeld, S. Hashmueli, G. Baum, E. Brill-Almon, G. Galili, O. Dym, S. A. Boldin-Adamsky, I. Silman, J. L. Sussman, A. H. Futerman and D. Aviezer, *Plant Biotechnol. J.*, 2007, **5**, 579–590.
- 113 V. Ferna, A. Rodriguez, A. Baly, L. Herrera, M. Izquierdo, A. Villar, Y. Valde, K. Cosme, M. L. Deler, M. Montane, E. Garcia, A. Ramos, A. Aguilar and E. Medina, *Science.*, 2004, **305**, 522–525.
- 114 K. Lavrsen, C. B. Madsen, M. G. Rasch, A. Woetmann, N. Ødum, U. Mandel, H. Clausen, A. E. Pedersen and H. H. Wandall, *Glycoconj. J.*, 2013, **30**, 227–236.
- 115 Y. Suzuki, M. Sutoh, S. Hatakeyama, K. Mori, H. Yamamoto, T. Koie, H. Saitoh, K. Yamaya, T. Funyu, T. Habuchi, Y. Arai, M. Fukuda, C. Ohyama and S. Tsuboi, *Int. J. Oncol.*, 2012, **40**, 1831–1838.
- 116 R. Jefferis, *Nat. Rev. Drug Discov.*, 2009, **8**, 226–34.
- 117 S. J. North, H.-H. Huang, S. Sundaram, J. Jang-Lee, a T. Etienne, A. Trollope, S. Chalabi, A. Dell, P. Stanley and S. M. Haslam, *J. Biol. Chem.*, 2010, **285**, 5759–5775.
- 118 M. Davies, J., Jiang, L., Pan, L. Z., LaBarre, M.J., Anderson, D., and Reff, *Biotechnol. Bioeng.*, 2001, **74**, 288–294.
- 119 P. Umaña, J. Jean-Mairet, R. Moudry, H. Amstutz and J. E. Bailey, *Nat. Biotechnol.*, 1999, **17**, 176–80.
- 120 F. M. Platt, G. R. Neises, G. Reinkensmeier, M. J. Townsend, V. H. Perry, R. L. Proia, B. Winchester, R. A. Dwek, T. D. Butters, E. F. Neufeld, E. Beutler, N. W. Barton, E. Beutler, F. M. Platt, G. R. Neises, R. A. Dwek, T. D. Butters, F. M. Platt, T. D. Butters, S. Yamanaka, K. Sango, M. Taniike and M. A. Fischl, *Science*, 1997, **276**, 428–31.
- 121 Z. J. Gartner and C. R. Bertozzi, *Proc. Natl. Acad. Sci.*, 2009, **106**, 1–5.
- 122 G. A. Lemieux and C. R. Bertozzi, *Chem. Biol.*, 2001, **8**, 265–275.
- 123 D. M. Carlson, *J. Biol. Chem.*, 1968, **243**, 616–626.
- 124 G. Spik, G. Strecker, B. Fournet, S. Bouquelet, J. Montreuil, L. Dorland, H. van Halbeek

- and J. F. Vliegenthart, *Eur. J. Biochem.*, 1982, **121**, 413–419.
- 125 D. Ungar, *Semin. Cell Dev. Biol.*, 2009, **20**, 762–769.
- 126 A. Dell, J. E. Oates, H. R. Morris and H. Egge, *Int. J. Mass Spectrom. Ion Phys.*, 1983, **46**, 415–418.
- 127 P. Babu, S. J. North, J. Jang-Lee, S. Chalabi, K. MacKerness, S. R. Stowell, R. D. Cummings, S. Rankin, A. Dell and S. M. Haslam, *Glycoconj. J.*, 2009, **26**, 975–986.
- 128 C. S. Chu, M. R. Niñonuevo and B. H. Clowers, *Proteomics*, 2009, **9**, 1939–1951.
- 129 L. Ruhaak, S. Miyamoto, K. Kelly and C. B. Lebrilla, *Anal. Chem.*, 2012, **84**, 396–402.
- 130 N. G. Karlsson, H. Nordman, H. Karlsson, I. Carlstedt and G. C. Hansson, *Biochem. J.*, 1997, **326**, 911–917.
- 131 H. Nordman, J. R. Davies, A. Herrmann, N. G. Karlsson, G. C. Hansson and I. Carlstedt, *Biochem. J.*, 1997, **326**, 903–910.
- 132 S. Ohara, K. Ishihara and K. Hotta, *Comp. Biochem. Physiol. -- Part B Biochem.*, 1993, **106**, 153–158.
