The importance of protein glycosylation: A Glycoscope into the reproductive process

Shathili Abdulrahman Mansour

Master of Research (Macquarie University)

Bachelor of Science, Molecular Biology and Genetics (University of Sydney)

This thesis is submitted for the degree of

Doctor of Philosophy

Department of Molecular Sciences

Faculty of Science and Engineering

Macquarie University, Sydney, Australia

2018

Contents

Dedicated to Huda my love,

and to my parents

Abstract

The aim of this thesis is to enhance our understanding of the role of glycosylation in the reproductive process in both male and female tracts from both a glycomics and glycoproteomics perspective. Within the scope of this thesis, protein macro-glycosylation (at the cellular/tissue level) was investigated by analysis of aberrant glycosylation in STZ- and high glucosamine - induced diabetic ovarian tissues of mice. The two diabetic models shared a relative decrease in protein N-linked oligomannose, and a decrease in sialylated, structures compared to non-diabetic mice. We are the first to report changes occurring in the glycan biosynthesis pathway in diabetes of any reproductive organ. The same level of analysis (macro-glycosylation) was investigated as a means of differentiating active sperm from inactive sperm. Active sperm were generated via two isolation protocols, namely Percoll density gradient centrifugation and "swim-up" collection and compared with inactive sperm. The most significant result was seen in the large individual biological variation within each donor's sperm populations. On the other hand, the micro-glycosylation (at the single glycoprotein level) focused on two functionally relevant reproductive proteins, namely: endometrial MUC1 and seminal Cluster of Differentiation 52 (CD52). MUC1 on the surface of mouse oviduct and uterine tissues exhibited no significant difference in O-glycan structures at the time of window of implantation (WOI) or in diabetes, although the carbohydrate/protein abundance increased on MUC1 on the uterine lining in the WOI. CD52 has a proposed immune-suppressive function in semen, but only lymphocyte CD52 has been fully characterized from the point of view of activity. Lymphocyte CD52 initiates immunosuppression by binding to the sialic acid-binding immunoglobulin-like lectin-10 (Siglec-10) receptor on activated T cells. As a prelude to understanding the role of CD52 glycosylation in reproduction we comprehensively defined the glycan structures on lymphocyte CD52. Sialylation was essential for the immunosuppression and enabled the identification of the most immune suppressive CD52 glycoforms. Overall, this thesis provides support for the inherent contribution of protein glycosylation, on a macro and micro level, in the process of reproduction in both male and female.

Statement

I hereby certify that this thesis is a presentation of my original research work carried out as part of the Doctorate of Philosophy program at Macquarie University. This thesis has not previously been submitted for a degree nor has it been submitted as part of the requirement for a degree to any other university or institution other than Macquarie University. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of personal assistance or advice. I consent to a copy of this thesis being available in the Macquarie University library for relevant consultation, loan and photocopying forthwith.

Shathili Abdulrahman Mansour (Shazly)

November 2018

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Prof Nicolle Packer, for her continuous support throughout the years, patience, immense knowledge and guidance, without whom this project would not have been possible. You are a very generous supervisor and I could not imagine a better mentor and supervisor for my PhD study. I would also like to thank Dr Robyn Peterson, for help with editing various parts of this thesis.

I thank my fellow mates in Nicki's analytical glycogroup. I would like to specify the following names: Dr Arun, Dr Mathew, Dr Edward, Dr Chris, and many others. Special thanks to my amazing friends Dr Jodie and Dr Zeynep; you made my journey enjoyable and always picked up the phone.

This work would not have been possible without the financial support from Centre of Excellence for Nanoscale Biophotonics and the scholarships provided by IMQRES and Saudi Arabian Cultural Mission (SACM).

I was also blessed with some great collaborators: Prof. Leonard and Dr. Esther from WEHI (Melbourne), Associate Prof. Mark Baker from University of Newcastle, and Prof. Jeremy Thompson and Dr. Hannah Brown from University of Adelaide. I thank you all for your contribution to my thesis.

I would also like to thank all the friends I have made during this amazing journey, with special thanks to David, Steph and Flora. We have shared countless coffee, lunch and gym sessions, and above all stimulative debates.

My gratitude is extended to my love Huda, the only one I relied on emotionally. You are the only one that witness my stressful moments when everyone else thought I was doing fine. I am glad to have you in my life.

Last but not least, massive thanks to my Mum and Dad; checking up on me every day has kept me happy throughout my studies from high school to PhD. I love you my dearest. Also, my gratitude is extended to my sisters and brothers and everyone that stood by my side during my amazing journey.

Published manuscripts from this thesis

1. A. V. Everest-Dass, E. S. X. Moh, C. Ashwood, **A. M. Shathili,** N. H. Packer. (2017). Human disease glycomics: recent advances in protein glycosylation analysis technology (part 1). *Expert Review of Proteomics*.15 (2):165-182.

2. A. V. Everest-Dass, E. S .X. Moh, C. Ashwood, **A. M. Shathili,** N. H. Packer. (2017). Human disease glycomics: recent advances in protein glycosylation analysis technology (part 2). *Expert Review of Proteomics*. 15(4): 341-352.

3. **A. M. Shathili**, H. M. Brown, A. V. Everest-Dass, T. C. Tan, L. M. Parker, J. G. Thompson, N. H. Packer. (2018). The effect of streptozotocin-induced hyperglycemia on *N*-and *O*-linked protein glycosylation in mouse ovary. *Glycobiology.* 1; 28(11):832-840.

Manuscripts in preparation from this thesis

4. D. Handler, F. Cheng, **A. M. Shathili**, S. Emery, G. Iniga, D. Pascovici, P. Haynes. A software package for the production of high stringency proteomics data (PeptideWitch). *Journal of Proteomics.* (**Manuscript in-preparation**)

5. **A. M. Shathili**, M. R. Baker, N. H. Packer Analysis of the N-glycome of Human spermatozoa demonstrates that large differences exists between men. F*ertility and Sterility*. (**Manuscript in-preparation**).

6. **A. M. Shathili**, E. Bandala Sanchez, E.Goddard-Borger, A. John, M. Thaysen-Andersen, L. C. Harrison, N. H. Packer. Specific Sialoforms Required for the Immunosuppressive Activity of Human Soluble CD52. *Frontier in Immunology*. ***Under publication embargo for the protection of intellectual property (IP).***

7. **A.M. Shathili,** D. Handler, N.H. Packer. Glyco-scope into the endometrium during the window of implantation: diabetes as a case study. *Trends in glycoscience and glycotechnology*. (**Invited review in preparation**).

Conference posters and presentations

Conference posters

1. 2016 Lorne Proteomics Symposium, Melbourne, Australia.” Insight into protein glycosylation in human capacitated sperm using PGC-LC-MS/MS”

2. 2016 Human proteome project international conference, Taipei, Taiwan. "Role of CD52 glycosylation in immunosuppression"

3. 2016 Centre of Excellence for nano-scale bio-photonics, Adelaide, Australia.” Insight into protein glycosylation in diabetic and high glucosamine mice ovaries using PGC-LC-MS/MS”

4. 2016 Bio-Focus Research Conference, Sydney, Australia. "The blastocyst experience through the oviduct and uterus of diabetic mice: changes in membrane proteomes, adhesins and MUC1 glycosylation”

5. 2017 Bio-Network "Killing it in Science" Research Symposium, Sydney Australia. "Human CD52 initiates its immunosuppressive activity via specific sialoforms”.

6. 2018 Lorne Proteomics Symposium, Melbourne, Australia. "Human CD52 initiates its immunosuppressive activity via specific sialoforms”.

Conference oral presentations

1. 2017 CNBP Conference, Melbourne, Australia. The blastocyst experience through the oviduct and uterus of diabetic mice: changes in membrane proteomes, adhesins and MUC1 glycosylation”.

2. 2018 Lorne Proteomics Symposium, Melbourne, Australia. "Human CD52 initiates its immunosuppressive activity via specific sialoforms”.

3. 2018 The 3rd Saudi scientific symposium (winner), Sydney, Australia. "Profiting from the sugar coating our cells: Immunology drugs as an example”.

4. 2018 Combio international conference, Sydney, Australia. "Human CD52 initiates its immunosuppressive activity via specific sialoforms”.

Abbreviations

ART: Assisted reproductive technology
Ag-PAGE: Agarose-polyacrylamide gel electrophoresis
Asn: Asparagine
AUC: Area under the curve
BCA: Bicinchoninic acid
C1q: Complement component 1q
CAM: Cell adhesion molecule
CD52: Cluster of Differentiation 52
CE: Capillary electrophoresis
CID: Collision-induced dissociation
CsCl: Cesium chloride
Da: Dalton
DAMP: Damage-associated molecular pattern
DDF: Detergent extraction fractionation
DTT: Dithiothreitol
EDTA: Ethylene-Diamine-Tetra-Acetic acid
EGFR: Epidermal growth factor receptor
ER: Endoplasmic reticulum
ESI: Electron spray ionization
ETD: Electron transfer dissociation
EThcD: Electron transfer/higher-energy collision dissociation
FA: Formic acid
FBS: Fetal bovine serum
FITC-PSA: Fluorescein isothiocyanate conjugate-*Pisum sativum* agglutinin
Fuc: Fucose
FUT: Fucosyltransferase
Gal: Galactose
GalNAc: *N*-acetylgalactosamine
GFAT: Glucosamine-fructose-6-phosphate amidotransferase enzyme
Glc: Glucose
GlcNAc: N-acetylglucosamine
GLUT2: Glucose transporter 2
HBP: Hexosamine biosynthesis pathway
Hex: Hexose
HMGB1: High-mobility group box 1
HPLC: High-performance liquid chromatography
IAA: Iodoacetamide
IFN: Interferon
IVF: *In vitro* fertilization
LC: Liquid chromatography
LDS: lithium dodecyl sulfate
LIF: Leukemia inhibitory factor
LNF-1: Lacto-*N*-fucopentaose

LTQ: linear ion trap quadrupole

MAL-I: *Maackia amurensis-1* lectin

MALDI: Matrix assisted laser desorption ionization

Man: Mannose

MAN: Mannosidase

MGAT-3: Mannosyl-glycoprotein-acetylglucosaminyltransferase

MS: Mass spectrometry

MUC: Mucin-type

MW: Molecular weight

m/z: Mass over charge

NaCl: Sodium chloride

NEU: Neuraminidase

NeuAc: *N*-acetylneuraminic acid / sialic acid

NeuGc: *N*-glycolylneuraminic acid/ sialic acid

P. aeruginosa: Pseudomonas aeruginosa

PAEP: Progestagen-associated endometrial protein

PBS: Phosphate buffered saline

PCR: Polymerase chain reaction

PDGF: Platelet derived growth factor

PG: Prostaglandin

PGC: Porous graphitized carbon

PGM: Porcine gastric mucin

PolySia: Polysialic acid

PBMC: Peripheral blood mononuclear cells

PTM: Post translational modification

PVDF: Polyvinylidene fluoride

PVP: Polyvinylpyrrolidone

Q-TOF: Quadropole-Time-of-flight

RSD: Relative standard deviation

SCX: Strong cation exchange

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

Ser: Serine

Siglec: sialic acid binding immunoglobulin type lectins

SMME: Supported Matrix Membrane Electrophoresis

SpC: spectral counts

STZ: Streptozotocin

TCA: Trichloroacetic acid

Thr: Threonine

Tris-HCl: Trisaminomethane-hydrocholride

WEHI: Walter and Elisa hall institute of medical research

WGA: Wheat germ agglutinin

WOI: Window of implantation

We stand by couples traumatized by infertility.

We hope these quotes help you in those days

that are tougher than others:

"We must accept finite disappointment, but

never lose infinite hope."

Dr Martin Luther King.Jr

"Miracles only happen to those who never

give up." - Ivankov Emporio (OnePiece)

Chapter 1

Introduction (Glycoscope into the male and female reproductive process)
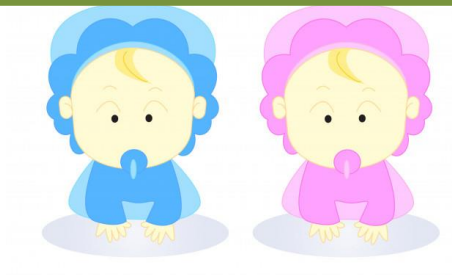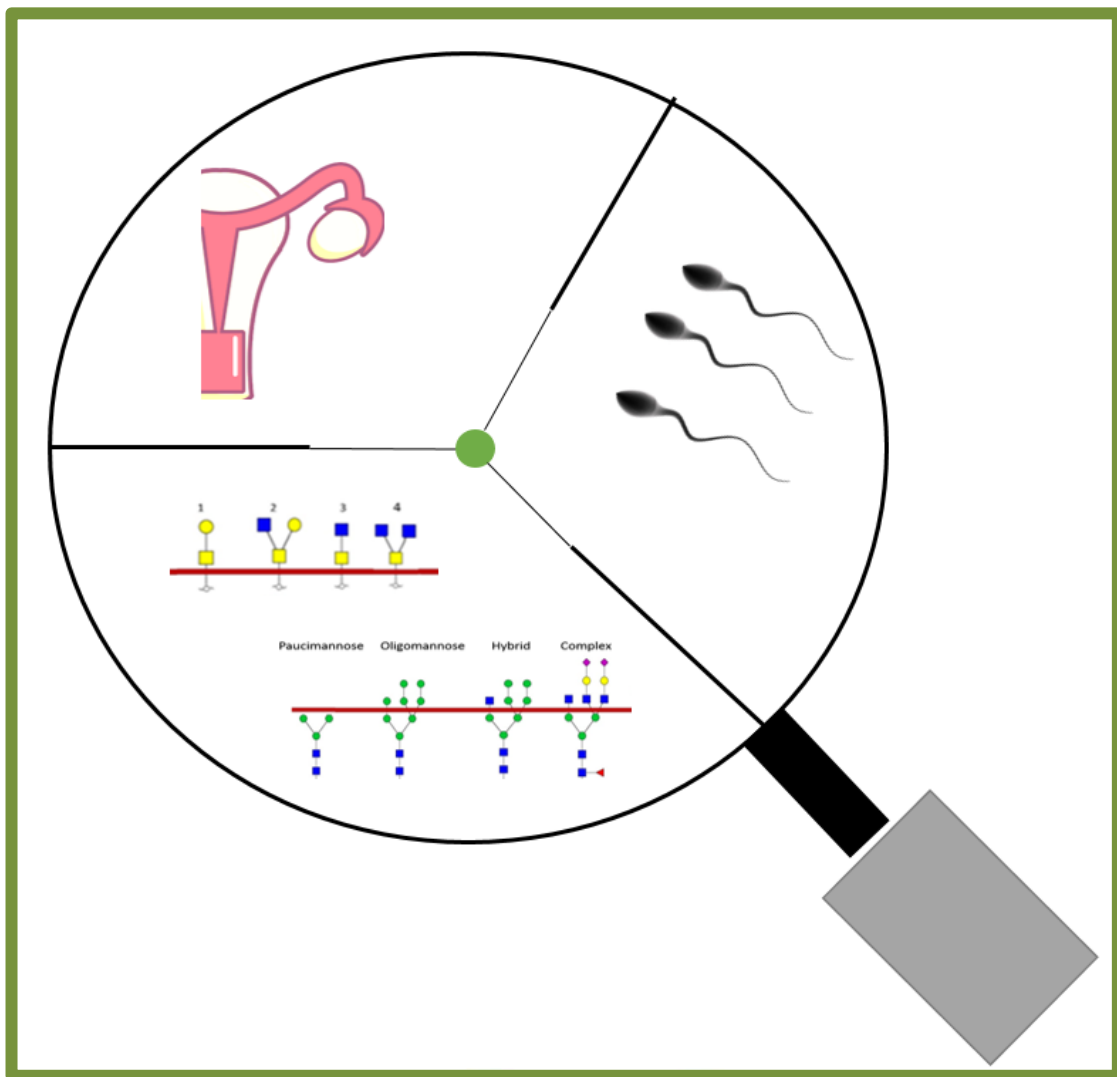
1.1 Protein glycosylation

The contribution of post-translational modifications (PTMs) of proteins to the overall functional complexity of organisms is well recognised [1]. Glycosylation is a predominant protein PTM in which carbohydrates (glycans) are added to the protein backbone. The presence of glycans influences the overall properties of a protein including mass, hydrophobicity, solubility and structural properties [2]. These glycans play important roles in many cellular processes including intercellular and intracellular interactions, and serve as markers for the cell physiological status [3]. Many monosaccharide building blocks are utilised for protein glycosylation including fucose (Fuc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), glucose (Glc), galactose (Gal), mannose (Man), *N*-acetylneuraminic acid (NeuAc) and *N*-glycolylneuraminic acid (NeuGc), the latter found predominantly in non-human mammals [4]. The elongation of glycans with the addition of monosaccharides is facilitated by glycosyltransferases. Glycosidases act by removing monosaccharaides to initiate further glycosyltransferase activity or to degrade glycans. The degree of glycan processing is dependent on numerous factors including primary amino acid structure and three-dimensional protein conformation [5]. There are several types of protein glycosylation seen in mammals and this thesis will focus on the two main types: *N*-linked and *O*-GalNac linked.

1.1.1 N-glycosylation

The process of *N*-glycosylation starts at the endoplasmic reticulum (ER) with the addition of a glycan to the amide nitrogen of the asparagine (Asn) in the Asn-X-serine (Ser)/ threonine (Thr) motif of a protein. The structure GlcNAc2Man3 (Figure 1.1) forms the basic core unit in the development of all *N*-linked glycans, and after removal or addition of monosaccharides in the Golgi, the structures then can be subdivided into oligomannose, hybrid, and complex glycan classes. There is a fourth class—namely paucimannose—that is widely expressed in invertebrates and plants, but has only recently been recognised in the human system [6]. Paucimannose

structures contain a truncated chitobiose *N*-glycan core displaying a composition less than or equal to GlcNAc2Man3 with or without core fucosylation (Figure 1.1) [7].

1.1.2 O-glycosylation

O-linked glycosylation is named according to the sugar attachment to the oxygen atom of Ser or Thr in a protein. Subtypes of *O*-linked glycosylation include α-linked *O*-Fuc and *O*-Man, β-linked *O*-GlcNAc, α- or β-linked *O*-Gal and *O*-Glc and mucin type (α-linked *O*-GalNAc) [8]. The orientation of the glycosidic bond between the carbon rings (up or down) defines their β or α linkage, respectively. In this thesis, the major mucin type α-*O*-linked glycosylation is investigated exclusively and is referred to as *O*-glycosylation.

Mucin type *O*-linked glycosylation is synthesised in the Golgi bodies where GalNAc is attached to the hydroxyl of Ser/Thr residues on the folded protein and is elongated one monosaccharide residue at a time [9]. *O*-linked glycans on mucins have mainly core structures 1, 2 ,3 and 4, while other core structures are less described (Figure 1.1). Mucins are proteins that make up the main constituent of the mucus layer which covers the surface of various cavities and internal organs, giving the mucus layer its sticky, viscous and lubricating properties. Glycans are responsible for approximately 40-80% of the molecular weight of mucins. The majority of the glycosylation on mucins occurs within the variable number tandem repeat (VNTR) region of the mucin protein and is predominantly comprised of *O*-linked glycans [10].

In humans, twenty different types of mucins are expressed and have been identified. Different tissues express different levels of different mucins resulting in altered glycosylation profiles across different tissues [11], and specific substructures on mucin glycans such as Lewis type antigens and sulphation further increases the diversity of the glycan epitopes displayed by mucins [9, 12]. The high amounts of *O*-glycans present on the mucins display a large variety of carbohydrate epitopes that can act as binding ligands or energy source for bacteria and viruses [13].
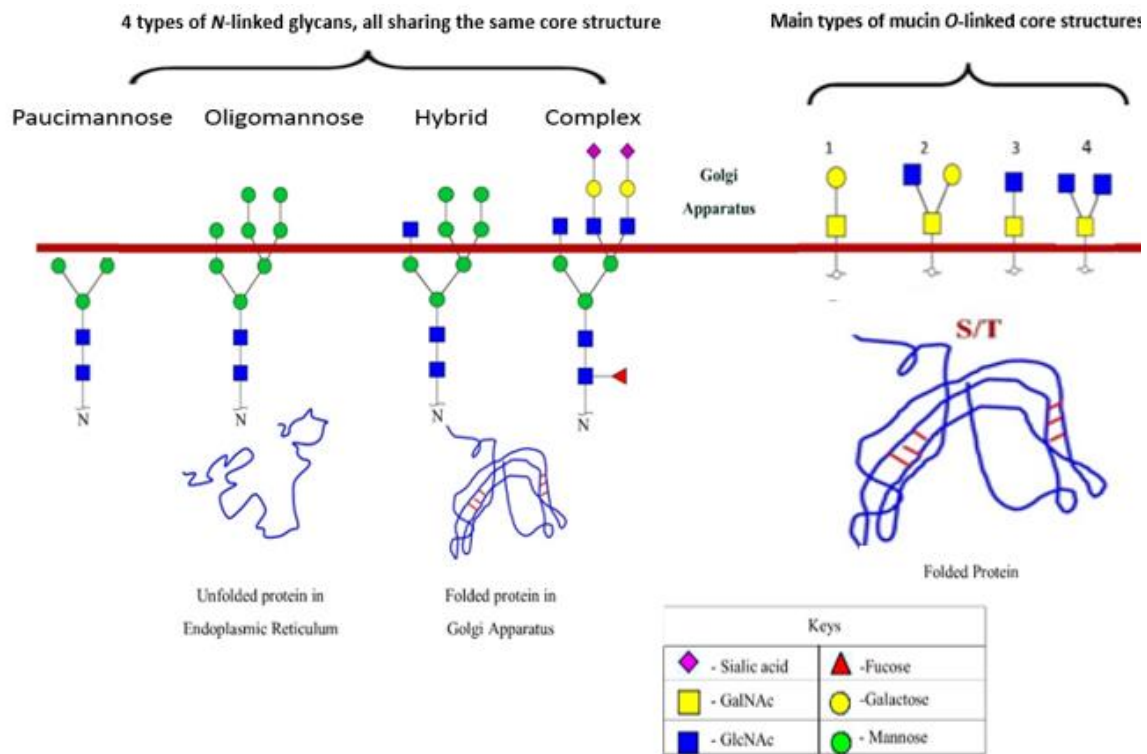
Figure 1.1: Four types of *N*-linked glycan classes are synthesised in eukaryotes (paucimannose, oligomannose, hybrid and complex). *N*-linked glycans are added to asparagine in the endoplasmic reticulum (ER) and undergo further modifications in the Golgi apparatus. *O*-linked glycans are added to serine (Ser) or threonine (Thr) residues in the Golgi apparatus. Four *O*-linked core structures that are typically found on the highly glycosylated mucin proteins are shown. GalNAc: *N*-acetylgalactosamine; GlcNAc: *N*-acetylglucosamine.

1.2 Biological roles of protein glycosylation from a reproductive perspective

There is a growing interest in the role of glycosylation in protein function, and glycosylation is being studied for its effects at the overall cell/tissue (macro) level as well as at the single protein (micro) level. At the macro-cellular level, the global glycosylation is seen to impact the physiological and interactive status of the cells [14, 15]. On the opposite end of the spectrum, aberrant glycosylation at the micro, single-protein level can influence protein structure, function, size, solubility, hydrophilicity and stability [2, 16].

In the case of a successful reproductive process, the importance of glycosylation has been implicated in several stages that mark the long journey of sperm to implanted blastocyst. For example, generally similar macro-cellular glycosylation can be seen at both the blastocyst and

uterine implantation sites of two animal species that can interbreed, namely the horse and donkey; this contrasts with the camel, which cannot interbreed with either horse or donkey, and shows a markedly different glycosylation profile of maternal (uterine implantation site) and foetal tissues (blastocyst), in particular in the abundance of Fuc and sialic acids (NeuAc or NeuGc) [17, 18]. Furthermore, glycosylation has been implicated in protection against vaginal infection, for example, where α1-2 Fuc epitopes on cervical mucin competitively bind to the yeast pathogen Candida, thereby reducing binding and infection of the vagina [19].

A role for glycosylation at the micro-single protein level in the reproductive tract is demonstrated in the protein glycodelin. Glycodelin, or PAEP (Progestagen-associated endometrial protein), is a glycoprotein with four distinct glycosylation expression profiles (glycoforms), each found at different locations in the reproductive system, namely in the amniotic fluid, follicular fluid, cumulus cells and seminal fluid. These glycoforms have distinct roles in sustaining a suitable pregnancy environment in the uterus including cell proliferation, differentiation, adhesion and motility [20]. For example, Glycodelin-S (seminal fluid) has roles in inhibiting premature capacitation through binding to sperm head. In the female reproductive tract, several glycoforms of glycodelin act to modulate sperm function such as Glycodelin-F (Follicular fluid), Glycodelin-A (Amniotic fluid) and Glycodelin-C (Cumulus cells). Glycodelin-F prevents premature acrosome reaction, whereas Glycodelin-A inhibits spermatozoa binding to zona pellucida, and Glycodelin-C stimulates the latter binding [21].

1.2.1 The female reproductive tract

In all mammalian systems, the ovulated oocyte must be fertilised by sperm inside the oviduct during the fertile window of the menstrual cycle. The blastocyst, an early form of embryo, is formed in the oviduct wall and then rolled towards the uterus. The blastocyst implants in uterine epithelial cells during the window of implantation (WOI), the latter is characterised by the presence of a receptive uterus. This implantation is an essential process involving synchronised changes on

the surface of the embryo as well as the uterine wall, resulting in significant molecular complexity [22].

1.2.1.1 Glyco-macroscope into the female reproductive tract

A glyco-macroscopic view encompasses the overall glycome of certain cells or tissues. In the context of the embryo implantation site, previous research has implicated the role of resident surface membrane glycoproteins in modulating the reception of blastocyst by the uterus. Ricketts and colleagues found changes in the surface labelling pattern of radio-iodinated surface proteins of the uterus at the window of implantation in rabbits [23]. Anderson and colleagues monitored the time of implantation using a lectin that binds to sialic acid (Wheat germ agglutinin: WGA) and found a decrease in sialic acid prior to blastocyst implantation that correlated with a change in morphology of the glycocalyx (a medium of glycolipids and glycoproteins) [24]. The decrease in charged components of the glycocalyx at the uterine surface is also seen in other attachment related surfaces such as sperm binding to egg [25, 26]. Sialic acid has also been implicated in immune cell activation by delivering inhibitory signals through binding to sialic acid binding immunoglobulin type lectins (SIGLEC) [27], and also plays several roles in white blood cell (leukocyte) trafficking, and as a ligand for microbes, all of these process have been documented with significance to the female reproductive tract. Structural proteomics using freeze-fracture techniques [28], has shown that integral membrane particles (pinopods) were seen to increase dramatically in the plasma membrane of endometrial epithelial cells during early pregnancy in the rat; these proteins were associated with a glycocalyx rod shaped component projecting into the lumen, possibly acting as a focal adhesive site for the blastocyst. Others have shown pinopods to be present throughout the luteal phase of the menstrual cycle [29].

1.2.1.2 Glyco-microscope into the female reproductive tract

Whilst a glyco-macroscopic view is the result of many glycoproteins, a glyco-microscopic view is manifested via investigating each glycoprotein individually. A glyco-microscopic view of the

female reproductive tract highlights glycosylation of individual proteins involved in uterine receptivity [22, 30-32].

The blastocyst's temporal and spatial management is due to a series of orchestrated events among several secreted glycoproteins including cytokines, chemokines, prostaglandins, growth factors and interferons, which provide the window of implantation [33]; and other glycoproteins that act as ligands, namely heparin sulphate, integrins, selectins, cadherins and, finally, mucins for the spatial tuning [34-37]. Moreover, the timing window seems to be controlled in part by PGs (prostaglandins) [33], whereas selectin-glycoprotein interactions direct spatial control by rolling the blastocyst (similar to leukocytes) over the oviduct, and ensure the blastocyst conforms to a polarized shape [36]. Chemokines and cytokines attract the blastocyst into the pinopods and both integrins and cadherins ensure temporal adhesiveness between endometrium and embryo [33]. The timing, function and clinical significance of pinopods has been strongly questioned in literature and the necessity of these structures for blastocyst implantation is not yet firmly established [29]. Above all, the mucin proteins, in particular MUC1, heavily decorated with *O*-glycans displayed as active ligands, appear pivotal to ensuring the blastocyst is implanted into the right location. MUC1 is the most abundant protein in the endometrium and has been the most studied factor in uterine receptivity, as described in more detail below.

1.2.1.3 The mucins of the female reproductive tract

Mucins have been seen in all reproductive organs, except the ovary, and the amount and phenotype of mucins change drastically during the menstrual cycle to facilitate sperm migration in the cervix [38]. There are many different MUC proteins expressed in the reproductive tract. By utilising Northern blot analysis, *in situ* hybridisation and specific immunohistochemistry to analyse biopsy samples of human reproductive tract epithelia including endocervix, ectocervix, vagina, endometrium and fallopian tube, different mucin gene and protein localisation was assigned for each tissue sample [39]. In this study, endocervical mucins were found to express the genes for *Muc1, Muc2, Muc5c, Muc5b* and *Muc6*. The ectocervical and vaginal tissues contained MUC1 and

MUC4. Endometrial mucins were found to be mainly MUC1 type and MUC6 at low level, and the fallopian tube contained MUC1 only. MUC6 immuno-reactivity expression was seen in scattered basalis glands from pre and postmenopausal women in the endometrium [39].

The endometrium (uterus MUC1 layer) microenvironments during the receptive phase are highly complex and constantly changing as implantation progresses [40]. MUC1 is a highly important factor for correct implantation where a very thick endometrial mucus is usually a sign of a non-conception cycle [37]. In mouse endometrium, MUC1 values increase in response to estrogen and down-regulate with progesterone release [41, 42], while mRNA levels (Northern blot) and protein levels (immune-fluorescence) of MUC1 in human endometrium in contrast have been shown to increase with ovarian progesterone release [37]. Knocking out *Muc1* and the removal of MUC1 (selective enzymatic removal of mucins from the apical surface of uterine epithelium) led to an increase of blastocyst binding in mice [22, 43, 44].

1.2.2 The male reproductive tract

Mammalian sperm contains a 20-60 nm thick glycocalyx with significant diversity in its glycoprotein make-up, described below [26]. Sperm glycans are dynamic throughout their journey from generation in testis to successful fertilisation of the oocyte. During spermatogenesis in the testis, sperm experience continuous changes to their surface glycome that are essential for the sperm survival in the female reproductive tract. Tecle and colleagues have recently published a comprehensive review summarising the various roles of glycosylation in sperm functionality [26]. This review elaborates that once spermatogenesis is complete, spermatozoa lose the ability to make new glycans but existing glycans are modified and elongated via the activity of glycosidases and glycosyltransferases in the lumen of the epididymis [45, 46]. Also, several glycoconjugates are incorporated into the sperm membrane along the male reproductive tract [47, 48]. These temporal and spatial modifications further diversify the sperm glycocalyx by adding a layer of sperm-coating antigens and other immune-modulatory glycoconjugates which is necessary for the sperm

passing through cervical mucus and uterotubal junction (path from the uterus to oviduct) as well as during capacitation [26, 49, 50]

1.2.2.1 Glyco-macroscope into the sperm

The sperm glycocalyx has several functions in the successful journey of a sperm. It facilitates the penetration of cervical mucus [51], protects the sperm from macrophages and neutrophils in the uterus [52, 53], and during capacitation, which is the physiological change that enables the sperm to fertilise the egg, specific monosaccharides (sialic acid) are shed [54]. During the journey inside the female reproductive tract, the sperm needs protections against humoral and cellular immunity which is partially provided by sperm glycans including sialic acid. This sialylation is then need to be removed during the process of capacitation [55, 56]. In fact, sperm has two sialidases - NEU1 and NEU3 - that have been implicated in capacitation *in vitro*, where their inhibition led to diminished binding of the sperm to the zona pellucida of the egg and infertile patients showed decrease in sialidase abundance by antibody staining [57]. Overall, many questions are still to be addressed on the role of specific glycan structures during sperm capacitation.

1.2.2.2 Glyco-microscope into the sperm

Less is known about the glycosylation of specific glycoproteins found in sperm (micro glycosylation) [26]. Several glycoproteins have been implicated in the successful fate of a sperm. This includes equatorin, hyaluronidase and Basigin, the latter has its *N*-glycans involved in spermatogenesis [58]. Other sperm glycoproteins have important functions inside the female reproductive tract such as glycodelin S, Beta-defensin 126 and Cluster of differentiation 52, all of which contributes to sperm protection against female immunity. Glycodelin S has two main functions, which includes preventing sperm premature capacitation in vagina by inhibiting the initial step required for capacitation (triggering calcium flux). Glycodelin S de-glycosylation abolishes its ability to inhibit capacitation [59]. The second function of glycodelin S is also through its heavy glycosylation that provide the protein with inhibitory effect of both innate and adaptive

22

immunity response [26]. Beta-defensin 126 is another protein with multiple functions in the female reproductive tract including allowing sperm passage through cervical mucus through its sialic acid [51], formation of a sperm reservoir in the oviduct, and sperm immune protection in the uterus [26].

1.2.2.3 Cluster of differentiation (CD52) is suspected to be immunosuppressive of female immunity

The sperm encounters female complement, antibodies and immune cells, such as macrophages and neutrophils [52]. The sialic acid displayed on the surface of sperm are recognised by Sialic acid-binding immunoglobulin-type lectin (Siglecs), which are known to inhibit the immune response [60, 61]. The sperm sialylation is provided by several immune-active integral membrane proteins such as Cluster of Differentiation 52 (CD52), where analysis of *N*-glycans on Male reproductive tract mrtCD52 showed heavily sialylated glycans with *N*-acetyllactosamine repeats (GalGlcNAc) [62, 63]. MrtCD52 is secreted in the epididymis and inserted into the sperm membrane with a speculated role in immunosuppression of female response against sperm during leukocyte-mediated immune reaction [64, 65] Subsequent work showed that the *N*-glycans of mrtCD52 directly bind the complement component 1q (c1q) which activates the classical complement pathway resulting in an innate immune response [66], suggesting that mrtCD52 exhibits immune-protective properties via its *N*-glycans that may play a role in sperm survival during a leukocytic-mediated immune reaction in the vagina of the female.

1.2.3 Addressing challenges to the reproductive process through protein glycosylation

Despite the established relationship between protein glycosylation and a successful reproductive process (discussed above), protein glycosylation presents a great tool for solving outstanding challenges in the reproductive field.

Infertility refers to any disease of the reproductive process that causes disability. There is a significant overall burden in global health caused by infertility that has persisted over the last 2

decades [67]. Previous research on reproduction has enabled the introduction of assisted reproductive technology (ART), which encompasses any techniques used to treat infertility by using both women and male reproductive processes, and is used as pivotal techniques all around the world. For example, the utilisation of *in vitro* fertilisation (IVF) technique has resulted in close to four million babies being born since its introduction in 1987 [68]. However, several issues contribute to the current poor rate of embryo implantation success, around 25% [69], including: the enigmatic fertility status of a sperm and the inadequate uterine receptivity which remains an elusive process responsible for approximately two-thirds of implantation failures [70]. Testing for uterine receptivity is likely to provide the next leap forward for infertility clinics. Such a test would provide clear information on whether inadequate uterine receptivity is the primary cause of unsuccessful IVF [33]. Many techniques have been utilised to treat the failure of implantation and few have proven to be successful including supplements with oral aspirin, Leukemia Inhibitory Factor and prostaglandin E [33].

Research into post-translational modification of proteins such as *O-* and *N*-glycosylation at the window of implantation, along with IVF supporting techniques, has the potential to further the success of future embryo implantation. In fact, some IVF clinics already utilise the importance of glycosylation in their IVF procedures such as endometrial injury (involves mucins - a highly glycosylated protein) and hyaluronic acid (long branched polysaccharide) enriched embryo transfer medium, with both techniques promising improvements in IVF success rate. Scratching the endometrium preceding ovarian stimulation has been shown to improve implantation in women with recurrent implantation failures for 70% of the 2062 participants, but recent meta-analysis questioned the general use of scratching in improving IVF success rates [71]. Hyaluronic acid has been suggested to promote cell-cell and cell-matrix adhesion when added as supplement in embryo culture medium [72]. Regarding the enigmatic fertility status of a sperm, whether the sperm glycome can be used to determine the fertility status of a spermatozoa remains a key question in the reproductive field.

1.2.3.1 Diabetes as a useful model to correlate protein glycosylation with uterine receptivity

One in seven pregnancies are complicated by diabetes worldwide, and Type 1 diabetes results in a ninefold increase in foetal loss [73, 74]. St Vincent's Hospital, Liverpool. UK, published a 10 years outcome of diabetic pregnancies type 1 to show 5.1 times the risk of perinatal mortality than infants in the general population [73]. Diabetes makes a useful model to study the role of glycosylation in the female reproductive tract for two main reasons: diabetes results in high failure rates of pregnancy and diabetes has been repeatedly shown to alter the glycome at both the macro- and micro- scopic levels in several organs/tissues[75]. We have recently published a review on aberrant glycosylation seen in several human diseases including diabetes at both the macro- and microscopic levels [75], Manuscript 1 (Page 42-55): Everest-Dass, A.V., et al., Human disease glycomics: technology advances enabling protein glycosylation analysis–Part 2. *Expert review of proteomics, 2018.* **15**(4): p. 341-352.

Regarding the effect of diabetes on the implantation process, diabetic non-obese mice have significantly higher loss of embryo compared to normal [76]. Immunohistochemical analysis has enabled the observation that diabetic mice are unable to express sufficient levels of cytokine and chemokine (IFN) mRNA needed to have receptive uterus in their uterine tissues [77]. Also, diabetic non-obese mice tend to express retarded pinopods, early matured with clustering of puffy microvillous tips on swelling cells and secretory droplets [78], and are believed to overexpress highly-glycosylated MUC1 at these implantation sites compared to normal -non-obese mice as seen by transmission electron microscopy [76]. Overall, diabetes has been shown to have major effects on the molecular players involved in embryo implantation and the uterine environment.

1.2.3.2 Assessing fertility status of sperm through protein glycosylation

New evidence has discredited the use of sperm parameters such as motility, concentration and morphology for infertility diagnosis [79]. Considerable variation in sperm parameters, such as count, morphology and motility, was detected in sperm longitudinally collected from 15 healthy

men over a 6-month period [80]. Whether there is a correlation between sperm glycocalyx and the quality of sperm is yet to be fully addressed.

1.3 Methods to study glycans

This section provides a brief overview on glycan analysis and some insights into experimental design used to study the roles of protein glycosylation in the reproductive process, which is a major focus of this thesis. We have recently published a review on technology advances enabling analysis of specific glycan structures on proteins [81], Manuscript 2 (Page 56-75): Everest-Dass, A.V., et al., Human disease glycomics: technology advances enabling protein glycosylation analysis–part 1. *Expert review of proteomics*, 2018. 15(2): p. 165-182.

1.3.1 Challenges with glycan analysis

Unlike polypeptides, glycans are not directly encoded in the genome. They are formed via successive activity of transferases and glycosidases, the former's activity is dependent on the pool of available nucleotide sugar donors [82]. Due to the wide combinatorial potential of the component monosaccharides, glycan analysis faces several challenges due to the complexity of glycans. This complexity is because of the large variety of glycan structures that can exist [83], variations in the occupancy and relative distribution of the glycans on potential glycosylation sites on a protein [84], and the fact that glycosylation is not template driven but a result of the concerted effort of cell-specific biosynthesis [85].

1.3.2 Overview on glycan analysis methods

There has been extensive method development in the field of glycomics that has enabled insights into glycan number, structure and function.

1.3.2.1 Lectin assays

Lectins are proteins that bind to carbohydrates and the recent development of multi-lectin-array assays has allowed for the simultaneous determination of several specific glycan motifs in the same biological context [86, 87]. There are many types of lectins that can recognise several glycan epitopes such as specific sialic acid and fucose linkages [88]. However, this technique cannot produce detailed structural information of the glycan structure.

1.3.2.2 Capillary Electrophoresis

Capillary electrophoresis (CE) is another technique used for glycan separation. This technique is able to separate and detect chemically derivatised glycans by UV or fluorescence while allowing for quantitation through comparison with given standard compounds [89]. However, the use of CE is troubled by the necessity to ensure CE compatible mass spectrometry (MS) methods [90]. Liquid chromatography (LC) has also been used as a sensitive and flexible technique by itself and coupled to MS in the field of both glycomics and glycoproteomics.

1.3.2.3 Liquid chromatography (LC)

Separation of biomolecules is routinely performed by high-performance liquid chromatography (HPLC), where glycans can be frequently resolved. HPLC provides a high resolving power and is compatible with other analytical techniques such as mass spectrometry [81, 91]. Two main LC matrices were utilised in this thesis namely Porous graphitized carbon (PGC) and reversed phase (RP) chromatography in order to perform glycomics and glycoproteomics studies respectively. Glycomics is the analysis of the glycans released from conjugates at a system wide level , whereas glycoproteomics focuses on the intact glycopeptides identifying the protein carrier, glycan attachment sites, composition and occupancy [92].

The use of PGC chromatography has advanced the field of glycomics, but our understanding of the way PGC retains glycans is still unclear. PGC has a hydrophobic, ionic and polar nature, all of

which can contribute to this retention process. However, PGC has proven to be superior to other conventional matrices in resolving released and reduced glycans [81]. Furthermore, PGC can be used in conjunction with specific glycosidase sequential treatments to aid in the identification of different glycan linkages by changes in chromatographic retention times [93]. Overall, PGC-ESI-MS/MS in negative mode provides detailed characterisation of the glycans that exist in tissues, thus allowing a comprehensive comparison between treatment and control states.

On the other hand, reversed phase (RP) chromatography is routinely used in glycoproteomics studies for the separation of glycopeptides/glycoproteins based on their protein hydrophobic nature while maintaining a relatively constant retention time for the same peptide backbone [94]. RP columns can be found as C4, C8 and C18, where the number of carbon reflects the degree of hydrophobicity. C18 columns are optimal for peptides and C4 columns are best utilised for the separation of large proteins.

1.3.2.4 Mass spectrometry

Glycan structures have multi-linkage branches that contain important and functionally-relevant information such bisecting, core-fucosylation and sialylation linkages. With mass spectrometry, glycan masses can be identified after releasing the glycan from the peptide backbone, and glycan structures can be verified based on the fragment masses resulting from tandem MS [95].

The MS experimental approach used in this thesis consists of three main components: generation of charged molecules via "ion source", separation of these molecules by "mass analyser" and final detection via "detector". The first ionisation process can be done by two different methods: matrix assisted laser desorption ionisation (MALDI) and electrospray ionisation (ESI), both known as soft ionisation methods for their ability to deal with large molecules keeping them intact for further subsequent analysis [96, 97]. In this thesis, ESI enabled ionisation of molecules from an in-line liquid phase, generating charged droplets which decrease in size due to the presence of inert gas at atmospheric pressure. The charged droplets then enter the mass analyser. Ionising glycans in

negative mode is preferable to positive mode since it provides more fragments, thus enabling better characterisation [95]. However, positive mode ionisation is preferable for glycopeptide analysis since it is efficient and sensitive for molecules with a basic arginine or lysine residue at the C-terminus of the peptide [73].

Mass analysers separate charged molecules in gas phase based on their mass-to-charge ratio (m/z). There are many types of mass analysers and our experiments utilise linear ion trap quadrupole (LTQ), Quadropole-Time-of-flight (Q-TOF) and Orbitrap analysers. These analysers provide the molecular mass information for any given analytes, but subsequent MS/MS fragmentation can elucidate the monosaccharide sequence, linkage position and branching for each specific glycan. This thesis has utilised Collision-induced dissociation (CID) as its main glycan MS/MS fragmentation tool. CID provides glycosidic cleavages between neighbouring residues with information on composition and sequence, as well as cross-ring cleavages with insights on different linkages types and internal fragments [98]. During this fragmentation event, glycans produce glycosidic cleavages as outlined in Domon and Costello (1988) [99]. In MS/MS, the fragment ions from the reducing end of a glycan are labelled A, B and C, where from the non-reducing end are X, Y and Z, as seen in Figure 1, from Manuscript 2 [81]. The information from these fragment ions often enables the full characterisation of the parent ion molecule. To our advantage, ion mode fragmentation provides specific fragment ions that manifest important structural information on the corresponding structure for each given parent ion, such as core-fucosylation [Z ion (m/z350) and Y ion (m/z368)] and presence of bisecting GlcNAc [Y ion (m/z670) and Z ion (m/z508)] [100, 101].

Electron transfer dissociation (ETD), fragmenting the peptide portion of a glycopeptide via electron transfer to positively charged amino acids, while keeping the attached glycan intact, can be used to achieve an insight into glycan site occupancy of a given glycopeptide; ETD is used mainly for site localisation analysis of *O*-glycopeptide [102, 103] .

1.3.3 Our general protocol for glycan analysis

Where described, this thesis has utilised PGC-LC-MS/MS as the method of choice to investigate the two main challenges addressed in this thesis: the aim to correlate protein glycosylation with uterine receptivity in diabetic tissues, and to determine the fertility status of sperm. The use of PGC chromatography enables the separation of isobaric glycan structural isomers, while the coupled electrospray ionization ESI-MS/MS in negative ion mode provides diagnostic fragment ions that allow identification of specific glycan structural features [104, 105]. Hence, this technique promises a more comprehensive analysis of the glycomes tested compared to other aforementioned techniques.