- 133 R. A. M. Stanley R A, Lee S P, *Biochim. Biophys. Acta - Gen. Subj.*, 1983, **760**, 262–269.
- 134 A. C. Bateman, R. Karamanska, M. G. Busch, A. Dell, C. W. Olsen and S. M. Haslam, *J. Biol. Chem.*, 2010, **285**, 34016–34026.
- 135 M. Nakano, R. Saldanha, A. Gobel, M. Kavallaris and N. H. Packer, *Mol. Cell. Proteomics*, 2011, **10**, 1–12.
- 136 N. Nagahori, T. Yamashita, M. Amano and S. I. Nishimura, *ChemBioChem*, 2013, **14**, 73–82.
- 137 N. Fujitani, J. Furukawa, K. Araki, T. Fujioka, Y. Takegawa, J. Piao, T. Nishioka, T. Tamura, T. Nikaido, M. Ito, Y. Nakamura and Y. Shinohara, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 2105–2110.
- 138 T. Satomaa, A. Heiskanen, M. Mikkola, C. Olsson, M. Blomqvist, M. Tiittanen, T. Jaatinen, O. Aitio, A. Olonen, J. Helin, J. Hiltunen, J. Natunen, T. Tuuri, T. Otonkoski, J. Saarinen and J. Laine, *BMC Cell Biol.*, 2009, **18**, 1–18.
- 139 A. Heiskanen, T. Hirvonen, H. Salo, U. Impola, A. Olonen, A. Laitinen, S. Tiitinen, S. Natunen, O. Aitio, H. Miller-Podraza, M. Wuhrer, A. M. Deelder, J. Natunen, J. Laine, P. Lehenkari, J. Saarinen, T. Satomaa and L. Valmu, *Glycoconj. J.*, 2009, **26**, 367–84.
- 140 M. R. Kudelka, A. Antonopoulos, Y. Wang, D. M. Duong, X. Song, N. T. Seyfried, A. Dell, S. M. Haslam, R. D. Cummings and T. Ju, *Nat. Methods*, 2016, **13**, 81–6.
- 141 S. A. Rahman, E. Bergstrom, C. J. Watson, K. M. Wilson, D. Ashford, J. R. Thomas, D. Ungar and J. E. Thomas-Oates, *J. Proteome Res.*, 2014, **13**, 1167–1176.

- 142 L. L. Manza, S. L. Stamer, A.-J. L. Ham, S. G. Codreanu and D. C. Liebler, *Proteomics*, 2005, **5**, 1742–1745.
- 143 J. Wiśniewski, A. Zougman, N. Nagaraj and M. Mann, *Nat. Methods*, 2009, **6**, 359–363.
- 144 A. L. Tarentino, C. M. Gómez and T. H. Plummer, *Biochemistry*, 1985, **24**, 4665–71.
- 145 Y. Matsushima and N. Fujii, *Bull. Chem. Soc. Jpn.*, 1957, **30**, 48–50.
- 146 B. Bayard and D. Roux, *FEBS Lett.*, 1975, **55**, 206–211.
- 147 R. P. Kozak, L. Royle, R. a Gardner, A. Bondt, D. L. Fernandes and M. Wuhrer, *Anal. Biochem.*, 2014, **453**, 29–37.
- 148 R. P. Kozak, L. Royle, R. a Gardner, D. L. Fernandes and M. Wuhrer, *Anal. Biochem.*, 2012, **423**, 119–128.
- 149 T. Patel, J. Bruce, A. Merry and C. Bigge, *Biochemistry*, 1993, **32**, 679–693.
- 150 A. H. Merry, D. C. a Neville, L. Royle, B. Matthews, D. J. Harvey, R. a Dwek and P. M. Rudd, *Anal. Biochem.*, 2002, **304**, 91–99.
- 151 J. J. Plantner and D. M. Carlson, *Anal. Biochem.*, 1975, **163**, 153–163.
- 152 R. L. Whistler and J. N. Bemiller, *Adv. Carbohydr. Chem.*, 1958, **13**, 289–329.
- 153 G. J. Rademaker, S. A. Pergantis, L. Blok-tip, J. I. Langridge, A. Kleen and J. E. Thomas-oates, *Anal. Biochem.*, 1998, **257**, 149–160.
- 154 Y. Huang, Y. Mechref and M. Novotny, *Anal. Chem.*, 2001, **73**, 6063–6069.
- 155 R. S. Zhou, H., Briscoe, A. C., Froehlich, J. W., Lee, *Anal. Biochem.*, 2012, **427**, 33–35.
- 156 C. a Retamal, P. Thiebaut and E. W. Alves, *Anal. Biochem.*, 1999, **268**, 15–20.