In our chosen technique, intact *N*-linked glycans are released by enzymatic cleavage with PNGase F - a known robust enzyme for the release of *N*-acetylglucosamine linked glycans from the asparagine residue. The *O*-linked glycans are then chemically released from the remaining protein portion via reductive β-elimination [106]. After clean-up by cation exchange chromatography, the sugar alditols are separated by using porous graphitised carbon (PGC) liquid chromatography. The oligosaccharide masses from the spectra are identified and then translated to possible carbohydrate compositions using GlycoMod (ExPasy, http://web.expasy.org/glycomod/). Glycan structures are then deduced from the tandem MS fragmentation ions, often by diagnostic fragment ions [101].

In this thesis, glycan analysis by PGC LC ESI MS/MS was utilised to characterise protein *N*- and *O*-glycosylation of components of both the male and female reproductive tracts. The investigation of female reproductive tract (specifically the mice ovary and endometrium) and human sperm glycosylation can further our understanding of the role of these PTM structures in a successful reproductive process, and hence initiate the understanding of the role of glycosylation in IVF and other relevant research. In addition, the glycosylation of CD52, a potential immunosuppressive sperm protein, was determined in the identification of the most active glycoform of CD52. This knowledge may feed into the discovery pipeline for the prevention of immunosuppression

involved in fertilisation. Overall, this thesis examines some of the significant roles of protein glycosylation in the female/male reproductive tract.

1.4 Aims

The overall aim of this thesis is thus to analyse protein glycosylation in several reproductive organs in both female (mouse) and male (human) tracts while alternating between the micro and macro effect of protein glycosylation. These glycans can serve as biomarkers of diseases as well as functional ligands for several mechanisms within the reproductive process. The thesis also uses diabetes as a model to investigate the role of protein glycosylation in the low embryo implantation rates in women with this disease whilst exploring the role of the glycosylation of an immunoreactive protein and the optimal protein glycosylation needed for active sperm in men.

The thesis aims specifically:

1. To investigate the macro and micro role of glycosylation in the female reproductive tract, using diabetes as a highly relevant model in both glycosylation and reproductions aspects. Specifically,

i. To understand the macro-effect of diabetes on protein glycosylation of mice ovary (Chapter 3).

ii. To compare the mucin glycans (micro-effect) presented in the unreceptive uterus/oviduct of diabetic mice to normal (non-diabetic) (Chapter 4).

2. To investigate the macro and micro role of glycosylation in the male reproductive system. Specifically,

i. To compare the glycome (macro-level) of active and inactive sperm isolated from different techniques (Chapter 5).

ii. To compare recombinant CD52 protein glycoforms (micro-level) and relate their glycosylation to their possible function in the female reproductive tract (Chapter 6).

1.5 References

1. Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications.* Nat Biotechnol, 2003. **21**(3): p. 255-61.
2. Dwek, R.A., *Glycobiology: toward understanding the function of sugars.* Chemical reviews, 1996. **96**(2): p. 683-720.
3. Varki, A. and J.B. Lowe, *Biological roles of glycans, in Essentials of Glycobiology,* A. Varki, et al., Editors. 2017: Cold spring Harbor (NY).
4. Rini, J.M. and J.D. Esko, *Glycosyltransferases and glycan-processing enzymes, IN Essenrial of Glycobiology,* A. Varki, et al, Editors. 2017: Cold Spring Harbor (NY).
5. Parker, B.L., et al., *Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity.* Journal of proteome research, 2013. **12**(12): p. 5791-5800.
6. Thaysen-Andersen, M., et al., *Human Neutrophils Secrete Bioactive Paucimannosidic Proteins from Azurophilic Granules into Pathogen-Infected Sputum.* The Journal of Biological Chemistry, 2015. **290**(14): p. 8789-8802.
7. Schachter, H., *Paucimannose N-glycans in Caenorhabditis elegans and Drosophila melanogaster.* Carbohydrate Research, 2009. **344**(12): p. 1391-1396.
8. Steen, P.V.d., et al., *Concepts and principles of O-linked glycosylation.* Critical reviews in biochemistry and molecular biology, 1998. **33**(3): p. 151-208.
9. Voynow, J.A. and B.K. Rubin, *Mucins, Mucus, and Sputum.* CHEST. **135**(2): p. 505-512.
10. Brockhausen, I. and P. Stanley, *O-GalNAc Glycans*, in *Essentials of Glycobiology*, A. Varki, et al, Editors. 2017: Cold Spring Harbor (NY).
11. Tian, E. and K.G. Ten Hagen, *Recent insights into the biological roles of mucin-type O-glycosylation.* Glycoconjugate journal, 2009. **26**(3): p. 325-334.
12. McGuckin, M.A., et al., *Mucin dynamics and enteric pathogens.* Nature Reviews Microbiology, 2011. **9**(4): p. 265.
13. Naughton, J., et al., *Interaction of microbes with mucus and mucins: recent developments.* Gut Microbes, 2014. **5**(1): p. 48-52.
14. Arnold, J.N., et al., *The impact of glycosylation on the biological function and structure of human immunoglobulins.* Annu. Rev. Immunol., 2007. **25**: p. 21-50.
15. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum.* Annual review of biochemistry, 2004. **73**(1): p. 1019-1049.
16. Bieberich, E., *Synthesis, processing, and function of N-glycans in N-glycoproteins*, in *Glycobiology of the Nervous System*2014, Springer. p. 47-70.
17. Aplin, J.D. and C.J. Jones, *Fucose, placental evolution and the glycocode.* Glycobiology, 2011. **22**(4): p. 470-478.
18. Jones, C.J. and J.D. Aplin, *Glycosylation at the fetomaternal interface: does the glycocode play a critical role in implantation?* Glycoconjugate journal, 2009. **26**(3): p. 359-366.
19. Domino, S.E., et al., *Cervical mucins carry α (1, 2) fucosylated glycans that partly protect from experimental vaginal candidiasis.* Glycoconjugate journal, 2009. **26**(9): p. 1125-1134.
20. Rutanen, E.-M., et al., *Progesterone-associated proteins PP12 and PP14 in the human endometrium.* Journal of steroid biochemistry, 1987. **27**(1-3): p. 25-31.
21. Uchida, H., et al., *Glycodelin in reproduction.* Reproductive medicine and biology, 2013. **12**(3): p. 79-84.
22. Carson, D.D., M.M. Desouza, and E.G.C. Regisford, *Mucin and proteoglycan functions in embryo implantation.* Bioessays, 1998. **20**(7): p. 577-583.
23. Ricketts, A.a., D. Scott, and D. Bullock, *Radioiodinated surface proteins of separated cell types from rabbit endometrium in relation to the time of implantation.* Cell and tissue research, 1984. **236**(2): p. 421-429.

24. Anderson, T.L., G.E. Olson, and L.H. Hoffman, *Stage-specific alterations in the apical membrane glycoproteins of endometrial epithelial cells related to implantation in rabbits.* Biology of reproduction, 1986. **34**(4): p. 701-720.

25. Chávez, D.J., *Possible involvement of D-galactose in the implantation process*, in *Trophoblast Invasion and Endometrial Receptivity*1990, Springer. p. 259-272.

26. Tecle, E. and P. Gagneux, *Sugar-coated sperm: Unraveling the functions of the mammalian sperm glycocalyx.* Molecular reproduction and development, 2015. **82**(9): p. 635-650.

27. Mahajan, V.S. and S. Pillai, *Sialic acids and autoimmune disease.* Immunological reviews, 2016. **269**(1): p. 145-161.

28. Murphy, C., et al., *A freeze-fracture electron microscopic study of tight junctions of epithelial cells in the human uterus.* Anatomy and embryology, 1982. **163**(4): p. 367-370.

29. Quinn, C.E., and Casper, R.F., *Pinopodes: a questionable role in endometrial receptivity.* Human Reproduction Update, 2009. **15**(2): p. 229–236.

30. Bentin-Ley, U., et al., *Presence of uterine pinopodes at the embryo–endometrial interface during human implantation in vitro.* Human Reproduction, 1999. **14**(2): p. 515-520.

31. Bentin-Ley, U., et al., *Isolation and culture of human endometrial cells in a three-dimensional culture system.* Journal of reproduction and fertility, 1994. **101**(2): p. 327-332.

32. Sarantis, L., D. Roche, and A. Psychoyos, *Displacement of receptivity for nidation in the rat by progesterone antagonist RU 486: a scanning electron microscopy study.* Human Reproduction, 1988. **3**(2): p. 251-255.

33. Achache, H. and A. Revel, *Endometrial receptivity markers, the journey to successful embryo implantation.* Human reproduction update, 2006. **12**(6): p. 731-746.

34. Farach, M.C., et al., *Differential effects of p-nitrophenyl-D-xylosides on mouse blastocysts and uterine epithelial cells.* Biology of reproduction, 1988. **39**(2): p. 443-455.

35. Lessey, B.A., et al., *Integrins as markers of uterine receptivity in women with primary unexplained infertility.* Fertility and sterility, 1995. **63**(3): p. 535-542.

36. Genbacev, O.D., et al., *Trophoblast L-selectin-mediated adhesion at the maternal-fetal interface.* Science, 2003. **299**(5605): p. 405-408.

37. Hey, N.A., et al., *The polymorphic epithelial mucin MUC1 in human endometrium is regulated with maximal expression in the implantation phase.* The Journal of Clinical Endocrinology & Metabolism, 1994. **78**(2): p. 337-342.

38. Gipson, I.K., et al., *The amount of MUC5B mucin in cervical mucus peaks at midcycle.* The Journal of Clinical Endocrinology & Metabolism, 2001. **86**(2): p. 594-600.

39. Gipson, I.K., et al., *Mucin genes expressed by human female reproductive tract epithelia.* Biology of reproduction, 1997. **56**(4): p. 999-1011.

40. Salamonsen, L.A., et al., *Proteomics of the human endometrium and uterine fluid: a pathway to biomarker discovery.* Fertility and sterility, 2013. **99**(4): p. 1086-1092.

41. Surveyor, G.A., et al., *Expression and steroid hormonal control of Muc-1 in the mouse uterus.* Endocrinology, 1995. **136**(8): p. 3639-3647.

42. Braga, V. and S.J. Gendler, *Modulation of Muc-1 mucin expression in the mouse uterus during the estrus cycle, early pregnancy and placentation.* Journal of cell science, 1993. **105**(2): p. 397-405.

43. DeSouza, M.M., et al., *MUC1/episialin: a critical barrier in the female reproductive tract.* Journal of reproductive immunology, 2000. **45**(2): p. 127-158.

44. Spicer, A.P., et al., *Delayed mammary tumor progression in Muc-1 null mice.* Journal of Biological Chemistry, 1995. **270**(50): p. 30093-30101.

45. Bernal, A., et al., *Presence and regional distribution of sialyl transferase in the epididymis of the rat.* Biology of reproduction, 1980. **23**(2): p. 290-293.

46. Tulsiani, D.R., *Glycan-modifying enzymes in luminal fluid of the mammalian epididymis: an overview of their potential role in sperm maturation.* Molecular and cellular endocrinology, 2006. **250**(1-2): p. 58-65.

47.   Kirchhoff, C. and G. Hale, *Cell-to-cell transfer of glycosylphosphatidylinositol-anchored membrane proteins during sperm maturation.* MHR: Basic science of reproductive medicine, 1996. **2**(3): p. 177-184.

48.   Sullivan, R., G. Frenette, and J. Girouard, *Epididymosomes are involved in the acquisition of new sperm proteins during epididymal transit.* Asian journal of andrology, 2007. **9**(4): p. 483-491.

49.   Rooney, I., et al., *Physiologic relevance of the membrane attack complex inhibitory protein CD59 in human seminal plasma: CD59 is present on extracellular organelles (prostasomes), binds cell membranes, and inhibits complement-mediated lysis.* Journal of Experimental Medicine, 1993. **177**(5): p. 1409-1420.

50.   Flickinger, C., et al., *Dynamics of a human seminal vesicle specific protein: Charakteristik eines Bläschendrüsen-spezifischen Proteins des Menschen.* Andrologia, 1990. **22**(S1): p. 142-154.

51.   Gilks, C.B., et al., *Histochemical changes in cervical mucus-secreting epithelium during the normal menstrual cycle.* Fertility and sterility, 1989. **51**(2): p. 286-291.

52.   Pandya, I.J. and J. Cohen, *The leukocytic reaction of the human uterine cervix to spermatozoa.* Fertility and sterility, 1985. **43**(3): p. 417-421.

53.   Thompson, L., et al., *The leukocytic reaction of the human uterine cervix.* American Journal of Reproductive Immunology, 1992. **28**(2): p. 85-89.

54.   Tollner, T.L., C.L. Bevins, and G.N. Cherr, *Multifunctional glycoprotein DEFB126—a curious story of defensin-clad spermatozoa.* Nature Reviews Urology, 2012. **9**(7): p. 365.

55.   Familiari, G. and P. Motta, *Morphological changes of mouse spermatozoa in uterus as revealed by scanning and transmission electron microscopy.* Acta biologica Academiae Scientiarum Hungaricae, 1980. **31**(1-3): p. 57-67.

56.   Focarelli, R., F. Rosati, and B. Terrana, *Sialylglycoconjugates release during in vitro capacitation of human spermatozoa.* Journal of andrology, 1990. **11**(2): p. 97-104.

57.   Ma, F., et al., *Sialidases on mammalian sperm mediate deciduous sialylation during capacitation.* Journal of Biological Chemistry, 2012. **287**(45): p. 38073-38079.

58.   Bi, J., et al., *Basigin null mutant male mice are sterile and exhibit impaired interactions between germ cells and Sertoli cells.* Developmental biology, 2013. **380**(2): p. 145-156.

59.   Yeung, W., et al., *Glycodelin: a molecule with multi-functions on spermatozoa.* Society of Reproduction and Fertility supplement, 2007. **63**: p. 143-151.

60.   Varki, A., *Letter to the glyco-forum: since there are PAMPs and DAMPs, there must be SAMPs? Glycan "self-associated molecular patterns" dampen innate immunity, but pathogens can mimic them.* Glycobiology, 2011. **21**(9): p. 1121-1124.

61.   Toshimori, K., et al., *Loss of sperm surface sialic acid induces phagocytosis: an assay with a monoclonal antibody T21, which recognizes a 54K sialoglycoprotein.* Archives of andrology, 1991. **27**(2): p. 79-86.

62.   Hasegawa, A. and K. Koyama, *Antigenic epitope for sperm-immobilizing antibody detected in infertile women.* Journal of Reproductive Immunology, 2005. **67**(1-2): p. 77-86.

63.   Koyama, K., A. Hasegawa, and S. Komori, *Functional aspects of CD52 in reproduction.* Journal of Reproductive Immunology, 2009. **83**(1-2): p. 56-59.

64.   Diekman, A.B., et al., *Evidence for a unique N-linked glycan associated with human infertility on sperm CD52: a candidate contraceptive vaccinogen.* Immunological reviews, 1999. **171**(1): p. 203-211.

65.   Parry, S., et al., *The sperm agglutination antigen-1 (SAGA-1) glycoforms of CD52 are O-glycosylated.* Glycobiology, 2007. **17**(10): p. 1120-1126.

66.   Hardiyanto, L., A. Hasegawa, and S. Komori, *The N-linked carbohydrate moiety of male reproductive tract CD52 (mrt-CD52) interferes with the complement system via binding to C1q.* Journal of Reproductive Immunology, 2012. **94**(2): p. 142-150.

67.   Inhorn, M. and F. Van Balen, *Infertility around the globe: New thinking on childlessness, gender, and reproductive technologies*2002: Univ of California Press.

68.   Adamson, G.D., et al., *The number of babies born globally after treatment with the assisted reproductive technologies (ART).* Fertility and sterility, 2013. **100**(3): p. S42.

69.   Khalil, M.B., et al., *Sperm capacitation induces an increase in lipid rafts having zona pellucida binding ability and containing sulfogalactosylglycerolipid.* Developmental biology, 2006. **290**(1): p. 220-235.

70.   Heger, A., M. Sator, and D. Pietrowski, *Endometrial receptivity and its predictive value for IVF/ICSI-outcome.* Geburtshilfe und Frauenheilkunde, 2012. **72**(8): p. 710.

71.   Potdar, N., T. Gelbaya, and L.G. Nardo, *Endometrial injury to overcome recurrent embryo implantation failure: a systematic review and meta-analysis.* Reproductive biomedicine online, 2012. **25**(6): p. 561-571.

72.   Adderson, E.E., et al., *Contributors*, in *Principles and Practice of Pediatric Infectious Diseases (Fourth Edition)* 2012, Content Repository Only!: London. p. vii-xx.

73.   Platt, M., et al., *St Vincent's Declaration 10 years on: outcomes of diabetic pregnancies.* Diabetic Medicine, 2002. **19**(3): p. 216-220.

74.   Garcia-Vargas, L., et al., *Gestational diabetes and the offspring: implications in the development of the cardiorenal metabolic syndrome in offspring.* Cardiorenal medicine, 2012. **2**(2): p. 134-142.

75.   Everest-Dass, A.V., et al., *Human disease glycomics: technology advances enabling protein glycosylation analysis–Part 2.* Expert review of proteomics, 2018. **15**(4): p. 341-352.

76.   Albaghdadi, A.J. and F.W. Kan, *Endometrial receptivity defects and impaired implantation in diabetic NOD mice.* Biology of reproduction, 2012. **87**(2).

77.   Pampfer, S., et al., *Increased cell death in rat blastocysts exposed to maternal diabetes in utero and to high glucose or tumor necrosis factor-alpha in vitro.* Development, 1997. **124**(23): p. 4827-4836.

78.   Adams, S.M., et al., *Manipulation of the follicular phase: Uterodomes and pregnancy-is there a correlation?* BMC Pregnancy and Childbirth, 2001. **1**(1): p. 2.

79.   Guzick, D.S., et al., *Sperm morphology, motility, and concentration in fertile and infertile men.* New England Journal of Medicine, 2001. **345**(19): p. 1388-1393.

80.   Poland, M.L., et al., *Variation of semen measures within normal men.* Fertility and sterility, 1985. **44**(3): p. 396-400.

81.   Everest-Dass, A.V., et al., *Human disease glycomics: technology advances enabling protein glycosylation analysis–part 1.* Expert review of proteomics, 2018. 15(2): p. 165-182.

82.   Varki, A. and N. Sharon, *Historical Background and Overview*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).

83.   Vliegenthart, J.F., *The complexity of glycoprotein-derived glycans.* Proceedings of the Japan Academy, Series B, 2017. **93**(2): p. 64-86.

84.   Parekh, R., et al., *Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG.* Nature, 1985. **316**(6027): p. 452.

85.   Zhang, P., et al., *Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs.* Drug discovery today, 2016. **21**(5): p. 740-765.

86.   Varki, A., et al., *Essentials of glycobiology. New York: Spring Harbor Laboratory Press*, 1999, Cold Spring Harbor.

87.   Roucka, M., K. Zimmermann, and M. Fido, *Glycosylation Pattern of Biotechnologically Produced Proteins-Lectin Array Technology as a Versatile Tool for Screening?* Medical Research Archives, 2018. **6**(3).

88.   Weis, W.I. and K. Drickamer, *Structural basis of lectin-carbohydrate recognition.* Annual review of biochemistry, 1996. **65**(1): p. 441-473.

89.   Suzuki, S., et al., *Analysis of sialo-N-glycans in glycoproteins as 1-phenyl-3-methyl-5-pyrazolone derivatives by capillary electrophoresis.* Journal of Chromatography A, 2001. **910**(2): p. 319-329.

90. Zaia, J., *Capillary Electrophoresis–Mass Spectrometry of Carbohydrates*, in *Capillary Electrophoresis of Biomolecules*2013, Springer. p. 13-25.

91. Campbell, M.P., et al., *GlycoBase and autoGU: tools for HPLC-based glycan analysis.* Bioinformatics, 2008. **24**(9): p. 1214-1216.

92. Thaysen-Andersen, M. and N.H. Packer, *Advances in LC–MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N-and O-glycoproteome.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2014. **1844**(9): p. 1437-1452.

93. Abrahams, J.L., M.P. Campbell, and N.H. Packer, *Building a PGC-LC-MS N-glycan retention library and elution mapping resource.* Glycoconjugate journal, 2018. **35**(1): p. 15-29.

94. Zhang, Y., H. Yin, and H. Lu, *Recent progress in quantitative glycoproteomics.* Glycoconjugate journal, 2012. **29**(5-6): p. 249-258.

95. Zaia, J., *Mass spectrometry and glycomics.* Omics: a journal of integrative biology, 2010. **14**(4): p. 401-418.

96. Karas, M., et al., *Matrix-assisted ultraviolet laser desorption of non-volatile compounds.* International Journal of Mass Spectrometry and Ion Processes, 1987. **78**(1): p. 53-68.

97. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

98. Mutenda, K.E. and R. Matthiesen, *Analysis of carbohydrates by mass spectrometry*, in *Mass spectrometry data analysis in proteomics*2007, Springer. p. 289-301.

99. Domon, B. and C.E. Costello, *A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates.* Glycoconjugate Journal, 1988. **5**(4): p. 397-409.

100. Nakano, M., et al., *Identification of glycan structure alterations on cell membrane proteins in desoxyepothilone B resistant leukemia cells.* Molecular & Cellular Proteomics, 2011. **10**(11): p. M111. 009001.

101. Everest-Dass, A.V., et al., *Structural feature ions for distinguishing N-and O-linked glycan isomers by LC-ESI-IT MS/MS.* Journal of The American Society for Mass Spectrometry, 2013. **24**(6): p. 895-906.

102. Hogan, J.M., et al., *Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation.* Journal of proteome research, 2005. **4**(2): p. 628-632.

103. Wang, B., et al., *Reliable determination of site-specific in vivo protein N-glycosylation based on collision-induced MS/MS and chromatographic retention time.* Journal of The American Society for Mass Spectrometry, 2014. **25**(5): p. 729-741.

104. Jensen, P.H., et al., *Structural analysis of N- and O-glycans released from glycoproteins.* Nat. Protocols, 2012. **7**(7): p. 1299-1310.

105. Harvey, D.J., et al., *Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry.* Analytical biochemistry, 2008. **376**(1): p. 44-60.

106. Jensen, P.H., et al., *Structural analysis of N-and O-glycans released from glycoproteins.* Nature protocols, 2012. **7**(7): p. 1299.

Human disease glycomics: technology advances enabling protein glycosylation analysis –

part 2 (Manuscript 1)

Contribution: The writing of the manuscript was a collaborative effort between all authors. A. M. Shathili was mainly responsible for reviewing the glycan features of diabetes section. Dr. A. Dass was the main author and contributed the cancer glycomics section. Dr. C. Ashwood was responsible for Neuronal disease in the glyco-context. Dr. E. Moh wrote the antibody glycan structural features. Prof. Nicolle H. Packer supervised, reviewed and edited the manuscript and Chapter.

Everest-Dass, A.V., et al., Human disease glycomics: technology advances enabling protein glycosylation analysis–Part 2. *Expert review of proteomics*, 2018. **15**(4): p. 341-352.

Human disease glycomics: technology advances enabling protein glycosylation analysis – part 2

Arun V Everest-Dass, Edward S X Moh, Christopher Ashwood, Abdulrahman M M Shathili & Nicolle H Packer

Taylor & Francis
Taylor & Francis Group

REVIEW

Check for updates

Human disease glycomics: technology advances enabling protein glycosylation analysis – part 2

Arun V Everest-Dass[a,b,c], Edward S X Moh[a,b], Christopher Ashwood[a,b], Abdulrahman M M Shathili[a,b] and Nicolle H Packer[a,b,c]

[a]Faculty of Science and Engineering, Biomolecular Discovery and Design Research Centre, Macquarie University, Sydney, Australia; [b]ARC Centre for Nanoscale BioPhotonics, Macquarie University, Sydney, Australia; [c]Institute for Glycomics, Griffith University, Gold Coast, Australia

ABSTRACT

Introduction: The changes in glycan structures have been attributed to disease states for several decades. The surface glycosylation pattern is a signature of physiological state of a cell. In this review we provide a link between observed substructural glycan changes and a range of diseases.
Areas covered: We highlight biologically relevant glycan substructure expression in cancer, inflammation, neuronal diseases and diabetes. Furthermore, the alterations in antibody glycosylation in a disease context are described.
Expert commentary: Advances in technologies, as described in Part 1 of this review have now enabled the characterization of specific glycan structural markers of a range of disease states. The requirement of including glycomics in cross-disciplinary *omics* studies, such as genomics, proteomics, epigenomics, transcriptomics and metabolomics towards a systems glycobiology approach to understanding disease mechanisms and management are highlighted.

1. Association of glycan structural motifs with disease states

Glycosylation of proteins has been known to alter in disease for decades. Given that almost all mammalian cells are covered with glycans, it is no surprise that protein glycosylation plays a role in many human diseases. Changes in cell-surface glycan patterns, often correlated to disease phenotype, can result in specific substructures being displayed at an increased or decreased frequency or even in conveying a change in net charge on the cell surface, both of which could have a significant impact on cell-to-cell interactions. A recent editorial has identified post-translational modifications, including glycosylation, as being useful in understanding dynamic disease mechanisms, when put in context with other omics strategies [1]. The technologies described in Part 1 of the review now enable the determination and relative quantitation of these detailed structural motifs, or glycotopes, that are displayed, usually on the surface of cells, in a multitude of disease states. The obvious differences in expressed glycan structures in diseased cells to those displayed on healthy cells offer opportunities for new biomarkers and therapeutic targets for numerous diseases. Some relevant examples are given here.

1.1. Cancer glycomics

The most studied glycosylation changes over the last decades have been in cancer and have given insight into the significance of glycome changes in most cancers. This has been made possible due to a combination of improved tumor cell biology techniques and advancements in glycomics analysis technologies. Glycan changes observed in cancer have revealed underlying changes in biosynthetic molecular pathways as well as aberrant glycotopes that correlate with more aggressive cancer cell and tumor features, including increased migration, invasion, and metastatic potential, thus providing novel targets for therapeutic intervention [2–7]. In addition to surface glycosylation changes, the secreted glycoproteins from these epithelial cancer cells also carry altered glycans into the blood stream and therefore provide a source of potential new serum biomarkers. The importance of these glycan terminal substructures (Figure 1) from various cancers has been recently reviewed in several publications [5,8–13], and are summarized and updated here.

The most commonly observed aberrant *N*- and *O*-linked protein glycosylation in cancer are changes in the terminal sialylation, fucosylation, and branching. In mucin type *O*-GalNAc glycosylation, the *O*-glycans are mostly observed to be truncated.

Sialylation is predominantly observed at the terminal or nonreducing ends of the glycans of glycoconjugates. The addition of sialic acid is carried out by more than 20 distinct Golgi-resident sialyltransferases (ST) that link sialic acids via their second carbon (C2) to the carbon atom at position C3 (ST3Gal I-VI), C6 (ST6Gal I,II and ST6GalNAc I-VI), or to C8 (ST8Sia I-VI) of a terminal monosaccharide residue, yielding α2,3, α2,6, or α2-8 linkages [14]. An overall increase in sialylation notably of α2–3 and α2,6 linkages on lactosamine

CONTACT Nicolle H Packer ✉ nicki.packer@mq.edu.au 🖂 Department of Chemistry and Biomolecular Sciences, Macquarie University, NSW-2109 North Ryde, Australia

Figure 1. Cell surface protein glycosylation changes, and associated glycosyltransferase genes, observed to alter in cancer. Adapted from Christiansen et al [8].

moieties has been associated with several cancers [11]. For example, melanoma cell lines have showed an increased α2,3 sialylation associated with cancer progression of four different melanoma cell lines (WM1552C, WM115, IGR-39, and WM26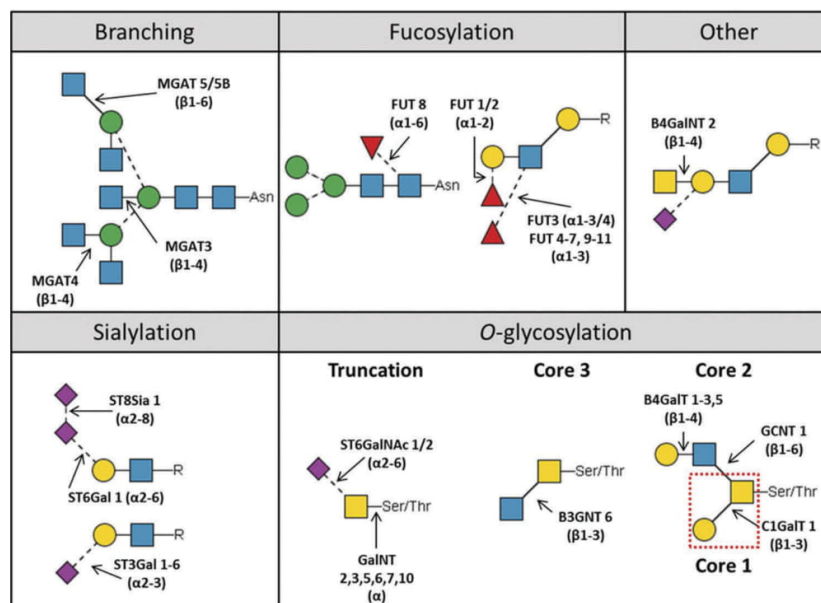6-4) when evaluated using lectins [15]. Similarly, Sethi et al. showed that α2,3 sialylation of N-glycans was abundantly identified in the metastatic (LIM1215) and aggressive (LIM2405) colorectal cancer cell lines [16]. In breast cancer, ST3Gal3 and ST3Gal1, the ST associated with the α2,3 sialylation of N- and O-glycans, respectively, were found to be upregulated [17,18]. On the other hand, Anugraham et al. used mass-spectrometry-based analysis to show that α2,6 sialylated N-glycans were abundant and unique in both ovarian cancer cell lines and serous ovarian cancer tissue [19,20]. Some other sialylated glycan antigens commonly associated with cancer are the structurally related sialylated Lewis epitopes such as sLewis[a] (NeuAc α2–3Galβ1–3(Fucα1–4)GlcNAc) and sLewis[x] (NeuAcα2–3Galβ1–4(Fucα1–3)GlcNAc) [21–23]. These terminal glycotopes have been suggested to play a crucial role in cancer metastasis through adhesion of the tumor cells to the endothelium [24–26]. Polysialic acid (PSA) is an α2-8 linked polymer of sialic acid and is a developmentally regulated glycan modification predominantly found in the central nervous system on neuronal cell adhesion molecule. PSA has now been found to be associated with several cancers in modulating cell–cell and cell–matrix adhesion, migration, invasion and metastasis and is strongly associated with poor clinical prognosis [11,27,28].

L-fucose or deoxyhexose is a common constituent of many N- and O-linked glycans produced by mammalian cells. For example, fucosylation plays important roles in blood group determination, immunological reactions, and signal transduction pathways [29,30]. Altered fucosyltransferases (FUTs)

expression have been extensively reported in various cancers [31,32]. In mammals, fucosylation is generally a non-extendable modification and can be subdivided into terminal fucosylation (Lewis blood-group antigens, such as Lewis[x], Lewis[y], Lewis[a] and Lewis[b], and the H-antigen) and core fucosylation. The sLewis antigens previously mentioned include the terminal addition of an α1–3 or α1–4 fucose to an α2-3 sialylated type 1 lactosamine or type 2 lactosamine unit, respectively. Increased expression of Lewis[x] antigens have been reported in colorectal, breast, and ovarian cancer cell lines [8]. They have generally been associated with higher malignancy and poor prognosis, as previously reported in colon and breast cancer [33–35], but interestingly, Rabassa et al. recently reported that Lewis[x] expression is associated with a better outcome in patients with head and neck squamous carcinoma [36]. Core fucosylation is the addition of α1-6 fucose residue to the innermost GlcNAc residue of the N-glycan chitobiose core by a single fucosyltransferase, FUT8. Core fucosylation of N-glycans is significantly upregulated in several cancers including liver, lung, colon and ovarian and breast cancers [10,37–39]. The increase of core fucosylation in certain serum glycoproteins can be used as a diagnostic marker; for example, core fucosylation of α-fetoprotein is an approved biomarker for the early detection of hepatocellular carcinoma (HCC) and is able to distinguish HCC from other liver diseases such as chronic hepatitis and liver cirrhosis [40].

The glycosyltransferase GnT-V catalyzes the addition of a GlcNAc residue to the α1-6 mannose branch of the N-glycan core through a β1-6 linkage [41]. Several studies have shown that the increased β1-6 branch and its corresponding MGAT5 gene is involved in cancer growth and metastasis of glioma, colon cancer, and gastric cancer cell lines [42–44]. Also tumors in GnT-V enzyme knockout mice showed decreased growth

and impaired metastasis [45]. This extended branch of β1-6 GlcNAc is further elongated by lactosamine chains to make polylactosamine structures that are high affinity ligands for galectins. Intriguingly, the action of GnT-V is inhibited by the presence of a bisecting GlcNAc [41], the bisecting GlcNAc is transferred to the trimannosyl core of complex or hybrid N-glycans on glycoproteins by the β1,4-N-acetylglucosaminyl-transferase III (GlcNAcT-III) or MGAT3. Nagae et al. recently used molecular conformational dynamics, crystallography, and NMR analysis to show the major conformational effect that bisecting GlcNAc has on N-glycans [46]. They suggested that the addition of this residue restricts the N-glycan confor-mation to a back-fold type, restricting the accessibility of other glycosyltransferases. Generally, bisecting GlcNAc has been suggested to reduce cancer growth and metastatic potential through the regulation of important cell membrane glycopro-teins such as EGFR, integrins, and cadherins [47–49]. Other studies have shown that MGAT3 is not always a tumor sup-pressing factor but can enhance cancer progression [16,50–52]. This is especially the case in high grade ovarian cancer, where an improved overall survival was observed with reduced MGAT3 expression, attributed to epigenetic regula-tion of this gene[53].

The GalNAc O-linked to serine or threonine is the initial sugar added in the O-glycan synthesis pathway and is com-monly referred to as the Thomsen-nouvelle antigen (Tn). This monosaccharide is usually further extended to a multitude of glycan structures [54]. The sialylation of this antigen (sialyl Tn) effectively truncates any further extension. Tn and sialyl Tn are well-established tumor markers and are mostly correlated with cancer invasion and metastasis [55]. The expression of sialyl Tn is usually low or absent in epithelial cells while is highly abundant in breast, colorectal, pancreatic, ovarian, and endo-metrial cancers [55,56]. Serum sialyl Tn levels have also been used as a prognostic indicator for epithelial ovarian cancer aggressiveness and metastatic potential [57]. There is a long known strong correlation between sialyl Tn expression and cancer progression but the specific mechanism and effects of sialyl Tn antigen on tumor cells are poorly understood.

1.2. Neuronal diseases in the glyco context

Many diseases that involve neuronal damage have altered glycan substructures, which may act as specific or broad markers for the disease. These glycan structures can be present on the cell surface or presented on proteins secreted by abnormal cells. Proteins secreted into the plasma are often the first place to look for biomarker discovery; however, the central nervous system fea-tures cerebrospinal fluid as a unique biological fluid for collection and analysis [58]. This fluid, while important functionally for human health, can also act as a potential source of biomarkers for diseases of the CNS and has been investigated as such [59,60].

In addition to proteins secreted by the brain into the CSF, the brain has its own specialized regions which, evidence suggests, possesses their own unique glycan profile [61]. The brain is a unique organ in terms of its structure and function and this is echoed by the glycans specific for this tissue region. Originally, glycoproteins highly abundant in the brain were studied such as CD24, identifying a range of glycotopes some

of which are rarely found elsewhere in the body (Table 1). Common glycan substructures such as Lewis^x, implicated in general inflammation and impaired neurite growth, are accompanied by rarer substructures such as Human Natural Killer 1, a key factor in normal dendritic spine maturation [62].

1.2.1. Impaired differentiation and regeneration of neuronal cells in disease

While these glycotopes are of interest based around brain devel-opment and function, more research focus has been oriented toward identifying glycan substructures correlated with disease-related impaired differentiation and regeneration of neuronal cells in the brain (Neuroglycomics). For example, Fogli et al., analyzing the protein glycosylation in the CSF, found four bian-tennary N-glycan diagnostic markers of leukodystrophies corre-lated with mutations of genes encoding the translation initiation factor, EIF2B [70]. Interestingly, these glycan changes were not echoed in the plasma of the patients, which further emphasizes CSF as an important biological fluid for biomarker discovery.

Impaired differentiation and regeneration mechanisms have ties to developmental disorders such as Amyotrophic Lateral Sclerosis and Huntington's disease. One key glycan of interest in this area is PSA and its role in peripheral nerve regeneration, where a long-term increase in PSA levels in regenerating nerves may favor selective motor target re-inner-vation [71]. The researchers suggested that controllable poly-sialyltransferase activity would be an interesting therapeutic approach to the treatment of these diseases.

1.2.2. Huntington's disease
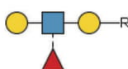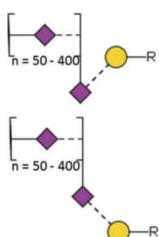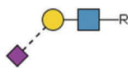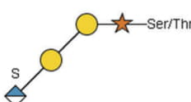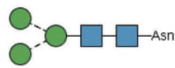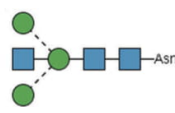
Investigating Huntington's disease, a progressive neurodegen-erative disease, Gizaw et al. performed, to date, the only com-prehensive glycomic profiling of a Huntington's disease mouse model [72]. Both tissue and plasma glycoproteins were analyzed from these mice, identifying an increase in core-fucosylated and bisecting GlcNAc N-glycans in the brain tissue of the Huntington's disease model mice compared to control mice. The core-fucosylation motif was found also to be increased in the serum of the Huntington's disease model mice (on biantenn-ary N-glycans) as well as having an increase in NeuGc. This study also emphasizes the need for biomarker research in neurode-generative diseases to move into the field of glycomics.

While there is still a need for further glycomic analysis of the cellular effects of Huntington's disease, Huntingtin, the protein encoded by the Huntington disease gene, has been studied in detail. Although it is an intracellular protein, and does not feature N-glycosylation, O-GlcNAcylation has been found to play a detrimental role in mutant Huntingtin toxicity in cell and fly models, by preventing the clearance of protein aggregates [73]. The prevention of protein aggregates by O-GlcNAc is not disease specific as this effect can also can be found in type II diabetes [74].

1.2.3. Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS), much like Huntington's dis-ease, is a neurodegenerative disease that primarily affects motor neurons; however, the exact mechanisms that cause this disease are not as well known. Edri-Brami et al. compared the serum of ALS patients with healthy donor controls and found a bi-

Table 1. Glycan substructures commonly observed in diseases of the central nervous system, R: *N*- or *O*-glycan scaffold.

Glycan substructure name	Substructure cartoon	Cell/tissue localization	Implicated in development/disease	Change in abundance
Lewis[X]		Neurons [63]	Impaired neurite growth [63]	Decrease
		CD24 from mouse brain [64]	Inflammation [66]	Increase
		Whole brain [65]		
Polysialic acid		Microglia [71]	Nerve regeneration [67]	Increase
Alpha 2-3 linked sialic acid		Whole brain [65]	N/A	N/A
Human Natural Killer 1		CD24 from mouse brain [64]	Axon outgrowth, regrowth of motor neurons [62]	Increase
Paucimannosylation		Neurons [68]	Neural stem cell proliferation [68]	Increase
Bisecting GlcNAc		CD24 from mouse brain [64]	Alzheimer's disease [69]	Increased
		CSF [69]	Brain developmental disorders [70]	Decreased

antennary glycan of IgG to be upregulated as well as decreased core-fucosylation [75], both of which were found to have functional effects through antibody-dependent cellular cytotoxicity. This demonstrates that profiling of IgG glycosylation is useful for discovering inflammation markers in plasma; however, the specificity of these IgG glycans to ALS has yet to be determined.

1.2.4. Pain
Peripheral nerve regeneration is one of the current chronic pain research areas and the European Union Seventh Framework Programme has identified glycomics as a valuable tool in identifying novel and prognostic biomarkers to predict chronic pain states [76]. As part of this framework programme and in investigating lower back pain, Freiden *et al*. [77] correlated plasma IgG glycan levels with lower back pain and found core-fucosylation to be at significantly lower levels in those suffering lower back pain. The mechanism is thought to be based on core-fucosylation acting as a safety switch to block antibody-dependent cell-mediated cytotoxicity, suggesting

core-fucosylation may be an important glycan substructure to monitor for inflammation.

1.2.5. Alzheimer's and Parkinson's diseases
Alzheimer's and Parkinson's disease are both neurodegenerative diseases that typically present in patients later in life. These diseases are therefore hypothesized to be age-related but the cause of Alzheimer's disease, unlike Parkinson's disease, has yet to be established. A review of the proteomic studies on cerebrospinal fluid biomarkers of Alzheimer's disease has recently been published and serves as a useful starting point upon which protein glycosylation can be considered [78]. Similar to low back pain, changes in the *N*-glycosylation of IgG in the plasma of patients with Parkinson's disease featured several glycans as novel biomarkers, with reduced relative abundance of sialylated mono-antennary *N*-glycans [79]. One difficulty in identifying biomarkers or treatment candidates for these diseases is the high correlation with age, where protein glycosylation is known to also alter with

human longevity [80]. O-glycosylation analysis has proven most fruitful in the area of Alzheimer's disease, identifying a highly sialylated O-glycan structure on amyloid-beta protein which was found to be highly abundant in patients' serum [81,82]. In addition, N-glycan profiling identified amyloid-beta to be more glycosylated than in controls, featuring higher relative abundances of bisecting GlcNAc and increased core-fucosylation. Research into treatment methods to follow on from these studies continues to look at enhancing glycosyl-transferase expression, particularly MGAT3 which is the enzyme responsible for adding a bisecting GlcNAc, a glycan feature that reportedly is protective in Alzheimer's disease [83,84].

1.3. Glycan features of diabetes

Diabetes is a complex process with several physiological effects in the human body such as increase in glucose levels and insulin resistance; the latter causing a spike in the pro-inflammatory cytokines released [85]. All of these physiological changes have been reported to alter the glycan bio-synthetic pathway and the glycome in general [86–90] (Table 2). In the case of inflammation, treating two human pancreatic ductal adenocarcinoma cell lines with several pro-inflammatory cytokines all resulted in higher staining for sialyl Lewis[x] and sialyl Lewis[a]. Differences in mRNA expression of ST and FUT, involved in the biosynthesis of these antigens, were determined [90]. TNF-α resulted in an increase in α2-3 ST and α1-2,3 fucosyltransferase mRNA expression which correlated with more sialyl Lewis[x] and Lewis[y] lectin staining using immunohistochemistry [90]. Other studies have also reported similar changes in α2-3 ST expression with the proinflammatory cytokine TNF-α [86,91,92]

Previous studies have utilized protein glycosylation as a biomarker for diabetes via detecting the NMR signal of GlcNAc residues on several acute phase glycoproteins such as α-acid glycoprotein, haptogloblin, α1-antitrypsin, α1-antic-hymotrypsin, and transferrin [94,95]. Since acute phase glyco-proteins are also a sign of general inflammation, detailed studies of each protein have been carried out over the years in order to allow for a higher level of accurate diagnosis. For instance, the study of glycans on α-acid glycoprotein showed an increase in α1-3 fucosylation using HPAEC-PAD and immu-nohistochemistry [96]. This was followed by a study using MALDI-TOF which showed α1-3 fucosylation changes on α-acid glycoprotein in type 2 diabetes [97].

1.3.1. Branched glycan structures

Glucose transporter 2 (GLUT2) is essential for sensing glucose changes in pancreatic B cells and maintaining appropriate insulin secretion. Recently, the elucidation of the molecular mechanism of GLUT2 function has revealed branched β1-4 GlcNAc linkage involvement in binding to Galectin-9 on the B cell surface [98]. A locus for diabetes onset 1q11.5 was seen to encode the gene for GnT-IVa which is the enzyme that catalyzes the transfer of GlcNAc to the core structure of N-linked oligosaccharides resulting

in multi-antennary structures [98]. A reduction in GnT-IVa expression in pancreatic B cells of islets from humans with type 2 diabetes was evident [99]. Interestingly, knockout GnT-IVa pancreatic B cells showed a significantly shortened half-life time of GLUT2 on the cell surface suggesting a role of branched glycan structures in controlling the stability and residency of the GLUT2 on the B cell surface. Mimicking Galectin-9 binding to GLUT2 resulted in reduction of B cell surface expression levels of GLUT2 [100,101]. Overall, branched glycan structures are important for GLUT2 function and consequently of maintaining appropriate insulin secretion from pancreatic cells.
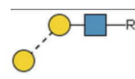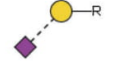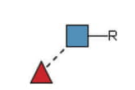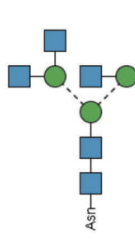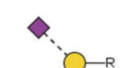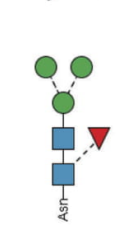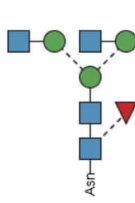
1.3.2. Sialic acid

The human abundant blood plasma chitrotriosidase (CHIT1) protein was extracted from 28 patients with type 2 diabetes and compared with 11 healthy individuals, in order to profile the glycosylation profile by lectin-ELISA [102]. The results showed significant decrease in SNA and MAA lectin binding, suggesting an overall decrease in sialic acid of both α2-6 and α2-3 linkages, respectively. A study of gestational diabetes mellitus has shown a more specific decrease in α2-6 sialic acid linked to the glycans on the pregnancy-related Glycodelin-A. This reduction in α2-6 sia-lic acid linkage, studied by MALDI-TOF and GC-MS, resulted in defective immune-suppressive activity by Glycodelin-A [103].