- 157 T. Mason, *Chem. Soc. Rev.*, 1997, **26**, 443–451.
- 158 D. J. Harvey, *J. Chromatogr. A*, 1996, **720**, 429–446.
- 159 J. Zaia, *Mass Spectrom. Rev.*, 2004, **23**, 161–227.
- 160 D. J. Harvey, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2011, **879**, 1196–1225.
- 161 I. Ciucanu and F. Kerek, *Carbohydr. Res.*, 1984, **131**, 209–217.
- 162 I. Ciucanu and C. Costello, *J. Am. Chem. Soc.*, 2003, **125**, 16213–16219.
- 163 A. Gunnarsson, *Glycoconj. J.*, 1987, **4**, 239–245.
- 164 A. Dell, M. E. Rogers, J. E. Thomas-Oates, T. N. Huckerby, P. N. Sanderson and I. A. Nieduszynski, *Carbohydr. Res.*, 1988, **179**, 7–19.
- 165 E. J. Bourne, M. Stacey and J. C. Tatlow, *J. Chem. Soc.*, 1949, 2976–2979.
- 166 L. Han and C. Costello, *Biochemistry*, 2013, **78**, 710–720.
- 167 P. Kang, Y. Mechref, I. Klouckova and M. V Novotny, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 3421–3428.

- 168 S. Hakomori, *J. Biochem.*, 1964, **55**, 205–208.
- 169 S. Zhou, N. Tello, A. Harvey, B. Boyes, R. Orlando and Y. Mechref, *Electrophoresis*, 2016, 1489–1497.
- 170 S. I. Snovida, J. M. Rak-Banville and H. Perreault, *J. Am. Soc. Mass Spectrom.*, 2008, **19**, 1138–1146.
- 171 J. A. Atwood, L. Cheng, G. Alvarez-Manilla, N. L. Warren, W. S. York and R. Orlando, *J. Proteome Res.*, 2008, **7**, 367–374.
- 172 Y. Mechref, Y. Hu, J. L. Desantos-Garcia, A. Hussein and H. Tang, *Mol. Cell. Proteomics*, 2013, **12**, 874–84.
- 173 P. Kang, Y. Mechref, Z. Kyselova, J. A. Goetz and M. V. Novotny, *Anal. Chem.*, 2007, **79**, 6064–6073.
- 174 G. Alvarez-Manilla, N. L. Warren, J. Atwood III, T. Abney, P. Azadi, M. Pierce and R. Orlando, *Glycobiology*, 2006, **16**, 677–687.
- 175 W. Zhang, H. Wang, H. Tang and P. Yang, *Anal. Chem.*, 2011, **83**, 4975–4981.
- 176 D. J. Harvey, L. Royle, C. M. Radcliffe, P. M. Rudd and R. A. Dwek, *Anal. Biochem.*, 2008, **376**, 44–60.
- 177 J. Erde, R. Loo and J. Loo, *J. Proteome Res.*, 2014, **13**, 1885–1895.
- 178 M. Karas, D. Bachmann and F. Hillenkamp, *Anal. Chem.*, 1985, **57**, 2935–2939.
- 179 M. Karas and F. Hillenkamp, *Anal. Chem.*, 1988, **60**, 2299–2301.
- 180 F. Hillenkamp and M. Karas, *Anal. Chem.*, 1991, **63**, 1193–1203.
- 181 K. Strupat, M. Karas and F. Hillenkamp, *Int. J. Mass Spectrom. Ion Process.*, 1991, **111**, 89–102.
- 182 R. Beavis, T. Chaudhary and B. Chaid, *Org. Mass Spectrom.*, 1992, **27**, 156–158.
- 183 T. W. Jaskolla, D. G. Papatotiriou and M. Karas, *J. Proteome Res.*, 2009, **8**, 3588–3597.
- 184 O. Vorm, P. Roepstorff and M. Mann, *Anal. Chem.*, 1994, **66**, 3281–3287.
- 185 M. Karas and R. Krüger, *Chem. Rev.*, 2003, **103**, 427–439.
- 186 R. Zenobi and R. Knochenmuss, *Mass Spectrom. Rev.*, 1998, **17**, 337–366.
- 187 E. Lehmann, R. Knochenmuss and R. Zenobi, *Rapid Commun. Mass Spectrom.*, 1997, **11**, 1483–1492.
- 188 M. Karas, M. Glückmann and J. Schäfer, *J. mass Spectrom.*, 2000, **35**, 1–12.
- 189 R. Knochenmuss and R. Zenobi, *Chem. Rev.*, 2003, **103**, 441–452.
- 190 R. Knochenmuss, *Analyst*, 2006, **131**, 966–86.