1.3.3. Fucosylation

Hepatocyte nuclear factor 1-α (HNF1α) has been shown to be a master transcriptional factor of a rare type of diabetes called maturity-onset diabetes of the young (MODY) and there is a strong correlation between fucosyltransferases (FUT6 and FUT8) and HNF1α expression in human plasma [82,104]. The relationship between MODY and HNF1α has been confirmed by a study that created a fucosylation index (DG9-glycan index) of fucosylated to non-fucosylated tri-antennary structures in plasma to differentiate control and other diabetes types [105]. On the other hand, the hyper-glycemic conditions seen in type 2 diabetes result in an overall increase in glucose and UDP-GlcNAc through the hexosamine biosynthetic pathway showing a direct link between protein glycosylation and type 2 diabetes [106,107]. A study of pyridylaminated oligosaccharides using RP-HPLC and MALDI-TOF/TOF of type 2 diabetic mice serum N-glycoproteins demonstrated a rise in core-fucosylation that correlated with increased mRNA expression of α1-6 FUT in the liver. The same study and others [97,108] have reported a small, though statistically significant, increase in core fucosylation in serum from human type 2 diabetes patients. However, another study showed an opposing effect in which the serum N-glycome of a large cohort (1161 individuals) of healthy and type 2 diabetic subjects showed a significant decrease of a specific core fucosylated biantennary N-glycan in type 2 diabetic subjects using DNA-sequencer-aided-fluorophore-assisted carbohy-drate electrophoresis [109].

Table 2. The involvement of glycan structures or glycan substructures in diabetes.

Glycan substructure	Observation	Importance	Method of detection	References
	Increase in di-galactose and decrease in α2-3 sialoglycans	Biomarkers for TNF-α induced insulin resistance in adipocytes	Quantitative glycoproteomics for glycosylation site determination of mouse adipocytes with or without TNF-α	[86]
	Increase in α1-3 fucosylation	Biomarker for increase of α-acid glycoprotein, a known acute phase protein	(HPAEC-PAD and immunohistochemistry) as well as (MALDI-TOF and sialidase treatment)	[96,97]
	β1-4 GlcNAc linkage on 1-3 mannose arm of the core structure forming branched structures	Essential for glucose transporter 2 (GLUT2) function. Defect in GLUT2 results in insulin intolerance	ESI-MS/MS of GLUT2. Knockout GnT-IVa in pancreatic B cells and mimicking GLUT2 binding partner (Galectin-9) both decreased GLUT2 life time.	[99,100]
	Decrease in sialic acid specific serum glycoproteins	Biomarker for chitrotiosidase (CHIT-1) and acute phase proteins	Lectin ELISA (SNA and MAA) of CHIT-1 human serum proteins	[102]
	Decrease in α2-6 sialic acid linkage in glycodelin-A	Possible biomarker for gestational diabetes mellitus	Glycodelin-A from amniotic fluid of diabetic pregnant women by lectin-binding, MALDI-TOF, and GC-MS	[103]
	Fucosylation index	Biomarker for maturity-onset diabetes of the young MODY	Lectin-assay	[82]
	Increase of core fucosylation on mice and human serum proteins	Biomarker for diabetic mice using serum proteins	Pyridylaminated oligosaccharides RP-HPLC, MALDI-TOF/TOF on type 2 diabetic mice serum samples and expression of liver α1,6 fucosyltransferase	[93]
	Decrease of di-antennary core fucosylation of human serum proteins	Biomarker for diabetic humans using serum proteins	DNA-sequencer fluorophore-assisted carbohydrate electrophoresis	[109]

Since all of the acute phase glycoproteins are found in blood, studies have been looking for easier obtained body fluids such as saliva, urine, and tears as diagnostic biomarker sources [110,111]. The investigation of basal tears in diabetes and diabetic retinopathy has shown a difference in glycan structures on the tear proteins from diabetic and diabetic

45

retinopathy patients. Those structures included low abundance N- and O-glycans including hybrid, complex and the di-sialylated core 2-mucin O-glycan [112].

1.4. Antibody glycan structural features

Antibodies play a huge role in all human diseases as a defense mechanism, but can sometimes be the cause of the problem as well, in cases such as autoimmune diseases. The five major types of antibodies in humans (IgG, IgA, IgM, IgE, and IgD, in order of abundance in serum) are all glycosylated proteins. While most of the scientific and commercial interest revolves around the specificity of the antibody–antigen targeting, the relationship between antibodies and their attached glycan counterparts have been shown to be implicated in a range of biological activities such as antibody–receptor binding, inflammation and immune response, and serum half-life [113–117]. The antibody glycans also sometimes act as a biomarker for different biological states such as inflammation (arthritis and vasculitis) and IgA nephropathy [113,115,118–120], but it is not yet known if the changes in the antibody glycan structures are the cause or the effect of the different biological states.

1.4.1. Immunoglobulin G glycosylation

The majority of high throughput analysis is based on IgG, partly due to the high abundance in human serum, ease of purification/isolation, and streamlined analytical methods available for IgG analysis [121–123]. Single monosaccharide differences of the N-linked glycans on the single glycosylation site, Asn297, have been reported to alter IgG functions [117,124–126]. Absence of core fucosylation and presence of the bisecting GlcNAc enhances antibody-dependent cellular cytotoxicity (ADCC) and affinity of IgG toward FcγRIII, whereas galactosylation of IgG1 was shown to be involved in anti-inflammatory activity by promoting association of FcγRIIB with the C-type lectin Dectin-1 [126], and enhancing FcγRIIIa binding affinity [127]. Recently, it was also shown that the sialic acid linkage on the IgG glycans could also result in a functional difference; bi-antennary α2-6 linked sialic acid on the glycan on Rituximab resulted in higher ADCC when compared with the α2-3 linked sialic acid form [128]. This study highlights the importance of in-depth glycan structural characterization, information that is more than just knowing the monosaccharide composition of the attached glycan. One recent high-throughput study, with over 5000 patient samples, showed that the relative abundance of several IgG glycan structures significantly changes with age. After a series of statistical modeling and refinements, it was found that the index of three specific glycan structures (core-fucosylated di-galactose, core-fucosylated bisecting, and core-fucosylated bisecting di-galactose glycans) provided a better correlation between chronological and biological age than current biomarkers of age such as telomere length [120].

Another reason that the effects of human IgG glycosylation have been explored more than the other antibody isotypes is the low complexity of the glycosylation on the protein. Most IgG antibodies only contain one N-glycosylation site, Asn297, on each heavy chain, which enables direct correlation between released glycan analysis and site-specific glycan analysis. With IgG, technological advances in glycoengineering have enabled the synthesis/production of homogenous, glycan-defined glycoproteins [129–131] for investigation of highly specific glycan mediated structure–function relationships [128,132]. However, the technology to control glycosylation down to the site-specific level on a multiply glycosylated protein is still unavailable. Regulation of site-specific glycosylation has been correlated with protein tertiary and quaternary structure, where certain glycosylation sites are more solvent exposed for better access to glycosylation enzymes [133,134]. For glycosylation sites that are equally accessible, it is not yet known if specific different glycosylation outcomes on a particular site could result in functional differences of the glycoprotein [135,136].

1.4.2. Glycosylation of other antibodies

The other four types of antibodies, IgA, IgM, IgD, and IgE, are all highly glycosylated with multiple N- (all) and O- (IgA1 and IgD) glycosylation sites [115]. The investigation of site-specific glycosylation changes affecting antibody function has been investigated since the 1990s, predominantly by glycosylation site knockouts (KO) using mutagenesis of the glycosylation sequon [137–145]. From the KO experiments on IgM and IgE, a single glycosylation site was identified to be crucial for effector function; Asn402 of IgM in complement activation and Asn394 of IgE in FcεR binding, both of which are homologous to the Asn297 glycosylation site on IgG. Interestingly, both of these glycosylation sites in IgM and IgE only display high mannose type glycans [145–149]. For IgA1, the consensus on the importance of glycosylation sites has not yet been reached, with evidence suggesting the two N-glycosylation site KOs either interfere with [138,140], or do not affect [139] IgA effector functions. These discrepancies could arise from different sources of IgA, different analytical methods, and/or different biological assays. To our knowledge, there has yet to be glycosylation differences relating to IgA2 function. For IgD, not much is known apart from the evidence that deglycosylated IgD impairs the IgD effector function [150], and Asn354 glycosylation is required for IgD secretion [137]. Mucin-type O-glycosylation is also present on IgA1 and IgD, located in the hinge region between the Fab and Fc portion of the antibody. Site-specific characterization of O-linked glycosylation is a difficult process, due to the lack of consensus sequon, heterogeneity of O-glycans, and clustering of multiple O-glycan sites amidst a short peptide region [151] so little work has been carried out because of this analytical difficulty that has yet to be resolved.

2. Expert commentary

Improved analytical technologies and innovative approaches have provided to a large extent the unraveling of the complex glycan code. This has enabled the identification of specific glycan substructures directly involved in disease manifestations or their progression in several areas including oncology,

neurobiology, diabetes, and immunology. Glycosylation changes observed in diseases are becoming recognized as increasingly valuable in understanding disease development, progression, and treatment.

Although genomic interpretations of the glycosylation machinery can provide a framework encompassing a cell's glycosylation profile, it lacks information about the complex interplay between transferases, glycosidases, nucleotide sugar donors, and epigenetic regulators that are involved in the generation of highly ordered, regulated, and conserved diverse glycan substructures observed in physiology and disease. Over the last decade many glycan-mediated disease associations have been suggested, some of which have been covered in this review. There is increased interest in new therapeutic interventions involving one or more of the many components of the glycosylation machinery, including the glycoengineering of recombinant antibodies. To fully understand the role of aberrant glycosylation changes observed in diseases, a systems-biology approach is required. High-throughput technologies that facilitate the analysis of complex glycan structures (as described in Part 1 of this review), detection methods to identify their interacting partners and innovative hyphenation of existing techniques are required to address these questions.

The simultaneous characterization of both the glycan (substructure) moieties and their protein carriers is also an added advantage especially in the discovery of biomarkers for diseases. This brings an added dimension and can increase the sensitivity associated with several disease biomarkers where both a protein and associated glycosylation change can be used to reduce the false positives. Similarly, a systems-glycobiology approach including the genomics, proteomics, epigenomics, transcriptomics, and metabolomics, all of which are involved in the cellular glycosylation machinery, will benefit the understanding and monitoring of disease mechanisms.

3. Five-year view: a speculative viewpoint on how the field will evolve in 5 years' time

Although several of the current disease biomarkers are glycoproteins (Table 3), diagnosis is often based on concentration of the glycoprotein. It is now clearly evident that incorporating the observed glycan changes in diseases with their protein carriers would dramatically improve the specificity and sensitivity of the current and future glycoprotein markers. With current advancements in analytical technology, particularly in liquid chromatography and mass spectrometry, glycomics can measure the macro- and micro-heterogeneity of these disease glycoprotein biomarkers both qualitatively and quantitatively from complex samples such as serum. The combination of measuring both protein and glycan alterations to glycoproteins in body fluids as diagnostics, and the engineering of therapeutic immunoglobulins to mimic the glycosylation changes that are observed in naturally occurring ADCC responses, in particular present attractive targets in this regard. The continued development and simplification of reliable, robust, and sensitive analysis workflows is essential to transform these technologies into a clinical setting.

The unique glycan substructure changes observed in diseases also provide interesting therapeutic targets for monoclonal antibody (mAb) production. Anti-glycan mAbs have a multifunctional role, as they can be used to detect/locate tumor distribution as well as to inhibit cell proliferation and cause cell death. In addition, although currently it is quite challenging to produce potent mAbs targeting tumor glycans, anti-glycan mAbs could be used as carrier molecules for nanoparticles loaded with drugs or imaging probes.

Table 3. List of current clinically relevant glycoprotein biomarkers.

Glycoprotein	Marker	Type of disease	Clinical application
α-Fetoprotein	Protein concentration and core fucosylation	Hepatoma, testicular cancers	Diagnosis, staging, detection, recurrence, and monitoring
Cancer antigen 125 (CA125, MUC16)	Protein concentration	Ovarian cancer	Detection, monitoring, and recurrence
Human epididymis protein 4 (HE 4)	Protein concentration	Ovarian cancer	Detection, monitoring, and recurrence
Cancer antigen 15-3 (MUC 1)	Sialyl Tn	Breast cancer	Monitoring
Carbohydrate antigen 19-9 (CA19-9)	Sialyl Lewis[a]	Pancreatic, ovarian cancers	Monitoring
Carcinoembryonic antigen (CEA)	Protein concentration	Gastric, colon, pancreatic, breast, and lung cancers	Detection, monitoring, and recurrence
Human epidermal growth factor receptor 2 (HER 2)	Protein concentration	Breast cancer	Choice of therapy
Human chorionic gonadotropin (hCG)	Protein concentration	Testicular, ovarian cancers	Diagnosis, staging, detection, recurrence, and monitoring
Fibrinogen degradation product (FDP)	Protein concentration	Bladder cancer	Detection and monitoring
Thyroglobulin	Protein concentration	Thyroid cancer	Monitoring and recurrence
Prostate-specific antigen (PSA)	Protein concentration	Prostate cancer	Detection, monitoring, and recurrence
Carbohydrate antigen 72–4 (CA72-4)	Sialyl Tn	Gastric cancer	Diagnosis, staging, detection, recurrence, and monitoring
Acute phase glycoproteins (α-acid glycoprotein, haptogloblin, α1-antitrypsin, α1-antichymotrypsin and transferrin)	Protein concentration	Inflammation such as in the case of diabetes	Detection and monitoring
Glucose transporter 2 (GLUT2)	Branching (Lower GLUT2 half-life resulted from GnT-IVa reduction in diabetes)	Type 2 diabetes	Detection and monitoring
Glycodelin-A	α2-6 sialic acid	Gestational diabetes mellitus	Detection and monitoring

47

This review does not aim to provide an exhaustive analysis of all glycan substructure-mediated changes in diseases, but rather to describe some crucial examples that demonstrate the possible impact of targeting protein glycosylation changes in disease. A major challenge from the beginning in glycomics has been the robust, reproducible, analytical task of deciphering the macro- and micro-heterogeneity changes at the glycoproteome level in an endogenous biological context. With the described (in Part 1) extraordinary analytical advances over the last decade, this now seems a possibility.

Key issues

- Changes in cell-surface glycan patterns in disease result in specific glycan substructures being displayed at an increased or decreased frequency on the cell surface, which could have a significant impact on cell-to-cell interactions and communications.
- The differences in glycan structures observed in diseased cells to those displayed on healthy cells offer opportunities for new biomarkers and therapeutic targets for numerous diseases.
- Glycan changes observed in cancer reveal aberrant glycotopes that correlate with more aggressive cancer cells and tumor features, including increased migration, invasion, and metastatic potential, thus providing novel targets for therapeutic intervention.
- Diseases that involve neuronal damage display unique glycotopes that may act as specific or broad markers for neuronal diseases.
- All of the physiological changes associated with diabetes have been reported to alter the glycan biosynthetic pathway and the glycome.
- Antibody glycosylation can differentiate between different biological states, such as potently altering the inflammatory profile and effector functions of the disease related antibodies.
- Combining glycomics with genomics, proteomics, epigenomics, transcriptomics and metabolomics will provide a systems-glycobiology approach to understanding many diseases.

Declartion of interest

Funding

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Zahedi RP. Joining forces: studying multiple post-translational modifications to understand dynamic disease mechanisms. Expert Rev Proteomics. 2016;13(12):1055–1057.
2. Kailemia MJ, Park D, Lebrilla CB. Glycans and glycoproteins as specific biomarkers for cancer. Anal Bioanal Chem. 2017;409(2):395–410.
3. Vankemmelbeke M, Chua JX, Durrant LG. Cancer cell associated glycans as targets for immunotherapy. Oncoimmunology. 2016;5(1):e1061177.
4. Nardy AF, Freire-de-Lima L, Freire-de-Lima CG, et al. The sweet side of immune evasion: role of glycans in the mechanisms of cancer progression. Front Oncol. 2016;6(54).
•• This review discusses the role of glycans in fundamental mechanisms controlling cancer development and progression, and their potential applications in oncology.
5. Taniguchi N, Kizuka Y. Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. Adv Cancer Res. 2015;126(11–51).
6. Adamczyk B, Tharmalingam T, Rudd PM. Glycans as cancer biomarkers. Biochim Biophys Acta. 2012;1820(9):1347–1353.
7. Lau KS, Dennis JW. N-glycans in cancer progression. Glycobiology. 2008;18(10):750–760.
8. Christiansen MN, Chik J, Lee L, et al. Cell surface protein glycosylation in cancer. Proteomics. 2014;14(4–5):525–546.
•• Comprehensive review of glycan changes observed in cancer.
9. Almeida A, Kolarich D. The promise of protein glycosylation for personalised medicine. Biochim Biophys Acta. 2016;1860(8):1583–1595.
10. Zhao YY, Takahashi M, Gu JG, et al. Functional roles of N-glycans in cell signaling and cell adhesion in cancer. Cancer Sci. 2008;99(7):1304–1310.
11. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015;15(9):540–555.
12. Oliveira-Ferrer L, Legler K, Milde-Langosch K. Role of protein glycosylation in cancer metastasis. Semin Cancer Biol. 2017;44:141–152.
13. Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. Annu Rev Pathol. 2015;10(473–510).
14. Dall'Olio F, Malagolini N, Trinchera M, et al. Sialosignaling: sialyltransferases as engines of self-fueling loops in cancer progression. Biochim Biophys Acta. 2014;1840(9):2752–2764.
15. Kolasinska E, Przybylo M, Janik M, et al. Towards understanding the role of sialylation in melanoma progression. Acta Biochim Pol. 2016;63(3):533–541.
16. Sethi MK, Thaysen-Andersen M, Smith JT, et al. Comparative N-glycan profiling of colorectal cancer cell lines reveals unique bisecting GlcNAc and alpha-2,3-linked sialic acid determinants are associated with membrane proteins of the more metastatic/aggressive cell lines. J Proteome Res. 2014;13(1):277–288.
17. Recchi MA, Hebbar M, Hornez L, et al. Multiplex reverse transcription polymerase chain reaction assessment of sialyltransferase expression in human breast cancer. Cancer Res. 1998;58(18):4066–4070.
18. Hebbar M, Krzewinski-Recchi MA, Hornez L, et al. Prognostic value of tumoral sialyltransferase expression and circulating E-selectin concentrations in node-negative breast cancer patients. Int J Biol Markers. 2003;18(2):116–122.
19. Anugraham M, Jacob F, Nixdorf S, et al. Specific glycosylation of membrane proteins in epithelial ovarian cancer cell lines: glycan structures reflect gene expression and DNA methylation status. Molecul Cell Prot. 2014;13(9):2213–2232.
20. Anugraham M, Jacob F, Everest-Dass AV, et al. Tissue glycomics distinguish tumour sites in women with advanced serous adenocarcinoma. Mol Oncol. 2017;11(11):1595–1615.
21. Trinchera M, Aronica A, Dall'Olio F. Selectin ligands sialyl-Lewis a and sialyl-Lewis X in gastrointestinal cancers. Biology. 2017;6(1).
22. Liang JX, Liang Y, Gao W. Clinicopathological and prognostic significance of sialyl Lewis X overexpression in patients with cancer: a meta-analysis. Onco Targets Ther. 2016;9:3113–3125.
23. Borzym-Kluczyk M, Radziejewska I. Changes of the expression of Lewis blood group antigens in glycoproteins of renal cancer tissues. Acta Biochim Pol. 2013;60(2):223–226.

24. Ugorski M, Laskowska A. Sialyl Lewis(a): a tumor-associated carbo-hydrate antigen involved in adhesion and metastatic potential of cancer cells. Acta Biochim Pol. 2002;49(2):303–311.

25. Kannagi R. Carbohydrate antigen sialyl Lewis a–its pathophysiolo-gical significance and induction mechanism in cancer progression. Chang Gung Med J. 2007;30(3):189–209.

26. Jeschke U, Mylonas I, Shabani N, et al. Expression of sialyl Lewis X, sialyl Lewis A, E-cadherin and cathepsin-D in human breast cancer: immunohistochemical analysis in mammary carcinoma in situ, inva-sive carcinomas and their lymph node metastasis. Anticancer Res. 2005;25(3A):1615–1622.

27. Wang X, Li X, Zeng YN, et al. Enhanced expression of polysialic acid correlates with malignant phenotype in breast cancer cell lines and clinical tissue samples. Int J Mol Med. 2016;37(1):197–206.

28. Elkashef SM, Allison SJ, Sadiq M, et al. Polysialic acid sustains cancer cell survival and migratory capacity in a hypoxic environment. Sci Rep. 2016;6(33026).

29. Miyoshi E, Moriwaki K, Nakagawa T. Biological function of fucosyla-tion in cancer biology. J Biochem. 2008;143(6):725–729.

30. Schneider M, Al-Shareffi E, Haltiwanger RS. Biological functions of fucose in mammals. Glycobiology. 2017;27(7):601–618.

31. Miyoshi E, Moriwaki K, Terao N, et al. Fucosylation is a promising target for cancer diagnosis and therapy. Biomolecules. 2012;2(1):34–45.

32. Moriwaki K, Miyoshi E. Fucosylation and gastrointestinal cancer. World J Hepatol. 2010;2(4):151–161.

33. Nakagoe T, Fukushima K, Nanashima A, et al. Expression of Lewis (a), sialyl Lewis(a), Lewis(x) and sialyl Lewis(x) antigens as prognos-tic factors in patients with colorectal cancer. Can J Gastroenterol. 2000;14(9):753–760.

34. Nakagoe T, Fukushima K, Itoyanagi N, et al. Expression of ABH/Lewis-related antigens as prognostic factors in patients with breast cancer. J Cancer Res Clin Oncol. 2002;128(5):257–264.

35. Koh YW, Lee HJ, Ahn JH, et al. Expression of Lewis X is associated with poor prognosis in triple-negative breast cancer. Am J Clin Pathol. 2013;139(6):746–753.

36. Rabassa ME, Pereyra A, Pereyra L, et al. Lewis x antigen is asso-ciated to head and neck squamous cell carcinoma survival. Pathol Oncol Res. 2017.

37. Liu YC, Yen HY, Chen CY, et al. Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. Proc Natl Acad Sci U S A. 2011;108 (28):11332–11337.

38. Potapenko IO, Haakensen VD, Luders T, et al. Glycan gene expres-sion signatures in normal and malignant breast tissue; possible role in diagnosis and progression. Mol Oncol. 2010;4(2):98–118.

39. Abbott KL, Nairn AV, Hall EM, et al. Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. Proteomics. 2008;8(16):3210–3220.

40. Sato Y, Nakata K, Kato Y, et al. Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. N Engl J Med. 1993;328(25):1802–1806.

41. Kizuka Y, Taniguchi N. Enzymes for N-glycan branching and their genetic and nongenetic regulation in cancer. Biomolecules. 2016;6 (2).

42. Yamamoto H, Swoger J, Greene S, et al. Beta1,6-N-acetylglucosa-mine-bearing N-glycans in human gliomas: implications for a role in regulating invasivity. Cancer Res. 2000;60(1):134–142.

43. Kim YS, Ahn YH, Song KJ, et al. Overexpression and beta-1,6-N-acetylglucosaminylation-initiated aberrant glycosylation of TIMP-1: a "double whammy" strategy in colon cancer progression. J Biol Chem. 2012;287(39):32467–32478.

44. Wang X, He H, Zhang H, et al. Clinical and prognostic implications of beta1, 6-N-acetylglucosaminyltransferase V in patients with gas-tric cancer. Cancer Sci. 2013;104(2):185–193.

45. Granovsky M, Fata J, Pawling J, et al. Suppression of tumor growth and metastasis in Mgat5-deficient mice. Nat Med. 2000;6(3):306–312.

46. Nagae M, Kanagawa M, Morita-Matsumoto K, et al. Atomic visuali-zation of a flipped-back conformation of bisected glycans bound to specific lectins. Sci Rep. 2016;6(22973).

47. Hoja-Lukowicz D, Przybylo M, Duda M, et al. On the trail of the glycan codes stored in cancer-related cell adhesion proteins. Biochim Biophys Acta. 2017;1861(1 Pt A):3237–3257.

48. Miwa HE, Song Y, Alvarez R, et al. The bisecting GlcNAc in cell growth control and tumor progression. Glycoconj J. 2012;29(8–9):609–618.

49. Song Y, Aglipay JA, Bernstein JD, et al. The bisecting GlcNAc on N-glycans inhibits growth factor signaling and retards mammary tumor progression. Cancer Res. 2010;70(8):3361–3371.

50. Yang X, Tang J, Rogler CE, et al. Reduced hepatocyte proliferation is the basis of retarded liver tumor progression and liver regeneration in mice lacking N-acetylglucosaminyltransferase III. Cancer Res. 2003;63(22):7753–7759.

51. Yang X, Bhaumik M, Bhattacharyya R, et al. New evidence for an extra-hepatic role of N-acetylglucosaminyltransferase III in the pro-gression of diethylnitrosamine-induced liver tumors in mice. Cancer Res. 2000;60(12):3313–3319.

52. Allam H, Johnson BP, Zhang M, et al. The glycosyltransferase GnT-III activates Notch signaling and drives stem cell expansion to pro-mote the growth and invasion of ovarian cancer. J Biol Chem. 2017;292(39):16351–16359.

53. Kohler RS, Anugraham M, Lopez MN, et al. Epigenetic activation of MGAT3 and corresponding bisecting GlcNAc shortens the survival of cancer patients. Oncotarget. 2016;7(32):51674–51686.

54. Springer GF. T and Tn, general carcinoma autoantigens. Science. 1984;224(4654):1198–1206.

55. Schultz MJ, Swindall AF, Bellis SL. Regulation of the metastatic cell phenotype by sialylated glycans. Cancer Metastasis Rev. 2012;31(3–4):501–518.

56. Pearce OM, Laubli H. Sialic acids in cancer biology and immunity. Glycobiology. 2016;26(2):111–128.

57. Kobayashi H, Terao T, Kawashima Y. Clinical evaluation of circulat-ing serum sialyl Tn antigen levels in patients with epithelial ovarian cancer. J Clin Oncol. 1991;9(6):983–987.

58. Kroksveen AC, Opsahl JA, Aye TT, et al. Proteomics of human cerebrospinal fluid: discovery and verification of biomarker candi-dates in neurodegenerative diseases using quantitative proteo-mics. J Proteomics. 2011;74(4):371–388.

59. Garden GA, Campbell BM. Glial biomarkers in human central ner-vous system disease. Glia. 2016;64(10):1755–1771.

60. Hoshi K, Matsumoto Y, Ito H, et al. A unique glycan-isoform of transferrin in cerebrospinal fluid: a potential diagnostic marker for neurological diseases. Biochim Biophys Acta. 2017;1861(10):2473–2478.

61. Ji IJ, Hua S, Shin DH, et al. Spatially-resolved exploration of the mouse brain glycome by tissue glyco-capture (TGC) and nano-LC/MS. Anal Chem. 2015;87(5):2869–2877.

62. Kizuka Y, Oka S. Regulated expression and neural functions of human natural killer-1 (HNK-1) carbohydrate. Cell Mol Life Sci. 2012;69(24):4135–4147.

63. Gouveia R, Schaffer L, Papp S, et al. Expression of glycogenes in differentiating human NT2N neurons. Downregulation of fucosyl-transferase 9 leads to decreased Lewisx levels and impaired neurite outgrowth. Biochimica Biophys Acta Genl Sub. 2012;1820 (12):2007–2019.

64. Bleckmann C, Geyer H, Reinhold V, et al. Glycomic analysis of N-linked carbohydrate epitopes from CD24 of mouse brain. J Proteome Res. 2009;8(2):567–582.

65. Hoffmann A, Nimtz M, Wurster U, et al. Carbohydrate structures of ß-trace protein from human cerebrospinal fluid: evidence for "brain-type" N-glycosylation. J Neurochem. 1994;63:2185–2196.

66. Telford JE, Doherty MA, Tharmalingam T, et al. Discovering new clinical markers in the field of glycomics. Biochem Soc Trans. 2011;39(1):327–330.

67. Schnaar RL, Gerardy-Schahn R, Hildebrandt H. Sialic acids in the brain: gangliosides and polysialic acid in nervous system develop-ment, stability, disease, and regeneration. Physiol Rev. 2014;94 (461–518).

68. Dahmen AC, Fergen MT, Laurini C, et al. Paucimannosidic glycoepitopes are functionally involved in proliferation of neural progenitor cells in the subventricular zone. Glycobiology. 2015;25(8):869–880.

49

69. Palmigiano A, Barone R, Sturiale L, et al. CSF N-glycoproteomics for early diagnosis in Alzheimer's disease. J Proteomics. 2016;131:29–37.

70. Fogli A, Merle C, Roussel V, et al. CSF N-glycan profiles to investigate biomarkers in brain developmental disorders: application to leukodystrophies related to eIF2B mutations. PloS One. 2012;7 (8):1–10.

71. Werneburg S, Mühlenhoff M, Stangel M, et al. Polysialic acid on SynCAM 1 in NG2 cells and on neuropilin-2 in microglia is confined to intracellular pools that are rapidly depleted upon stimulation. Glia. 2015;63(7):1240–1255.

72. Gizaw ST, Koda T, Amano M, et al. A comprehensive glycome profiling of Huntington's disease transgenic mice. Biochimica Biophys Acta Genl Sub. 2015;1850(9):1704–1718.

73. Portelius E, Brinkmalm G, Pannee J, et al. Proteomic studies of cerebrospinal fluid biomarkers of Alzheimer's disease: an update. Expert Rev Proteomics. 2017;14(11):1007–1020.

74. Kumar A, Singh PK, Parihar R, et al. Decreased O-linked GlcNAcylation protects from cytotoxicity mediated by huntingtin exon1 protein fragment. J Biol Chem. 2014;289(19):13543–13553.

75. Edri-Brami M, Rosental B, Hayoun D, et al. Glycans in sera of amyotrophic lateral sclerosis patients and their role in killing neuronal cells. PloS One. 2012;7(5):1–15.

76. Kringel D, Lötsch J. Pain research funding by the European Union Seventh Framework Programme. Eur J Pain. 2015;19(5):595–600.

77. Freidin MB, Keser T, Gudelj I, et al. The association between low back pain and composition of IgG glycome. Sci Rep. 2016 May 27;6: 26815. doi:10.1038/srep26815. PMID:27229623.

78. Marsh SA, Powell PC, Dell'italia LJ, et al. Cardiac O-GlcNAcylation blunts autophagic signaling in the diabetic heart. Life Sci. 2013;92 (11):648–656.

79. Russell AC, Šimurina M, Garcia MT, et al. The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. Glycobiology. 2017;27(5):501–510.

80. Ruhaak LR, Uh HW, Beekman M, et al. Decreased levels of bisecting GLcNAc glycoforms of IgG are associated with human longevity. PloS One. 2010;5(9):1–8.

81. Chen CC, Engelborghs S, Dewaele S, et al. Altered serum glycomics in Alzheimer disease: a potential blood biomarker? Rejuvenation Res. 2010;13(4):439–444.

82. Majidi S, Fouts A, Pyle L, et al. Can Biomarkers Help Target Maturity-Onset Diabetes of the Young Genetic Testing in Antibody-Negative Diabetes? Diabetes Technol Ther. 2018 Feb;20(2):106–112. doi: 10.1089/dia.2017.0317. PMID:29355436.

83. Fiala M, Mahanian M, Rosenthal M, et al. MGAT3 mRNA: a biomarker for prognosis and therapy of Alzheimer's disease by vitamin D and curcuminoids. J Alzheimers Dis. 2011;25(1):135–144.

84. Fiala M, Porter V. Delineating Alzheimer's disease progression with MGAT3, a biomarker for improved prognosis and personalized therapy. Biomark Med. 2011;5(5):645–647.

85. Hotamisligil G, Shargill N, Spiegelman B. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. Science. 1993;259(5091):87–91.

86. Parker BL, Thaysen-Andersen M, Fazakerley DJ, et al. Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF-alpha-induced insulin resistance in adipocytes. Mol Cell Proteomics. 2016;15(1):141–153.

87. Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. Cell. 2006;126(5):855–867.

88. Rellier N, Ruggiero D, Lecomte M, et al. Advanced glycation end products induce specific glycoprotein alterations in retinal microvascular cells. Biochem Biophys Res Commun. 1997;235 (2):281–285.

89. Rellier N, Ruggiero-Lopez D, Lecomte M, et al. In vitro and in vivo alterations of enzymatic glycosylation in diabetes. Life Sci. 1999;64 (17):1571–1583.

90. Bassagañas S, Allende H, Cobler L, et al. Inflammatory cytokines regulate the expression of glycosyltransferases involved in the biosynthesis of tumor-associated sialylated glycans in pancreatic cancer cell lines. Cytokine. 2015;75(1):197–206.

91. Majuri ML, Pinola M, Niemelä R, et al. α2, 3-Sialyl and α1, 3-fucosyl-transferase-dependent synthesis of sialyl Lewis x, an essential oligosaccharide present on L-selectin counterreceptors, in cultured endothelial cells. Eur J Immunol. 1994;24(12):3205–3210.

92. Majuri ML, Niemelä R, Tiisala S, et al. Expression and function of α2, 3-sialyl-and α, 3/l, 4-fucosyltransferases in colon adenocarcinoma cell lines: role in synthesis of E-selectin counter-receptors. Int J Cancer. 1995;63(4):551–559.

93. Itoh N, Sakaue S, Nakagawa H, et al. Analysis of N-glycan in serum glycoproteins from db/db mice and humans with type 2 diabetes. Am J Physiol Endocrinol Metab. 2007;293(4):E1069–1077.

94. Schmidt MI, Duncan BB, Sharrett AR, et al. Markers of inflammation and prediction of diabetes mellitus in adults (atherosclerosis risk in communities study): a cohort study. The Lancet. 1999;353 (9165):1649–1652.

95. Bell JD, Brown JC, Nicholson JK, et al. Assignment of resonances for 'acute-phase'glycoproteins in high resolution proton NMR spectra of human blood plasma. FEBS Letters. 1987;215(2):311–315.

96. Poland DCW, Schalkwijk CG, Stehouwer CDA, et al. Increased α3-fucosylation of α1-acid glycoprotein in type I diabetic patients is related to vascular function. Glycoconj J. 2001;18(3):261–268.

97. Higai K, Azuma Y, Aoki Y, et al. Altered glycosylation of α1-acid glycoprotein in patients with inflammation and diabetes mellitus. Clinica Chimica Acta. 2003;329(1–2):117–125.

98. Ohtsubo K. Glyco-predisposing factor of diabetes. In: Suzuki T, Ohtsubo K, Taniguchi N, editors. Sugar chains: decoding the functions of glycans. TokJapan, yo: Springer; 2015. p. 209–218.

99. Ohtsubo K, Chen MZ, Olefsky JM, et al. Pathway to diabetes through attenuation of pancreatic beta cell glycosylation and glucose transport. Nat Med. 2011;17(9):1067–1075.

100. Ohtsubo K, Takamatsu S, Minowa MT, et al. Dietary and genetic control of glucose transporter 2 glycosylation promotes insulin secretion in suppressing diabetes. Cell. 2005;123(7):1307–1321.

101. Ohtsubo K, Takamatsu S, Gao C, et al. N-glycosylation modulates the membrane sub-domain distribution and activity of glucose transporter 2 in pancreatic beta cells. Biochem Biophys Res Commun. 2013;434(2):346–351.

102. Żurawska-Płaksej E, Kratz EM, Ferens-Sieczkowska M, et al. Changes in glycosylation of human blood plasma chitotriosidase in patients with type 2 diabetes. Glycoconj J. 2016;33(1):29–39.

103. Lee C-L, Chiu PCN, Pang P-C, et al. Glycosylation failure extends to glycoproteins in gestational diabetes mellitus: evidence from reduced α2-6 sialylation and impaired immunomodulatory activities of pregnancy-related glycodelin-A. Diabetes. 2011;60(3):909–917.

104. Yamagata K, Oda N, Kaisaki PJ, et al. Mutations in the hepatocyte nuclear factor-1[alpha] gene in maturity-onset diabetes of the young (MODY3). Nature. 1996;384(6608):455–458.

105. Thanabalasingham G, Huffman JE, Kattla JJ, et al. Mutations in HNF1A result in marked alterations of plasma glycan profile. Diabetes. 2013;62(4):1329–1337.

106. Marshall S, Bacote V, Traxinger R. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. J Biol Chem. 1991;266(8):4706–4712.

107. Buse MG. Hexosamines, insulin resistance, and the complications of diabetes: current status. Am J Physiol Endocrinol And Metabolism. 2006;290(1):E1–E8.

108. McMillan DE. Elevation of glycoprotein fucose in diabetes mellitus. Diabetes. 1972;21(8):863–871.

109. Testa R, Vanhooren V, Bonfigli AR, et al. N-Glycomic changes in serum proteins in type 2 diabetes mellitus correlate with complications and with metabolic syndrome parameters. PloS One. 2015;10 (3):e0119983.

110. Rao PV, Laurie A, Bean ES, et al. Salivary protein glycosylation as a noninvasive biomarker for assessment of glycemia. J Diabetes Sci Technol. 2015;9(1):97–104.

111. Brownlee M, Vlassara H, Cerami A. Measurement of glycosylated amino acids and peptides from urine of diabetic patients using affinity chromatography. Diabetes. 1980;29(12):1044–1047.

112. Nguyen-Khuong T, Everest-Dass AV, Kautto L, et al. Glycomic characterization of basal tears and changes with diabetes and diabetic retinopathy. Glycobiology. 2015;25(3):269–283.

113. Epp A, Sullivan KC, Herr AB, et al. Immunoglobulin glycosylation effects in allergy and immunity. Curr Allergy Asthma Rep. 2016;16(11):79.

114. K-T S, Anthony R. Antibody glycosylation and inflammation. Antibodies. 2013;2(3):392.
    • This review discusses the contribution of the Fc glycan to IgG antibody effector functions and *in vivo* regulation of the antibody glycosylation.

115. Maverakis E, Kim K, Shimoda M, et al. Glycans in the immune system and the altered glycan theory of autoimmunity: a critical review. J Autoimmun. 2015;57:1–13.

116. Jefferis R. Glycosylation as a strategy to improve antibody-based therapeutics. Nat Rev Drug Discov. 2009;8(3):226–234.

117. Hudak JE, Bertozzi CR. Glycotherapy: new advances inspire a reemergence of glycans in medicine. Chem Biol. 2014;21(1):16–37.

118. Takahashi K, Raska M, Stuchlova Horynova M, et al. Enzymatic sialylation of IgA1 O-glycans: implications for studies of IgA nephropathy. PloS One. 2014;9(2):e99026.

119. Gornik O, Lauc G. Glycosylation of serum proteins in inflammatory diseases. Dis Markers. 2008;25(4–5):267–278.

120. Krištić J, Vučković F, Menni C, et al. Glycans are a novel biomarker of chronological and biological ages. Journals Gerontol Ser A: Biol Sci Med Sci. 2014;69(7):779–789.

121. Dotz V, Haselberg R, Shubhakar A, et al. Mass spectrometry for glycosylation analysis of biopharmaceuticals. In: TrAC trends in analytical chemistry. 2015;1-9. Available from: https://www.sciencedirect.com/science/article/pii/S0165993615001995

122. Jansen BC, Reiding KR, Bondt A, et al. MassyTools: a high-throughput targeted data processing tool for relative quantitation and quality control developed for glycomic and glycoproteomic MALDI-MS. J Proteome Res. 2015;14(12):5088–5098.

123. Pucic M, Knezevic A, Vidic J, et al. High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. Molecul Cell Prot. 2011;10(10):M111.010090.

124. Ratner M. Genentech's glyco-engineered antibody to succeed Rituxan. Nat Biotechnol. 2014;32(1):6–7.

125. Ferrara C, Grau S, Jager C, et al. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcgammaRIII and antibodies lacking core fucose. Proc Natl Acad Sci U S A. 2011;108(31):12669–12674.

126. Karsten CM, Pandey MK, Figge J, et al. Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of Fc[gamma] RIIB and dectin-1. Nat Med. 2012;18(9):1401–1406.

127. Kurogochi M, Mori M, Osumi K, et al. Glycoengineered monoclonal antibodies with homogeneous glycan (M3, G0, G2, and A2) using a chemoenzymatic approach have different affinities for fcγriiia and variable antibody-dependent cellular cytotoxicity activities. PloS One. 2015;10(7):e0132848.

128. Lin CW, Tsai MH, Li ST, et al. A common glycan structure on immunoglobulin G for enhancement of effector functions. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(34):10611–10616.

129. Goodfellow JJ, Baruah K, Yamamoto K, et al. An endoglycosidase with alternative glycan specificity allows broadened glycoprotein remodelling. J Am Chem Soc. 2012;134(19):8030–8033.

130. Zou G, Ochiai H, Huang W, et al. Chemoenzymatic synthesis and Fcgamma receptor binding of homogeneous glycoforms of antibody Fc domain. Presence of a bisecting sugar moiety enhances the affinity of Fc to FcgammaIIIa receptor. J Am Chem Soc. 2011;133(46):18975–18991.

131. Zha D. Glycoengineered Pichia-based expression of monoclonal antibodies. Methods Mol Biol. 2013;988:31-43. doi: 10.1007/978-1-62703-327-5_3.

132. Dashivets T, Thomann M, Rueger P, et al. Multi-angle effector function analysis of human monoclonal IgG glycovariants. PloS One. 2015;10(12):e0143520.

133. Lee LY, Lin CH, Fanayan S, et al. Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. Front Immunol. 2014;5(404).

134. Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology. 2012;22 (11):1440–1452.

135. S-I. N, Itoh A, Nakakita Y, et al. Cooperative interactions of oligosaccharide and peptide moieties of a glycopeptide derived from IgE with galectin-9. J Biol Chem. 2016;291 (2):968–979.

136. Sterner E, Flanagan N, Gildersleeve JC. Perspectives on anti-glycan antibodies gleaned from development of a community resource database. ACS Chem Biol. 2016;11(7):1773–1783.

137. Gala FA, Morrison SL. The role of constant region carbohydrate in the assembly and secretion of human IgD and IgA1. J Biol Chem. 2002;277(32):29005–29011.

138. Chuang PD, Morrison SL. Elimination of N-linked glycosylation sites from the human IgA1 constant region: effects on structure and function. J Immunol. 1997;158(2):724–732.

139. Mattu TS, Pleass RJ, Willis AC, et al. The glycosylation and structure of human serum IgA1, Fab, and Fc Regions and the role of N-glycosylation on Fcα receptor interactions. J Biol Chem. 1998;273(4):2260–2272.

140. Carayannopoulos L, Max EE, Capra JD. Recombinant human IgA expressed in insect cells. Proc Natl Acad Sci U S A. 1994;91 (18):8348–8352.

141. Ubelhart R, Bach MP, Eschbach C, et al. N-linked glycosylation selectively regulates autonomous precursor BCR function. Nat Immunol. 2010;11(8):759–765.

142. Wright JF, Shulman MJ, Isenman DE, et al. C1 binding by mouse IgM. The effect of abnormal glycosylation at position 402 resulting from a serine to asparagine exchange at residue 406 of the mu-chain. J Biol Chem. 1990;265(18):10506–10513.

143. Muraoka S, Shulman MJ. Structural requirements for IgM assembly and cytolytic activity. Effects of mutations in the oligosaccharide acceptor site at Asn402. J Immunol. 1989;142(2):695–701.

144. Cals MM, Guenzi S, Carelli S, et al. IgM polymerization inhibits the Golgi-mediated processing of the mu-chain carboxy-terminal glycans. Mol Immunol. 1996;33(1):15–24.

145. Shade KT, Platzer B, Washburn N, et al. A single glycan on IgE is indispensable for initiation of anaphylaxis. J Exp Med. 2015;212 (4):457–467.

146. Moh ES, Lin CH, Thaysen-Andersen M, et al. Site-specific N-glycosylation of recombinant pentameric and hexameric human IgM. J Am Soc Mass Spectrom. 2016;27(7):1143–1155.

147. Pabst M, Kuster SK, Wahl F, et al. A microarray-matrix-assisted laser desorption/ionization-mass spectrometry approach for site-specific protein N-glycosylation analysis, as demonstrated for human serum immunoglobulin M (IgM). Molecul Cell Prot. 2015;14(6):1645–1656.

148. Loos A, Gruber C, Altmann F, et al. Expression and glycoengineering of functionally active heteromultimeric IgM in plants. Proc Natl Acad Sci U S A. 2014;111(17):6263–6268.

149. Plomp R, Hensbergen PJ, Rombouts Y, et al. Site-specific N-glycosylation analysis of human immunoglobulin E. J Proteome Res. 2013;13(2):536–546.