- 191 R. Knochenmuss, a Stortelder, K. Breuker and R. Zenobi, *J. Mass Spectrom.*, 2000, **35**, 1237–45.

- 192 H. Ehring, M. Karas and F. Hillenkamp, *Org. mass Spectrom.*, 1992, **27**, 472–480.
- 193 T. W. Jaskolla and M. Karas, *J. Am. Soc. Mass Spectrom.*, 2011, **22**, 976–988.
- 194 J. S. Klassen, S. G. Anderson, A. T. Blades and P. Kebarle, *J. Phys. Chem.*, 1996, **100**, 14218–14227.
- 195 M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson and M. B. Alice, *J. Chem. Phys.*, 1968, **49**, 2240–2249.
- 196 C. M. Whitehouse, R. N. Dreyer, M. Yamashita and J. B. Fenn, *Anal. Chem.*, 1985, **57**, 675–679.
- 197 J. B. Fenn, M. Mann, C. K. A. I. Meng, S. F. Wong and C. M. Whitehouse, *Science.*, 1989, **246**, 64–71.
- 198 T. G. I, *Proc. R. Soc. London A*, 1964, 383–397.
- 199 J. Zeleny, *Phys. Rev.*, 1914, **III**, 69–91.
- 200 J. V Iribarne and B. A. Thomson, *J. Chem. Phys.*, 1976, **64**, 2287–2294.
- 201 J. Amster, *J. mass Spectrom.*, 1996, **31**, 1325–1337.
- 202 A. G. Marshall, C. L. Hendrickson and G. S. Jackson, *Mass Spectrom. Rev.*, 1998, **17**, 1–35.
- 203 S. A. Goudsmit, *Phys. Rev.*, 1948, **74**, 622–623.
- 204 J. A. Hipple, J. A. Thomas, *Phys. Rev.*, 1949, **75**, 1616–1616.
- 205 W. C. Wiley and I. H. McLaren, *Rev. Sci. Instrum.*, 1955, **26**, 1150–1157.
- 206 J. Tabet and R. J. Cotter, *Anal. Chem.*, 1984, **56**, 1662–1667.
- 207 J. G. Boyle and C. M. Whitehouse, *Anal. Chem.*, 1992, **64**, 2084–9.
- 208 R. J. Cotter, *Anal. Chem.*, 1992, **64**, 1027–1039.
- 209 B. A. Mamyryn, *Int. J. Mass Spectrom.*, 2001, **206**, 251–266.
- 210 A. Makarov, *Anal. Chem.*, 2000, **72**, 1156–1162.
- 211 Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman and R. G. Cooks, *J. Mass Spectrom.*, 2005, **40**, 430–443.
- 212 A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat and S. Horning, *Anal. Chem.*, 2006, **78**, 2113–2120.
- 213 Thermoscientific, 2016, <http://planetorbitrap.com/orbitrap-fusion#tab:sche>.
- 214 R. B. Cody, R. C. Burnler and B. S. Frelser, *Anal. Chem.*, 1982, **54**, 96–101.
- 215 B. Domon and C. Costello, *Glycoconj. J.*, 1988, **5**, 397–409.
- 216 H. Steen and M. Mann, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 699–711.
- 217 P. Roepstorff, *Biomed. Mass Spectrom.*, 1984, **11**, 601.
- 218 K. Biemann, *Biomed. Environ. Mass Spectrom.*, 1988, **16**, 99–111.

- 219 C. Wang, W. Fan, P. Zhang, Z. Wang and L. Huang, *Proteomics*, 2011, **11**, 4229–4242.
- 220 W. Morelle and J.-C. Michalski, *Nat. Protoc.*, 2007, **2**, 1585–1602.
- 221 G. J. Rademaker, S. a Pergantis, L. Blok-Tip, J. I. Langridge, a Kleen and J. E. Thomas-Oates, *Anal. Biochem.*, 1998, **257**, 149–160.
- 222 D. R. Mould and B. Meibohm, *Biodrugs*, 2016, **30**, 275–293.
- 223 G. Kohler and C. Milstein, *Nature*, 1975, **256**, 495–497.
- 224 A. Planinc, J. Bones, B. Dejaegher, P. Van Antwerpen and C. Delporte, *Anal. Chim. Acta*, 2015, **921**.
- 225 <http://www.wiley.com/legacy/products/subject/life/elgert/CH04.pdf>.
- 226 C. J. Janeway, P. Travers, M. Walport and M. J. Schlomchik, *Immunobiology: The Immune System in Health and Diseases*, Garland Science, New York, 5th Ed., 2001.
- 227 R. R. Porter, *Biochem. J.*, 1959, **73**, 119–127.
- 228 D. N. Forthal, *Microbiol. Spectr.*, 2014, **2**, 1–17.
- 229 D. D. Chaplin, *J. Allergy Clin. Immunol.*, 2010, **125**.
- 230 K. Hajela, *Biochem. Educ.*, 1991, **19**, 50–57.
- 231 S. T. Lee and F. Paraskevas, *Cellular Immunol.*, 1978, **40**, 141–153.
- 232 D. L. Nelson, B. M. Bundy, T. D. West and W. Strober, *Cellular Immunol.*, 1976, **23**, 89–98.
- 233 U. J. E. Seidel, P. Schlegel and P. Lang, *Front. Immunol.*, 2013, **4**, 1–8.
- 234 E. Uçaktürk, *J. Sep. Sci.*, 2012, **35**, 341–350.
- 235 W. Wang, S. Singh, D. L. Zeng, K. King and S. Nema, *J. Pharm. Sci.*, 2007, **96**, 1–26.
- 236 W. Zhang and M. J. Czupryn, *Biotechnol. Prog.*, 2002, **18**, 509–513.
- 237 G. R. S. Hartig, T. T. Tran and M. L. Smythe, *Protein Sci.*, 2005, **14**, 474–482.
- 238 J. T. Pelton and L. R. McLean, *Anal. Biochem.*, 2000, **277**, 167–176.
- 239 S. R. Brych, Y. R. Gokarn, H. Hultgen, R. J. Stevenson, R. Rajan and M. Matsumura, *J. Pharm. Sci.*, 2010, **99**, 764–781.
- 240 J. Bongers, J. J. Cummings, M. B. Ebert, M. M. Federici, L. Gledhill, D. Gulati, G. M. Hilliard, B. H. Jones, K. R. Lee, J. Mozdzanowski, M. Naimoli and S. Burman, *J. Pharm. Biomed. Anal.*, 2000, **21**, 1099–1128.
- 241 N. Lundell and T. Schreitmuller, *Anal. Biochem.*, 1999, **266**, 31–47.
- 242 E. Tamizi and A. Jouyban, *Eur. J. Pharm. Biopharm.*, 2016, **98**, 26–46.
- 243 M. C. Manning, K. Patel and R. T. Borchardt, *Pharm. Res.*, 1989, **6**, 903–918.
- 244 N. Jenkins, L. Murphy and R. Tyther, *Mol. Biotechnol.*, 2008, **39**, 113–118.
- 245 R. M. Daniel, M. Dines and H. H. Petach, *Biochem. J.*, 1996, **317**, 1–11.

- 246 D. Chelius, D. S. Render and P. V. Bondarenko, *Anal. Chem.*, 2005, **77**, 6004–6011.
- 247 J. Lin, Q. Tan and S. Wang, *J. Sep. Sci.*, 2011, **34**, 1696–1702.
- 248 T. S. Raju, *Curr. Opin. Immunol.*, 2008, **20**, 471–478.
- 249 R. Jefferis, *Trends Pharmacol. Sci.*, 2009, **30**, 356–362.
- 250 Y. Kanda, T. Yamada, K. Mori, A. Okazaki, M. Inoue, K. Kitajima-Miyama, R. Kuni-Kamochi, R. Nakano, K. Yano, S. Kakita, K. Shitara and M. Satoh, *Glycobiology*, 2007, **17**, 104–118.
- 251 T. S. Mattu, R. J. Pleass, A. C. Willis, M. Kilian, M. R. Wormald, A. C. Lellouch, P. M. Rudd, J. M. Woof and R. A. Dwek, *J. Biol. Chem.*, 1998, **273**, 2260–2272.
- 252 Y. Kaneko, F. Nimmerjahn and J. V Ravetch, *Science.*, 2006, **313**, 670–3.
- 253 B. J. Scallon, S. H. Tam, S. G. McCarthy, A. N. Cai and T. S. Raju, *Mol. Immunol.*, 2007, **44**, 1524–1534.
- 254 E. S. X. Moh, C.-H. Lin, M. Thaysen-Andersen and N. H. Packer, *J. Am. Soc. Mass Spectrom.*, 2016, 1143–1155.
- 255 M. Fellenberg, H. N. Behnken, T. Nagel, A. Wiegandt, M. Baerenfaenger and B. Meyer, *Anal. Bioanal. Chem.*, 2013, **405**, 7291–7305.
- 256 M. Schubert, M. J. Walczak, M. Aebi and G. Wider, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 7096–100.
- 257 A. Wiegandt and B. Meyer, *Anal. Chem.*, 2014, **86**, 4807–4814.