150. Amin AR, Tamma SM, Oppenheim JD, et al. Specificity of the murine IgD receptor on T cells is for N-linked glycans on IgD molecules. Proc Natl Acad Sci U S A. 1991;88(20):9238–9242.

151. Hoffmann M, Marx K, Reichl U, et al. Site-specific O-glycosylation analysis of human blood plasma proteins. Molecul Cell Prot. 2016;15(2):624–641.

51

Human disease glycomics: technology advances enabling protein glycosylation analysis –

part 1 (Manuscript 2)

Contribution: The writing of the manuscript was a collaborative effort between all authors. A. M. Shathili was mainly responsible for reviewing *O*-glycosylation, NMR and intact glycoprotein analysis. Dr. A. Dass was the main author and contributed the introduction, structural determination of protein glycosylation and detection of glycoconjugates sections. Dr. C. Ashwood was responsible for fragmentation nomenclature. Dr. E. Moh wrote the detection of glycoconjugates section. Prof. Nicolle H. Packer supervised, reviewed and edited the manuscript and Chapter.

Everest-Dass, A.V., et al., *Human disease glycomics: technology advances enabling protein glycosylation analysis–part 1*. Expert review of proteomics, 2018. 15(2): p. 165-182.

Human disease glycomics: technology advances enabling protein glycosylation analysis – part 1

Arun V Everest-Dass, Edward S X Moh, Christopher Ashwood, Abdulrahman M M Shathili & Nicolle H Packer

Taylor & Francis
Taylor & Francis Group

REVIEW

Check for updates

Human disease glycomics: technology advances enabling protein glycosylation analysis – part 1

Arun V Everest-Dass[a,b,c], Edward S X Moh[a,c], Christopher Ashwood[a,c], Abdulrahman M M Shathili[a,c] and Nicolle H Packer[a,b,c]

[a]Biomolecular Discovery and Design Research Centre, Faculty of Science and Engineering, Macquarie University, Sydney, Australia; [b]Institute for Glycomics, Griffith University, Gold Coast, Australia; [c]ARC Centre for Nanoscale BioPhotonics, Macquarie University, Sydney, Australia

ABSTRACT

Introduction: Protein glycosylation is recognized as an important post-translational modification, with specific substructures having significant effects on protein folding, conformation, distribution, stability and activity. However, due to the structural complexity of glycans, elucidating glycan structure-function relationships is demanding. The fine detail of glycan structures attached to proteins (including sequence, branching, linkage and anomericity) is still best analysed after the glycans are released from the purified or mixture of glycoproteins (glycomics). The technologies currently available for glycomics are becoming streamlined and standardized and many features of protein glycosylation can now be determined using instruments available in most protein analytical laboratories.

Areas covered: This review focuses on the current glycomics technologies being commonly used for the analysis of the microheterogeneity of monosaccharide composition, sequence, branching and linkage of released *N*- and *O*-linked glycans that enable the determination of precise glycan structural determinants presented on secreted proteins and on the surface of all cells.

Expert commentary: Several emerging advances in these technologies enabling glycomics analysis are discussed. The technological and bioinformatics requirements to be able to accurately assign these precise glycan features at biological levels in a disease context are assessed.

1. Glycosylation

All cells in nature are coated with glycans that are essential for biological processes and communication. These glycans are mainly distributed as a large array of glycoconjugates covalently bound to proteins and lipids, which include glycoproteins, proteoglycans, and glycolipids [1,2].

The thick layer of glycoconjugates on the cell surface is commonly referred to as the glycocalyx. These extensive, complex glycan structures were classically thought to only provide structural roles, but now it is understood that glycans participate in fundamental inter- and intracellular functions, such as protein quality control, adhesion, motility, endocytosis, and signal transduction [1,3]. Glycans are also known to affect cellular processes important in development, cell proliferation, differentiation, and morphogenesis [2,4]. Glycosylation changes are often associated with disease states, for example, cancer cells frequently display differently expressed glycans compared to those from normal cells [5]. Additionally, the evolved glycan differences between species and phylogeny are important markers for the immune system in discriminating between self and nonself [6].

Monosaccharides, the basic building blocks of glycans are linked in either linear or branched forms by glycosidic bonds [2]. Several sequential enzymatic actions of glycosidases and glycosyltransferases assemble diverse glycan structures from these basic units [7]. They can be further modified by chemical substitutions such as phosphorylation, sulfation, and acetylation. Mammalian glycans are assembled from only 10 core monosaccharides: fucose (Fuc), galactose (Gal), glucose (Glc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), glucuronic acid (GlcA), iduronic acid (IdoA), mannose (Man), sialic acid (SA), and xylose (Xyl) . Their assembly is aided through an estimated action of 700 proteins involved in the glycosylation machinery to generate the full diversity of mammalian glycans (estimated to be ~7,000 protein-attached structures) [8,9]. This non-template-driven process produces heterogeneous glycan structures differing in sequence, branching, linkage, and anomericity.

Eukaryotic protein glycosylation is usually via *N*- and *O*-linkages. *N*-glycans are formed by the covalent linkage of oligosaccharide chains to the asparagine amino acid residue of the protein, commonly involving a GlcNAc residue and the consensus peptide sequence Asn-X-Ser/Thr (where X is any amino acid except proline). The *N*-linked glycans biosynthesis is initiated when the nascent protein enters the endoplasmic reticulum (ER). The *N*-glycans are initially synthesized as a lipid-linked oligosaccharide (LLO) precursor, and the glycans are transferred from LLO to a nascent polypeptide chain during translation and the block of sugars ($Glc_3Man_9GlcNAc_2$) is transferred to the amino group in the asparagine side chain. A series of processing reactions further trims the *N*-glycan in the ER. Interestingly, these initial steps are highly conserved across all eukaryotes [10]. The initial trimming of the $Glc_3Man_9GlcNAc_2$ moiety is through the removal of the Glc residues by glucosidases that reside in the lumen of the ER. Most

glycoproteins exiting the ER carry nine or eight Man residues, depending on the action of the ER α-mannosidase I. In multi-cellular organisms, α1–2 mannosidases IA, IB, and 1C in the *cis*-Golgi further trim the *N*-glycan to give $Man_5GlcNAc_2$. The multitude of glycans ($Man_{5-9}GlcNAc_2$) through these trimming steps generate the so-called oligomannose-type *N*-glycans. Further processing of the $Man_5GlcNAc_2$ gives rise to the complex and hybrid types in the medial-Golgi by the action of an *N*-acetylglucosaminyltransferase (GlcNAcT-I), which adds a GlcNAc residue to the Man α1–3 in the core of $Man_5GlcNAc_2$. The addition of this residue enables the action of α-mannosidase II to remove the terminal α1–3 and α1–6 Man residues. Hybrid-type *N*-glycans are formed due to the incomplete action of this α-mannosidase II. Complex-type *N*-glycans initiated by antennae or branches by the addition of terminal GlcNAc residues by various *N*-acetylglucosaminyltranferases in the Golgi can create a plethora of branched structures. Further capping (Gal, Fuc, SAs) or modification of the core glycan (Fuc) occurs in the *trans*-Golgi. Thus the *N*-glycan heterogeneity arises from the combinatorial enzyme processing in the Golgi apparatus. For more detailed description of the glycosyltransferases, glycosidases, and the biosynthetic pathway, readers are encouraged to access the freely accessible online resource 'Essentials of Glycobiology' [10].

An *O*-glycan is commonly linked to the polypeptide mainly through a GalNAc residue (commonly referred to as mucin-type) to the hydroxyl group of serine or threonine amino acid residue [11]. Their biosynthesis follows protein *N*-glycosylation and folding. Unlike *N*-glycans, the synthesis of *O*-glycans is through stepwise addition of single monosaccharide residues. The initial step is the addition of the GalNac residue which is catalyzed by a polypeptide-*N*-acetyl-galactosaminyltransferase (ppGalNAcTs). There have been about 20 different ppGalNAcTs reported to initiate this process with varying tissue expression and differential specificities [12]. The GalNAc residues are further extended generating eight different cores; cores 1–4 are the most common in humans, while cores 5–8 have extremely restricted occurrences [13]. The *N*- and *O*-glycan cores are usually further extended into a variety of terminal glycans by the addition and substitution of various monosaccharides such as GlcNAc, Gal, GalNAc, Fuc, and SA. Sulfation of Gal and GlcNAc residues causes further diversification. The mature glycan now comprises of several biologically relevant determinants, such as the blood group antigens. They are often found to be the major regulation and recognition factor of the glycoconjugate [14].

Several other *O*-linked modifications of eukaryotic proteins have also been reported such as *O*-fucosylation, *O*-mannosylation, *O*-galactosylation, and *O*-GlcNAcylation which are distinct to mucin-type *O*-glycans. These other types of *O*-glycosylation are comparatively rare, require less building blocks, and are found on specific proteins. For example, *O*-GlcNAc is a dynamic protein modification found on specific cytoplasmic, nuclear, and mitochondrial proteins with key roles in cellular physiology and progression of many diseases [15]. Despite the importance of this modification, *O*-GlcNAc is difficult to detect by standard biochemical methods due to its lability in mass spectrometers [16].

The limited development of glycomics characterization compared to that of genomics and proteomics is usually attributed to the analytical challenges associated with the complete elucidation of diverse heterogeneous glycan structures. On the other hand, we are increasing our understanding of glycan-associated biological processes such as host–pathogen interactions and the immune response. Part 1 of this review summarizes methodological developments and technical innovations that have advanced our ability to structurally characterize glycans. While Part 2 gives examples of the application of these technologies to determine glycosylation changes occurring in a range of human diseases.

2. Detection of glycoconjugates

The various ways of detecting glycans on glycoconjugates include chemical reactions, metabolic labeling, and recognition by specific lectins or antibodies. Classic colorimetric chemical methods for detection of glycans on proteins include periodic acid–Schiff's (PAS) stain that requires the oxidation of vicinal hydroxyl groups on the monosaccharides; Alcian Blue, which is a charge-based stain typically used to visualize proteoglycans and mucopolysaccharides [17,18]; and the phenol and concentrated sulfuric acid (PSA) stain that is used to detect oligosaccharides, polysaccharides, and their derivatives at submicro amounts of sugars [19]. These colorimetric methods provide a quick look at the rough abundances of glycans (compared to glycoprotein standards) or distribution of carbohydrates across a histological preparation. However, detailed information of the specific glycoconjugate cannot be revealed using these methods alone.

Lectins and glycan-specific antibodies are used in glycan analysis to discriminate specific structural features of oligosaccharides such as α2-6-linked SA (*Sambucus nigra* lectin), sialyl-LewisX (E-selectin), and polysialic acids (mAb-735). These lectins (Table 1) and glycan-specific antibodies have been used in glycan visualization on histological slides, protein blots, and ELISA-type assays as a primary glycoconjugate-binding protein [20–22]. Broad specificity lectins are also used for enrichment of glycoproteins from a complex protein mixture for detailed analysis [23–25]. However, it is important to note that lectins and glycan-specific antibodies only react to the glyco-epitope that they target and do not necessarily reflect the whole glycoconjugate structure.

Glycan metabolic labeling using a chemical reporter such as an azide is another powerful tool to detect and track the glycosylation state of live cells [26] and animals [27,28] or specific glycoproteins [29]. Feeding nonnatural, acetylated monosaccharides containing the azide functional group (*N*-azidoacetylmannosamine, *N*-azidoacetylgalactosamine, *N*-azidoacetylglucosamine, and 6-azidofucose) triggers the cellular salvage pathway to convert them into azide-labeled nucleotide sugar substrates [30]. The glycan biosynthetic machinery then incorporates the azide-labeled sugars into the glycoconjugates and provides a unique functional group for specific targeting using strained-alkyne or phosphine chemical probes by biocompatible copper-free click chemistry. Apart from direct glycan imaging, glycan metabolic labeling also provides new tools for glycoprotein enrichment [31], glycoproteomics [32], and studies on *O*-GlcNAcylation [33].

Table 1. Commercially available lectins and their glycan structure specificity as described in the literature.

Lectin	Glycan affinity
Mannose-binding lectins	
Con A (Concanavalin A)	Branched α-mannosidic structures; high-mannose type, hybrid type and biantennary complex type N-?glycans
GNA (*Galanthus nivalis* lectin)	α1–3 and α1–6-linked high-mannose structures
LCH (Lentil lectin)	Fucosylated core region of bi- and triantennary complex type N-?glycans
Galactose/N-acetylgalactosamine-binding lectins	
RCA (*Ricinus communis* agglutinin)	Galβ1-4GlcNAcβ1-R
ECL (*Erythrina cristagalli* lectin)	Galβ1-4GlcNAcβ1-R
PNA (Peanut agglutinin)	Galβ1-3GalNAcα1-Ser/Thr (T-Antigen)
AIL (*Artocarpus integrifolia* lectin/ Jacalin)	(Neu5Ac)Galβ1-3GalNAcα1-Ser/Thr (T-Antigen)
VVL (*Vicia villosa* lectin)	GalNAcα-Ser/Thr (Tn-Antigen)
N-acetylglucosamine-binding lectins	
WGA (Wheat germ agglutinin)	GlcNAcβ1-4GlcNAcβ1-4GlcNAc Neu5Ac
PHA (Phytohemagglutinin)	GlcNAc
Sialic acid - binding lectins	
SNA (*Sambucus nigra* lectin)	Neu5Acα2-6Gal(NAc)
MAL (*Maackia amurensis* leukoagglutinin)	Neu5Ac/Gcα2-3Galβ1-4GlcNAcβ1
Fucose-binding lectins	
UEA (*Ulex europaeus* agglutinin)	Fucα1-2Gal
AAL (*Aleuria aurantia* lectin)	Fucα1-2Galβ1-4(Fucα1-3/4)Galβ1-4GlcNAc GlcNAcβ1-4(Fucα1-6)GlcNAc

Use of azide-labeled nucleotide sugars has also been applied to therapeutic monoclonal antibodies as a novel method to create site-specific antibody-drug conjugates [34]. While glycan metabolic labeling provides great potential for visualizing the glycan landscape on cells, the underlying structural specifics of the glycome is not revealed and requires orthogonal methodologies for the investigation of the structural glycome.

3. Structural determination of protein glycosylation

The complexity associated with protein glycosylation analysis is due to the macro- and microheterogeneity of the oligosaccharides attached to the proteins. The microheterogeneity results from the range of glycan structures that can be attached at each glycosylation site on a protein, while the degree of site glycosylation confers the macroheterogeneity. The three usual approaches to study protein glycosylation are (a) characterization of intact glycoproteins, (b) characterization of protease-digested glycopeptides, and (c) structural analysis of glycans released from proteins.

3.1. Characterization of intact glycoproteins

Mass spectrometric (MS) characterization of intact glycoproteins is complicated by the extensive heterogeneity of the attached glycan moieties. The presence of these carbohydrate moieties further decreases the efficiency of ionization. The MS characterization of intact glycoproteins has been reported using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF), nanoelectrospray ionization (nESI), and capillary electrophoresis (CE) coupled to MS.

MALDI-based analysis of intact glycoproteins has been achieved with limited success for small proteins such as ribonuclease B, ovalbumin, bovine fetuin, and recombinant human erythropoietin [35–37]. These studies have revealed the significance of the proper matrix and optimum MS parameters that influence the detection of various glycoforms [36–38]. Novel approaches using in-source decay of intact glycoproteins induced by hydrogen radical transfer from the matrix produced several intense c-, y-, and z-type ions that enabled the site-specific assignment of mucin-type *O*-glycans on 17–18.5 kD MUC1 hexarepeat domains [39]. Analysis of large glycoproteins (~>100 kDa) with multiple glycosylation sites by MALDI-TOF-MS produces only an average molecular mass and is unable to resolve the molecular ions of the different glycoforms [40,41].

Glycoprotein analysis of intact glycoproteins by ESI-MS analyses is challenging due to the less-efficient desolvation resulting from the glycan heterogeneity and adduct formation in conventional ESI. The degree of glycosylation also limits detection, as glycans cover large surface areas of proteins thereby reducing efficient ionization. The resulting decreased charge state reduces the range of m/z analyzed by ESI-MS instruments. These technical difficulties are to some extent overcome by the application of nanoelectrospray as shown by Wilm and Mann [42] in the analysis of ovalbumin glycoforms. The coupling of nano-ESI with high-resolution mass analyzers such as time-of-flight (TOF) analyzers have also produced well-resolved glycoforms of bovine α1-acid glycoprotein [43] and cellulases purified from *Trichoderma reesi* [44]. Similarly, Nagy et al. showed the high-resolution α1-acid glycoprotein glycoforms by ESI-Fourier transform ion cyclotron resonance (FTICR) MS [45]. Heck and coworkers recently demonstrated the glycosylation analysis of native human erythropoietin using high-resolution native MS. The work elegantly demonstrated the characterization of site-specific glycans with minimal sample preparation and analysis time required to quantify glycan composition without ionization bias [46].

CE-MS at present provides great success for resolving the glycoforms of highly glycosylated proteins. Several studies have shown the application of CE-MS to almost completely resolving the various glycoforms of biologically relevant glycoproteins such as human plasma antithrombin [47] and recombinant erythropoietin [48,49]. The high resolving power of CE in analyzing glycoforms has potential for high-throughput screening of recombinant glycoproteins. One limitation is that there is less information obtained about the structural features of the attached glycans by this type of analysis. However, the integration of CE-MS with other orthogonal methodologies can mitigate this issue. For example, Takur et al. [50] demonstrated the characterization of 60 glycoforms of recombinant human chorionic gonadotrophin using CE coupled to a high-resolution FTICR MS; subsequent analysis of the tryptic glycopeptides enabled site-specific glycan variant identification.

3.2. Characterization of intact glycopeptides

Glycoproteins can be enzymatically digested to generate peptides, which overcomes the problem of mass-limited resolution of intact glycoprotein analysis in MS methods. Intact glycopeptide analysis (glycoproteomics) has the benefit of

allowing the identification of the site-specific heterogeneity of glycoproteins. There are several approaches to study glycopeptides by MS; typically, the sample is digested with specific endoproteinases (trypsin, chymotrypsin, Asp-N, Glu-C, and Lys-C) and the resulting peptides are analyzed by tandem MS [51]. Thaysen-Andersen et al. have recently reviewed in detail the advances in LC-MS/MS-based glycoproteomics [52], and will not be discussed in detail here. Similar to intact glycoprotein analysis by MS, intact glycopeptide analysis has its own challenges. Glycopeptides are often of low abundance compared to the non-glycosylated peptides and are represented with lower signal in the MS spectrum of the total peptide digest due to site-specific heterogeneity occurring at specific sites results in the signal intensities of the glycopeptides being relatively low. For example, human erythrocyte CD59 has over 100 different glycans on a single glycosylation site [53]. Furthermore, glycopeptides ionize poorly relative to the non-glycosylated peptides [54].

To alleviate these shortcomings, glycopeptides are usually enriched prior or during MS analysis [55]. Lectin affinity chromatography is a commonly used tool to remove non-glycosylated peptides and concentrate on the retained glycopeptides [24,56]. Typically, broadly selective lectins such as Con A are used [57,58] or alternatively, sequential or multi-lectin columns are used for the selective enrichment of glycopeptides [59,60]. Hydrophilic interaction chromatography (HILIC) is gaining importance in the enrichment of glycopeptides and relies on the added hydrophilicity conferred to glycopeptides due to the presence of the attached glycans. HILIC materials such as cellulose, sepharose, silica, aminopropyl, and zwitterionic types have been used for glycopeptide enrichment prior to MS analysis [61–64]. Boronic acid-based stationary phases also possess the potential to enrich for glycopeptides. Boronic acid is able to bind reversibly to vicinal cis-diol OH groups that are abundantly present in glycans. Boronic acid has recently been developed as the functionalized group on the surface of polymer particles or magnetic beads for chromatographic enrichment of both N- and O-glycopeptides [65–68].

Franc et al. showed that site-specific heterogeneity can also be determined by off-line separation and characterization [69]. They demonstrated this by simple fractionation of the hinge-region O-glycopeptides from IgA1 by reversed-phase liquid chromatography using a microgradient device and identified the glycopeptides by MALDI-TOF/TOF. This allowed the identification of isomeric O-glycoforms having the same molecular mass, but a different glycosylation pattern.

Recent developments in commercial availability of hybrid mass spectrometers with high-resolution mass analyzers have tremendously improved glycopeptide detection and analysis [70–72]. Researchers have attempted to use several fragmentation options such as higher-energy collision dissociation (HCD)/collision-induced dissociation (CID) and ETD sequentially to combine fragmentation information to decipher glycan- and peptide-specific information [73,74]. A data-dependent product ion trigger emanating from the high abundant oxonium ions, commonly observed in HCD fragmentation of glycopeptides, is used to activate subsequent ETD events [75]. This approach dedicates more instrument time for detailed ETD analysis of glycopeptides. The Heck lab recently introduced a hybrid dissociation method that combines ETD and HCD, called EThcD. After an initial electron-transfer dissociation step, all ions including the unreacted precursor ions are subjected to a supplemental energy which yields b/y- and c/z-type fragment ions in a single spectrum [76,77]. This unique hybrid fragmentation has already shown immense promise in glycopeptide analysis [72,78]. Although these novel methods provide valuable insight into site-specific occupation of glycan structures, very little information is currently available on the glycan structural features other than glycan composition. The glycan fragment ions for terminal substructures such as sialic acids or Lewis antigens can be detected but without any linkage information.

A proteomic strategy used to enrich and identify the sites of glycosylation on glycopeptides is by the selective covalent attachment of the glycopeptides to hydrazide resin after periodate oxidation of the carbohydrate moieties [79]. The glycosylation site is identified by releasing the attached N-glycans with the enzyme PNGaseF, and using MS to identify the deglycosylated peptides by the increased mass shift of the modified asparagine of $+1Da$ in water or $+3Da$ in $H_2^{18}O$. However, this approach is restricted to N-linked glycopeptide site identification and any information on the previously attached glycan moiety is lost. A related approach for identifying O-glycosylation sites is to release the glycans from the resin-bound O-linked glycopeptides using ammonia or ethylamine after PNGaseF release of the N-glycans [80,81]. The peptide integrity is maintained and the O-glycosylated serine or threonine site is labeled with a mass increment of 1Da (for ammonia) and 27Da (for ethylamine), that can be deduced by MS analysis of the deglycosylated peptides.

Currently the information garnered from intact glycopeptide analysis yields limited information on the fine structure of the attached glycan moiety, providing monosaccharide compositions but lacking detailed characterization of the terminal structural motifs. Nevertheless, qualitative and quantitative information on both the micro-and macroheterogeneity of glycoproteins can be obtained.

3.3. Structural analysis of glycans released from proteins

The characterization of N- and O-glycans released from proteins (glycomics) still remains the best currently available approach for the determination of the complete detailed glycan structural heterogeneity including monosaccharide composition, sequence, branching, and linkages as displayed on the surface of a cell. The remainder of this review (Part 1) will thus focus on the state-of-the-art and recent advances made in the characterization of the glycan structural motifs presented on the glycans released from proteins, and then give some examples in Part 2 of where this level of detail is required for an understanding of the difference between health and disease.

4. Characterization of the structure of glycan structural motifs

4.1. Release of N- and O-glycans from glycoproteins

The removal of N-linked glycans from proteins is facilitated by the use of the enzyme endoglycosidase peptide-N-glycosidase

F (PNGase F). This endoglycosidase cleaves *N*-glycans by hydrolysis of the glycosidic bond between the asparagine residue of a glycosylated protein and the reducing end GlcNAc of the attached oligosaccharide. *N*-glycans that contain an α1-3-linked Fuc attached to the reducing end GlcNAc (found in plants and insects) cannot be released by PNGase F but are released from glycopeptides by another endoglycosidase known as PNGase A [82]. There have been some recent advancements to this routine enzymatic method, with one approach employing ultra-high-pressure cycling, which subjects proteins to pressures of up to 30,000 psi and where glycoproteins become sufficiently denatured to be deglycosylated in as little as 20 min with PNGase F remaining active under these conditions [83]. Another method utilizes microwave radiation to assist the enzymatic cleavage of glycans from monoclonal antibodies in less than 10 min, whereas other glycoprotein standards required up to 1 h [84]. Recently, commercial companies such as New England Biolabs (Rapid PNGase F) and Bulldog Bio (PNGase F Prime) claim to have recombinantly engineered their enzymes for increased activity and wider coverage to enable rapid high-throughput analysis. A recombinant rice amidase called PNGase Ar (New England Biolabs) is able to remove glycans from intact plants glycoproteins and insect cells. Yan et al. showed that PNGase Ar was as effective in releasing *N*-glycans from *C. elegans* when compared to the use of hydrazine and far superior to traditional PNGase F and A release in this organism [85].

Chemical release methods are predominantly used to release the *O*-glycans from proteins, due to the lack of a general *O*-endoglycosidase. The only available endoglycosidase, *O*-glycanase, only cleaves the Core 1 type *O*-glycan (Galβ-3GalNAc) without any other modifications. Reductive β-elimination, using a strong alkaline reaction to cleave the glycoside-peptide bond, followed by reduction of the reducing terminus with $NaBH_4$ to prevent alkaline degradation of the released oligosaccharide, is the most widely used chemical method to release *O*-glycans [86]. An alternative approach is the glycan-releasing reaction of anhydrous hydrazine in which *N*- and/or *O*-glycans can be released under different conditions; *N*-glycans are released from glycoproteins using more vigorous conditions (85–100°C, 5–16 h) [87] whereas *O*-glycans are released and recovered under milder conditions (60°C, 6 h) [88]. Hydrazinolysis of *O*-glycans produces intact glycans with free reducing termini in high yield, but a major concern with this method is the occurrence of undesirable peeling of the reducing terminus and the restricted availability of the reagents. Other nonreductive chemical methods of release of *O*-glycans using dimethylamine [89] and ammonia [90] have also been described. These reactions claim to be quick, have low level of peeling, and provide free reducing ends that may be labeled.

The Cummings lab recently described a simple oxidative strategy using household bleach to release all types of free reducing *N*-glycans and *O*-glycan-acids from glycoproteins [91]. They showed by controlled treatment with sodium hypochlorite (NaClO) in commercial bleach, *N*- and *O*-glycans can be selectively released from glycoproteins. The further demonstrated the applicability of this method in releasing large quantities of glycans for chromatographic separation, MS sequencing and their use in glycan arrays for functional glycomics.

4.2. Nuclear magnetic resonance

Complete and unambiguous assignment of glycans including monosaccharide constituents, anomericity, linkages, and modifications is still best determined by nuclear magnetic resonance (NMR) [92]. In particular, NMR has gained significant importance in detecting relevant features in glycan-receptor interaction studies [93]. Recently, saturation transfer difference (STD)-NMR has been shown to be a robust and powerful approach to revealing the fine details of the molecular recognition between glycans and their binding receptors. For example, this technique provided valuable information for the design of inhibitors of Rhesus rotavirus and *Vibrio cholera* infections by enabling the detailed characterization of the molecular binding of various SA derivatives to Rotavirus VP8 subunit and *Vibrio cholera* sialidase, the causative agents of these two infections, respectively [94,95]. However, the major drawback of NMR analysis is the requirement of a large quantity of highly purified glycan, restricting its applicability as usually not enough sample can be obtained from biological sources [96].

4.3. Microarray-based analysis of glycans

Microarray-based technology is increasingly used in glycomics to identify glycan-binding partners and has been applied to the determination of glycan structural motifs. Glycan microarrays consist of immobilized glycans on glass slides and are used to characterize interactions between glycans and carbohydrate-binding proteins of interest [97,98,99]. Lectin and antibody-lectin sandwich microarrays have also been used to study glycosylation profiles. A typical lectin microarray consists of lectins immobilized as spots, either by activated aldehyde or epoxy groups, or adsorbed non-covalently on nitrocellulose and hydrogel slides. Glycoproteins, glycans, or whole cells are probed against the lectin microarray to identify specific glycan-binding structural motifs. Detection is usually carried out by fluorescent detection of the labeled ligands or through specific antibodies directed against the glycoprotein of interest [20,100]. Patwa et al. demonstrated the application of lectin microarray in the profiling of serum glycans [101]. Using evanescent-field fluorescence detection, weak lectin–glycan interactions under equilibrium conditions were measured at very low background, since the excitation wavelength of the evanescent field does not extend beyond 200 nm. Fluorescently labeled standard glycoproteins of known glycosylation were probed against 39 different lectins and analyzed quantitatively by this approach [102].

Lectin microarray profiling has multiple advantages over other glycan profiling techniques in terms of high throughput, low sample processing, and speed; however, it has also certain limitations, such as the low binding of lectins contributing to ligand loss during washing steps, low reproducibility, and limited glycan motif specificities. To the latter point, several studies have suggested that lectins have varying affinities toward several glycan motifs rather than absolute specificity for one structure [21,23].

4.4. Capillary electrophoresis

CE is a valuable separation system, widely used in the analysis of released glycans [103]. Commonly the glycans analyzed by CE are labeled by a charged chromophore or fluorophore such as 8-aminopyrene-1,3,6-trisulfonic acid, trisodium salt (APTS). A typical CE system consists of a silica separation capillary, a high voltage power supply, a detector, and an autosampler. The silica capillary contains surface silanol groups (Si-OH) that become ionized under electric potential. Usually, a very small amount of glycan sample analyte (1–50 nl) is required to separate well-resolved glycans in short run times [104]. Some drawbacks of this method for determining glycan structures include the need to pre-fractionate and label glycans and the high sample concentration required for detection post separation.

High isomeric selectivity can be obtained in CE analysis of fluorescently labeled glycans through the use of borate buffer that complexes with glycans. In noncomplexing buffers, the separation is usually attributed to size, charge, and to a lesser extent the shape of the glycan moiety [105]. Glycan peaks from CE are usually identified by (a) correlation of retention time with Glc units of a maltose ladder as reference; (b) comparison with glycan standards; (c) structure deduction through a series of exoglycosidase digestions; and (d) by coupling with detectors such as that used in DNA sequencers, laser-induced fluorescence (LIF) and MS [105].

4.5. Liquid chromatographic separation of glycans

High-performance liquid chromatography (HPLC) is an established separation technique for biomolecules and is frequently applied to the separation of glycans. Although HPLC has lower throughput capabilities than CE, it has a higher resolving power and better multiplexing with other analytical devices, such as easy front-end coupling with other chromatographic methods and back-end coupling to fluorescent detectors and MS. The various chromatographic techniques used to separate a complex mixture of underivatized glycans include high-pH anion exchange chromatography (HPAEC), size exclusion chromatography (SEC), HILIC, reversed phase (RP) and reversed-phase ion-pairing (RP-IP) chromatography, and porous graphitized carbon (PGC) chromatography. Similar to CE, an HPLC strategy can also be used for fluorescence-based quantitation and characterization of labeled released glycans. An internal dextran ladder standard combined with a bioinformatic platform such as GlycoBase/AutoGU [106] allows confirmation of glycan linkages and overall structure by retention time calibration and is available in a commercial package (Waters Corporation) [107].

Though most of these stand-alone separation methods have been extensively used to characterize and quantify glycans from complex mixtures using comparison with glycan standards and using retention time as the main parameter for identification, they have difficulty identifying novel or previously uncharacterized structures. Coupling these separation techniques to other detection methods such as MS mitigates this issue to an extent.

4.5.1. High-pH anion exchange chromatography

Similar to HPAEC-pulsed amperometric detection (PAD) analysis of monosaccharides, underivatized glycans can also be separated by HPAEC. The major drawback that arises through this type of analysis is the high amount of salt needed for elution and separation, which is incompatible with subsequent MS analysis. Although developments of desalting cartridges online prior to ESI-MS have been described for the analysis of N-glycans [108], this separation method is less favored. Maier et al. recently demonstrated the detection of low abundant N-glycans from various sources using a prototype 1 mm I.D. size HPAEC CarboPac PA200 column coupled to PAD detection or on-line ion trap (IT) MS for glycan structural assignment and quantification [109].

4.5.2. RP chromatography

Free glycans are too hydrophilic to be adsorbed by RP matrices. Reductive amination of glycans with a hydrophobic tag is used to allow the retention of glycans on RP columns. Several tags have been used, such as 2-aminopyridine (2-AP), 2-aminobenzamide (2-AB), 2-aminobenzoic acid (2-AA), and others [110,111]. The RP retention is due to the hydrophobicity of the tag, while resolution is based on the features of the glycan [112].

Permethylated N-glycans have also been separated by RP on a C18 column chip and analyzed by online MS [104,113]. This technique is promising, as permethylated glycans produce tandem mass spectra with high structural information. Some associated drawbacks are the under-permethylation of the glycans, possible desialylation of the acidic glycans and artifacts such as epimerization of the reducing end [114].

An additional strategy to retain glycans on RP is by ion pairing (IP). In RP-IP, charged additives in the eluting buffer are used as IP agents to increase the retention of oppositely charged analytes. RP-IP coupled to MS analysis has been used to characterize acidic glycans, especially negatively charged glycosaminoglycans [115,116].

4.5.3. Hydrophilic interaction chromatography

HILIC offers separation on the basis of glycan hydrophilicity, that is, dependent on properties such as size, charge, composition, structure, and linkage. HILIC is normal-phase chromatography, where an aqueous phase is usually used as the eluting solvent with an amide-derivatized silica as stationary phase.

Both charged and uncharged glycans can be separated by HILIC using a gradient from high to low organic content. The high reproducibility of retention times has allowed the high-throughput analysis of glycans using retention time libraries and standards. A high-throughput HILIC automated workflow in a robotic 96-well plate format with a workflow from sample processing through to data interpretation has been developed by Royle et al. [117]. These studies used fluorescently labeled glycans (2-aminobenzamide (2AB)) and detected them by fluorimetry. Wuhrer et al. showed that the structures of both labeled and underivatized glycans can be analyzed by the coupling of HILIC to ESI-MS [118].

There are several HILIC phases with interesting specificities and selectivities available for enriching hydrophilic molecules.

An interesting comparison test study was conducted by Ruhaak et al. to compare different HILIC phases (cellulose, sepharose, diol-bonded silica beads, and Biogel P-10) before subjecting the samples to a 48-channel multiplexed CGE-LIF analysis [119]. APTS-labeled glycans were purified from the reaction mixture that contained unused label, detergents, and proteins. Cellulose and Biogel P10 desalting gave the best results.

4.5.4. PGC chromatography

Porous glassy carbon as a chromatographic support matrix was modified by Knox and Gilbert to produce PGC that displayed good stability and chromatographic performance [120]. Earlier studies using PGC for chromatography suggested a reversed-phase-type behavior due to the proportional increase of retention of increasing hydrocarbons, but this does not explain the unique properties of isomeric and charge resolution. The retention mechanisms of glycans on PGC are still only vaguely comprehended with hydrophobic, ionic, and polar retention effects on graphite as known contributors. PGC shows superior resolution of native nonreduced and reduced glycans compared to conventional phases. Lately, the separation of permethylated glycans by PGC has also been reported [113,121]; other advancements include the packing of PGC into nanoscale chromatography chips for nano-LC-MS-based analysis of glycans [122].

The preparative sequential elution of neutral and acidic glycans using specific additives to the mobile phase was shown by Packer et al. [123], while Pabst and Altmann studied the influence of ionic strength, pH, and temperature on the retention of glycans by PGC [124]. They showed that when ionic strength is reduced while pH is maintained constant, acidic N-glycans are retained longer. Increased pH also increased retention time of acidic N-glycans with the neutral glycans unaffected, and a temperature increase resulted in stronger retention of all N-glycans. PGC has been shown to have remarkable selectivity in separating structurally similar glycan isomers. Several studies on both N- and O-glycans have described the baseline resolution of many structural glycan isomers [125,126].

An analytical development that has increased the sensitivity and improved reproducibility of nano-flow LC is the microfluidic chip [127]. The LC chip consists of a trapping column, switching valve, and LC column integrated and interfaced directly to the inlet of an ESI-based mass spectrometer. The chromatographic packing materials include reversed-phase, HILIC and PGC. These configurations have been used in several glycomic studies including the characterization of native milk oligosaccharides, permethylated glycans derived from blood serum glycoproteins from control individuals and late-stage breast cancer patients, and salivary glycoproteins [128]. A variant of this microfluidic chip incorporates an immobilized PNGase F reactor that readily releases glycans from intact IgG monoclonal antibodies in 6 s to claim a digestion efficiency of approximately 98% [129]. Unfortunately, this microfluidic chromatographic chip is no longer available from the manufacturer.

4.6. Glycosidase digestions

Structural elucidation of glycans by the enzymatic cleavage of specific exoglycosidases and endoglycosidases is a valuable strategy that complements CE, HPLC, and MS analyses [130]. Exoglycosidases are routinely used to confirm terminal glycan epitopes such as Lewis fucosylation linkages and SA linkages. An inherent disadvantage with this strategy is the multiple experimental analyses required, especially if sample has limited or low-abundance glycans.

4.7. MS analysis of released glycans

The coupling of MS with glycan separation methodologies has become the most favored and powerful technique for structural analysis of glycans. Some advantages of MS-based glycan analysis include high sensitivity, low sample requirement, and analytical versatility. MS analysis of peptides is relatively well defined and performed at high throughput, while, until reliable purpose-specific software is developed, glycan analysis by tandem MS data is particularly complex and labor-intensive to interpret. The tandem MS behavior is dependent on various factors such as product ion patterns, the type of derivative, and the fragmentation method.

The two main MS ionization techniques used currently for glycan analysis are MALDI and ESI.

4.7.1. Matrix-assisted laser desorption ionization

MALDI-MS-based structural characterization of glycans is an extensively used technique. Detailed information about the application of MALDI to glycan analysis, including matrices that are of particular use for carbohydrates, can be found in a series of comprehensive reviews by Harvey [131,132]. The MALDI-MS sample preparation is quite simple; the sample is mixed with a matrix solution and spotted onto a target plate. Crystals of the matrix-sample mixture are formed as the sample spots dry. Their strong optical absorption in either UV or IR range can rapidly and efficiently absorb the laser irradiation. A pulsed beam laser causes the matrix to transfer protons to the analyte molecules resulting in desorption and ionization of the sample molecules are extracted into the mass analyzer in the positive ionization mode. There are a wide variety of matrices available for glycan analysis, though the earliest developed matrix, 2,5-dihydroxybenzoic acid (DHB), remains the most popular. MALDI predominantly produces $[M+Na]^{1+}$ ions from positive mode ionization of glycans; other cations can also be introduced by doping the matrix with appropriate salts. Negative ions have also been reported, mainly from the use of β-carbolines as matrix.

A major advantage of MALDI is that contaminants and relatively high salt concentrations do not affect the sensitivity compared to other ionization methods. In addition, the glycan profile obtained is less complex because of the mainly singly charged ion formation. Coupling MALDI to a TOF/TOF analyzer can aid in accurate mass fragment analysis. Sialylated glycans however are quite labile when generated in the vacuum MALDI source without cooling of the newly formed ions, and have been reportedly lost in varying amounts in reflectron TOF, TOF/TOF, and qTof instruments [133,134]. To overcome

this problem, SAs can be permethylated, or the ions analyzed after they dissipate energy in an intermediate pressure MALDI source, prior to TOF analysis [135,136].

A recent development that promises to impact glycan analysis by MALDI-MS is the specific derivatization of the SA residues [137,138]. The reagent 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM) has been used in an esterification reaction with methanol that produces α2,3-linked SA lactones, whereas α2,6-linked SAs are esterified. This differential labeling creates a distinguishable 32 Da mass difference between α2,3- and α2,6-linked sialylated structures. Furthermore, these derivatized SAs were much more stable than their natural counterparts in MALDI analysis.

4.7.2. Electrospray ionization

Ionization by electrospray is obtained by introducing the liquid analyte by forced flow through a capillary with a fine nozzle. The high voltage applied to the nozzle creates a spray of charged droplets that enter the mass analyzer. These droplets are usually dried by high temperature, and by the flow of a drying gas, typically nitrogen. The electric charge density of the droplets increases as their size decreases, and eventually the mutual charge repulsion overcomes the surface tension (Rayleigh limit [43,139]) of the droplets resulting in their division [86].

ESI is usually referred to as a soft ionization technique, since it imparts relatively low energy to the analyte molecules during desolvation. Therefore, it is useful in the analysis of glycans with unstable modifications such as sialylation and sulfation [139]. Unlike MALDI-MS, ESI-MS is sensitive to salts and other contaminants, and therefore is usually best used after chromatography using volatile buffers. ESI usually produces multiply charged glycan ions that are advantageous for MS analyzers with limited mass range, albeit increasing the complexity of interpretation of the resultant spectra.

Depending on the ion source conditions and additives to solvents, ESI produces various types of ions. In the positive ion mode, both $[M + nH]^{n+}$ and their sodium adducts are formed under mild ion source conditions [140]. Under negative ion mode, glycans form $[M - nH]^{n-}$ ions or adducts with various anions. Harvey showed that nitrate adducts were particularly effective in the analysis of underivatized large N-glycans [141].

4.7.3. Mass analyzers

MS measurements are carried out in the gas phase on ionized analytes. The mass analyzer is used to separate the ions and is fundamental to the performance of the MS. The important features of mass analysis are sensitivity, accuracy, resolution, and the data acquisition speed. The various mass analyzers used in the analysis of glycans are IT, TOF, quadrupole, and FTICR analyzers. Each analyzer is different in design and capability, and can be used alone or in tandem with another.

Though IT have limited resolution, their MS^n fragmentation capabilities make them attractive analyzers when coupled to ESI. Modern Q-TOFs and Orbitraps possess higher resolution and their very high mass accuracy enables detailed characterization of glycans, for example, the phosphate and sulfate modifications could be distinguished natively without further

sample processing [142]. Due to the increase in core facilities for proteomics and the desire to also perform glycan analysis, glycomics workflows that are compatible with instrument set-ups for proteomics have become increasingly popular. A strategy by Hsiao et al. [143] accomplishes delineation of permethylated glycan isomers using acidic RP C18 nanoLC separation. The use of high-resolution mass analyzers also allows for multiplexing of mass tags such as aminoxyTMT [144] and stable-isotope labeling [145].

5. Recent advances in glycan MS/MS analysis

The topic of tandem MS (MS/MS or molecular fragmentation) for glycan analysis has been recently reviewed [146] but in the context of glycan biomarkers of disease the potential of MS is the ability to validate not only the composition and sequence of all glycans in a complex mixture, but also to determine the branching and linkage structural glycoepitopes that are well-known to alter in many human diseases. This information is derived from tandem mass spectra of glycan analytes.

5.1. Fragmentation nomenclature

Given that fragmentation patterns of protein glycosylation can be quite complex, a systematic nomenclature of carbohydrate fragmentations was developed by Domon and Costello for FAB-MS/MS spectra and has since been adopted widely across the glycomics field [147]. Domon and Costello identified A, B, C, X, Y, and Z fragments; however, additional fragmentation pathways have since been characterized (Figure 1).

Spina et al. expanded upon fragmentation pathways defined by Domon and Costello to include a previously uncharacterized fragmentation mechanism in MALDI-TOF/TOF, denoting the fragments generated from this method as E, F, and G fragments which, in combination with existing A fragments, now allow identification of the linkage positions present and discrimination between isomeric sugars by MALDI [148]. Harvey et al. also identified D, E, and F fragments (denoted as D, E_2, F_2 for simplicity) as structurally informative features [149]. One such fragment, D, gives composition of the 6-antenna and identifies the presence of a bisecting GlcNAc.

The identification of these structurally important differentiating glycan structural fragments (examples in Table 2) has allowed glycobiologists to characterize glycan epitopes based on which fragments are present and to associate these detailed differences in glycan structure with many human diseases.

Despite the numerous structural fragments that can be generated from tandem MS of glycans, different fragmentation methods and glycan analyte forms can give varying fragment populations which impact on the level of structural detail that can be obtained from tandem MS.

Glycan fragmentation has been frequently studied in both CID and electron activated dissociation (ExD) modes. Each of these areas has had improvements in fragmentation with Orbitrap-based HCD and FTICR-based electron capture dissociation and electron detachment dissociation [151,152]. Another variable in fragmentation pattern and therefore structural information includes the sample preparation and sample
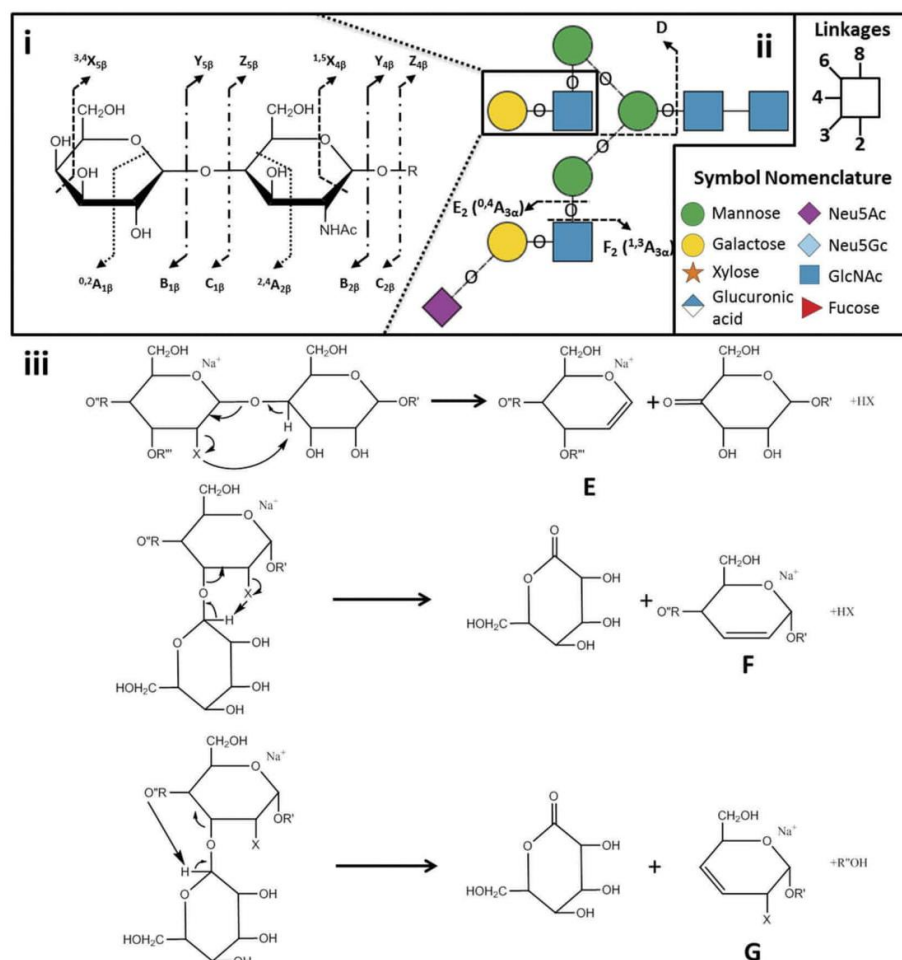
Figure 1. Fragmentation cleavages and nomenclature applicable to studying glycans. (i) Initial fragmentation nomenclature adapted from Domon and Costello [147]. (ii) Expansion of known carbohydrate fragments to include fragments indicative of antennae composition and glycan structure, adapted from Harvey et al. [149]. iii) Most recently identified fragmentation mechanisms involving gas phase rearrangement of glycan fragments, adapted from Spina et al. [148].

state such as permethylated glycans and reduction of the reducing end of native glycans. Both these variables impact the structural information obtained with tandem MS and have their own advantages and disadvantages.

5.2. Collision-induced dissociation
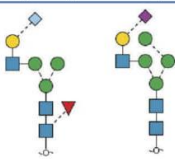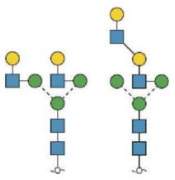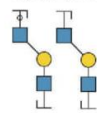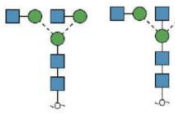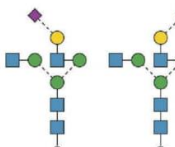
CID (also known as collisionally activated dissociation) is the most commonly used fragmentation type for glycan analysis probably due to its availability across a wide range of mass analyzers provided by instrument manufacturers. This fragmentation method, as the name suggests, involves collision of molecular ions with neutral and nonreactive gases such as nitrogen, argon, and helium which generates structurally informative fragment ions. This technique results in the detection of all previously described fragments (A, B, C, D, E, F,G, X, Y, and Z) (Figure 1). The CID glycan spectra are dependent on the instrument type, and especially on the collisional energy [104] since the collision results in fragmentation of chemical

bonds based on their lability. The resulting fragment ions are separated according to their m/z ratio and detected.

IT CID is one of the most frequently utilized tandem MS instruments used to characterize glycans partly due to its ability to perform MS^n, allowing glycans to be fragmented and structurally important ions to be further fragmented to confirm or ascertain more information. By using the MS in negative ion mode, more informative spectra can be generated through the generation of product ions consistent with glycosidic bond cleavage (B, C, Y, and Z fragments) and cross-ring cleavages (A and X fragments) [141]. Recent improvements in MS instruments containing IT have resulted in a wider variety of CID fragmentation methods such as being able to perform both beam-type and trap-type CID [153,154]. These improvements have resulted in a more versatile toolset for characterizing glycan structure and reporting of many different fragmentation patterns for even the same glycan.

Beam-type CID is a feature of triple quadrupole or QTOF-based instruments. These fragmentation techniques are

Table 2. Glycan structural information derived from tandem MS with examples of fragments that give this information.

Fragmentation information contribution	Example	Glycan-specific examples	Types of fragments	Example diagnostic ions
Composition	Identifying monosaccharide composition for isobaric glycans (both have same mass of 1730.6239 Da)		B, C, Y, and Z fragments	Neu5Ac (291 and 308 Da): Neu5Gc (306 and 324 Da):
Sequence	Presence of poly-*N*-acetyllactosamine chains		A, B, C, X, Y, and Z fragments	Poly-N-acetyllactosamine repeat (550 and 567 Da):
Branching	Number of antennae and/or bisecting		A, B, C, X, Y, and Z fragments	D-221 ion (presence of bisecting GlcNAc (509 Da)) [149]:
Linkages	α2,3 vs. α2,6 sialic acid linkage		A, B, C, E, F, G, X, Y, and Z fragments	Diagnostic ion for the presence of an α2,6 Neu5Ac isomer (307 Da) [150] -CO_2

denoted as being tandem-in-space due to the absence of a trapping mechanism which is commonly found in IT and FTICR mass analyzers which perform fragmentation tandem-in-time [155].

5.3. Electron activated dissociation

(ExD) of glycans contrasts with CID by providing fragmentation patterns based around cross-ring cleavages rather than glycosidic bond cleavage. ExD recently has been utilized to characterize and differentiate glycan isomers based on their linkage. In one example, CID and ExD were used in a complementary fashion to determine the composition, sequence, branching, and linkages of a permethylated disialylated biantennary *N*-glycan [152].

5.4. Standardization

Given the diverse types of glycomics analyses, the MIRAGE (Minimum Information Required for A Glycomics Experiment) initiative was developed to address the standardization of a wide range of glycosylation analysis challenges from sample preparation through to publication of data, including many aspects in between [156]. As a result of the glycomics community accepting these guidelines, it is hoped that the evaluation and reproduction of glycomics experiments will be improved as a result. Currently, three specific MIRAGE guidelines have been published on sample preparation [157], MS-based glycoanalytic data [158], and glycan microarrays [159].

5.5. Software tools

To complement the broad range of analytical techniques for quantitation of glycans, there is an accompanying variety of data analysis software being developed for analyzing data generated using these experimental techniques. For manual analysis of the data generated by these techniques, vendor-specific software is often used; however, as glycomics experiments become high throughput and more complicated, software packages dedicated to glycomics (varying from open-source to commercially available) are increasingly being developed. There are a few free software and databases available for glycomics, for example, ExPASy has provided a separate tab for glycomics (www.expasy.org/glycomics) that includes both internal and external links to various databases and tools, and a new website Glyconnect https://glyconnect.expasy.org/browser has been developed to integrate information on protein glycosylation structure and function from a range of bioinformatic resources. The UniCarb KnowledgeBase

(UniCarbKB) [160] is based on a curated database of glycan structures from glycoproteins as reported in scientific literature. An international collaborative effort has also been undertaken to provide a centralized resource for depositing glycan structures, compositions and topologies, and to retrieve accession numbers for each of these registered entries. This repository, called GlyTouCan [161,162], is available at http://glytoucan.org/. The emergence of these databases has been discussed in more detail elsewhere [163–166].

Software-assisted interpretation of glycan spectra can be performed by tools such as GlycoWorkBench [167], SimGlycan [168], Cartoonist [169], and MultiGlycan [170] among others [171] and UnicarbDB [172,173] collects annotated experimental MS/MS data on released glycans. Currently, there is little software available for high-throughput identification and quantitation of glycans and this remains an area that could be further improved; however, advances in proteomics software (Skyline [174] and MaxQuant [175]) could serve as useful starting platforms to apply to glycomics quantitation.

Although this review focuses predominantly on the structural characterization of glycans released from proteins, important intact glycopeptide analysis tools are briefly discussed here. GlycoMod [176], GlycoPep DB [177], and GlycoPep ID [178] are some of the freely accessible web-based tools for glycopeptide analysis. NetOGlyc and NetNGlyc are popular web-based tools for the prediction of potential glycan sites in the mammalian and human proteome [179,180]. GlycoMiner [181], Byonic [182], GlypID [183], and Protein Prospector [184] are promising software tools where sophisticated algorithms have been developed to characterize glycopeptides. Tsai et al. [171] have recently reviewed, in detail, the various features available in these and similar glycopeptide analysis software tools.

5.6. MALDI-MS tissue imaging of released glycans

Since its introduction, MS imaging (MSI) has provided unique advantages in the analysis of tissue specimens [185]. A broad spectrum of analytes ranging from proteins, peptides, glycans, lipids, small molecules, pharmaceutical compounds, endogenous, and exogenous metabolites can be analyzed in situ from tissue sections by this technique [186–194]. The data from such analysis is a pictograph that can be overlaid on the actual tissue sample and the location of the signal corresponds to the analyte's molecular mass. Although there are several types of MSI, the majority of analysis are performed using the MALDI-MSI approach. Any tissue section can be used for such analysis; typically these sections are covered with a suitable organic matrix material that co-crystallizes with the sample molecules. A laser beam irradiation ablates the matrix surface with the desorption and formation (for the most part) of singly protonated molecular species ([M + H]$^+$). These resulting ions traverse through a TOF analyzer, and their mass-to-charge (m/z) is determined. Using different matrices and technology (such as MALDI-TOF or MALDI-FT-ICR), the analyte classes can be chosen [195]. Typically a raster-scanned image of a tissue section provides a spatial resolution ranging from approximately 200 μm down to 20 μm generating a mass spectrum for every individually measured spot [196]. Recent technological improvements have contributed

to imaging phospholipids, neuropeptides, and drug compounds at a pixel size between 5 and 10 μm [196–200].

MALDI-MSI of released N-glycans has been recently described by a few groups ([191,201–204]. Unique glycan m/z were clearly able to discriminate between cancer and noncancerous tissues [202,205,206]. MALDI-MSI was also used to image released glycan masses from Tissue MicroArrays (TMAs) of liver cancer in which many small tumor tissue samples of different patients were assembled on a single slide, demonstrating the high-throughput capability. On-tissue SA derivatization that differentiated between the linkage isomers of SA and also stabilized these acidic residues has also been successfully reported [207].

The correlation of MALDI MSI molecular information with other traditional histology of the same tissue section is a potential advantage offered by this technique. Furthermore, the label-free and multiplex analysis offers the simultaneous analysis of hundreds to thousands of molecules in a single analysis.

5.7. Ion-mobility MS

A rapidly developing MS-based technique is ion-mobility spectrometry (IMS). Through this setup, gas-phase ion analytes are subjected to a series of collisions under low-electric field, and the ions thereby separate due to their collisional cross sections and mass-to-charge values; a separation relative to the overall shape of the molecule. This technique in glycan analysis has found favorable application to isomer separation and characterization. One limitation of tandem-MS techniques is the dependence on separation of the isomers before fragmentation. This challenge appears to have been recently addressed by separation and identification of isomeric glycans by selected accumulation – trapped ion mobility MS (TIMS). In this case, a mixture of isomeric permethylated tetrasaccharides was separated using TIMS and characterized using ExD [208]. While the MS setup is largely experimental at this stage, it has potential for the discrimination of glycan isomers at the level of a single monosaccharide linkage difference. Other IMS studies clearly showed the advantage offered by the technique in the separation of branched oligosaccharides from linear oligosaccharides [209]. Three distinct permethylated Man$_5$ N-glycan isomers from ovalbumin were separated and reproducibly characterized by Plasencia et al. [210]. The group of Peter-Katalinic have used computer-assisted assignment of N- and O-free and amino acid-linked glycans. Unique patterns of glycans specific to their m/z values and drift times were observed and recorded by IMS-MS for the de novo identification of the human urinome [211]. Hinneburg et al. recently demonstrated separation of different N-acetylneuraminic acid linkages on N-glycans in a site-specific manner on individual glycopeptides using IMS analysis of diagnostic fragment ions [212].

A major advantage of IMS glycan analysis is its potential application to rapid clinical diagnosis. As early as 2012, Isailovic et al. used a combination of IMS and multivariate analysis with principal component analysis (PCA) to characterize serum N-linked glycans from 81 individuals (28 with cirrhosis of the liver, 25 with liver cancer, and 28 apparently healthy). Supervised PCA analysis of the combined ion-mobility profiles of different mass-to-charge ratios of glycan ions improved the

delineation of diseased states with an analysis time of 2 min per sample [213].

6. Expert commentary

The intrinsic complexity of glycan moieties has been a major factor in the progression of glycomics. The advancements in proteomics and genomics analysis were primarily due to the characterization of their respective linear codes by DNA sequencers and MS sequencing, but in glycomics, no single technique is yet capable of total and unbiased analysis.

The currently available technologies of glycan analysis require the release of glycans from their carrier proteins if the detailed structural glycan data in terms of composition, sequence, branching, and linkage are needed to be known. Determination of the heterogeneity of glycosylation at a site-specific level is currently reliant on glycopeptide analysis that provides important glycan composition and occupation data, but only partial glycan structural detail. The new derivatization methods on linkage-specific SAs (α2, 3 and α2, 6) are compatible with glycopeptides [214], and structural modifications on the N-glycan core such as core fucosylation and bisecting GlcNAc can now be identified from glycopeptide tandem MS fragmentation [51,215]; however, structural features such as antennae branching, linkage of outer arm fucosylation, Lewis epitopes, and polylactosamine extensions at particular protein sites in a complex mixture still require improvements in the analytical technologies currently available. The presence or addition of specific structural glycan features on proteins has been shown to be critical to many biological functions, but detailed analysis of the glycan heterogeneity present at a particular site is still required. Clearly, as far as can be seen, MS analysis is the core technology that will solve these shortcomings and we are seeing important steps in intact glycoprotein MS, ion mobility, and software development for spectral interpretation that promise to simplify the analysis of both the correct glycan structure and the specific site heterogeneity.

The recent years have brought about many inventive approaches and resourceful methods that enable the unraveling of the glycan code. To a large extent, glycan analysis complexities have been reduced and the role of specific glycan epitopes have been identified as a consequence.

7. Five-year view: a speculative viewpoint on how the field will evolve in 5 years time

In the field of proteomics, a more developed area of biomolecular MS, there have been recent developments in the availability of data through open-access platforms such as PRIDE. These platforms need to be developed so they can be used for glycomics analysis, providing open access to data and improving data sharing that will stimulate further method development in glycomics. Glycomics standardization initiatives such as MIRAGE are facilitating these developments; however, open-source and freely available informatics tools are required for this aim to be accomplished.

Analytical methodology for glycomic and glycoproteomic analyses will continue to improve. Since MS-based glycomic analyses require expensive equipment and a fair amount of expertise, other alternative technologies such as arrays and novel antibody-based assays will also continue to promote glycobiological knowledge. The MALDI-MSI application to glycan analysis is still in its primary stages of development, but shows enormous potential in determining the topological distribution of glycans in tissue. This is highly relevant especially in understanding tumor biology and for rapid classification of cancer types or subtypes.

The next major challenge is to assign function to the glycosylation differences that we are observing in every biological perturbation that we study. The problem is that there is no single answer and that every molecular interaction is being found to involve unique glycan structure(s), often at a particular site, density, or neighborhood. It is to be hoped that as more molecular biologists and medical researchers discover that glycosylation is involved in the fine-tuning of their particular biological system, there will be more emphasis on the determination of these precise molecular structures that modify proteins and that are integral to just about every biological system. And this review does not even address the challenges of analyzing glycan structures attached to other conjugates such as glycolipids, proteoglycans, peptidoglycans, and lipopolysaccharides!

Key issues

- Glycans are involved in cellular processes important in development, cell proliferation, differentiation and morphogenesis. The complexity associated with protein glycosylation analysis is due to the macro- and microheterogeneity of the oligosaccharides attached to the proteins.
- The limited development of glycomics compared to that of genomics and proteomics is usually attributed to the analytical challenges associated with the complete elucidation of diverse heterogeneous glycan structures.
- Mass spectrometric characterization of intact glycoproteins is complicated by the extensive heterogeneity of the attached glycan moieties. The presence of these carbohydrate moieties further decreases the efficiency of ionization.
- Similarly, glycopeptides also ionize poorly relative to the non-glycosylated peptides. To alleviate these shortcomings, glycopeptides are usually enriched prior or during mass spectrometry analysis.
- The characterization of N- and O- glycans released from proteins is currently the best approach to determine glycan structural heterogeneity as displayed on the surface of a cell, including monosaccharide composition, sequence, branching and linkages.
- High-performance liquid chromatography is an established separation technique applied routinely in the separation of glycans.
- The coupling of mass spectrometry with glycan separation methodologies has become the most favored and powerful technique for structural analysis of glycans. Some advantages of MS based glycan analysis include high sensitivity, low sample requirement and analytical versatility.

- Glycan mass spectrometry fragmentation spectra provide valuable information in the elucidation of their structure. Different fragmentation methods and glycan analyte forms can give varying fragment populations impacting the level of structural detail obtained.
- MALDI-MS imaging is a new promising technique that offers label-free multiplex analysis of several glycan species capable of delineating tissues by their spatial localisation.
- Ion mobility MS is another rapidly advancing technology that can be used to characterize and identify glycans from complex mixtures.
- Glycomics standardization initiatives are facilitating regularization of data reporting, however open-source, reliable and freely available informatics tools are required for this aim to be accomplished.
- Assigning function to the observed glycosylation differences is an enormous challenge that requires several orthogonal approaches and multi-disciplinary studies.

Funding

Declaration of interest

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Brown JR, Crawford BE, Esko JD. Glycan antagonists and inhibitors: a fount for drug discovery. Crit Rev Biochem Mol Biol. 2007;42 (6):481–515.
2. Varki A, Lowe JB. Biological roles of glycans. In: Varki A, Cummings RD, Esko JD, editors. Essentials of glycobiology. New York (NY): Cold Spring Harbor; 2009. p. 75–88.
3. Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. 1993;3(2):97–130.
•• A comprehensive review on several glycan mediated roles. The review takes into account the several theories proposed for the different biological roles of glycans, and continues to explain that, while all of the theories are correct, exceptions to each can also be found.
4. Hardy MR, Townsend RR, Lee YC. Monosaccharide analysis of glycoconjugates by anion exchange chromatography with pulsed amperometric detection. Anal Biochem. 1988;170(1):54–62.
5. Dube DH, Bertozzi CR. Glycans in cancer and inflammation–potential for therapeutics and diagnostics. Nat Rev Drug Discov. 2005;4 (6):477–488.
6. Marth JD, Grewal PK. Mammalian glycosylation in immunity. Nat Rev Immunol. 2008;8(11):874–887.
7. Yarema KJ, Bertozzi CR. Characterizing glycosylation pathways. Genome Biol. 2001;2(5):REVIEWS0004.
8. Cummings RD. The repertoire of glycan determinants in the human glycome. Mol Biosyst. 2009;5(10):1087–1104.
9. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol. 2012;13(7):448–462.
10. Stanley P, Taniguchi N, Aebi M. N-glycans. In: Varki, A, Cummings, RD, editors. Essentials of glycobiology. 3rd ed. New York (NY):Cold Spring Harbor; 2015. p.99–111.
11. Schachter H. The joys of HexNAc. The synthesis and function of N- and O-glycan branches. Glycoconj J. 2000;17(7–9):465–483.
12. Brockhausen I, Schachter H, Stanley P, et al. O-GalNAc glycans. In: Varki A, Cummings RD, editors. Essentials of glycobiology. nd. New York (NY): Cold Spring Harbor; 2009.
13. Brockhausen I, Stanley P. O-GalNAc glycans. In: Varki A, Cummings RD, editors. Essentials of glycobiology. rd. New York (NY): Cold Spring Harbor; 2015. p. 113–123.
14. Varki A. Biological roles of glycans. Glycobiology. 2017;27(1):3–49.
15. Hart GW, Slawson C, Ramirez-Correa G, et al. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. Annu Rev Biochem. 2011;80:825–858.
16. Zachara N, Akimoto Y, Hart GW, et al. The O-GlcNAc modification. In: Varki A, Cummings RD, editors. Essentials of glycobiology. rd. New York (NY): Cold Spring Harbor; 2015. p. 239–251.
17. Kilcoyne M, Gerlach JQ, Farrell MP, et al. Periodic acid-Schiff's reagent assay for carbohydrates in a microtiter plate format. Anal Biochem. 2011;416(1):18–26.
18. Dong W, Matsuno YK, Kameyama A. A procedure for Alcian blue staining of mucins on polyvinylidene difluoride membranes. Anal Chem. 2012;84(20):8461–8466.
19. Dubois M, Gilles K, Hamilton JK, et al. A colorimetric method for the determination of sugars. Nature. 1951;168(4265):167.
20. Ribeiro JP, Mahal LK. Dot by dot: analyzing the glycome using lectin microarrays. Curr Opin Chem Biol. 2013;17(5):827–831.
•• A review focused on lectin microarray arrays and recent technological advancements.
21. Gupta G, Surolia A, Sampathkumar SG. Lectin microarrays for glycomic analysis. Omics. 2010;14(4):419–436.
22. Sun Y, Cheng L, Gu Y, et al. A human lectin microarray for sperm surface glycosylation analysis. Mol Cell Proteomics. 2016;15 (9):2839–2851.
23. Lee A, Nakano M, Hincapie M, et al. The lectin riddle: glycoproteins fractionated from complex mixtures have similar glycomic profiles. Omics. 2010;14(4):487–499.
24. Drake PM, Schilling B, Niles RK, et al. A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma. Anal Biochem. 2011;408(1):71–85.
25. Lee LY, Hincapie M, Packer N, et al. An optimized approach for enrichment of glycoproteins from cell culture lysates using native multi-lectin affinity chromatography. J Sep Sci. 2012;35(18):2445–2452.
26. Baskin JM, Prescher JA, Laughlin ST, et al. Copper-free click chemistry for dynamic in vivo imaging. Proc Natl Acad Sci U S A. 2007;104(43):16793–16797.
27. Chang PV, Prescher JA, Sletten EM, et al. Copper-free click chemistry in living animals. Proc Natl Acad Sci U S A. 2010;107(5):1821–1826.
28. Laughlin ST, Baskin JM, Amacher SL, et al. In vivo imaging of membrane-associated glycans in developing zebrafish. Science. 2008;320(5876):664–667.
29. Belardi B, De La Zerda A, Spiciarich DR, et al. Imaging the glycosylation state of cell surface glycoproteins by two-photon fluorescence lifetime imaging microscopy. Angew Chem Int Ed Engl. 2013;52(52):14045–14049.
30. Laughlin ST, Bertozzi CR. Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation. Nat Protoc. 2007;2(11):2930–2944.
• This detailed protocol illustrates the syntheses of the azido sugars *N*-azidoacetylmannosamine (ManNAz), *N*-azidoacetylgalactosamine (GalNAz), *N*-azidoacetylglucosamine (GlcNAz) and 6-

66

azidofucose (6AzFuc), and the detection reagents phosphine-FLAG and phosphine-FLAG-His6 for cellular metabolic labeling and subsequent glycan-profiling and visualization.

31. Roper SM, Zemskova M, Neely BA, et al. Targeted glycoprotein enrichment and identification in stromal cell secretomes using azido sugar metabolic labeling. Proteomics Clin Appl. 2013;7(5–6):367–371.

32. Woo CM, Felix A, Zhang L, et al. Isotope-targeted glycoproteomics (IsoTaG) analysis of sialylated N- and O-glycopeptides on an Orbitrap Fusion Tribrid using azido and alkynyl sugars. Anal Bioanal Chem. 2017;409(2):579–588.

33. Palaniappan KK, Hangauer MJ, Smith TJ, et al. A chemical glycoproteomics platform reveals O-GlcNAcylation of mitochondrial voltage-dependent anion channel 2. Cell Rep. 2013;5(2):546–552.

34. Agarwal P, Bertozzi CR. Site-specific antibody-drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. Bioconjug Chem. 2015;26(2):176–192.

35. Sottani C, Fiorentino M, Minoia C. Matrix performance in matrix-assisted laser desorption/ionization for molecular weight determination in sialyl and non-sialyl oligosaccharide proteins. Rapid Commun Mass Spectrom. 1997;11(8):907–913.

36. Gimenez E, Benavente F, Barbosa J, et al. Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 2007;21(16):2555–2563.

37. Gimenez E, Benavente F, Barbosa J, et al. Ionic liquid matrices for MALDI-TOF-MS analysis of intact glycoproteins. Anal Bioanal Chem. 2010;398(1):357–365.

38. Liu Z, Schey KL. Optimization of a MALDI TOF-TOF mass spectrometer for intact protein analysis. J Am Soc Mass Spectrom. 2005;16(4):482–490.

39. Hanisch FG. Top-down sequencing of O-glycoproteins by in-source decay matrix-assisted laser desorption ionization mass spectrometry for glycosylation site analysis. Anal Chem. 2011;83(12):4829–4837.

40. Lee BS, Krishnanchettiar S, Lateef SS, et al. Characterization of oligosaccharide moieties of intact glycoproteins by microwave-assisted partial acid hydrolysis and mass spectrometry. Rapid Commun Mass Spectrom. 2005;19(18):2629–2635.

41. Sanz-Nebot V, Benavente F, Gimenez E, et al. Capillary electrophoresis and matrix-assisted laser desorption/ionization-time of flight-mass spectrometry for analysis of the novel erythropoiesis-stimulating protein (NESP). Electrophoresis. 2005;26(7–8):1451–1456.

42. Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. Anal Chem. 1996;68(1):1–8.

43. Karas M, Bahr U, Dulcks T. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius J Anal Chem. 2000;366(6–7):669–676.

44. Hui JP, White TC, Thibault P. Identification of glycan structure and glycosylation sites in cellobiohydrolase II and endoglucanases I and II from *Trichoderma reesei*. Glycobiology. 2002;12(12):837–849.

45. Nagy K, Vekey K, Imre T, et al. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry of human alpha-1-acid glycoprotein. Anal Chem. 2004;76(17):4998–5005.

46. Yang Y, Liu F, Franc V, et al. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. Nat Commun. 2016;7:13397.

•• **This study provides an integrative approach, combining two advanced mass spectrometry-based methods, high-resolution native mass spectrometry, and middle-down proteomics, to analyze the micro-heterogeneity of intact glycoproteins such as human erythropoietin and the human plasma properdin.**

47. Demelbauer UM, Plematl A, Kremser L, et al. Characterization of glyco isoforms in plasma-derived human antithrombin by on-line capillary zone electrophoresis-electrospray ionization-quadrupole ion trap-mass spectrometry of the intact glycoproteins. Electrophoresis. 2004;25(13):2026–2032.

48. Balaguer E, Demelbauer U, Pelzing M, et al. Glycoform characterization of erythropoietin combining glycan and intact protein

analysis by capillary electrophoresis - electrospray - time-of-flight mass spectrometry. Electrophoresis. 2006;27(13):2638–2650.

49. Neususs C, Demelbauer U, Pelzing M. Glycoform characterization of intact erythropoietin by capillary electrophoresis-electrospray-time of flight-mass spectrometry. Electrophoresis. 2005;26(7–8):1442–1450.

50. Thakur D, Rejtar T, Karger BL, et al. Profiling the glycoforms of the intact alpha subunit of recombinant human chorionic gonadotropin by high-resolution capillary electrophoresis-mass spectrometry. Anal Chem. 2009;81(21):8900–8907.

51. Wuhrer M, Catalina MI, Deelder AM, et al. Glycoproteomics based on tandem mass spectrometry of glycopeptides. J Chromatogr B Analyt Technol Biomed Life Sci. 2007;849(1–2):115–128.

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

•• **This review focuses on the system-wide site-specific analysis of protein *N*- and *O*-linked glycosylation and the recent advances in LC-MS/MS-based glycoproteomics. A general discussion of experimental designs in glycoproteomics and sample preparation prior to LC-MS/MS is also covered.**

53. Rudd PM, Elliott T, Cresswell P, et al. Glycosylation and the immune system. Science. 2001;291(5512):2370–2376.

54. Toyama A, Nakagawa H, Matsuda K, et al. Deglycosylation and label-free quantitative LC-MALDI MS applied to efficient serum biomarker discovery of lung cancer. Proteome Sci. 2011;9:18.

55. Ito S, Hayama K, Hirabayashi J. Enrichment strategies for glycopeptides. Methods Mol Biol. 2009;534:195–203.

56. Kubota K, Sato Y, Suzuki Y, et al. Analysis of glycopeptides using lectin affinity chromatography with MALDI-TOF mass spectrometry. Anal Chem. 2008;80(10):3693–3698.

57. Catala C, Howe KJ, Hucko S, et al. Towards characterization of the glycoproteome of tomato (*Solanum lycopersicum*) fruit using Concanavalin A lectin affinity chromatography and LC-MALDI-MS/MS analysis. Proteomics. 2011;11(8):1530–1544.

58. Lee YC, Srajer Gajdosik M, Josic D, et al. Plasma membrane isolation using immobilized concanavalin A magnetic beads. Methods Mol Biol. 2012;909:29–41.

59. Qiu R, Regnier FE. Use of multidimensional lectin affinity chromatography in differential glycoproteomics. Anal Chem. 2005;77(9):2802–2809.

60. Durham M, Regnier FE. Targeted glycoproteomics: serial lectin affinity chromatography in the selection of O-glycosylation sites on proteins from the human blood proteome. J Chromatogr A. 2006;1132(1–2):165–173.

61. Wada Y, Tajiri M, Yoshida S. Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. Anal Chem. 2004;76(22):6560–6565.

62. Hagglund P, Bunkenborg J, Elortza F, et al. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. J Proteome Res. 2004;3(3):556–566.

63. Calvano CD, Zambonin CG, Jensen ON. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. J Proteomics. 2008;71(3):304–317.

64. Mysling S, Palmisano G, Hojrup P, et al. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. Anal Chem. 2010;82(13):5598–5609.

65. Lee JH, Kim Y, Ha MY, et al. Immobilization of aminophenylboronic acid on magnetic beads for the direct determination of glycoproteins by matrix assisted laser desorption ionization mass spectrometry. J Am Soc Mass Spectrom. 2005;16(9):1456–1460.

66. Monzo A, Bonn GK, Guttman A. Boronic acid-lectin affinity chromatography. 1. Simultaneous glycoprotein binding with selective or combined elution. Anal Bioanal Chem. 2007;389(7–8):2097–2102.

67. Ozohanics O, Turiak L, Drahos L, et al. Comparison of glycopeptide/glycoprotein enrichment techniques. Rapid Commun Mass Spe. 2012;26(2):215–217.

68. Desaire H. Glycopeptide analysis: recent developments and applications. Mol Cell Proteomics. 2013 Apr;12(4):893-901.

69. Franc V, Rehulka P, Raus M, et al. Elucidating heterogeneity of IgA1 hinge-region O-glycosylation by use of MALDI-TOF/TOF mass spectrometry: role of cysteine alkylation during sample processing. J Proteomics. 2013;92:299–312.

70. Zhao P, Viner R, Teo CF, et al. Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. J Proteome Res. 2011;10 (9):4088–4104.

71. Darula Z, Sherman J, Medzihradszky KF. How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. Mol Cell Proteomics. 2012;11(7):O111 016774.

72. Yu Q, Wang B, Chen Z, et al. Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. J Am Soc Mass Spectrom. 2017;28 (9):1751–1764.

73. Saba J, Dutta S, Hemenway E, et al. Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. Int J Proteomics. 2012;560391:2012.

74. Alley WR Jr., Mechref Y, Novotny MV. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. Rapid Commun Mass Spe. 2009;23(1):161–170.

75. Singh C, Zampronio CG, Creese AJ, et al. Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. J Proteome Res. 2012;11(9):4517–4525.

76. Frese CK, Altelaar AF, Van Den Toorn H, et al. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem. 2012;84(22):9668–9673.

77. Mommen GP, Frese CK, Meiring HD, et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A. 2014;111 (12):4507–4512.

78. Marino F, Bern M, Mommen GPM, et al. Extended O-GlcNAc on HLA Class-I-bound peptides. J Am Chem Soc. 2015;137(34):10922–10925.

79. Nilsson J, Ruetschi U, Halim A, et al. Enrichment of glycopeptides for glycan structure and attachment site identification. Nat Methods. 2009;6(11):809–811.

80. Hanisch FG, Jovanovic M, Peter-Katalinic J. Glycoprotein identification and localization of O-glycosylation sites by mass spectrometric analysis of deglycosylated/alkylaminylated peptide fragments. Anal Biochem. 2001;290(1):47–59.

81. Rademaker GJ, Pergantis SA, Blok-Tip L, et al. Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. Anal Biochem. 1998;257(2):149–160.

82. Tretter V, Altmann F, Marz L. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1-3 to the asparagine-linked N-acetylglucosamine residue. Eur J Biochem/FEBS. 1991;199(3):647–652.

83. Szabo Z, Guttman A, Karger BL. Rapid release of N-linked glycans from glycoproteins by pressure-cycling technology. Anal Chem. 2010;82(6):2588–2593.

84. Zhou H, Briscoe AC, Froehlich JW, et al. PNGase F catalyzes de-N-glycosylation in a domestic microwave. Anal Biochem. 2012;427 (1):33–35.

85. Yan S, Vanbeselaere J, Wols F, et al. Core richness of N-glycans of Caenorhabditis elegans: a case study on chemical and enzymatic release. Anal Chem. 2018 Jan 2;90(1):928-935.

86. Carlson DM. Oligosaccharides isolated from pig submaxillary mucin. J Biol Chem. 1966;241(12):2984–2986.

87. Patel T, Bruce J, Merry A, et al. Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. Biochemistry. 1993;32(2):679–693.

88. Alley WR Jr., Madera M, Mechref Y, et al. Chip-based reversed-phase liquid chromatography-mass spectrometry of permethylated N-linked glycans: a potential methodology for cancer-biomarker discovery. Anal Chem. 2010;82(12):5095–5106.

89. Zauner G, Koeleman CA, Deelder AM, et al. Mass spectrometric O-glycan analysis after combined O-glycan release by beta-elimination and 1-phenyl-3-methyl-5-pyrazolone labeling. Biochim Biophys Acta. 2012;1820(9):1420–1428.

90. Wang C, Fan W, Zhang P, et al. One-pot nonreductive O-glycan release and labeling with 1-phenyl-3-methyl-5-pyrazolone followed by ESI-MS analysis. Proteomics. 2011;11(21):4229–4242.

91. Song X, Ju H, Lasanajak Y, et al. Oxidative release of natural glycans for functional glycomics. Nat Methods. 2016;13(6):528–534.

92. Battistel MD, Azurmendi HF, Yu B, et al. NMR of glycans: shedding new light on old problems. Prog Nucl Magn Reson Spectrosc. 2014;79:48–68.

93. Gimeno A, Reichardt NC, Canada FJ, et al. NMR and molecular recognition of N-glycans: remote modifications of the saccharide chain modulate binding features. ACS Chem Biol. 2017;12(4):1104–1112.

94. Haselhorst T, Wilson JC, Thomson RJ, et al. Saturation transfer difference (STD) 1H-NMR experiments and in silico docking experiments to probe the binding of N-acetylneuraminic acid and derivatives to Vibrio cholerae sialidase. Proteins. 2004;56 (2):346–353.

95. Haselhorst T, Blanchard H, Frank M, et al. STD NMR spectroscopy and molecular modeling investigation of the binding of N-acetylneuraminic acid derivatives to rhesus rotavirus VP8* core. Glycobiology. 2007;17(1):68–81.

96. Mulloy B, Dell A, Stanley P, et al. Structural analysis of glycans. In: Varki A, Cummings RD, editors. Essentials of glycobiology. rd. New York (NY): Cold Spring Harbor; 2015. p. 639–652.

97. Oyelaran O, Gildersleeve JC. Glycan arrays: recent advances and future challenges. Curr Opin Chem Biol. 2009;13(4):406–413.

98. Liang CH, Wu CY. Glycan array: a powerful tool for glycomics studies. Expert Rev Proteomics. 2009;6(6):631–645.

99. Day CJ, Paton AW, Higgins MA, Shewell LK, Jen FE, Schulz BL, Herdman BP, Paton JC, and Jennings MP. Structure aided design of a Neu5Gc specific lectin. Sci rep. 2017;7(1):1495.28473713, PMID 28473713.

100. Feizi T, Fazio F, Chai W, et al. Carbohydrate microarrays - a new set of technologies at the frontiers of glycomics. Curr Opin Struct Biol. 2003;13(5):637–645.

101. Patwa TH, Zhao J, Anderson MA, et al. Screening of glycosylation patterns in serum using natural glycoprotein microarrays and multi-lectin fluorescence detection. Anal Chem. 2006;78(18):6411–6421.

102. Kuno A, Uchiyama N, Koseki-Kuno S, et al. Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. Nat Methods. 2005;2(11):851–856.

103. Zamfir A, Peter-Katalinic J. Capillary electrophoresis-mass spectrometry for glycoscreening in biomedical research. Electrophoresis. 2004;25(13):1949–1963.

104. Bielik A, Zaia J. Historical overview of glycoanalysis. In: Li J, editor. Functional glycomics. New York, USA: Humana Press; 2010. p. 9–30.

105. Zaia J. Capillary electrophoresis-mass spectrometry of carbohydrates. Methods Mol Biol. 2013;984:13–25.

106. Campbell MP, Royle L, Radcliffe CM, et al. GlycoBase and autoGU: tools for HPLC-based glycan analysis. Bioinformatics. 2008;24 (9):1214–1216.

107. Lauber MA, Yu YQ, Brousmiche DW, et al. Rapid preparation of released N-glycans for HILIC analysis using a labeling reagent that facilitates sensitive fluorescence and ESI-MS detection. Anal Chem. 2015;87(10):5401–5409.

108. Guignard C, Jouve L, Bogeat-Triboulot MB, et al. Analysis of carbohydrates in plants by high-performance anion-exchange chromatography coupled with electrospray mass spectrometry. J Chromatogr A. 2005;1085(1):137–142.

109. Maier M, Reusch D, Bruggink C, et al. Applying mini-bore HPAEC-MS/MS for the characterization and quantification of Fc N-glycans

from heterogeneously glycosylated IgGs. J Chromatogr B Analyt Technol Biomed Life Sci. 2016;1033–1034:342–352.

110. Pabst M, Kolarich D, Poltl G, et al. Comparison of fluorescent labels for oligosaccharides and introduction of a new postlabeling purification method. Anal Biochem. 2009;384(2):263–273.

111. Marino K, Bones J, Kattla JJ, et al. A systematic approach to protein glycosylation analysis: a path through the maze. Nat Chem Biol. 2010;6(10):713–723.

112. Bleckmann C, Geyer H, Geyer R. Nanoelectrospray-MS(n) of native and permethylated glycans. Methods Mol Biol. 2011;790:71–85.

113. Huang Y, Zhou S, Zhu J, et al. LC-MS/MS isomeric profiling of permethylated N-glycans derived from serum haptoglobin of hepatocellular carcinoma (HCC) and cirrhotic patients. Electrophoresis. 2017 Sep;38(17):2160-2167.

114. Ashline DJ, Lapadula AJ, Liu YH, et al. Carbohydrate structural isomers analyzed by sequential mass spectrometry. Anal Chem. 2007;79(10):3830–3842.

115. Kuberan B, Lech M, Zhang L, et al. Analysis of heparan sulfate oligosaccharides with ion pair-reverse phase capillary high performance liquid chromatography-microelectrospray ionization time-of-flight mass spectrometry. J Am Chem Soc. 2002;124(29):8707–8718.

116. Thanawiroon C, Rice KG, Toida T, et al. Liquid chromatography/mass spectrometry sequencing approach for highly sulfated heparin-derived oligosaccharides. J Biol Chem. 2004;279(4):2608–2615.

117. Royle L, Campbell MP, Radcliffe CM, et al. HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. Anal Biochem. 2008;376(1):1–12.

118. Wuhrer M, Koeleman CAM, Hokke CH, et al. Nano-scale liquid chromatography-mass spectrometry of 2-aminobenzamide-labeled oligosaccharides at low femtomole sensitivity. Int J Mass Spectrom. 2004;232(1):51–57.

119. Reusch D, Haberger M, Kailich T, et al. High-throughput glycosylation analysis of therapeutic immunoglobulin G by capillary gel electrophoresis using a DNA analyzer. MAbs. 2014;6(1):185–196.

120. Gilbert MT, Knox JH, Kaur B. Porous glassy carbon, a new columns packing material for gas chromatography and high-performance liquid chromatography. Chromatographia. 1982;16(1):138–146.

121. Costello CE, Contado-Miller JM, Cipollo JF. A glycomics platform for the analysis of permethylated oligosaccharide alditols. J Am Soc Mass Spectrom. 2007;18(10):1799–1812.

122. Hua S, An HJ, Ozcan S, et al. Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. Analyst. 2011;136(18):3663–3671.

123. Packer NH, Lawson MA, Jardine DR, et al. A general approach to desalting oligosaccharides released from glycoproteins. Glycoconj J. 1998;15(8):737–747.

124. Pabst M, Altmann F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. Anal Chem. 2008;80 (19):7534–7542.

125. Jensen PH, Karlsson NG, Kolarich D, et al. Structural analysis of N- and O-glycans released from glycoproteins. Nat Protoc. 2012;7 (7):1299–1310.

126. Pabst M, Bondili JS, Stadlmann J, et al. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. Anal Chem. 2007;79 (13):5051–5057.

127. Hua S, Lebrilla C, An HJ. Application of nano-LC-based glycomics towards biomarker discovery. Bioanalysis. 2011;3(22):2573–2585.

128. Ruhaak LR, Miyamoto S, Lebrilla CB. Developments in the identification of glycan biomarkers for the detection of cancer. Mol Cell Proteomics. 2013;12(4):846–855.

129. Bynum MA, Yin H, Felts K, et al. Characterization of IgG N-glycans employing a microfluidic chip that integrates glycan cleavage, sample purification, LC separation, and MS detection. Anal Chem. 2009;81(21):8818–8825.

130. Greenwell P. Blood group antigens: molecules seeking a function?. Glycoconj J. 1997;14(2):159–173.

131. Harvey DJ. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: an update for 2011–2012. Mass Spectrom Rev. 2017;36(3):255–422.
   •• A continuing widespread review about the analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry.

132. Harvey DJ. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: an update for 2009–2010. Mass Spectrom Rev. 2015;34(3):268–422.

133. Nie H, Li Y, Sun XL. Recent advances in sialic acid-focused glycomics. J Proteomics. 2012;75(11):3098–3112.

134. O'Connor PB, Mirgorodskaya E, Costello CE. High pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry for minimization of ganglioside fragmentation. J Am Soc Mass Spectrom. 2002;13(4):402–407.

135. Harvey DJ. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: an update covering the period 1999–2000. Mass Spectrom Rev. 2006;25(4):595–662.

136. Zaia J. Mass spectrometry and the emerging field of glycomics. Chem Biol. 2008;15(9):881–892.

137. Reiding KR, Blank D, Kuijper DM, et al. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. Anal Chem. 2014;86 (12):5784–5793.

138. Bladergroen MR, Reiding KR, Hipgrave Ederveen AL, et al. Automation of high-throughput mass spectrometry-based plasma N-glycome analysis with linkage-specific sialic acid esterification. J Proteome Res. 2015;14(9):4080–4086.

139. Cech NB, Enke CG. Practical implications of some recent studies in electrospray ionization fundamentals. Mass Spectrom Rev. 2001;20 (6):362–387.

140. Dell A. Preparation and desorption mass spectrometry of permethyl and peracetyl derivatives of oligosaccharides. Methods Enzymol. 1990;193:647–660.

141. Harvey DJ. Fragmentation of negative ions from carbohydrates: part 1. Use of nitrate and other anionic adducts for the production of negative ion electrospray spectra from N-linked carbohydrates. J Am Soc Mass Spectrom. 2005;16(5):622–630.

142. Bossio RE, Marshall AG. Baseline resolution of isobaric phosphorylated and sulfated peptides and nucleotides by electrospray ionization FTICR MS: another step toward mass spectrometry-based proteomics. Anal Chem. 2002;74(7):1674–1679.

143. Hsiao CT, Wang PW, Chang HC, et al. Advancing a high throughput glycotope-centric glycomics workflow based on nanoLC-MS(2)-product dependent-MS(3) analysis of permethylated glycans. Mol Cell Proteomics. 2017;16(12):2268–2280.

144. Zhou S, Hu Y, Veillon L, et al. Quantitative LC-MS/MS glycomic analysis of biological samples using aminoxyTMT. Anal Chem. 2016;88(15):7515–7522.

145. Hu Y, Desantos-Garcia JL, Mechref Y. Comparative glycomic profiling of isotopically permethylated N-glycans by liquid chromatography/electrospray ionization mass spectrometry. Rapid Commun Mass Spe. 2013;27(8):865–877.

146. Leymarie N, Zaia J. Effective use of mass spectrometry for glycan and glycopeptide structural analysis. Anal Chem. 2012;84(7):3040–3048.

147. Domon B, Costello CE. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. Glycoconjugate Journal. 1988;5(4):397–409.
   •• Describes the fragmentation nomenclature for glycans.

148. Spina L, Romeo DI, Garozzo G, et al. New fragmentation mechanisms in matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry of carbohydrates. Rapid Communications in Mass Spectrometry. 2004;18:392–398.

149. Harvey DJ, Royle L, Radcliffe CM, et al. Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. Anal Biochem. 2008;376:44–60.
   •• Review of N-glycan fragmentation features.