- 258 H. N. Behnken, M. Fellenberg, M. P. Koetzler, R. Jirmann, T. Nagel and B. Meyer, *Anal. Bioanal. Chem.*, 2012, **404**, 1427–1437.
- 259 J. Keeler, http://www-keeler.ch.cam.ac.uk/lectures/understanding/chapter_2.pdf, 2004.
- 260 J. Keeler, http://www-keeler.ch.cam.ac.uk/lectures/understanding/chapter_3.pdf, 2004.
- 261 J. Keeler, http://www-keeler.ch.cam.ac.uk/lectures/understanding/chapter_4.pdf, 2004.
- 262 J. Keeler, http://www-keeler.ch.cam.ac.uk/lectures/understanding/chapter_5.pdf, 2004.
- 263 J. Keeler, http://www-keeler.ch.cam.ac.uk/lectures/understanding/chapter_7.pdf.
- 264 C. W. N. Damen, W. Chen, A. B. Chakraborty, M. van Oosterhout, J. R. Mazzeo, J. C. Gebler, J. H. M. Schellens, H. Rosing and J. H. Beijnen, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 2021–2033.
- 265 K. Masuda, Y. Yamaguchi, K. Kato, N. Takahashi, I. Shimada and Y. Arata, *FEBS Lett.*,

- 2000, **473**, 349–357.
- 266 S. Krapp, Y. Mimura, R. Jefferis, R. Huber and P. Sonderrmann, *J. Mol. Biol.*, 2003, **325**, 979–989.
- 267 <http://glycosciences.de/database/structure/>.
- 268 D. J. Prockop, *Science.*, 1997, **276**, 71–74.
- 269 M. F. Pittenger, A. M. Mackay, S. Beck, R. K. Jaiswal, R. Douglas, J. D. Mosca, M. a. Moorman, D. W. Simonetti, S. Craig and D. Marshak, *Science.*, 1999, **284**, 143–147.
- 270 J. B. Noronha-Matos and P. Correia-de-Sá, *J. Cell. Physiol.*, 2016, 1852–1861.
- 271 G. D’Ippolito, P. C. Schiller, C. Ricordi, B. a Roos and G. a Howard, *J. Bone Miner. Res.*, 1999, **14**, 1115–1122.
- 272 O. Lindvall, Z. Kokaia and A. Martinez-Serrano, *Nat. Med.*, 2004, **10**, 42–50.
- 273 S. P. Bruder, D. J. Fink and A. I. Caplan, *J. Cell. Biochem.*, 1994, **56**, 283–294.
- 274 A. P. Beltrami, L. Barlucchi, D. Torella, M. Baker, F. Limana, S. Chimenti, H. Kasahara, M. Rota, E. Musso, K. Urbanek, A. Leri, J. Kajstura, B. Nadal-Ginard and P. Anversa, *Cell*, 2003, **114**, 763–776.
- 275 A. Leri, J. a N. Kajstura and P. Anversa, *Physiol Rev*, 2005, **85**, 1373–1416.
- 276 A. Uccelli, L. Moretta and V. Pistoia, *Nat Rev Immunol*, 2008, **8**, 726–736.
- 277 F. P. Barry and J. M. Murphy, *Int. J. Biochem. Cell Biol.*, 2004, **36**, 568–584.
- 278 M. Mimeault, R. Hauke and S. K. Batra, *Clin Pharmacol Ther*, 2007, **82**, 252–264.
- 279 V. Sueblinvong, B. T. Suratt and D. J. Weiss, *Clin. Chest Med.*, 2007, **28**, 361–379.
- 280 J. Jaipaew, P. Wangkulangkul, J. Meesane, P. Raungrut and P. Puttawibul, *Mater. Sci. Eng. C*, 2016, **64**, 173–182.
- 281 A. Menssen, T. Häupl, M. Sittlinger, B. Delorme, P. Charbord and J. Ringe, *BMC Genomics*, 2011, **12**, 461.
- 282 T. Nakamura, S. Shiojima, Y. Hirai, T. Iwama, N. Tsuruzoe, A. Hirasawa, S. Katsuma and G. Tsujimoto, *Biochem. Biophys. Res. Commun.*, 2003, **303**, 306–312.
- 283 H.-K. Lee, B.-H. Lee, S.-A. Park and C.-W. Kim, *Proteomics*, 2006, **6**, 1223–9.
- 284 M. Scheideler, C. Elabd, L.-E. Zaragosi, C. Chiellini, H. Hackl, F. Sanchez-Cabo, S. Yadav, K. Duszka, G. Friedl, C. Papak, A. Prokesch, R. Windhager, G. Ailhaud, C. Dani, E.-Z. Amri and Z. Trajanoski, *BMC Genomics*, 2008, **9**, 340.