150. Wheeler SF, Harvey DJ. Negative ion mass spectrometry of sialy-lated carbohydrates: discrimination of N-acetylneuraminic linkages by MALDI-TOF and ESI-TOF mass spectrometry. Anal Chem. 2000;72(20):5027–5039.

151. Hu Q, Noll RJ, Li H, et al. The Orbitrap: a new mass spectrometer. J Spectrom. 2005;40(4):430–443.

152. Yu X, Jiang Y, Chen Y, et al. Detailed glycan structural characterization by electronic excitation dissociation. Anal Chem. 2013;85 (21):10017–10021.

153. Olsen JV, Schwartz J, Griep-Raming J, et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol Cell Proteomics. 2009;8(12):2759–2769.

154. Löner C, Blackstock W, Gunaratne J. Enhanced performance of pulsed q collision induced dissociation-based peptide identification on a dual-pressure linear ion trap. J Am Soc Mass Spectrom. 2012;23(1):186–189.

155. Johnson JV, Yost R, Kelley PE, et al. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. Anal Chem. 1990;62(20):2162–2172.

156. York WS, Agravat S, Aoki-Kinoshita KF, et al. MIRAGE: the minimum information required for a glycomics experiment. Glycobiology. 2014;24(5):402–406.

157. Struwe WB, Agravat S, Aoki-Kinoshita KF, et al. The minimum information required for a glycomics experiment (MIRAGE) project: sample preparation guidelines for reliable reporting of glycomics datasets. Glycobiology. 2016;26(9):907–910.

158. Kolarich D, Rapp E, Struwe WB, et al. The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting mass-spectrometry-based glycoanalytic data. Mol Cell Proteomics. 2013;12:991–995.
•• **Glycan standardization initiatives to improving the standards for reporting MS-based glycoanalytic data.**

159. Liu Y, McBride R, Stoll M, et al. The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting glycan microarray-based data. Glycobiology. 2016;27(4):280–284.

160. Campbell MP, Peterson R, Mariethoz J, et al. UniCarbKB: building a knowledge platform for glycoproteomics. Nucleic Acids Res. 2014;42(Database issue):D215–221.

161. Aoki-Kinoshita K, Agravat S, Aoki NP, et al. GlyTouCan 1.0–the international glycan structure repository. Nucleic Acids Res. 2016;44(D1):D1237–1242.

162. Tiemeyer M, Aoki K, Paulson J, et al. GlyTouCan: an accessible glycan structure repository. Glycobiology. 2017;27(10):915–919.

163. Campbell MP, Peterson RA, Gasteiger E, et al. Navigating the glycome space and connecting the glycoproteome. Methods Mol Biol. 2017;1558:139–158.

164. Lisacek F, Mariethoz J, Alocci D, et al. Databases and associated tools for glycomics and glycoproteomics. Methods Mol Biol. 2017;1503:235–264.

165. Barnett CB, Aoki-Kinoshita KF, Naidoo KJ. The glycome analytics platform: an integrative framework for glycobioinformatics. Bioinformatics. 2016;32(19):3005–3011.

166. Lutteke T. Handling and conversion of carbohydrate sequence formats and monosaccharide notation. Methods Mol Biol. 2015;1273:43–54.

167. Ceroni A, Maass K, Geyer H, et al. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. J Proteome Res. 2008;7:1650–1659.

168. Meitei NS, Apte A, Snovida SI, et al. Automating mass spectrometry-based quantitative glycomics using aminoxy tandem mass tag reagents with SimGlycan. J Proteomics. 2015;127:211–222.

169. Goldberg D, Sutton-Smith M, Paulson J, et al. Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. Proteomics. 2005;5(4):865–875.

170. Yu CY, Mayampurath A, Hu Y, et al. Automated annotation and quantification of glycans using liquid chromatography-mass spectrometry. Bioinformatics. 2013;29(13):1706–1707.

171. Tsai PL, Chen SF. A brief review of bioinformatics tools for glycosylation analysis by mass spectrometry. Mass Spectrom (Tokyo). 2017;6(Spec Iss):S0064.

172. Campbell MP, Nguyen-Khuong T, Hayes CA, et al. Validation of the curation pipeline of UniCarb-DB: building a global glycan reference MS/MS repository. Biochim Biophys Acta. 2014;1844(1 Pt A):108–116.

173. Hayes CA, Karlsson NG, Struwe WB, et al. UniCarb-DB: a database resource for glycomic discovery. Bioinformatics. 2011;27(9):1343–1344.

174. MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966–968.

175. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nature Protocols. 2016;11(12):2301–2319.

176. Cooper CA, Gasteiger E, Packer NH. GlycoMod–a software tool for determining glycosylation compositions from mass spectrometric data. Proteomics. 2001;1(2):340–349.

177. Ep G, Rebecchi KR, Dalpathado DS, et al. GlycoPep DB: a tool for glycopeptide analysis using a "Smart Search". Anal Chem. 2007;79 (4):1708–1713.

178. Irungu J, Go EP, Dalpathado DS, et al. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. Anal Chem. 2007;79(8):3065–3074.

179. Hansen JE, Lund O, Tolstrup N, et al. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J. 1998;15(2):115–130.

180. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002;7:310–322.

181. Ozohanics O, Krenyacz J, Ludanyi K, et al. GlycoMiner: a new software tool to elucidate glycopeptide composition. Rapid Commun Mass Spe. 2008;22(20):3245–3254.

182. Bern M, Kil YJ, Becker C. Byonic: advanced peptide and protein identification software. Curr Protoc Bioinform. 2012;13(20):1–14.

183. Wu Y, Mechref Y, Klouckova I, et al. Mapping site-specific protein N-glycosylations through liquid chromatography/mass spectrometry and targeted tandem mass spectrometry. Rapid Commun Mass Spe. 2010;24(7):965–972.

184. Chalkley RJ, Baker PR. Use of a glycosylation site database to improve glycopeptide identification from complex mixtures. Anal Bioanal Chem. 2017;409(2):571–577.

185. Caprioli RM, Farmer TB, Gile J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal Chem. 1997;69(23):4751–4760.

186. Wang J, Qiu S, Chen S, et al. MALDI-TOF MS imaging of metabolites with a N-(1-naphthyl) ethylenediamine dihydrochloride matrix and its application to colorectal cancer liver metastasis. Anal Chem. 2015;87(1):422–430.

187. Uzbekova S, Elis S, Teixeira-Gomes AP, et al. MALDI mass spectrometry imaging of lipids and gene expression reveals differences in fatty acid metabolism between follicular compartments in porcine ovaries. Biology. 2015;4(1):216–236.

188. Touboul D, Brunelle A. MALDI mass spectrometry imaging of lipids and primary metabolites on rat brain sections. Methods Mol Biol. 2015;1203:41–48.

189. Quiason CM, Shahidi-Latham SK. Imaging MALDI MS of dosed brain tissues utilizing an alternative analyte pre-extraction approach. J Am Soc Mass Spectrom. 2015;26(6):967–973.

190. Patel E, Cole LM, Bradshaw R, et al. MALDI-MS imaging for the study of tissue pharmacodynamics and toxicodynamics. Bioanalysis. 2015;7(1):91–101.

191. Gustafsson OJ, Briggs MT, Condina MR, et al. MALDI imaging mass spectrometry of N-linked glycans on formalin-fixed paraffin-embedded murine kidney. Anal Bioanal Chem. 2015;407(8):2127–2139.

192. Barry JA, Groseclose MR, Robichaud G, et al. Assessing drug and metabolite detection in liver tissue by UV-MALDI and IR-MALDESI mass spectrometry imaging coupled to FT-ICR MS. Int J Mass Spectrom. 2015;377:155–448.

193. Zemski Berry KA, Gordon WC, Murphy RC, et al. Spatial organization of lipids in the human retina and optic nerve by MALDI imaging mass spectrometry. J Lipid Res. 2014;55(3):504–515.

194. Rebours V, Le Faouder J, Laouirem S, et al. In situ proteomic analysis by MALDI imaging identifies ubiquitin and thymosin-

beta4 as markers of malignant intraductal pancreatic mucinous neoplasms. Pancreatology. 2014;14(2):117–124.

195. Norris JL, Caprioli RM. Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research. Chem Rev. 2013;113(4):2309–2342.

196. Aichler M, Walch A. MALDI imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. Lab Invest. 2015;95(4):422–431.

197. Rompp A, Spengler B. Mass spectrometry imaging with high resolution in mass and space. Histochem Cell Biol. 2013;139 (6):759–783.

198. Rompp A, Guenther S, Schober Y, et al. Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging. Angew Chem Int Ed Engl. 2010;49 (22):3834–3838.

199. Rompp A, Guenther S, Takats Z, et al. Mass spectrometry imaging with high resolution in mass and space (HR(2) MSI) for reliable investigation of drug compound distributions on the cellular level. Anal Bioanal Chem. 2011;401(1):65–73.

200. Anderson DM, Ablonczy Z, Koutalos Y, et al. High resolution MALDI imaging mass spectrometry of retinal tissue lipids. J Am Soc Mass Spectrom. 2014;25(8):1394–1403.

201. Toghi Eshghi S, Yang S, Wang X, et al. Imaging of N-linked glycans from formalin-fixed paraffin-embedded tissue sections using MALDI mass spectrometry. ACS Chem Biol. 2014;9(9):2149–2156.

202. Drake RR, Powers TW, Jones EE, et al. MALDI mass spectrometry imaging of N-linked glycans in cancer tissues. Adv Cancer Res. 2017;134:85–116.

203. Briggs MT, Kuliwaba JS, Muratovic D, et al. MALDI mass spectrometry imaging of N-glycans on tibial cartilage and subchondral bone proteins in knee osteoarthritis. Proteomics. 2016;16(11–12):1736–1741.

204. Powers TW, Neely BA, Shao Y, et al. MALDI imaging mass spectrometry profiling of N-glycans in formalin-fixed paraffin embedded clinical tissue blocks and tissue microarrays. PloS One. 2014;9(9): e106255.

205. Heijs B, Holst S, Briaire-De Bruijn IH, et al. Multimodal mass spectrometry imaging of N-glycans and proteins from the same tissue section. Anal Chem. 2016;88(15):7745–7753.

206. Everest-Dass AV, Briggs MT, Kaur G, et al. N-glycan MALDI imaging mass spectrometry on formalin-fixed paraffin-embedded tissue enables the delineation of ovarian cancer tissues. Mol Cell Proteomics. 2016;15(9):3003–3016.

207. Holst S, Heijs B, De Haan N, et al. Linkage-specific in situ sialic acid derivatization for N-glycan mass spectrometry imaging of formalin-fixed paraffin-embedded tissues. Anal Chem. 2016;88(11):5904–5913.

208. Pu Y, Ridgeway ME, Glaskin RS, et al. Separation and identification of isomeric glycans by selected accumulation-trapped ion mobility spectrometry-electron activated dissociation tandem mass spectrometry. Anal Chem. 2016;88(7):3440–3443.

209. Liu Y, Clemmer DE. Characterizing oligosaccharides using injected-ion mobility/mass spectrometry. Anal Chem. 1997;69(13):2504–2509.

210. Plasencia MD, Isailovic D, Merenbloom SI, et al. Resolving and assigning N-linked glycan structural isomers from ovalbumin by IMS-MS. J Am Soc Mass Spectrom. 2008;19(11):1706–1715.

211. Vakhrushev SY, Langridge J, Campuzano I, et al. Ion mobility mass spectrometry analysis of human glycourinome. Anal Chem. 2008;80 (7):2506–2513.

212. Hinneburg H, Hofmann J, Struwe WB, et al. Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. Chem Commun (Camb). 2016;52 (23):4381–4384.

213. Isailovic D, Plasencia MD, Gaye MM, et al. Delineating diseases by IMS-MS profiling of serum N-linked glycans. J Proteome Res. 2012;11(2):576–585.

214. De Haan N, Reiding KR, Haberger M, et al. Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. Anal Chem. 2015;87(16):8284–8291.

215. Moh ES, Lin CH, Thaysen-Andersen M, et al. Site-specific N-glycosylation of recombinant pentameric and hexameric human Ig M. J Am Soc Mass Spectrom. 2016;27(7):1143–1155.

Chapter 2

Materials & Methods

Flow chart for released glycan analysis
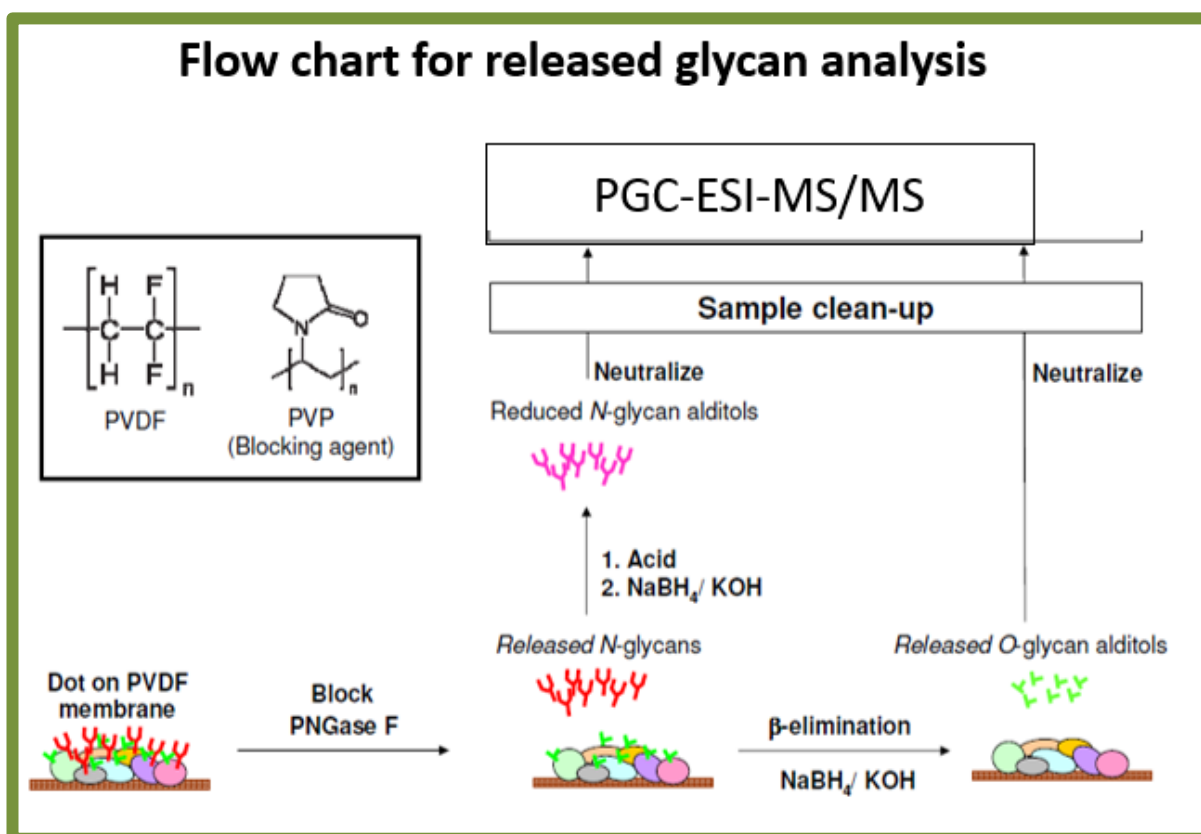
2.1 Materials & Methods

All chemicals and reagents were purchased from Sigma-Aldrich unless otherwise specified.

2.2 *N*- and *O*-linked glycan analysis mass spectrometry

This method was used throughout the thesis

2.2.1 N- and O-linked glycan release for mass spectrometry analysis

Samples (20 µg of proteins) were dot-blotted onto a Polyvinylidene fluoride (PVDF) membrane and *N*-glycans were released by an overnight incubation with 2.5 U *N*-glycosidase F (PNGase F, Flavobacterium meningosepticum, Roche, Australia) at 37°C, then *O*-glycans by β-elimination (0.5mM NaBH$_4$, 50mM KOH) as described previously (1). A low binding Eppendorf tube was used to collect released *N*-glycans after incubation with 100mM ammonium acetate for 1 h. Samples were then reduced by the addition of 1 M NaBH$_4$ in 50 mM KOH for 3 h at 50°C. The reduction was stopped by glacial acetic acid.

2.2.2 Desalting and enrichment of released glycans

Released glycans were desalted on a strong cation exchange chromatography column (Dowex resin AG-50W-X8, Bio-Rad). Residual borate ions were removed with a minimum of three methanol washes while drying in a vacuum concentrator after each wash. A PGC packing (Extract-Clean carbon SPE cartridge, Grace) C18 reversed phase platform (Thermo Fisher Scientific (analytical scale), strataX, Phenomenex (preparative scale)) was used for further purification. Bound glycans are eluted with 40% acetonitrile with 0.1% v/v trifluoroacetic acid. Eluted glycans were then dried down and re-suspended in deionized water. For *O*-glycan release, reductive β-elimination was performed using 0.5 M NaBH$_4$ in 50 mM KOH for 16 h at 50°C. The reaction was stopped by the addition of glacial acetic acid. Released *O*-glycans were desalted and purified on a PGC column following the aforementioned protocol for *N*-glycans.

2.2.3 Neuraminidase treatment of released glycans

A pooled mixture of released *N*-glycans from control and diabetic ovarian proteins (20ug) was used for neuraminidase treatment to determine sialic acid linkages. Glycans (from 3 ug proteins) were released and desalted as mentioned earlier. Glycans were treated with α2-3 neuraminidase (1mU/Prozyme) or broad specificity α2-3,6,8 neuraminidase (1mU/Prozyme) and incubated at 37°C overnight. Treated glycans were dried and solubilised in water for future MS analysis.

2.2.4 Mass spectrometry analysis of released glycans

Analysis of glycans was performed using a PGC column (5 μ particle size, 180 μ internal diameter x 10 cm, Hypercarb KAPPA Capillary Column, Thermo Fisher Scientific) by online porous graphitised column-liquid chromatography-electrospray ionisation tandem mass spectrometry (PGC-LC-ESI-MS/MS) on a ThermoFisher Velos LTQ mass spectrometer. The column was equilibrated with 10 mM ammonium bicarbonate (Sigma Aldrich) and samples were separated on a 0-70% v/v acetonitrile gradient in 10 mM ammonium bicarbonate over 75 min, with a flow rate of 4 µl/min. Capillary voltage for ESI was set at 3.2 kV, full auto gain control of 80000, scanning for ion masses of m/z between 600-2000 and mass spectra were acquired in negative ion mode. For MS/MS experiments, Collison Energy (CID) fragmentation was carried out by 30 normalised-Collison Energy for the 5 top intense ions. Identified glycan masses (*m/z*) were assigned possible monosaccharide composition using GlycoMod (Expasy, http://web.expasy.org/glycomod/) (2). Analysis of MS/MS spectra was carried out using the Thermo Xcalibur Qual browser software (4.0). Possible terminal glycan structures were assigned based on diagnostic fragment ions and fragment ion mass differences (162Da for hexose, 203Da for N-acetylhexosamine (HexNac), 146Da for fucose, 291Da for N-acetylneuraminic acid (NeuAc), with a 1 Da mass error tolerance). Diagnostic ions include: 368 and 350 *m/z* for detecting core-fucosylation, 670 and 508 *m/z* for the presence of bisecting GlcNAc and others seen in previous literature (3). Relative abundance was

determined on MS signal acquisition by calculating the area under the peak from the extracted ion chromatograms (EICs) of the precursor *m/z* as a percentage of the total peak area.

2.2.5 Statistical analysis

The average relative abundance for each glycan structure from all biological replicates was measured. The average of each *N*-glycan type was grouped by calculating the standard deviation from the three biological replicates. Paired Student t-tests were used to detect any statistical differences between the relative abundance of *N*- and *O*-glycans between samples of interest, where $p < 0.05$ was considered a significant value.

2.3 References

1.   Jensen PH, Karlsson NG, Kolarich D, Packer NH. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protocols* (2012) 7(7):1299-310.
2.   Cooper CA, Gasteiger E, Packer NH. Predicting glycan composition from experimental mass using GlycoMod. *Handbook of Proteomic Methods*. Springer (2003). p. 225-31.
3.   Everest-Dass AV, Abrahams JL, Kolarich D, Packer NH, Campbell MP. Structural feature ions for distinguishing N-and O-linked glycan isomers by LC-ESI-IT MS/MS. *Journal of The American Society for Mass Spectrometry* (2013) 24(6):895-906.

Chapter 3

Effects of diabetes on ovarian protein *N-* and *O-*glycosylation in two diabetic mice models.
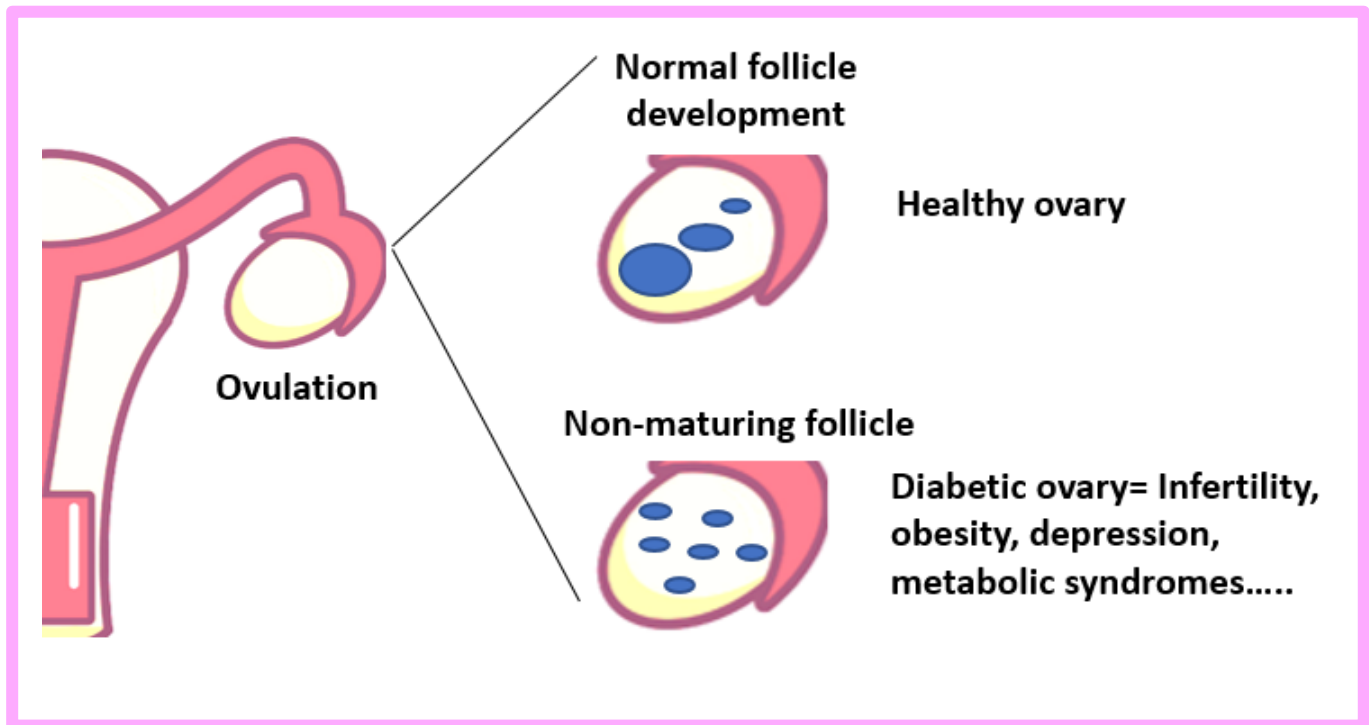
As previously discussed in Chapter 1 (1.2.3.1), diabetes has drastic impacts on protein glycosylation as well as on a successful reproductive process. The research in Chapter 3 investigated protein macro-glycosylation in the ovary of two common diabetic mouse models. We used ovaries because it is the first organ in a successful reproductive journey where eggs are formed and stored. By testing the relatively easy to collect organ (ovary), we aimed to find out if glycosylation is changed in the female reproductive tissues in diabetes before we move to other female organs in the next Chapter of this thesis (Chapter 4).There are two commonly used models for diabetes research using mice: Streptozotocin (STZ) induced and high-glucosamine injected. STZ is a chemical inducer of Type 1 diabetes [1], whereas feeding a high-glucosamine diet has been shown to induce type 2 diabetes by altering the hexosamine biosynthetic pathway that is heavily associated with protein *N*- and *O*-glycosylation [2]. The effect of diabetes on the glycosylation of ovarian proteins has not been investigated previously. In part 1 of this Chapter, *N*- and *O*-glycans of Type 1 diabetic (STZ induced) mice ovarian proteins were analysed using PGC-LC-MS/MS, and the results were then compared with the profile of liver tissue as a non-reproductive organ from the same mice, Manuscript 3 (Pages 88-102): Shathili, A.M., et al., The effect of streptozotocin-induced hyperglycemia on N-and O-linked protein glycosylation in mouse ovary. *Glycobiology*, 2018. 28(11): p. 832-840.The inflammatory status of the STZ ovary samples was also probed using qPCR to detect proinflammatory markers to investigate any relationship between the observed changes in glycosylation and ovarian inflammation. In addition to the work presented in the manuscript, the mRNA expression of several glycosyltransferases/sidases was also measured to identify any correlations between glycosylation enzyme transcriptional levels and glycosylation structure profiles. In part 2, the *N*- and *O*-glycosylation of ovarian proteins of mice injected a high-glucosamine diet were analysed using the same methods to investigate the changes caused by altering the hexosamine biosynthetic pathway in a Type 2 diabetes model.

3.1 Introduction

3.1.1 Impact of diabetes on the female reproductive system

Diabetes was illustrated as a useful model to investigate the female reproductive process. Many reproductive processes such as ovulation, menstruation and implantation display hallmark signs of inflammation in diabetes [3]. Irregular ovarian function, with impacts on the menstrual cycle, has been associated with both types of diabetes. The effect of type 1 diabetes on menstruation is first illustrated by ovarian hypofunction (failure to develop first menstrual cycle) [4, 5]. Type 2 diabetes has also been linked with menstrual irregularities given the role of insulin in ovarian physiology such as energy production, oocyte maturation and cell growth and differentiation [6, 7]. Furthermore, hyperglycaemia (high-blood glucose) exposure has been shown to be toxic to early embryo development resulting in lower implantation rates and higher embryopathy rates (abnormal development of an embryo) [8].

3.1.2 Streptozotocin (STZ): a chemical inducer of hyperglycemia (type 1 diabetes)

Streptozotocin (STZ) is a selective chemical toxin that destroys the insulin producing pancreatic B-cells by blocking Glucose transport (GLUT2) transporter system, causing DNA methylation and increasing free radical generation, which induces type 1 diabetes [9]. STZ fed mice have been repeatedly shown to express high levels of pro-inflammatory cytokines such as TNF-α and IL6 [10-12]. *In vitro* changes in TNF-α, IL-1B and IL6 proteins have been linked to alteration in the monosaccharide building process, namely to the enzymes (glycosyltransferases) responsible for the addition of terminal glycan-epitopes including terminal fucose and sialic acid, for example, TNF-α resulted in an increase in α-2,3 sialyltransferase and α-1,2,3 fucosyltransferase mRNA expression [13, 14]. Others have used STZ affected kidney tissues to demonstrate a link between aberrant *N*- and *O*-glycosylation and diabetes progression [15, 16]. At the transcriptional level, a comprehensive study on the expression of various glycosyltransferases/ glycosidases in STZ affected rat kidney was performed using RNA-seq, showing an increase in the sialyltransferase

and Mannosyl-glycoprotein-acetylglucosaminyltransferase-4 (MGAT-4) (responsible for glycan structure branching by regulating the formation of tri-antennary complex glycans) [17].

Although STZ has been extensively used over the last three decades as diabetes inducer, the associated toxicity and physiochemical characteristics associated with STZ– treated animals remain major obstacles for research using STZ as diabetes model [18]. For example, non-specific toxicity on other organs is associated with STZ treatment [19]. Also, STZ successful treatment on animals is dependent on age, animal species, gender and many other factors [20]. However, research has found that paying attention to several factors such as appropriate use of STZ, animal-related characteristics and the target blood glucose level will enable the use of this model as a representative to hyperglycaemia (Type 1 diabetes) [18].

3.1.3 High-glucosamine feeding: a model of hyperglycemia (and possible type 2 diabetes)

Whereas STZ is a known chemical inducer of hyperglycaemia by inducing toxicity to the pancreatic β cells which produce insulin, feeding a high-glucosamine diet is a different diabetic model used to specifically address one aspect of hyperglycemia - the hexosamine biosynthesis pathway (HBP) [21]. Many of the damaging effects of hyperglycaemia are facilitated by increased flux through the HBP (Figure 3.1). HBP has been shown to be highly involved in the pathogenesis of type 2 diabetes as well as fat-induced insulin resistance [22-26]. For example, elevated free fatty acid levels have been directly linked to increased HBP influx which overall induces skeletal muscle insulin resistance [26]. In fact, activating the HBP (glucosamine infusion) has been shown to contribute to insulin resistance (type 2 diabetes) by inducing chronic hyperglycaemia [27].

Glucosamine (GlcN) has been shown to activate the hexosamine pathway (hyperglycaemia mimetic) where it is converted to GlcN-6 phosphate, which is an intermediate of the pathway that bypasses the rate limiting glucosamine-fructose-6-phosphate amidotransferase enzyme (GFAT) [28]. High-glucosamine feeding increases UDP-GlcNAc levels, the donor for *N-* and *O-*glycoproteins glycosylation as well as nuclear and cellular *O*-GlcNAcylation (Figure 3.1) [26].

The latter has been linked to regulation of insulin activity via disrupting Ser/Thr phosphorylation [21, 29].
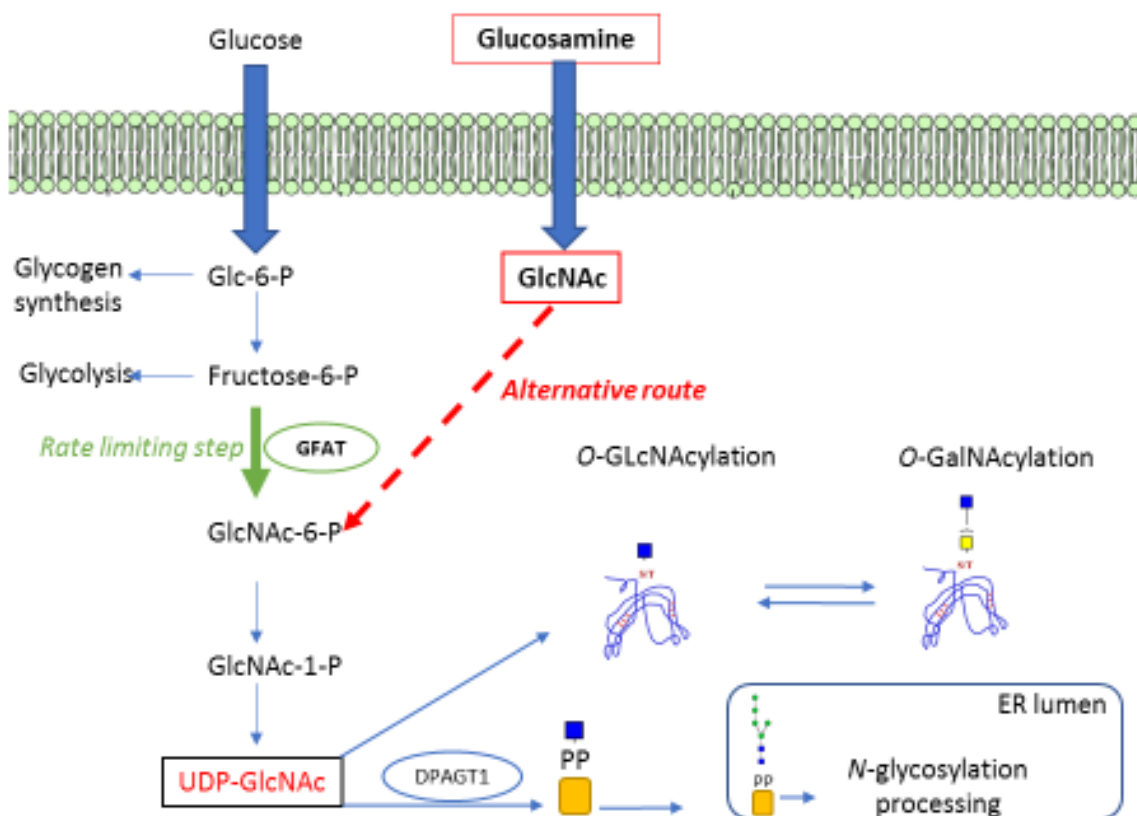
Figure 3.1: Hexosamine Biosynthetic Pathway (HBP) pathway showing glucose pathway as well as glucosamine alternative pathway; the latter acts through bypassing the rate limiting step of converting Fructose-6-P to GlcNAc-6-P (GFAT). UDP-*N*-acetylglucosamine-dolichyl-phosphate *N*-acetylglucosaminephosphotransferase (DPAGT1) catalyses the initial step in *N*-linked glycosylation pathway. This figure also shows the route in which a high-glucosamine diet may influence both *N*- and *O*-protein glycosylation. Inspired from [30].

The effect of hyperglycaemia on protein glycosylation has been investigated in some tissues and cell lines [31]. In one study, GlcN (ultimate glucosamine) has been used to reprogram the cell's biosynthetic pathways via feeding the mice GlcN enriched water, resulting in an increase in growth and body weight. The oral introduction of GlcN increased *N*-glycan branching on hepatic glycoproteins in mice, where cultured cells exhibited enhanced glucose uptake and lipid synthesis. The study showed the effect of high-glucosamine in reprogramming the cells to enhance nutrient uptake via altering the *N*-glycan branching [32]. To the best of our knowledge, the relationship

between diabetes and protein glycosylation remain to be addressed in any reproductive tissues including the ovary. Also, any differential changes on ovarian protein glycosylation caused by type 1 and type 2 diabetes remains unexplored.

The overall aims of the work reported in this Chapter were:

i. To understand the macro-effect of hyperglycemia on protein glycosylation of STZ affected mouse ovaries, and correlating any changes to the inflammatory status of the ovaries.

ii. To compare the macroscopic protein glycosylation of ovaries from STZ and high-glucosamine injected diabetic mice models

iii. To profile the protein glycosylation of the diabetic mouse liver, as an example of a non-reproductive organ, and compare to diabetic ovary glycosylation.

3.2 Part 1: The effect of STZ-induced hyperglycemia on N-and O-linked protein glycosylation (Manuscript 3)

Contribution: The writing of the manuscript, mass spectrometry analysis of ovary/liver glycosylation, and qPCR experiment and analysis were done by A. M. Shathili. A. M. Shathili also travelled to Adelaide to assist T.C.Y Tan in mice handling and sacrifice; Tan performed the animal handling and organ extraction. Dr Hannah Brown supervised, provided technical support and edited the manuscript. Dr A. Everest-Dass assisted with initial glycan analysis and reviewed the manuscript. Dr L. M. Parker assisted with qPCR experimental planning and reviewed the manuscript. Prof. J. G. Thomson supervised and reviewed the manuscript. Dr Robyn Peterson edited and reviewed this Chapter. Prof. Nicolle H. Packer supervised, reviewed and edited the manuscript and Chapter.

Shathili, A.M., et al., The effect of streptozotocin-induced hyperglycemia on N-and O-linked protein glycosylation in mouse ovary. *Glycobiology*, 2018. 28(11): p. 832-840.

Chapter 4

The effect of diabetes on protein and glycan complement of the oviduct and receptive uterus: the blastocyst journey through the female reproductive tract
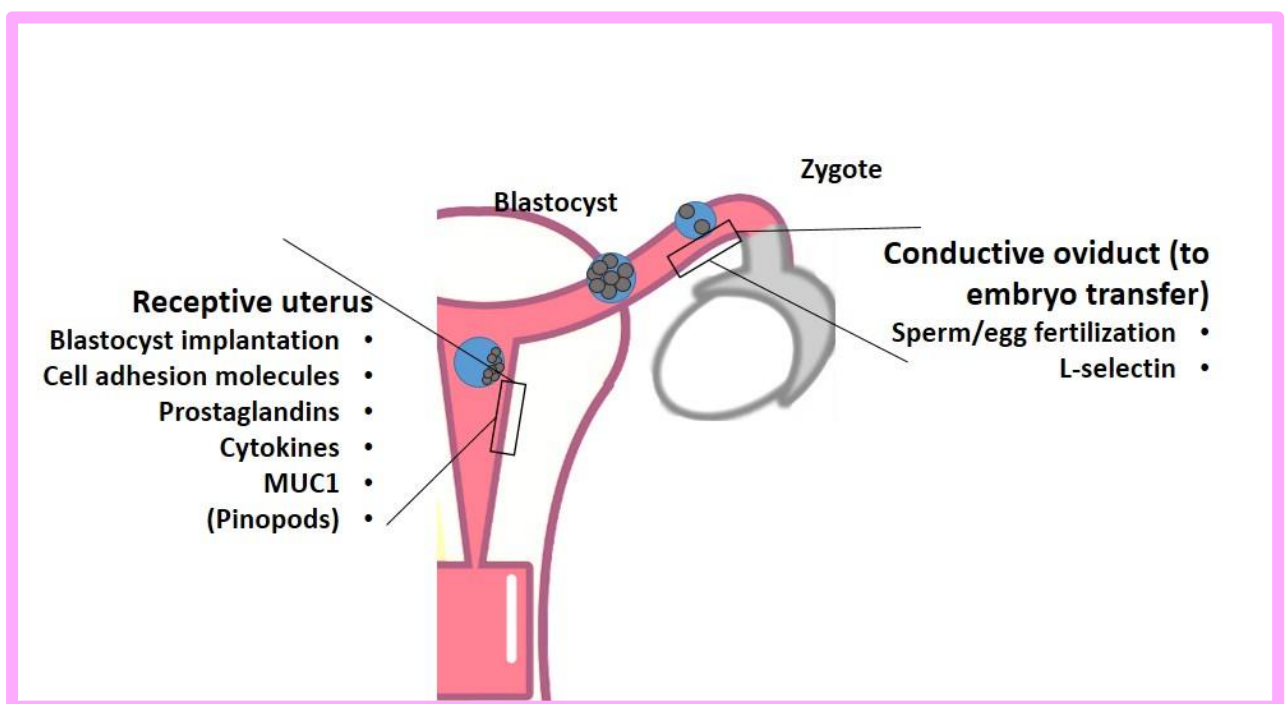
Since our results in Chapter 3 showed that diabetes resulted in significant alteration in protein macro glycosylation of a female reproductive tissue (ovary) in Streptozotocin (STZ) -induced diabetic mice, we investigated whether STZ -induced diabetic mice also exhibit aberrant proteins and protein glycosylation in their oviduct and uterus during blastocyst implantation. The process of blastocyst implantation is managed by a series of orchestrated events among several glycoproteins found on the surface of oviduct and uterine tissues, and a change in the molecular surface (proteome and/or glycome) of the oviduct and uterine tissues would affect the blastocyst journey through the female reproductive tract.

This Chapter provides a glyco-microscopic view of the surface of the oviduct and uterus that the blastocyst moves over after fertilisation of the egg, in diabetic mice. STZ chemically induces type 1 diabetes by selectively destroying insulin-producing cells (β cells) in the pancreas. In Part 1, shotgun proteomics is used to investigate the changes to the oviduct and uterine membrane proteome caused by STZ treatment, before and post embryo implantation. Part 2 is an investigation of the glycosylation presented by the mucins on the surface of the oviduct and uterus in diabetic mice since these heavily glycosylated proteins have been shown to play an important role in blastocyst implantation [2]. MUC1 is the major mucin found on the lining of the uterus so in order to compare techniques for mucin analysis, porcine gastric mucin (PGM) was used as a suitable analogue for method development. The samples were from oviduct and uterus of control and STZ -induced diabetic mice, 1.5 and 3.5 days post coitus in order to profile the uterine mucin *O*-glycosylation occurring before and after embryo implantation.

4.0 Introduction

As previously described in Chapter 1, in all mammalian systems, the new embryo undergoes a well described journey once sperm and egg are fused. After attachment to the oviduct wall, the embryo undergoes its early divisions as it progresses along the oviduct, and enters the uterine cavity as an early blastocyst, and this process occurs in a limited time period known as the window of implantation (WOI). The endometrium needs to be receptive for implantation to occur. In mice, the zygote forms in the conductive oviduct around 1.5 days post coitus and then the developed blastocyst implants in the receptive uterus around 3.5 days post coitus (Figure 4.1).
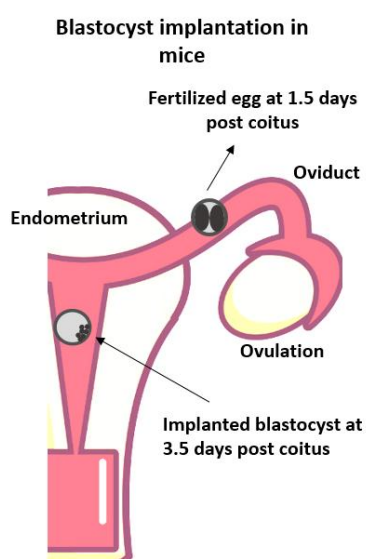
Figure 4.1: Mice blastocyst implantation in the uterus. A drawing of the female productive tract showing fertilized egg located in oviduct at 1.5 days post coitus. Blastocyst implants in uterus at 3.5 days post coitus. This figure is adapted from [1].

Briefly, in mice, prostaglandins (PGs) and ovarian hormones (progesterone and estrogen) seem to control the timing window [2]. Upon fertilisation, the blastocyst starts rolling over the oviduct via tethering to L-selectin ligands (sialyl Lewis x) found in the surface of the oviduct [3], then MUC1 ensures blastocyst implants into the right location on the surface of the endometrial epithelial cells. Chemokines and cytokines direct the blastocyst onto the right location marked with a reduction in MUC1 expression, while adhesion molecules (integrin and cadherins) ensure temporal adhesiveness between endometrium and embryo [2] (Figure 4.2). Mice and human differ in their mucin expression during implantation. Mice and other rodents, unlike human, exhibit a decrease in MUC1 expression during the window of implantation [4-8] whereas humans MUC1 protein expression has been shown to increase during the same receptive window [8].
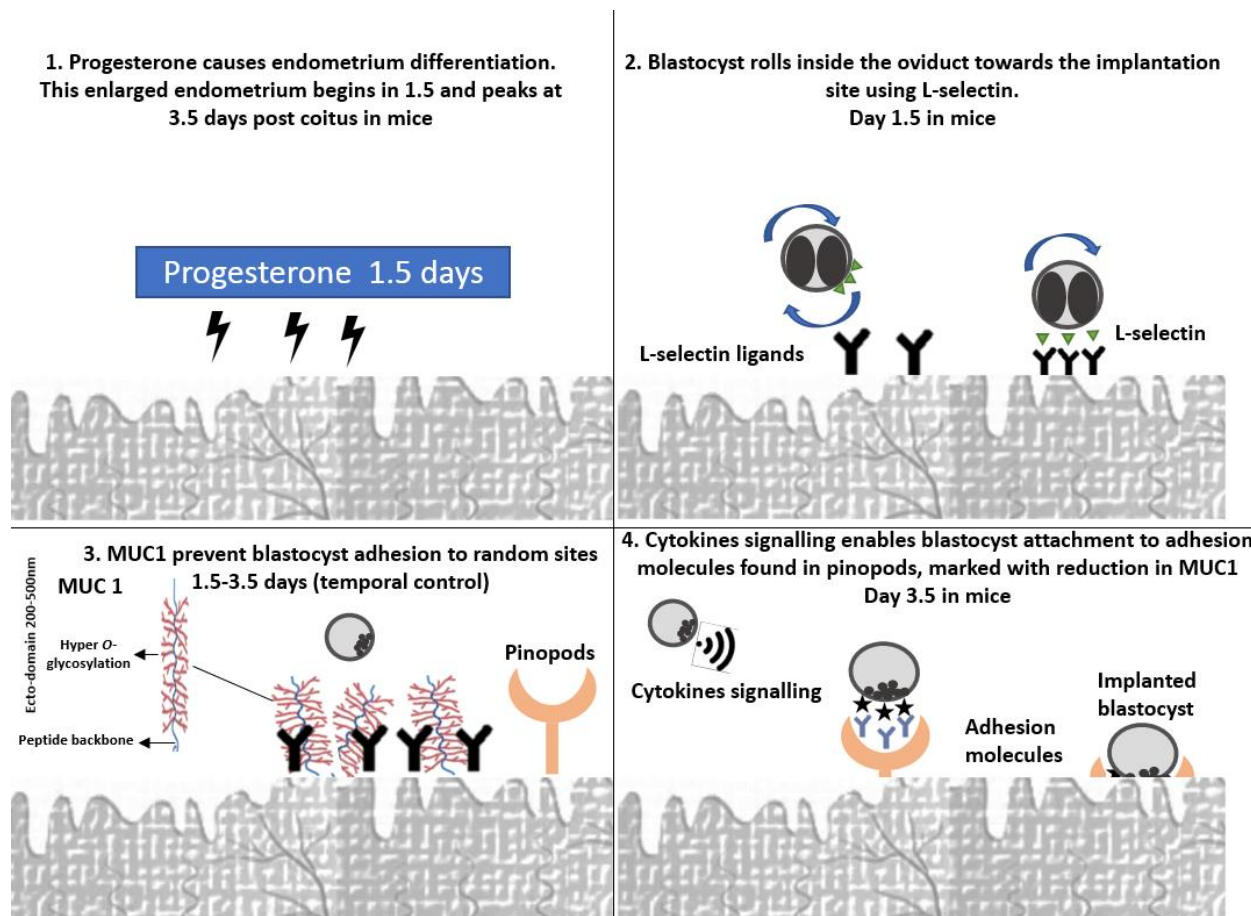
Figure 4.2: Mice blastocyst implantation in the uterus. Endometrial differentiation starts due to increase in progesterone levels, starts around 1.5 days and peaks at 3.5 days post coitus (1). Fertilised egg begins rolling in oviduct tube towards uterus using L-selectin glycan ligands (sialyl Lewis x) found on oviduct surface (2). Mucin (MUC1) competes with mature blastocyst and prevents its adherence to the uterus surface (3). Cytokines and chemokines lead the blastocyst to pinopods, free of MUC1, while adhesion molecules (Cadherin, IntegrinB3) ensure firm attachment to pinopods (4). This figure is adapted from [9] and from [2].