- 285 H. Hamouda, M. Ullah, M. Berger, M. Sittlinger, R. Tauber, J. Ringe and V. Blanchard, *Stem Cells Dev.*, 2013, **22**, 3100–13.
- 286 S. James, J. Fox, F. Afsari, J. Lee, S. Clough, C. Knight, J. Ashmore, P. Ashton, O. Preham, M. Hoogduijn, R. D. A. R. Ponzoni, Y. Hancock, M. Coles and P. Genever,

- Stem Cell Reports*, 2015, **4**, 1004–1015.
- 287 K. M. Wilson, J. E. Thomas-Oates, P. G. Genever and D. Ungar, *Front. Cell Dev. Biol.*, 2016, **4**, 52.
- 288 A. Paszkiewicz-Gadek, H. Porowska, D. Lemancewicz, S. Wolczynski and A. Gindzienski, *Int. J. Mol. Med.*, 2006, **17**, 669–674.
- 289 S. F. Kuan, J. C. Byrd, C. Basbaum and Y. S. Kim, *J. Biol. Chem.*, 1989, **264**, 19271–7.
- 290 P. Delannoy, I. Kim, N. Emery, C. De Bolos, A. Verbert, P. Degand and G. Huet, *Glycoconj. J.*, 1996, **13**, 717–726.
- 291 N. Dilulio and V. P. Bhavanandan, *Glycobiology*, 1995, **5**, 195–199.
- 292 E. Reynders, F. Foulquier, E. L. Teles, D. Quelhas, W. Morelle, C. Rabouille, W. Annaert and G. Matthijs, *Hum. Mol. Genet.*, 2009, **18**, 3244–3256.
- 293 S. Wopereis, S. Grunewald, E. Morava, J. M. Penzien, P. Briones, M. T. Garcia-Silva, P. N. M. Demacker, K. M. L. C. Huijben and R. A. Wevers, *Clin. Chem.*, 2003, **49**, 1839–1845.
- 294 S. Lewis, *Am. J. Physiol. Physiol.*, 2000, **278**, 867–874.
- 295 I. Kątnik-Prastowska, J. Lis and A. Matejuk, *Glycoconj. J.*, 2014, **31**, 623–36.
- 296 B. F. Martin, *J. Anat.*, 1972, **112**, 433–55.
- 297 R. M. Hicks, *Biol. Rev.*, 1975, **50**, 215–246.
- 298 G. Apodaca, *Traffic*, 2004, **5**, 117–128.
- 299 C. L. Varley, M. A. E. Garthwaite, W. Cross, J. Hinley, L. K. Trejdosiewicz and J. Southgate, *J. Cell. Physiol.*, 2006, **208**, 407–417.
- 300 J. Taganna, A. R. de Boer, M. Wuhrer and J. Bouckaert, *Biochem. Soc. Trans.*, 2011, **39**, 349–354.
- 301 R. E. Hurst, *World J. Urol.*, 1994, **12**, 3–10.
- 302 S. Stojnev, A. Ristic-petrovic, L. J. Velickovic, M. Krstic and D. Bogdanovic, 2014, **7**, 4945–4958.
- 303 R. M. Hicks, *J. Cell Biol.*, 1965, **26**, 25–48.
- 304 B. Kachar, F. Liang, U. Lins, M. Ding, X. R. Wu, D. Stoffler, U. Aebi and T. T. Sun, *J. Mol. Biol.*, 1999, **285**, 595–608.
- 305 B. Rosenburg, L. Van Camp, J. E. Trosko and V. H. Mansour, *Nature*, 1969, **222**, 385–386.
- 306 J. Vergara, W. Longley and J. D. Robertson, *J Mol Biol*, 1969, **46**, 593–596.
- 307 T. Walz, M. Häner, X. R. Wu, C. Henn, a Engel, T. T. Sun and U. Aebi, *J. Mol. Biol.*, 1995, **248**, 887–900.

- 308 J. Yu, J. H. Lin, X. R. Wu and T. T. Sun, *J. Cell Biol.*, 1994, **125**, 171–182.
- 309 X.-R. Wu and T.-T. Sun, *J. Cell Sci.*, 1993, **106**, 31–43.
- 310 J. Yu, M. Manabe, X. R. Wu, C. Xu, B. Surya and T. T. Sun, *J. Cell Biol.*, 1990, **111**, 1207–1216.
- 311 X. R. Wu in *sod.*, *J. Biol. Chem.*, 1990, **265**, 19170–19179.