This Chapter aims to determine any differences in the oviduct and uterus membrane proteome due to diabetes in the receptive window of implantation. In the same diabetic model, changes in oviduct and uterine mucin glycosylation (mainly MUC1) between pre-receptive and receptive uterus were also determined using mass spectrometry after comparing different mucin analysis protocols.

The overall aims of this Chapter are thus:

Part 1: Investigate changes in oviduct and uterus membrane proteomes due to diabetes during the journey from conception through the oviduct to the uterus.

Part 2: Profile the *O*-glycomes found in the surface of oviduct and uterine MUC1 in the same diabetic mouse before and after embryo implantation, after optimisation of mucin analysis.

4.0.1 Samples from mouse female reproductive tract

T.C.Y Tan (University of Adelaide), with the help of A. M. Shathili performed the animal handling and sacrificing. A. M. Shathili provided assistance in feeding mice and regular cleaning of mice cages. Also, A. M. Shathili provided assistance during mice STZ-injection, mating, glucose reading and sacrifice. Briefly, mice were purchased from Animal Resources Centre, Perth, maintained in 14 hr/ 10 hr light/dark conditions and given water and rodent chow *ad libitum*. All experiments were approved by The University of Adelaide Animal Ethics Committee (M-2013-233/172) and conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes. Eight-week-old female C57BL/6 mice were treated with STZ (180 mg/kg) dissolved in citrate buffer (pH 4.5). Blood glucose was monitored every two days by tail prick (Accu-Chek, Roche, Sydney, NSW, Australia). When blood glucose reached 14 mmol/L (approximately one week following injection), mice were available for experiments. Both control (n=6) and STZ -induced diabetic mice (n=6) were equally divided into day 1.5 post-coitus (n=3) and day 3.5 post-coitus (n=3). The days were counted as follows: female mice were first administered (i.p.) 5 IU equine chorionic gonadotrophin (eCG; Folligon, Intervet, Boxmeer, The Netherlands) in order to induce oestrus. Forty-six hours post-eCG injection, mice were administered 5 IU human chorionic gonadotrophin (hCG; hCG/Pregnyl; Merck, Kilsyth, VIC, Australia) in order to induce ovulation and to start progesterone production, and mated with males of the same strain. The following morning is considered day 0.5 post-coitus, and mice were culled and oviduct and uterus collected and snap frozen at 1.5 and 3.5 days post-coitus.

4.1 Part 1: The proteome of oviduct and receptive uterus

The dynamic nature of the receptive human uterus during the WOI has been previously studied using proteomics approaches [10]. Using mass spectrometry (MS) based techniques, some have implicated the Jnk and EGF signalling pathways, involved in stress, growth and differentiation, as key regulators of protein expression between pre-receptive and receptive phases in human uterus using 2D differential in-gel electrophoresis [11]. Others have shown receptive human uterus to over express proteins involved in the process of destabilizing microtubules and adhesion molecules of embryo compared to pre-receptive uterus [12]. Specifically, the cell adhesion molecule (CAM) family members have been implicated in mediating cell to cell adhesion between embryo and receptive uterus [2]. This protein family, which is comprised of four types of glycoproteins: integrins, cadherins, selectins and immunoglobulins, all rely on glycans interaction in their adhesion function.

On the other hand, diabetes results in many changes at the proteome level with many alterations in biological process linked to diabetes, such as metabolic pathways, fatty acid metabolic processes, oxidation reductions, the ubiquitin-proteasome pathway, protein folding and many others [13-19]. Furthermore, human diabetes has been reported to result in overexpression of integrin presented in small human blood vessels. Immunostaining and mRNA expression have shown an increase in integrin *in vivo* of human diabetic retinopathy (B1 subunit) [20]. As a result, in the retinal microvessels, diabetes affects the interaction of vascular endothelial cells with their basement membranes in the direction of firmer cell-matrix adhesion, possibly leading to slow cell migration [20]. The effect of diabetes on the process of embryo implantation has been shown to affect the rate of foetal loss in human that increases by ninefold due to diabetes [21, 22]. In fact, there is a significantly higher loss of embryo in diabetic non-obese mice compared to normal mainly due to implantation factors [23]. For example, in the uterus of diabetic mice, there was an insufficient expression of the cytokines and chemokines (INF-γ) required for directing the blastocyst into pinopods [23]. Few proteomics studies have been concerned with the effect of

diabetes on uterine tissues. One proteomics analysis of uterine lymphocytes of non-obese diabetic mice (2DE/MS) detected 24 differentially expressed proteins compared to wild type mice; results were confirmed by Western blot. Twenty proteins were down-regulated in diabetic and 4 were up-regulated. The up-regulation was related to the overall protein functions of cell movement, cell cycle control and metabolism [24]. To the best of our knowledge, no study has focused on the proteome of oviduct during blastocyst formation in diabetes models. Overall, diabetes has been demonstrated to have a comprehensive and profound effect on the physiology of the whole human body, and we show that this includes changes in the uterus proteome.

4.1.1 Materials and methods

Protein sample preparation for proteomics experiments

Tissue from mice uterus and oviduct from four states (diabetic/non-diabetic 1.5 and 3.5 days post coitus) was excised and ground with liquid nitrogen by the use of mortar and pestle. Ground tissues were then solubilised using homogenising buffer (1% w/v SDS with protease inhibitors) in order to extract soluble proteins. Soluble proteins were then subjected to acetone precipitation in 1:9 ratio and incubated at 20°C overnight. Acetone precipitated protein pellets were solubilised with 1% w/v SDS. After quantification using bicinchoninic acid (BCA) assay, 25ug of oviduct and 50ug of uterus protein (optimised to obtain a similar number of protein IDs) was then run on a 10% w/v SDS gel (BioRad Laboratories, Australia). Protein separated lanes were excised into six segments, chopped and digested using trypsin (1:100ug) with incubation overnight at 37ºC degrees. Digestion was stopped using 2% v/v TFA and peptide samples were then zip-tipped using manually packed styrene-divinylbenzene (SVD) stage tips. Eluted peptides were dried and reconstituted in 0.1% v/v formic acid , then separated on a nanoLC column (ThermoFisher) in a 85 min gradient (5–40% v/v acetonitrile, 0.1% v/v formic acid for 80 min followed by 90% v/v acetonitrile, and 0.1% v/v formic acid for 5 min) at 300nl/min flow rate. The peptides were ionized on a Q Exactive mass spectrometer (ThermoFisher) as per following MS parameters. Briefly, the

electrospray was maintained at 2.0 kV voltage with MS scan range between m/z 350 and 1800 Da, 35,000 resolution and an automatic gain control (AGC) target of 1×10^6 ions in full MS; and MS/MS scans were carried out at 17,500 resolution with an AGC target of 2×10^5 ions. Precursor ion detection for MS/MS was made using a data-dependent "Top 10" method operating with CID fragmentation, with an isolation width of 2.0 Da and normalised collision energy of 35%.

Protein database searching and analysis

Raw MzXML files were analysed using Proteome Discoverer and a standard mouse (*Mus musculus*) library. Proteome Discoverer was used with the following settings: spectrum selector with 2-5 charges and 200-2000 kDa window; Mascot search based on 3 missed cleavages, precursor tolerance of 20 ppm, fragment tolerance of 0.1 Da and dynamic modifications including oxidation, acetyl (K) and deamidated (N) and one static modification which is carbamidomethyl (C); and percolator (a tool able to discriminate correct from incorrect peptide spectrum matching (PSM)) with strict/relaxed false discovery rate (FDR) of 0.05. In order to produce highly stringent protein identifications for subsequent analysis, a newly in-house developed module called PeptideWitch, developed by D. Handler (Manuscript 4: A software package for the production of high stringency proteomics data (PeptideWitch)), was used. As previously reported in another paper, the production of high stringency datasets can be achieved by following the Scrappy rules – that is, the summing of protein spectral counts with no values missing from any replicate, the conversion of spectral counts into normalised abundance factors and comparisons being conducted on these highly stringent IDs [25]. PeptideWitch facilitates this process with an updated engine; the source code can be found in (https://bitbucket.org/peptidewitch/peptidewitch). PeptideWitch covers the typical spread of statistical measures that are standard within the field of proteomics – the production of high stringency/low FDR data, Volcano Plots, Venn Diagrams, as well as introducing a few novel tests. The improvements to the previous software include intra-replicate PCA analysis, P-value histograms and binned natural log NSAF histograms for each replicate state. The graphical nature

of PeptideWitch outputs allowed for the quick and reproducible verification of highly-stringent protein identifications from the complicated biological tissue time-point samples.

4.1.2 Results

4.1.2.1 High stringency proteomic comparative analysis of control, diabetic oviduct and pre-receptive and receptive uterine tissues

"The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD011847".

Shotgun proteomic analysis was conducted on non-diabetic control and diabetic oviduct and uterine total tissues (from STZ -induced diabetic mouse model) during pre-receptive and receptive phases of WOI. High stringency protein identifications were obtained from both the oviduct and uterine tissues and were quantified by employing a series of rules based on previously published work [25]. The rules are as follows: a valid protein identity was produced where the sum of the spectral counts (SpC) across all replicates was greater or equal to five SpC and, in addition, where the protein was identified in all three replicates. 'Unique' proteins were assigned under strict conditions, namely a protein must be found in a minimum of two biological replicates within a single state while missing from the other state in two or more replicates, with minimum sum SpC of five. One separate biological sample from mouse testis was analysed between the two tissue samples as a control to reflect on differences in peptide spectral matching (PSM) and spectral counting arising from different tissue types (Tables 4.1). Table 4.1 lists results on low stringency peptide identifications and peptide spectral matches whilst Table 4.2 gives an overview of the highly stringent peptide identifications. Although few differences in peptide spectral matching and spectral counting between oviduct and uterus were observed at low stringency (Table 4.1), the use of testis tissues as control run after oviduct and before uterine tissue peptide samples have allowed us to exclude the possibility of technical errors causing differential identities; although only one sample was used, this control confirms that the

differences seen are tissue specific (Table 4.2). In the high stringency analysis, Table 4.2, both oviduct and uterine tissues showed peptide SpC with relative standard deviation (±%RSD) averaging 2.23 and 1.38 respectively, demonstrating the robustness of our obtained data.

Oviduct	PSM – 1	PSM - 2	PSM - 3	Average	StDev	SpC – 1	SpC - 2	SpC - 3	Average	StDev
C 1.5	44706	49830	40713	45083	3.38	15871	16625	14858	38274.3	1.87
D 1.5	39884	45459	37822	41055	3.21	13910	15846	14316	36096.3	2.32
C 3.5	41612	44738	41589	42646.33 33	1.42	14689	16458	15687	38510.3	1.89
D 3.5	42948	46565	41005	43506	2.16	14985	16497	13485	37213	3.35
Uterus	PSM – 1	PSM - 2	PSM - 3	Average	StDev	SpC – 1	SpC - 2	SpC - 3	Average	StDev
C 1.5	64872	61877	64205	63651.33	0.82	18374	17569	18358	18100.33	0.85
D 1.5	67032	57074	67739	63948.33	3.11	20489	17046	16997	18177.33	3.67
C 3.5	73961	65246	77239	72148.67	2.86	18203	17207	18374	17928.00	1.17
D 3.5	71479	56882	61328	63229.67	3.94	17996	17343	18210	17849.67	0.84
C (testis)	59142					10216				

Table 4.1: Low stringency peptide spectral matching (PSM) and confident spectral counts (SpC) of uterus and oviduct tissues from 1.5 days control, 3.5 days control, 1.5 days diabetes and 3.5 days diabetes with three biological replicates each. One testis sample was run on the MS between the two tissues and was used as control for technical variation.

Mice model	High stringency peptide spectral count (3 biological replicates)			Average number of peptide (±%RSD)	Number of R.I. proteins common to three replicates
Oviduct					
C 1.5	13934	14579	13195	1.66	1717
D 1.5	12538	13814	12480	1.94	1588
C 3.5	12963	14655	13673	2.06	1685
D 3.5	13317	14222	11704	3.25	1618
Uterus					
C 1.5	16017	16070	16619	0.68	1760
D 1.5	17376	15202	14096	3.58	1803
C 3.5	16289	15822	16166	0.50	1722
D 3.5	15677	15813	16364	0.76	1717

Table 4.2: Highly stringent peptide identifications of uterine and oviduct tissues from 1.5 days control, 3.5 days control, 1.5 days diabetes and 3.5 days diabetes with three biological replicates each, 1 replicate wildcard included. The analysis was performed using PeptideWitch.

A series of descriptive statistical analyses were used for confirmation of data reproducibility. In addition to the production of high stringency data, PeptideWitch was used to produce several valuable measures of inter-replicate variability and overall quality. A Venn diagram was produced to compare highly stringent protein identifications between the samples (Figure 4.3 and 4.4, A). Volcano plots were produced in order to visually display the relationship between

fold change and t test significance (Figure 4.3 and 4.4, B). Intra-replicate PCA plots display in two dimensions the amount of variation between all replicates; in low stringency test states with a lower overall RSD% have more of their variability explained in the first axis of differentiation compared with replicates of higher RSD% (Figure 4.3 and 4.4, C). PCA plots examine the level of variability on the natural log spectral abundance factors. Reproducibility can also be examined by the lnNSAF histograms, where deviations from a bell shape inform us as to the distribution of abundance factors (Figure 4.3 and 4.4, D). Combined, these specific statistical measures highlighted the strong reproducibility among our biological replicates.
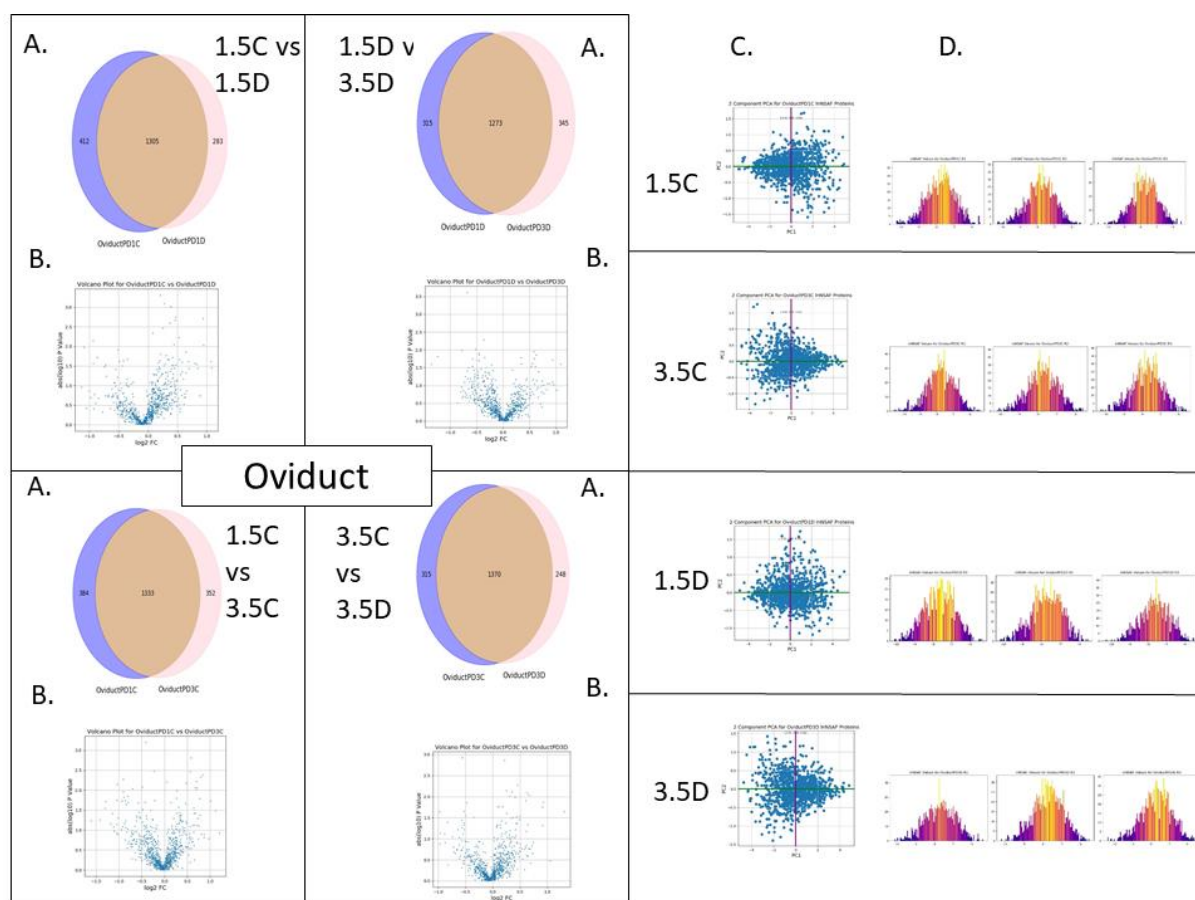
Figure 4.3: Descriptive statistical analyses for testing data reproducibility obtained for oviduct tissues via PeptideWitch. A: a Venn diagram was produced to compare highly stringent protein identifications between each two test states. B: volcano plots visually display the relationship between fold change and t test significance. C: intra-replicate PCA plots display in two dimensions the amount of variation between all replicates. D: lnNSAF histograms showing deviations from a bell shape as a distribution of abundance factors. Combined, these specific statistical measures highlighted the strong reproducibility among our biological replicates.
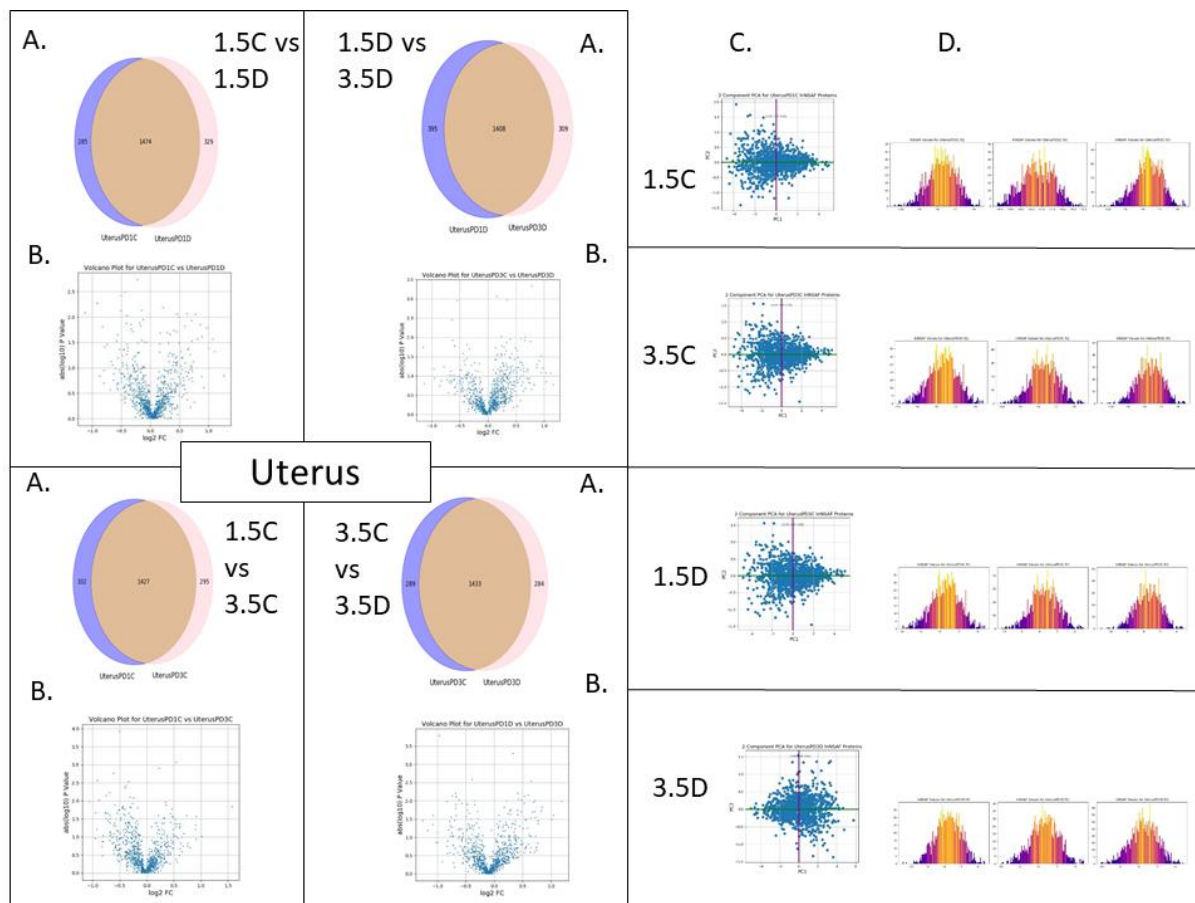
Figure 4.4: Descriptive statistical analyses for testing data reproducibility obtained for uterine tissues via PeptideWitch. A: a Venn diagram was produced to compare highly stringent protein identifications between each two test states. B: volcano plots visually display the relationship between fold change and t test significance. C: intra-replicate PCA plots display in two dimensions the amount of variation between all replicates. D: lnNSAF histograms showing deviations from a bell shape as a distribution of abundance factors. Combined, these specific statistical measures highlighted the strong reproducibility among our biological replicates.

4.1.2.2 Protein network and functional analysis revealed several altered biological processes in oviduct and uterine tissues due to both WOI and diabetes.

The analysis of altered biological processes due to WOI and diabetes in mice generated several differentially expressed as well as unique proteins between the different states. To provide a better view on molecular pathways and biological process altered in our comparisons, DAVID online software was used to perform functional annotation by describing a protein's biological identity, molecular functions and roles, subcellular location and others. DAVID enables the

rapid annotation of protein identifiers by grouping differentially expressed and unique proteins into classes of biological process and pathways according to shared categorical data for protein domain and biochemical pathway membership [26]. Mice oviduct and uterine tissues had very different responses to WOI and diabetic treatments, which can be explained by the difference in protein IDs found in both tissues (Figure 4.5. and 4.6).

Figure 4.5: A Venn diagram is produced to compare highly stringent protein identifications between oviduct and uterus tissues. A: indicates differences in 1.5 days and B: shows 3.5 days post coitus.
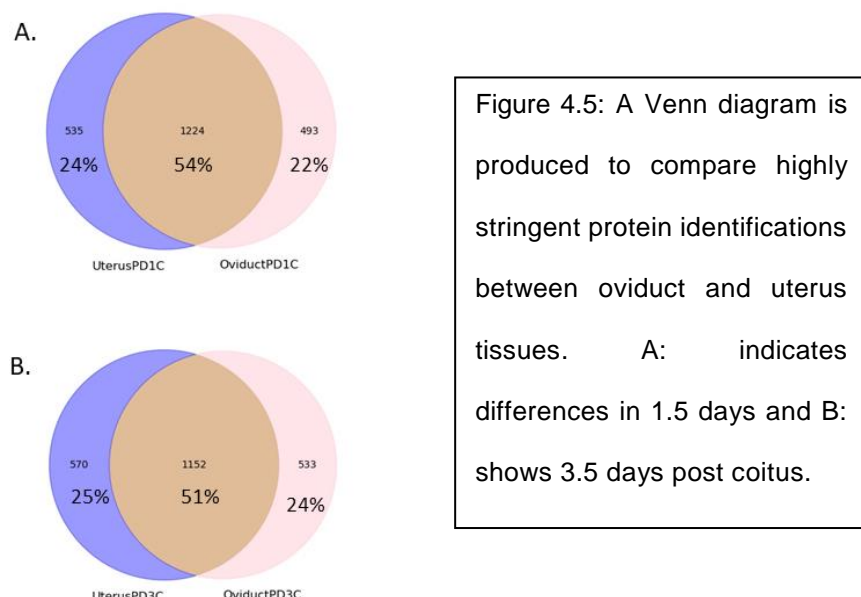
Several biological process and pathways were seen to be significantly altered in both tissues, including fatty acid metabolism, oxidation-reduction, response to endoplasmic reticulum (ER) stress, ER mediated transport and many others (Table 4.3 and 4.4).

Oviduct	Term	Count	Benjamini
Down-regulated in receptive diabetic (1.5 days)			
GOTERM_BP_DIRECT	GO:0006412~translation	24	0.000
GOTERM_BP_DIRECT	GO:0006631~fatty acid metabolic process	13	0.004
GOTERM_BP_DIRECT	GO:0098609~cell-cell adhesion*	13	0.019
KEGG_PATHWAY	mmu01100:Metabolic pathways	53	0.001
Up-regulated in receptive diabetic (1.5 days)			
COG_ONTOLOGY	Posttranslational modification, protein turnover, chaperones	8	0.033
KEGG_PATHWAY	mmu04510:Focal adhesion*	15	0.006
KEGG_PATHWAY	mmu04512:ECM-receptor interaction*	10	0.012
Down-regulated in receptive (1.5 days) non-diabetic mice			
GOTERM_BP_DIRECT	GO:0006412~translation	22	0.001
GOTERM_BP_DIRECT	GO:0006397~mRNA processing	18	0.007
GOTERM_BP_DIRECT	GO:0015031~protein transport	23	0.042
GOTERM_BP_DIRECT	GO:0008380~RNA splicing	14	0.043
KEGG_PATHWAY	mmu04510:Focal adhesion*	16	0.004
KEGG_PATHWAY	mmu04512:ECM-receptor interaction*	9	0.029

KEGG_PATHWAY	mmu04151:PI3K-Akt signaling pathway	18	0.034
KEGG_PATHWAY	mmu03015:mRNA surveillance pathway	9	0.039
KEGG_PATHWAY	mmu03040:Spliceosome	11	0.039
KEGG_PATHWAY	mmu04611:Platelet activation	10	0.049
Up-regulated in receptive (1.5 days) non-diabetic mice			
GOTERM_BP_DIRECT	GO:0055114~oxidation-reduction process	37	0.000
GOTERM_BP_DIRECT	GO:0034976~response to endoplasmic reticulum stress	11	0.001
GOTERM_BP_DIRECT	GO:0006412~translation	23	0.001
GOTERM_BP_DIRECT	GO:0015031~protein transport	29	0.001
GOTERM_BP_DIRECT	GO:0006810~transport	59	0.001
GOTERM_BP_DIRECT	GO:0008152~metabolic process	24	0.002
GOTERM_BP_DIRECT	GO:0006457~protein folding	12	0.003
GOTERM_BP_DIRECT	GO:0050821~protein stabilization	12	0.003
GOTERM_BP_DIRECT	GO:0006886~intracellular protein transport	15	0.012
GOTERM_BP_DIRECT	GO:0045454~cell redox homeostasis	8	0.023

Table 4.3: Analysis of the differentially expressed and unique proteins from comparisons of uterine tissues between 1.5 days C (control) VS 1.5 days D (diabetes), 3.5 days C VS 3.5 days D and 1.5 days C VS 3.5 days C using DAVID software. The table contains term, count (protein #) and Benjamin values, the latter reflecting the significance of these biological process with values < 0.05.

Uterus	Term	Count	Benjamini
Down-regulated in receptive diabetic (3.5 days)			
GOTERM_BP_DIRECT	GO:0098609~cell-cell adhesion*	16	0.000
GOTERM_BP_DIRECT	GO:0006412~translation	18	0.042
KEGG_PATHWAY	mmu03013:RNA transport	13	0.044
Up-regulated in receptive diabetic (3.5 days)			
GOTERM_BP_DIRECT	GO:0098609~cell-cell adhesion*	14	0.001
GOTERM_BP_DIRECT	GO:0008152~metabolic process	22	0.002
GOTERM_BP_DIRECT	GO:0006888~ER to Golgi vesicle-mediated transport	10	0.002
GOTERM_BP_DIRECT	GO:0055114~oxidation-reduction process	26	0.003
GOTERM_BP_DIRECT	GO:0008380~RNA splicing	13	0.039
KEGG_PATHWAY	mmu03040:Spliceosome	12	0.025
Down-regulated in receptive (3.5 days) non-diabetic mice			
GOTERM_BP_DIRECT	GO:0055114~oxidation-reduction process	31	0.004
GOTERM_BP_DIRECT	GO:0070527~platelet aggregation	7	0.035
KEGG_PATHWAY	mmu04145:Phagosome	17	0.002
KEGG_PATHWAY	mmu01100:Metabolic pathways	53	0.025
Up-regulated in receptive (3.5 days) non-diabetic mice			
GOTERM_BP_DIRECT	GO:0006397~mRNA processing	15	0.048
KEGG_PATHWAY	mmu03013:RNA transport	12	0.045

Table 4.4: Analysis of the differentially expressed and unique proteins from comparisons of Oviduct tissues between 1.5 days C (control) VS 1.5 days D (diabetes), 3.5 days C VS 3.5 days D and 1.5 days C VS 3.5 days C using DAVID

software. The table contains term, count (protein #) and Benjamin values, the latter reflecting the significance of these biological process with values < 0.05.

As we were specifically interested in the effect of diabetes on the embryo implantation environment, membrane proteins were analysed separately to show differences in biological processes directly impacting the embryo adhesion event. Membrane proteins were grouped via the use of Panther classification system. In the WOI (from 1.5 to 3.5 days post coitus), the oviduct membrane proteome showed a significant reduction in four transport processes including: vesicle mediated transport, intracellular mediated transport, retrograde transport (endosome to Golgi) and, small GTPase mediated signal transduction, which regulates vesicle mediated transport. This implies that all these processes are essential for the zygote interaction with a conductive oviduct at 1.5 days. Diabetes resulted in a significant reduction in the process involving vesicle mediated transport (Figure 4.6).

Oviduct

Figure 4.6: Analysis of membrane proteomes of oviduct tissues. DAVID was used to group differentially expressed and unique proteins into groups according to published literature. Heat maps for differentially expressed biological and pathway processes due to WOI and diabetes. Heatmaps are generated via PpetideHnadler. Heatmaps are giving two shades of colours (red and blue). Colour schemes are as follow: dark blue (purple) means unique to the first state, dark red means unique to second state, grey means not present. For each comparison, a red square implies a significant overexpression and a blue square reflects a significant reduction in the comparison between states.

129

In the receptive uterus, the analysis of the membrane proteome of the receptive uterus showed a reduction in oxidative phosphorylation processes compared to pre-receptive uterus. In diabetes, there was a reduction in autophagosome assembly in the receptive uterine membrane proteome and an increase in ER Golgi vesicle mediated transport (Figure 4.7).
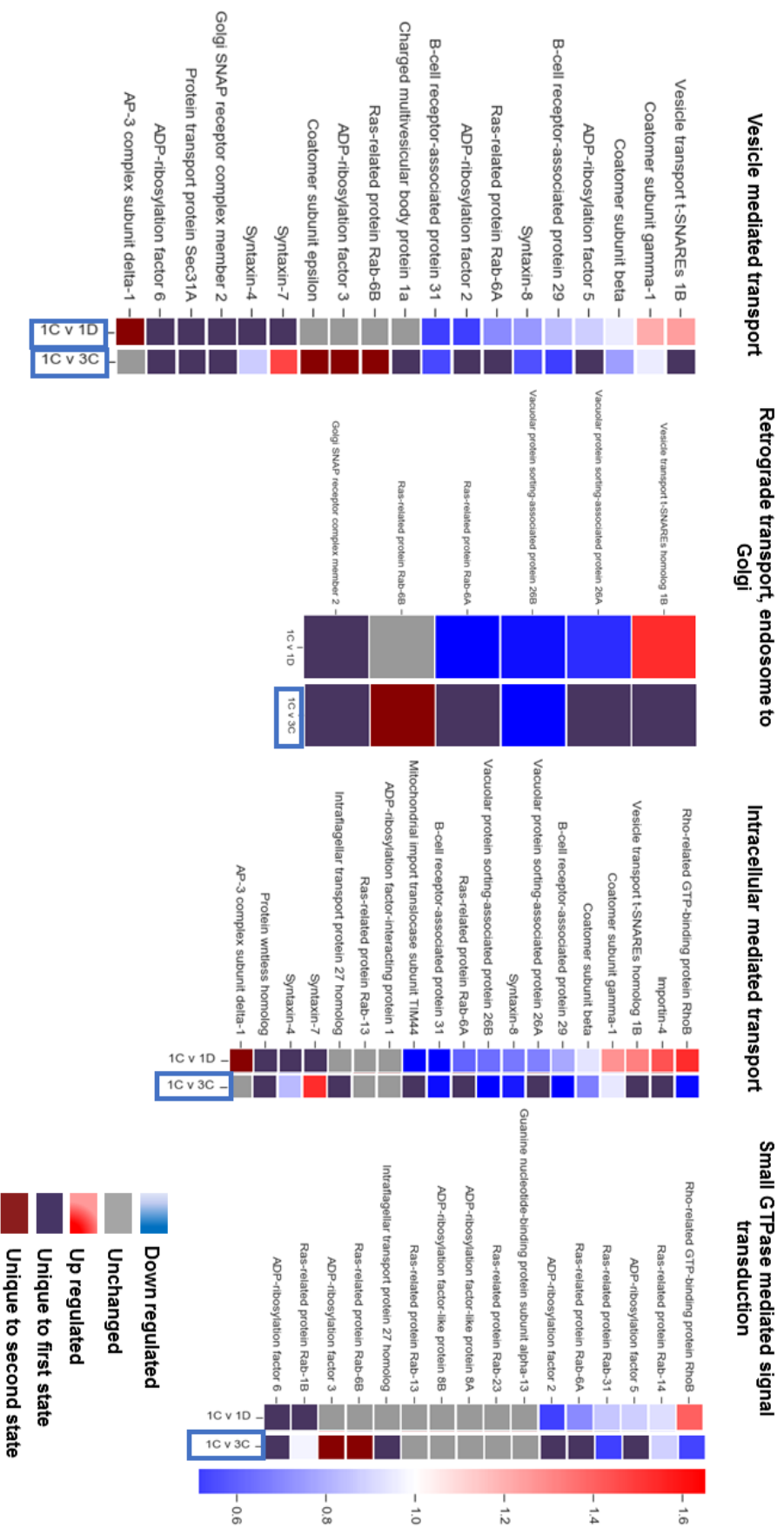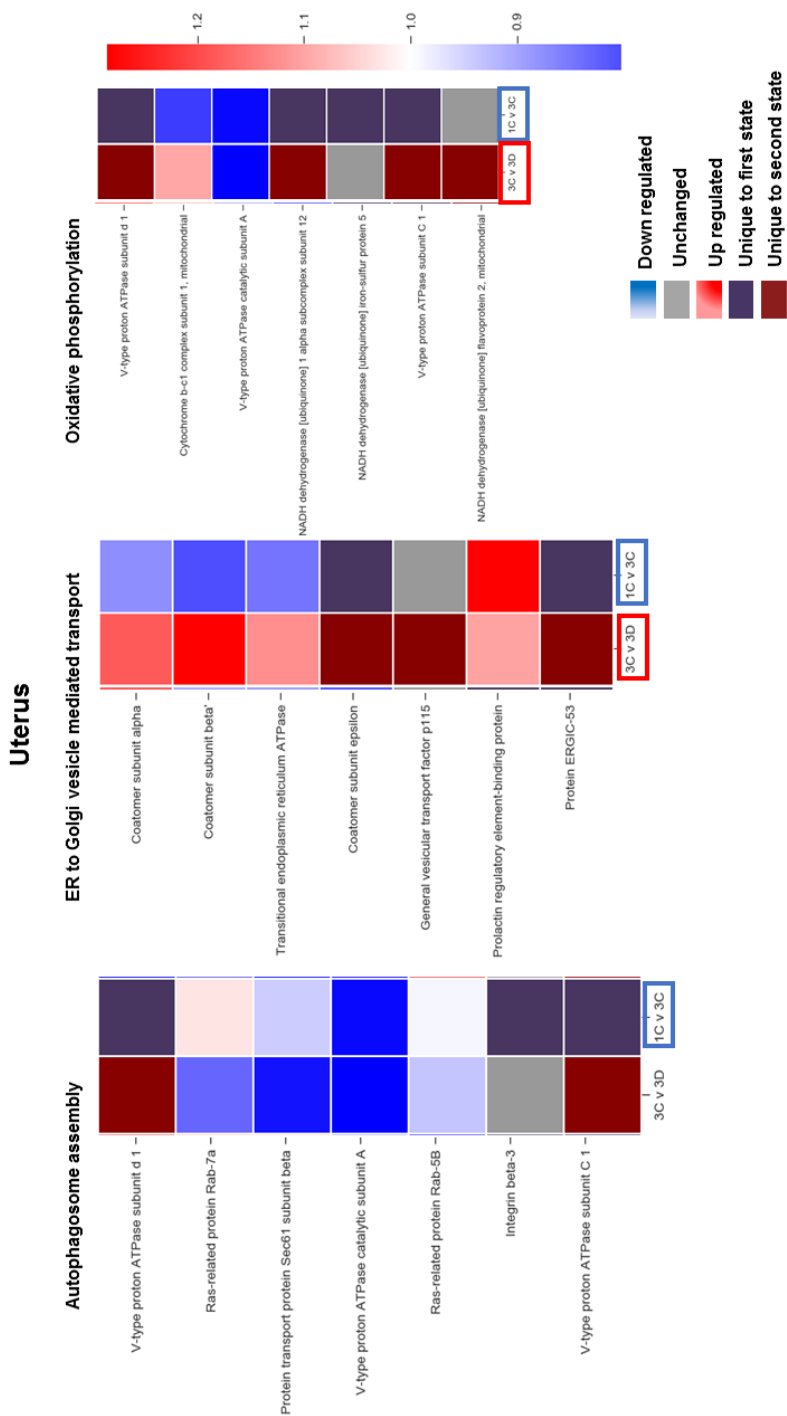

Figure 4.7: Analysis of membrane proteomes of uterus tissues. DAVID was used to group differentially expressed and unique proteins into groups according to published literature. Heat maps for differentially expressed biological and pathway processes due to WOI and diabetes. Heatmaps are generated via PpetideHnadler. Heatmaps are giving two shades of colours (red and blue). Colour schemes are as follow: dark blue (purple) means unique to the first state, dark red means unique to second state, grey means not present. For each comparison, a red square implies a significant overexpression and a blue square reflects a significant reduction in the comparison between states.

As for proteins involved in embryo adhesion in the uterus, two CAM proteins (integrinB3 and cadherins) were observed in our analysis, with integrinB3 seen highly expressed in both oviduct and receptive uterus and cadherin showing no changes in the WOI. With respect to proteins

controlling growth and proliferation: Epidermal growth factor receptor (EGFR) and Platelet derived growth factor (PDGFR) showed significant increase in the receptive uterus, while these same proteins exhibited a significant decrease in receptive diabetic uterus (Figure 4.8).
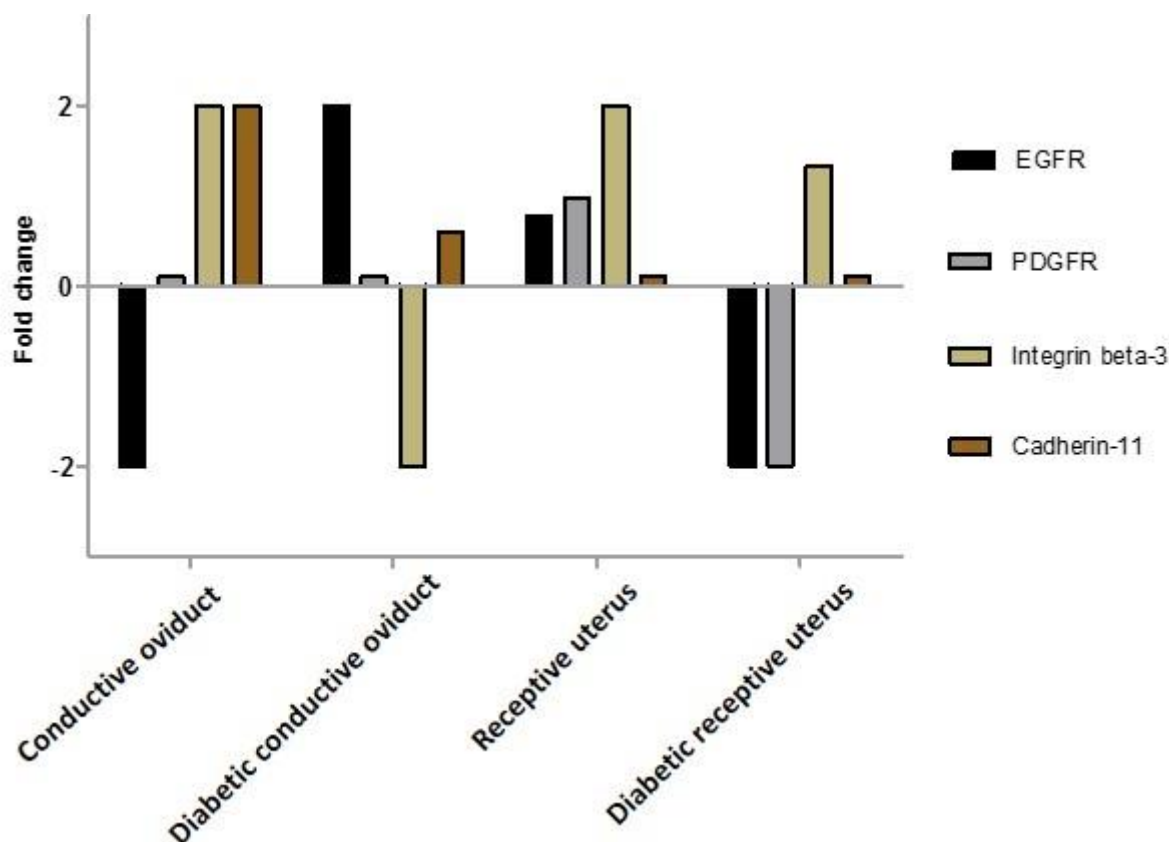
Figure 4.8: The expression of four proteins involved in growth and adhesion, namely: Epidermal Growth Factor Receptor (EGFR), Platelet Derived Growth Factor Receptor (PDGFR), IntegrinB3 and Cadherin 11. The analysis was performed for oviduct and receptive uterus and their diabetic counterparts.

4.1.3 Discussion

In our study, the use of MS based technologies has enabled the establishment of the proteomic landscape for the process of blastocyst implantation. Our proteomics results serve as a useful starting point upon which researchers can investigate protein expressions in both oviduct and uterine tissues during the process of blastocyst implantation. Also, the effect of diabetes in the proteomes of these tissues allow further investigations into molecular players involved in the onset of this disease. Mice oviduct and uterine tissues reported significant changes in several biological

and pathway processes in the WOI and the effect of diabetes on these. Oviduct was considered conductive to embryo transfer at 1.5 days and post conductive at 3.5 days post coitus. Uterus was considered pre-receptive at 1.5 days and receptive at 3.5 days post coitus.

Little similarity was observed between oviduct and uterus at the proteome level in both the WOI and diabetes, reflecting that different tissues exhibit different proteomes [27](Figure 4.5). Our proteomics analysis was facilitated via the use of a newly developed in-house statistical package called PeptideWitch to provide statistical validation for the replicates of our test states and to produce high-stringency shotgun data. PeptideWitch covers the typical spread of statistical measures that are standard within the field of proteomics – the production of high stringency/low FDR data, Volcano Plots, Venn Diagrams, as well as introducing a few novel tests. These include intra-replicate PCA analysis, P-value histograms and binned natural log NSAF histograms for each replicate state. The graphical nature of PeptideWitch outputs allowed for the quick and reproducible verification of highly-stringent protein identifications from the complicated biological tissue time-point samples.

4.1.3.1 The physiology of the oviduct and uterus at the time of initial blastocyst attachment to the endometrium (WOI)

The use of DAVID for functional and pathway annotations facilitated grouping of differentially expressed and unique proteins into biological processes. At the whole proteome level, conductive oviduct exhibited many changes including an increase in the response to ER stress, which has been reported as a result of the accumulation of unfolded or misfolded proteins in the ER lumen as a response to ER stress stimuli [28]. Previous reports on receptive uterus have reported an alteration in ER stress response with no reported studies carried out on oviduct tissues, to the best of our knowledge [11].