- 312 X. R. Wu, J. H. Lin, T. Walz, M. Häner, J. Yu, U. Aebi and T. T. Sun, *J. Biol. Chem.*, 1994, **269**, 13716–13724.
- 313 P. Hu, F. M. Deng, F. X. Liang, C. M. Hu, A. B. Auerbach, E. Shapiro, X. R. Wu, B. Kachar and T. T. Sun, *J. Cell Biol.*, 2000, **151**, 961–971.
- 314 M. E. Kreft, S. Hudoklin, K. Jezernik and R. Romih, *Protoplasma*, 2010, **246**, 3–14.
- 315 C. H. Birder, L. A., Ruggieri, M., Takeda, M., Van Koeveringe, G., Veltkamp, S., Korstanje, C., Parsons, B., Fry, *Neurourol. Urodyn.*, 2012, **31**, 293–299.
- 316 J. N'Dow, N. Jordan, C. Robson, D. Neal and J. Pearson, *J. Urol.*, 2005, **173**, 2025–2031.
- 317 S. Kaur, N. Momi, S. Chakraborty, D. G. Wagner, A. J. Horn, S. M. Lele, D. Theodorescu and S. K. Batra, *PLoS One*, 2014, **9**, e92742.
- 318 S. M. A. McGuckin, MD, Walsh M D, Walsh B G, Hohn W ,Thong P L, Devine R A, Gardiner M L, *Br. J. Urol.*, 1994, **73**, 256–262.
- 319 A. Halim, J. Nilsson, U. Rüetschi, C. Hesse and G. Larson, *Mol. Cell. Proteomics*, 2012, **11**, M111.013649.
- 320 B. Xie, G. Zhou, S.-Y. Chan, E. Shapiro, X.-P. Kong, X.-R. Wu, T.-T. Sun and C. E. Costello, *J. Biol. Chem.*, 2006, **281**, 14644–14653.
- 321 T. M. Hooton, R. Besser, B. Foxman, T. R. Fritsche and L. E. Nicolle, *Clin. Infect. Dis.*, 2004, **39**, 75–80.
- 322 R. Kucheria, P. Dasgupta, S. H. Sacks, M. S. Khan and N. S. Sheerin, *Postgrad. Med. J.*, 2005, **81**, 83–6.
- 323 X. R. Wu, T. T. Sun and J. J. Medina, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 9630–9635.
- 324 E. D. Lobban, B. a Smith, G. D. Hall, P. Harnden, P. Roberts, P. J. Selby, L. K. Trejdosiewicz and J. Southgate, *Am. J. Pathol.*, 1998, **153**, 1957–1967.
- 325 C. Varley, G. Hill, S. Pellegrin, N. J. Shaw, P. J. Selby, L. K. Trejdosiewicz and J. Southgate, *Exp. Cell Res.*, 2005, **306**, 216–229.
- 326 M. A. Knowles, A. Finesilver, A. E. Harvey, R. J. Berry and R. M. Hicks, *Cancer Res*, 1983, **43**, 374–385.

- 327 S. C. Baker, S. Shabir and J. Southgate, *Mol. Pharm.*, 2014, **11**, 1964–1970.
- 328 W. R. Cross, I. Eardley, H. J. Leese and J. Southgate, *Am. J. Physiol. Renal Physiol.*, 2005, **289**, F459–F468.
- 329 K. Benson, S. Cramer and H.-J. Galla, *Fluids Barriers CNS*, 2013, **10**, 5.
- 330 E. Fromter and J. Diamond, *Nat. new Biol.*, 1972, **235**, 9–13.
- 331 P. C. Rubenwolf, N. T. Georgopoulos, L. A. Kirkwood, S. C. Baker and J. Southgate, *PLoS One*, 2012, **7**.
- 332 B. Srinivasan, A. R. Kolli, M. B. Esch, H. E. Abaci, M. L. Shuler and J. J. Hickman, *J. Lab. Autom.*, 2015, **20**, 107–126.
- 333 R. Shahani, C. Streutker, B. Dickson and R. J. Stewart, *Urology*, 2007, **69**, 810–812.
- 334 S. C. Baker, J. Stahlschmidt, J. Oxley, J. Hinley, I. Eardley, F. Marsh, D. Gillatt, S. Fulford and J. Southgate, *Acta Neuropathol. Commun.*, 2013, **1**, 64.
- 335 S. C. Baker, S. Shabir, N. T. Georgopoulos and J. Southgate, *Am. J. Pathol.*, 2016, **186**, 1267–1277.
- 336 Y. Miura and T. Endo, *Biochim. Biophys. Acta - Gen. Subj.*, 2015, **1860**, 1608–1614.