Increase in protein transport and decrease in oxidative phosphorylation

Associated with ER stress, the analysis of oviduct membrane proteome (i.e what the zygote interacts with) in conductive oviduct exhibited a significant increase in biological processes

involved in vesicle mediated transport, which occurs in the ER, compared to post-conductive oviduct. On the other hand, receptive uterine membrane proteome exhibited a significant decrease in oxidative phosphorylation compared to pre-receptive uterus. Alteration in the oxidative phosphorylation process of receptive uterus has been previously suggested as an early sensor of *in vitro* embryo implantation, where the higher metabolism and molecular transport in the uterus requires more energy and affects the oxidative status at the embryo implantation surface [29].

Overexpression of cell adhesion molecules (Cadherin and Integrin)

Receptive uterine tissue has been previously shown to have increased expression levels of adhesion molecules compared to pre-receptive uterine tissue in humans [12]. Analysis of our conductive oviduct and uterine tissue implicated two main cell adhesion molecules (CAM), namely: cadherin and integrin (Figure 4.8). Cadherin is a type of CAM that is important for adhering cells to each other. Defective preimplantation development was observed from a mutation in the E-cadherin gene in mouse embryo implantation [30]. Although E-cadherin mRNA levels have been shown to be regulated according to the menstrual cycle [31], several studies have failed to detect these variations at the protein level using immunohistochemical studies in uterus [32-34]. Our study has aligned with previous findings by reporting unchanged protein levels of cadherins in receptive uterus using MS techniques. Another important CAM protein in embryo implantation is integrin. The importance of integrin is highlighted in its absence in some diseases that promote infertility such as luteal deficiency, endometriosis and others [2]. There are several integrin proteins found in uterine tissues but mainly integrinB3 is shown highly expressed during the WOI in rabbit uterus [35]. The same subtype of integrin (integrinB3) was seen with significantly higher expression in receptive uterus in our mice model.

Increase in two main growth and proliferation receptors in receptive uterus

Other researchers have pointed out growth and proliferation as key regulators of protein expression between pre-receptive and receptive uterus [11]. Our receptive uterus showed an increase in two main proteins involved in growth and proliferation, namely: Epidermal growth factor receptor (EGFR) and Platelet derived growth factor receptor (PDGFR) (Figure 4.8).

Generally, EGF stimulates cell proliferation via binding to its EGFR. In endometrium, EGF has been associated with hormonal regulation of CAM molecules such as integrinB3 [36]. In our study, receptive uterus was seen with an increase in EGFR as well as integrinB3. The other growth regulator (PDGFR) has been reported to be highly regulated with Leukemia inhibitory factor (LIF) expression in receptive uterus [2]. LIF contributes to blastocyst development and implantation through controlling proliferation and differentiation, and LIF mRNA expression was, in a previous report, detected to peak at the receptive window in biopsies of fertile women [37]. In our study, conductive oviduct showed unique expression of CAM proteins (cadherin and integrinB3) and lack of EGFR expression.

Overall, the physiology of mice oviduct membrane proteome, seemed to change due to zygote presence towards a higher capacity to transport and deliver proteins to their proper locations. The membrane proteome of uterus exhibited different changes, probably due to different tissue function, with a decrease in the oxidative phosphorylation process in the WOI as the only significantly down-regulated biological process, and also showed an increase in expression of cell adhesion molecules (integrinB3) and growth factors (EGFR and PDGFR).

4.1.3.2 Effect of diabetes on zygotes and blastocyst adherence environment

Our diabetic oviduct and uterine whole proteome showed a series of significant changes in many biological processes that have been previously linked with diabetes, such as metabolic pathways, fatty acid metabolic processes, oxidation reduction and the spliceosome (Table 4.3 and table 4.4) [13, 14][15][16, 17][18].

Decreased autophagosome activity and increased oxidative phosphorylation in diabetic tissues

In our study, from the membrane proteome viewpoint, diabetes resulted in the alteration of several biological processes in both oviduct and uterus. In conductive oviduct, a significant decrease in vesicle mediated transport was observed, a phenomenon that has been repeatedly linked with diabetes and metabolic diseases [38]. Regarding the uterus, our diabetic receptive uterine membrane proteome showed a significant decrease in auto phagosome assembly proteins.

Autophagy is a catabolic process involved in maintenance of cellular homeostasis by delivering cytoplasmic constituents to the lysosomes for degradation. Some have reported an inhibition in auto phagosome protein recruitments in diabetic patients with prolonged exposure to highly elevated insulin levels [39].

Furthermore, our diabetic uterine membrane proteome exhibited a significant increase in ER to Golgi vesicle transport and oxidative phosphorylation. Diabetes is known to cause mitochondrial dysfunction while generating considerable reactive oxygen species [40]. Mitochondria have the capacity to invoke adaptive mechanisms that have evolved to defend against oxidative stress [41]. One putative mechanism is a physiologic or "mild" uncoupling of oxidative phosphorylation, which would reduce superoxide generation by reducing mitochondrial membrane potential [42]. Interestingly, in our dataset, oxidative phosphorylation is significantly reduced in receptive membrane proteome, but significantly upregulated in diabetic receptive uterus. As a result, oxidative phosphorylation represents a strong candidate as a biomarker for diabetic uterus failure of embryo implantation and, therefore, an unreceptive uterus.

Little change in the expression of cell adhesion molecules (Cadherins and Integrin)

Regarding the effect of diabetes in CAM and growth factors in oviduct and receptive uterus, diabetic receptive uterus expressed a decrease in the receptors of two main growth factors (PDGFR and EGFR), without observed changes in CAM proteins. Cadherin and integrinB3 were seen with unchanged levels in expression in diabetic tissues compared to control tissues. Although some previous research has shown a significant decrease in integrinB3 in infertile women during the receptive window [43], others have reported no temporal relationship between integrinB3 expression and pinopod formation in both normal and infertile women regardless of WOI, which keeps the functional significance of integrinB3 yet to be established [44].

Decrease in two main growth and proliferation receptors in diabetic receptive uterus

Regarding growth factors, as previously mentioned, PDGFR regulates LIF expression which is very important for embryo implantation [11]. In fact, female mice with LIF gene deficiency were

unable to have a successful embryo implantation in the endometrium, which was rescued by LIF supplementation to the gestational carrier [45]. Further work is needed to see whether there is a link low PDGFR expression to LIF expression levels and subsequently the low successful rates of embryo implantation in diabetic uterus.

Overall, several changes were observed in diabetic receptive uterus including increase in oxidative phosphorylation and vesicle transport and a decrease in autophagosome assembly, no change in the expression of CAM proteins (IntegrinB3 and Cadherin) and a decrease in growth factors (EGFR and PDGFR). Interestingly, both oviduct and uterus have shown differential response in their membrane proteome to diabetes which reflects the molecular complexity and specificity of this disease as well as the difference in the proteomes of these two tissues despite the close proximity.

4.2 Part 2: The glycome of oviduct and receptive uterus

Using freeze-fracture technique, pinopods have been seen [46] to increase dramatically in the plasma membrane of endometrial epithelial cells during early pregnancy in the rat; these proteins were associated with the glycocalyx (a medium of glycolipids and glycoproteins). Pinopods are rod shaped component projecting into the lumen, possibly acting as a focal adhesive site for the blastocyst. In fact, cell surface glycoconjugates have been previously implicated in cell-cell adhesion in several biological systems. Milk contains abundant and structurally diverse glycan structures, and these glycans has been previously investigated in cellular binding inhibition studies including microbial binding to the gut lining [47, 48]. One of seven tested milk oligosaccharides, named Lacto-N-fucopentasaccharide I[LNF-1=Fucα(1-2)-Galβ(1-3)-GlcNAcβ(1-3)-Galβ(1-4)-Glc], resulted in significant reduction of mouse blastocyst attachment to uterine epithelial monolayers [49]. In fact, fucosylated epitopes have been reported previously in the uterus of pregnant domestic animals [50, 51] and women [52]. However, no complete inhibition of blastocyst attachment was observed in the LNF-1 experiments, thus showing the need to further research the role of the uterine glycome surface.

The glycome of a receptive uterus has also been studied in monkeys during different phases of implantation by Niklaus [53], using different lectins. Endometrial tissues (displaying uterine mucins) from different reproductive cycle phases showed a significant reduction in WGA (GlcNAc + sialic acid) binding at receptive state compared to non-receptive pre-ovulatory state. Anderson [54] has also demonstrated the decrease of WGA binding in a receptive uterus in rabbit compared to non-receptive. Both binding experiments showed a reduction in WGA binding after sialidase treatment suggesting that WGA reactivity is related to sialic acid instead of GlcNAc epitope. As a result, both rabbit and monkeys share the decrease in sialic acid at receptive phase. Many other carbohydrate components have been reported to exhibit stage-specific alterations prior to implantation (Table 4.7).

Observation/scope	Physiology of a receptive endometrium	Technique used	Ref
The presence of sulphated and sialomucin	The presence of sulphated and sialomucin (8.4% and 6.2% respectively)	Scraping of 10 endometrium, from women made receptive via synthetic progesterone. CsCl density for mucin extraction	[55]
Binding of ConA (Mannose 1-3); SBA (terminal GalNAc); WGA (GlcNAc-1-4GalNAc + sialic acid): LCA (Core fucose); PHA (Gal, GlcNAc and Mannose)	All lectins showed some degree of binding with PHA given a variable cycle dependent activity	48 human endometrial biopsy during pregnancy with lectin staining performed on paraffin sections	[56]
Activity of antibodies specific to MUC1 (PankoMab) and Thomsen-Friedenreich F (Nemod TF) epitope	Up regulation of TF epitope colocalised with overexpression of MUC1,	Human endometrial tissue was obtained from 54 premenopausal patients and was immunohistochemically analysed with monoclonal antibodies against MUC1, TF epitope	[57]
Cationic Ferritin and Ruthenium Red staining	Decrease in negative charge during receptive phase	The use of negative charging labels in pregnant rat uterine epithelial cells	[58]
Increase binding of FBP (fucose) and SBA (terminal α-GalNAc) lectins	Increase in fucose and terminal α-GalNAc	Staining of rats uterus surface during implantation and post insemination	[59]
Loss of PNA (Gal-β(1-3)-GalNAc) binding at secretory (receptive) phase and restoration post sialidase treatment	Increase in Gal-β(1-3)-GalNAc sialylation during receptive phase	Mice endometrial tissue fixed in formalin and processed in paraffin wax prior to lectin staining	[60]
Expression of Lewis a, sialyl Lewis a, Lewis b, Lewis x, sialyl Lewis x and lewis Y	Lewis a, sialyl Lewis a and Lewis b were detected at apical surface of a minority of glandular epithelial endometrial cells. Lewis x, sialyl Lewis x and lewis y were present in cytoplasm in both phases with Lewis x decreases at secretory phase	Lectin staining and microscopy of mice endometrial tissues using several antibodies specific to Lewis structures	[61, 62] [63]

Table 4.7 Reported glycosylation seen in endometrium at receptive stage. All examples are performed in human tissues except shown otherwise.

Most of these observations have used lectin staining which does not give quantitative information or definitive evidence of the glycan structures. In contrast, glycan analysis by mass spectrometry

has the ability to show more detailed structural information on the glycan markers for uterine receptivity, such as the type and abundance of glycan classes, expressed structural epitopes and sites of attachment. To the best of our knowledge, no MS analysis of uterine glycosylation has been performed to date. Only cervical mucus has been analysed using MS based techniques [64]. Glycans analysed from cervical mucins by MS demonstrated glycosylation changes between before and after ovulation such as an increase in core type 2 *O*-glycans and neutral fucosylated structures, with a relative decrease in sialylated structures [64].

The importance of the highly glycosylated MUC1 in embryo implantation

Previously published data has implicated *O*-glycosylated mucin proteins in the success of embryo implantation since one of the most important phenomena observed in a receptive uterine endometrium is the decrease in MUC1 expression in mice [2]. Mice, unlike human, exhibit a reduction in MUC1 expression during blastocyst implantation [4-8],but mice uterine MUC1 is significantly increased during the window of implantation in diabetes mellitus [23]. The reported increase in MUC1 expression in diabetic mice endometrium is a probable cause for the low embryo implantation rates seen in diabetes mellitus [23]. Due to the mucin's large size and high glycosylation of serine and threonine repeats, our sample preparation for shotgun proteomics in Part 1 of this Chapter was not optimal for mucin identification. However, in our proteomics comparison, we were able to detect several altered protein functions involved in mucin production such as reduction-oxidation process, ER stress and vesicle mediated transport (Table 4.4 and 4.5) [65-67]. Although our analysis did not detect the MUC1 protein, several unique proteins in our comparison between states are highly involved in mucin functionality such as EGFR, PDGFR and proteins involved in vesicle meditated transport (Table 4.6).

Process/protein	Relationship to MUC1 in other tissues	Conductive oviduct (1.5 days)	Receptive uterus (3.5 days)
Vesicle mediated transport proteins	Mucus secretion [68]	Decrease compared to post-conductive	Increase in diabetes when its expected to decrease (Figure 4.7)
Platelet-derived growth factor receptor	Required for normal development of the mucosa lining [69]	Unchanged	Decrease in diabetes
Epidermal growth factor receptor	Phosphorylates MUC1 and increases MUC1 interactions [70]	Increase in diabetes	Decrease in diabetes

Table 4.6: The differential expression of biological process and proteins involved in MUC1 expression and function due to WOI or diabetes in mouse.

Overall, our data supports a possible alteration in diabetic uterus of the highly glycosylated MUC1, a protein with a fundamental role in the correct embryo implantation process [71] [72] that requires a highly spatial and temporal regulation of cellular events occurring in the uterine surfaces [73]. One of these cellular events is the regulation of the mucin layer expression on the surface of uterus [74]. Mucins are large, sticky, extensively O-glycosylated and have been seen in most reproductive organs, with reported changes in their amount and phenotype during the menstrual cycle [71]. MUC1 was found to be the most abundant mucin protein expressed in oviduct and uterus [75]. MUC1 is a large glycoprotein (>250KDa) that obstructs interaction through steric hindrance and hence the inhibition of blastocyst implantation is proportional to the length of MUC1 ectodomain. As illustrated in Brayman [72], the proline residues and glycosylation give rise to a rigid extended MUC1 structure that projects 200-500 nm above the cell surface, much further than the distance spanned by most cell-surface proteins including integrins (Figure 4.2) [76] [77]. In mice, and in more localised regions around pinopods in humans, MUC1 experiences a transcriptional and translational reduction at the surface of pinopods, at embryo anchor sites, and at the implantation moment [74].

MUC1 abundance during the window of implantation

The endometrium (MUC1 layer) microenvironments during the receptive phase are highly complex and constantly changing among molecular players as implantation progresses [78]. As previously mentioned, unlike human and rabbits, MUC1 is down regulated before implantation in

the receptive endometrium of mice [4, 5], rats [6] and pigs [7]. Interestingly, in rabbit, direct studies of implantation utilising affinity purified rabbit antibody to the peptide CT-1 of MUC1 (immune-fluorescence localisation) have shown that MUC1 is absent at the interface between trophectoderm (blastocyst) and uterine epithelium [79]. Scanning electron microscopy combined with immuno-histochemistry has succeeded in precisely showing MUC1 missing from the surface of uterine pinopods [80]. Indeed, human *in vitro* implantation models indicate MUC1 loss at the site of embryo attachment [81]. This *in vitro* three-dimensional system showed MUC1 staining to be absent from epithelial cells beneath and in the surrounding vicinity of the attached embryo, where its expression was unaffected at greater distances [72, 82].

MUC1 glycoforms during the window of implantation

The importance of different MUC1 glycoforms in the reproductive tract is seen in the different patterns of expression of MUC1 glycoforms observed between pre-receptive and receptive implantation phases on the luminal (surface) epithelium [83, 84]. In human, although MUC1 protein expression has been shown to not change on glandular (secretory) epithelium, antibody CT-1, directed against the cytoplasmic domain found in all cell-associated MUC1 species, and two MUC1-ectodoamin epitopes, HMFG-1 and HMFG-2, of biopsy samples have observed differences in their immunohistochemistry morphology [84]. Luminal epithelium is responsible for embryo attachment and HMFG-1 ectodomain expression was shown to decrease in luminal epithelium during the receptive phase; and its reactivity was seen to be restored by keratanases and neuraminidase treatment suggesting a decrease in keratan sulphate and, interestingly, sialic acid-D-Galactose in MUC1 during receptive implantation phase compared to pre-receptive with no differences in MUC1 glycoforms seen in glandular epithelium [85]. Although derived from glandular epithelium, MUC1 covered luminal epithelium has acquired distinctive glycosylation patterns. Overall, the use of specific antibodies has determined MUC1 epithelial glycoforms to be distributed in a regional and cycle-dependent fashion [84]. The glycosylation found on MUC1 glycoforms is yet to be studied in mice.

Diabetes and MUC1 glycosylation

A thick endometrium is indicative of non-conception cycle [4, 5, 8, 86], and some have reported thick uterine lining in diabetic mice during the receptive window [23]. STZ-induced diabetic mice have been shown to exhibit reduced embryo implantation rates, retarded pinopods and above all overexpression of the highly O-glycosylated MUC1 mucin protein at the receptive phase compared to normal non-obese mice, the latter seen by transmission electron microscopy (Table 4.8) [87-89][23]. The high expression of MUC1 is believed to cause steric hindrance of the binding between blastocyst and cell adhesion molecules such as cadherins and integrins (Figure 4.8) [90]. Diabetes has also been repeatedly shown to alter protein glycosylation, (reviewed in paper 1 and Chapter 3).

Mice model	1.5 days (post coitus)		3.5 days (post coitus)	
	Oviduct	Uterus	Oviduct	Uterus
Presence of embryo	Two-cell zygote	No	No	Multi-cellular Blastocyst
Hormones	Estrogen (proliferation)		Progesterone (differentiation)	
tissue state	Conductive to embryo transfer	Pre-receptive	Post-conductive	Receptive
Docking sites		No		Present
Docking sites in diabetes		No		Retarded
MUC1 expression	High	High	low	Low
MUC1 expression in diabetes	N/A	High	N/A	High

Table 4.8: Details on mice oviduct and uterine tissue samples used in this work. Two time points, 1.5 and 3.5 days post coitus, of control and STZ-induced diabetic mice oviduct and uterine lining were analysed.

Thus, the increase of MUC1 expression in receptive STZ-induced diabetic mice uterus is established and we have investigated the MUC1 glycosylation patterns as well, which would influence the size, hydrophilicity and above all binding ligands (glycan epitopes) of MUC1 on the surface of oviduct and uterus [2]. Unlike mice, human receptive uterine tissues do not experience an overall reduction in MUC1 expression but may rely on specific changes in MUC1 glycosylation patterns [4, 5, 8]. Given the established overexpression of MUC1 expression in diabetic mice [23], changes in MUC1 glycosylation patterns in diabetic mice may also occur in diabetic human receptive uterus, and be related to the the low successful embryo implantation rate in diabetic individuals. In fact, diabetes has been repeatedly shown to alter the glycosylation of several proteins [91]. Overall, the effect of diabetes in mice, and human on MUC1 glycosylation is still unknown.

Challenges of studying mucins and their O-glycans

Mucus is composed primarily of water (95%), but also contains salts, lipids such as fatty acids, phospholipids and cholesterol [1]. However, the main component that is responsible for its viscous and elastic gel-like properties is the glycoprotein mucins.

Mucins are made up of a family of 20 molecules and 8 possible core *O*-linked glycan structures [92]. Mucins are highly *O*-glycosylated with repeats of serine and threonine residues forming a large number of glycans in very close proximity [93]. Jensen *et al* has summarised the reasons that make *O*-linked glycans more difficult to analyse than *N*-linked protein glycans [94]. First of all, there is the absence of a known amino acid consensus glycosylation sequence. Secondly, there is no universal enzyme to cleave the *O*-glycans from protein substrates. The commercially available *O*-glycanase only cleaves the structure Gal-GalNAc and leaves other structures alone, hence resistance to *O*-glycanase does not mean the absence of *O*-linked glycans [95]. Finally, mucins are very heterogeneous in *O*-glycan structures and their possible glycosylation sites [94]. As mucin proteins differ in their number of repeat sequences, large size and heterogeneous glycosylation (hydrophilicity), this results in inherent difficulties in analysing these molecules.

Mucins have been reported to have numerous functions in several diseases such as cystic fibrosis, asthma, cancer and others [96-98]. As a result, a reliable analytical protocol is pivotal to investigate the importance of these molecules and their wide involvement in biology, including in the context of this thesis, diabetes.

Common mucin isolation protocols.

Due to the difficulties with working with mucins using conventional separation techniques, several techniques have been tested for the purification of these proteins. Properties that make mucins hard to study (high molecular weight and high glycan content) are beneficial for their purification from other proteins. There are varieties of methods that can be applied to study mucins, yet only few are reproducible, low cost and can be used in high-throughput studies. Several techniques have been used routinely for mucin purification such as Cesium chloride (CsCl) density gradient

centrifugation [99], Agarose-Polyacrylamide electrophoresis (Ag-PAGE) [100, 101] and Supported Matrix Membrane Electrophoresis (SMME) [102].

Traditionally, CsCl density gradient has been well-established for the extraction of mucins. Following mucin solubilisation with guanidine hydrochloride, CsCL density gradient centrifugation has been applied to purify mucins from proteins and DNA [99]. The high-molecular weight of mucins prevents their migration in regular polyacrylamide gels. Ag-PAGE gels contain large pores via mixing agarose and polyacrylamide, thus allowing mucins to enter the gel [103]. Gels can then be transferred onto a PVDF membrane via a vacuum blotter, and the mucin glycans typically released via β-elimination of the *O*-linked glycans directly from the immobilized mucin on the PVDF [100, 101].

A relatively new membrane electrophoresis technique has been developed for the purification of mucins called Supported Matrix Membrane Electrophoresis (SMME). SMME has promised simplicity and rapid characterisation [102]. SMME requires fixing mucins on a PVDF membrane and adding water-soluble polymer (such as Polyvinylpyrrolidone (PVP)) in the fixation process. By this method MUC1 has been detected in several cancer cell lines and *O*-glycans were released for analysis; SMME was also able to detect MUC1 using antibodies to demonstrate that the migration pattern and glycan profile was different between different cell lines [104]. This part 2 of the Chapter aims to compare any differences in oviduct and uterine mucin glycosylation (mainly MUC1) between pre-receptive and receptive using mass spectrometry after comparing different mucin isolation protocols.

4.2.1 Materials and methods

4.2.1.1 Testing different mucin isolation techniques

For all protocols, as an example of highly glycosylated mucin proteins, porcine gastric mucin (PGM) was reduced with fresh 10 mM DTT (Sigma Aldrich, Australia) and incubated at 37˚C for

5 hours with slow mixing (300 rpm). Alkylation was performed with 25 mM IAA (Sigma Aldrich, Australia) at room temperature overnight in the dark.

PVDF and in-solution O-linked glycan release for mass spectrometry analysis

PGM (20 ug) was dot-blotted onto a PVDF membrane using a vacuum blotter and *N*-glycans were removed by an overnight incubation with 2.5 U *N*-glycosidase F (PNGase F, *Flavobacterium meningosepticum*, Promega, Australia) at 37°C, then *O*-glycans were released by β-elimination by the addition of 1 M $NaBH_4$ in 50 mM KOH for 3 h at 50°C as described previously [105].

For in-solution release, PGM (20 ug) was kept in-solution while incubated with 2.5 U *N*-glycosidase F (PNGase F, *Flavobacterium meningosepticum*, Promega, Australia) at 37°C to remove *N*-glycans. Proteins were acetone precipitated (1:9 ratio) by incubation at 20°C overnight. Protein pellets were washed with 300 ul of 60% v/v methanol at -20°C repeated twice; the acetone/methanol extracts contained the released *N*-glycans. Pellet was then subjected to β-elimination using 0.5 M $NaBH_4$ in 50 mM KOH for 16 h at 50°C, reduction was neutralised by the addition of 2ul of glacial acetic acid. Effervescence will be observed upon the addition of acid. Samples were desalted and cleaned as described in section 2.2.

CsCL density gradient centrifugation

PGM (100ug) was added to a Cesium Chloride (CsCL-Biochemicals, Australia) at a concentration of 1.4 g/ml. Samples were added onto the CsCl in polyallomer tubes (12ml) and balanced for subsequent ultracentrifugation. Ultracentrifugation was performed using Beckman SW41 Ti rotor at 36000 rpm for 72 hours. A fraction collector (GE-HealthCare, Australia) was used to carefully remove fractions at a speed of 1ml/min (from bottom). The density of each fraction was measured. All fractions were spotted onto a PVDF membrane and stained with Alcian blue to detect negatively charged (sialylated/sulphated)-mucin rich fractions. Briefly, dry membranes were immersed into 0.1% w/v Alcian blue 8GX dissolved in 10% v/v acetic acid and 25% v/v ethanol, while gently shaken for 2 min [106]. Membrane was washed with methanol for a few min to determine the spots positive for Alcian blue staining. Mucin fractions, stained positive with Alcian blue and had 1.3 g/ml density, were dialysed overnight in 1000ml milliQ-water to remove access

CsCL salts. Samples were then spotted onto PVDF membrane using a vacuum blotter and used for subsequent *O*-glycan release as described in section 2.2.1.

Agarose-Polyacrylamide (Ag-PAGE)

Ag-PAGE gels were made by mixing two solutions of agarose and acrylamide using a special in-house apparatus in a 70˚C oven [107]. The upper chamber solution contained 0.5% w/v agarose while the lower chamber solution contained 2% w/v agarose: 3% w/v acrylamide: 2% v/v glycerol. TEMED (0.07% w/v) and 2ul ammonium persulphate (40% w/v) were added into lower solution before mixing began. The mixed solution was slowly poured into gel cassettes before incubating for 30 min at room temperature followed by storage at 4˚C. After adding lithium dodecyl sulfate (LDS) sample loading buffer to PGM (100ug), the gel was run at 10mA for 2 hours. Gels were stained with Coomassie Blue G-250 overnight and de-stained with 3% v/v acetic acid. Proteins were transferred from the gel to a PVDF membrane, without any staining, via semi-dry transfer technique (Trans-Blot SD semi-Dry Transfer cell (BioRad, Australia)). The semi-dry transfer apparatus was first made wet with ion trap solution (10x 3 M Tris base) in order to trap charged particles between the PVDF membrane and blot papers. PVDF membrane wetted with methanol, and then added to cathode buffer solution (10x 250mM Tris base, 400mM α-amino-n-hexanoic acid) for 10 min. Two blot papers (BioRad, Australia) were soaked in cathode buffer solution for 10 min. Two more blot papers were then soaked in anode buffer solution (10x 250mM Tris base) for 10 min. The gel was soaked in cathode buffer solution for 5 min. Papers soaked in ion trap solution were stacked first, then papers from anode followed by PVDF, then gel and finally cathode paper; while making sure to remove any air bubbles formed between each layer. Ice was placed on top of apparatus to keep cool while blotting at 5mA for 80 min. Alcian blue was used to stain PVDF membrane while Coomasie blue was used to stain gels before and after transfer. Stained spots on the PVDF membrane were used for subsequent *O*-glycan release as described in section 2.2.1.

Supported matrix membrane electrophoresis (SMME)

SMME was performed as described in a recent publication by Kameyama [108]. PVDF membrane wetted with methanol and then immersed into 0.25%w/v hydrophilic polymers (PEG/PVA 3:2) for 30 min. Excess polymers were removed by membrane filtration. Membranes were then placed into the SMME apparatus using 0.5%v/v pyridine as running buffer (Figure 4.9). 10ug PGM was run onto the starting line with constant current 1mA for 90 min. Mucin fractions were stained with Alcian blue and used for subsequent *O*-glycan release as described in section 2.2.1.
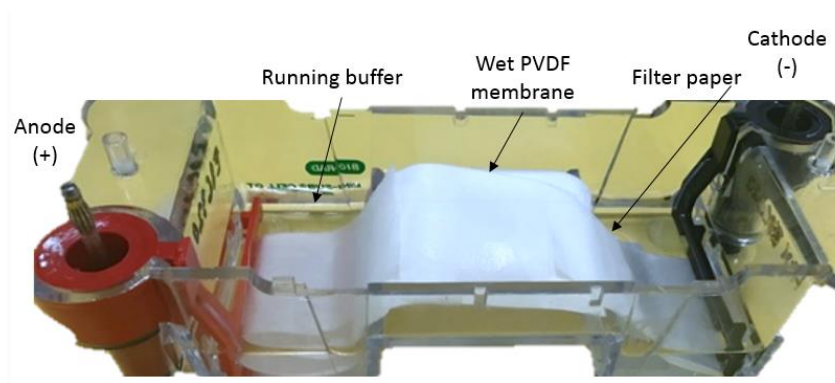

Figure 4.9: An image of supported molecular matrix electrophoresis (SMME) apparatus. The image shows two electrical poles (cathode and anode), PVDF membrane, filter paper and running buffer (0.5% pyridine).

Desalting and carbon cleaning of released glycans

Only released *O*-glycans were desalted on a strong cation exchange (SCX) chromatography column and enriched using a Porous Graphitised Carbon (PGC) column as described in section 2.2.2.

Mass spectrometry analysis of released glycans

Analysis of released *O*-glycans was performed using a PGC column coupled to a ThermoFisher Velos LTQ mass spectrometer as described in in section 2.2.4. No analysis was performed for released *N*-glycans in this Chapter.

4.2.1.2 The analysis of oviduct O-glycome and uterine mucin layers

Mucin preparation for glycomics experiments

For uterus, the endometrial epithelial surface was scraped to remove the glycocalyx away from cells, to collect mucins using a blunt instrument (forceps) to collect mucins. For oviduct, tissue was homogenised using a Polytron homogeniser in 10mM Tris buffer (pH 7.5) for 3 min (30 sec on:10 sec off). For both endometrial uterine scrapings and homogenised oviduct cells, samples were acetone precipitated overnight in 1:9 ratio. Acetone pellets from all samples weighed about 50ug measured by subtracting the weight of the tube containing acetone pellets from the tare weight. The epithelial scrapings and oviduct acetone pellets were placed in at least five volumes of ice-cold 6 M guanidinium chloride extraction buffer and dispersed with the Polytron for 3 min (30 sec on:10 sec off). Samples were gently stirred overnight at 4°C, then centrifuged (23,000 × g, 4°C, 45 min). Supernatants were poured off and termed "guanidinium chloride soluble" mucins. Extractions steps were repeated two more times and all extracts were pooled together. Guanidinium chloride (6 M) reduction buffer containing 10 mM DTT was added to the extracts and incubated for 5 h at 37°C. Iodoacetamide (IAA) (25 mM) was added and left to incubate overnight in the dark at room temperature. Supernatant was termed the "reduced/alkylated guanidinium chloride insoluble" mucins [99]. Both fractions were mixed and dot blotted onto PVDF membrane using vacuum blotter.

Periodic acid-Schiff and Direct Blue staining

For two PVDF membranes, 5ug mucins protein from all uterine mice samples were immobilised using a vacuum blotter. One membrane was oxidized in 1% w/v periodic acid in 3% v/v acetic acid for 30 min before staining with Schiff's reagent for 15 min (Sigma-Aldrich, Australia). The second membrane wetted using methanol, then stained with Direct Blue (10% v/v acetic acid, 40% v/v ethanol, 0.1% w/v direct blue), and de stained (10% v/v acetic acid, 40% v/v ethanol). Spot intensities were measured using image G (USA National Institute of Health). Each class (control/diabetic pre-receptive and receptive) was prepared as 3 biological replicates and two

technical replicates. Protein intensities were normalised from the intensity of empty spots containing no proteins. Human saliva, purified as per [109], was used as a control to reflect the accuracy of our staining methods.

O-linked glycan release for mass spectrometry analysis

Reduced and alkylated mucin samples were then spotted onto PVDF membrane using a vacuum blotter. Blotted samples were treated with 2.5 U *N*-glycosidase F (PNGase F, *Flavobacterium meningosepticum*, Roche, Australia) at 37°C, then *O*-glycans by β-elimination (0.5mM NaBH$_4$, 50mM KOH) as described in section 2.2.1.

Released *O*-glycans were desalted and enriched using SCX and PGC column respectively (details found in sections 2.2.2).

Mass spectrometry analysis of released O-glycans

Released *O*-glycans were analysed using a PGC column coupled to a ThermoFisher Velos LTQ mass spectrometer and statistical analysis was performed using Student t-tests as described in sections 2.2.4 and 2.2.5.

4.2.2 Results

4.2.2.1 Comparison between different mucin analysis protocols

The overall results from most of the techniques tested showed Porcine Gastric Mucin (PGM) to be carrying around 85% non-charged glycan structures. In-solution β-elimination release allowed the characterisation of the least number (23 *O*-glycan strtucres) of individual glycan strctures, while PVDF and all other techniques (CsCL density gradient, Ag-PAGE and SMME) released 27 or 28 *O*-glycan structures (Figure 4.10).
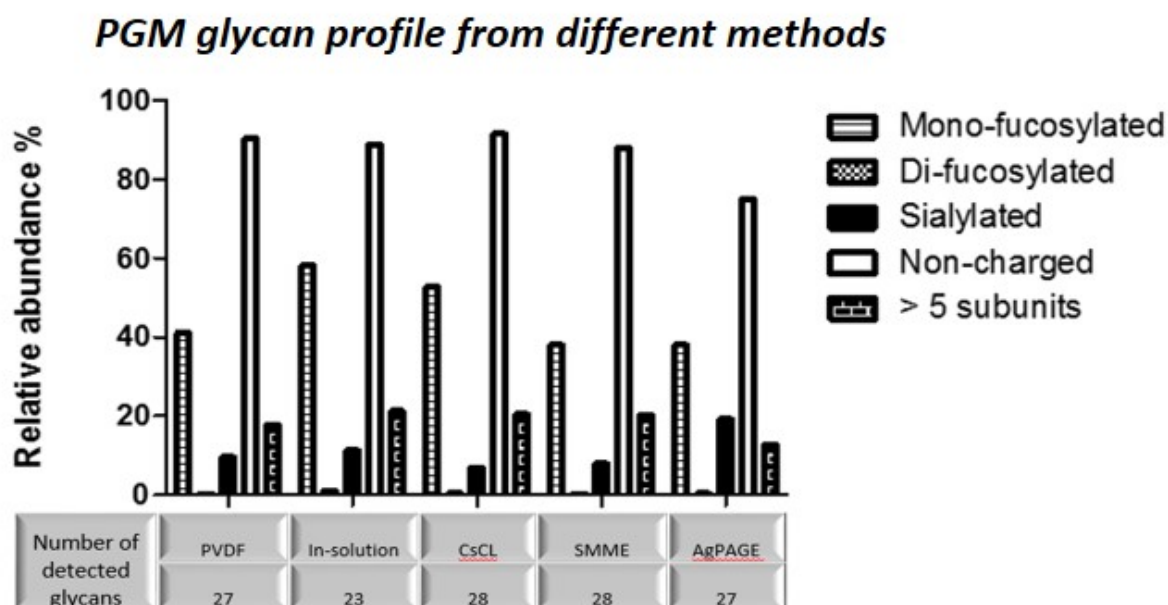
Figure 4.10: *O*-linked released glycans from PGM using five different techniques: PVDF (dot blotting), in-solution, CsCL density gradient, SMME and AgPAGE. The abundance of each glycan class is calculated from the EIC measured for each individual glycan mass in that class. The figure also provides the overall number of individual glycans (including isomers) detected by each technique.

There are reported advantages and disadvantages for each of the mucin analysis protocols (Table 4.9). Due to high purity (only MUC1) [2][9], low amount of samples in oviduct and uterine mucin proteins and large cohort (total of 24 biological samples), PVDF blotting was chosen over CsCl density gradient, Ag-PAGE and SMME for the work presented in the remaining part of this Chapter. See Supplementary figures in chapter 4 for details on the comparison of mucin analysis protocols.

Protocols	Advantages	Disadvantages
PVDF blotting (vacuum blot)	• High signal to noise ratio. • Little loss of sample material.	• Not applicable with complex mucin mixtures.
In-solution	• Little loss of sample material.	• Low signal to noise ratio. • Not applicable with complex mucin mixtures. • The only method lacking in number of detected *O*-glycan structures.
Cesium chloride density gradient centrifugation (CsCL)	• Isolation of different mucin proteins or glycoforms. • Optimal for purifying mucins from complex mixtures and tissues.	• Requires large amount of sample. • Not required with relatively pure mucin samples to avoid sample loss. • Loss of sample during dialysis with mucin sticking to the dialysis bags.
Supported matrix membrane electrophoresis (SMME)	• Most resolution of PGM mucin proteins or glycoforms.	• Irreproducibility with the loading PGM into the polymer matrix (fixation process) • Irregularities due to basic/hand-made apparatus • Few spots can be made per each run; difficult for large cohort
Agarose-Polyacrylamide (Ag-PAGE)	• Isolation of different glycoforms. • Purifying mucins from complex mixture.	• Requires large amount of sample. • Hard to accurately reproduce gel casting and membrane transfer of large mucin proteins. • Not suitable for large cohort of samples.

Table 4.9: Comparison of mucin (PGM) analysis protocols: PVDF (dot blotting), in-solution, CsCL density gradient, SMME and AgPAGE. Observed advantages and disadvantages are shown with details found in Supplementary Figure 4.

When comparing PVDF and in-solution glycan release technique, the main difference between the two profiles was the signal to noise ratio generated from the same amount of PGM amounts such that the in-solution technique release gave a higher signal to noise ratio. Also, PVDF release generated more *O*-glycan individual structures (27) compared to in-solution (23) (Figure 4.11).
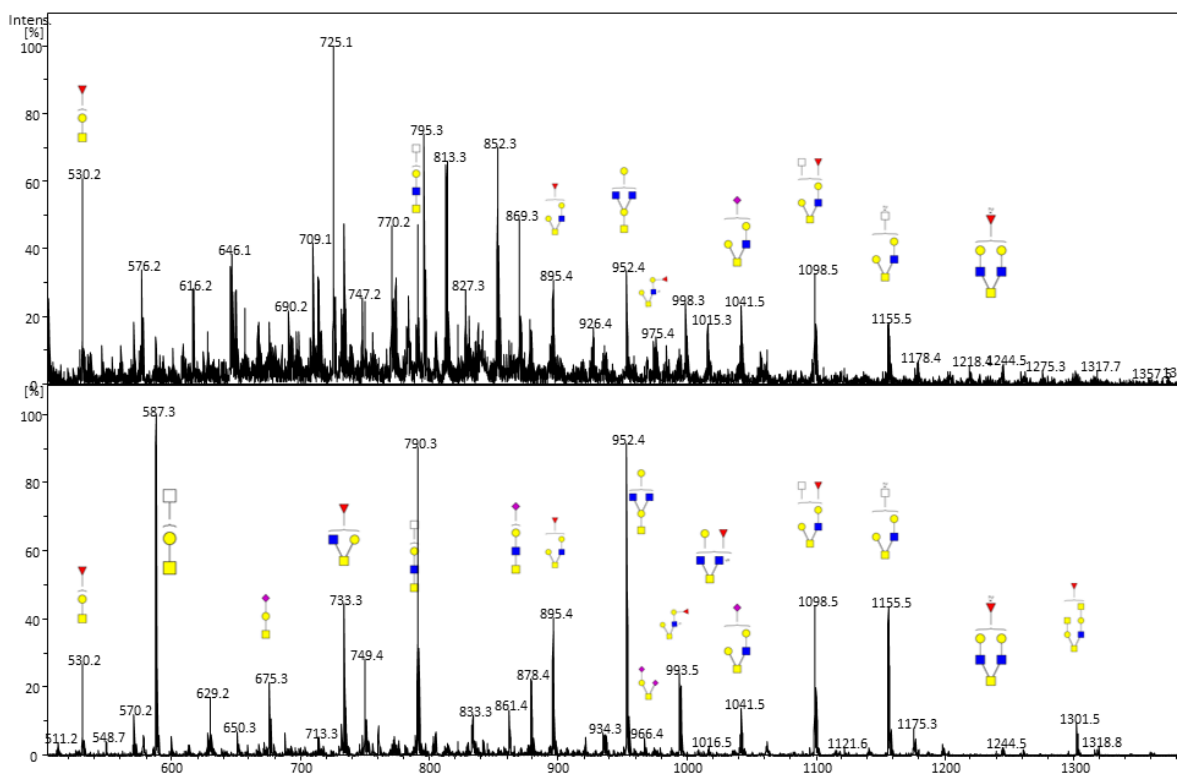
Figure 4.11: Summed MS of *O*-glycans released from PGM via in-solution (top) and PVDF dot-blotting (bottom) methods. The x-axis represents mass over charge with each signal characterized by its assigned *O*-glycan structure.

Our comparison between in-solution and PVDF mucin glycan release has answered a very important question: What is the best approach to release *O*-glycans from purified mucins? Our results show that blotting mucin using a vacuum blotter before β-elimination provides a clean *O*-glycan profile compared to in-solution release. Also, the PVDF approach has enabled the characterisation of more *O*-glycan structures than in-solution release (4 more glycans), mainly due to the higher signal-to-noise MS ratio, so was the method of choice for the following mucin glycosylation analyses.

4.3.2.2 A micro-glycomic view of the uterine and oviduct O-glycome surfaces

Using PGC-ESI-MS/MS, the analysis of *O*-glycans in oviduct and uterine scraping showed no significant difference between 1.5 and 3.5 days post coitus. Also, there was no significant change in the individual *O*-glycan structures between control and diabetic samples of oviduct and uterine surfaces (Figure 4.12, A and B).
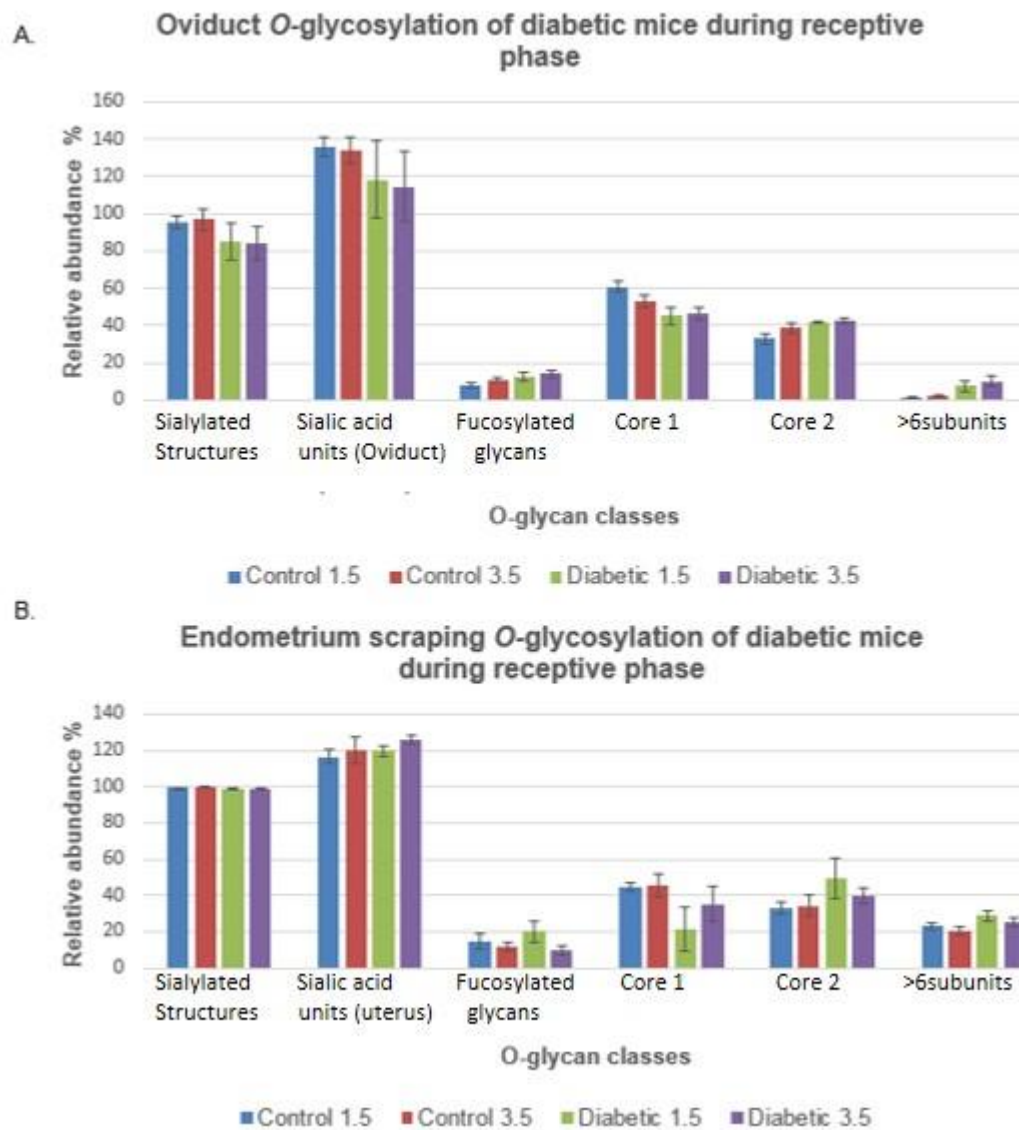
Figure 4.12: Analysis of MUC1 glycosylation in conductive/receptive and diabetic oviduct and uterine tissues. A: *O*-linked released glycans from control and diabetic mice oviduct. The abundance of each class is relative to the EIC measured for all glycans found in this class compared to all other glycans. SDs are shown (n=3). B: *O*-linked released glycans from control and diabetic mice uterine scraping. The abundance of each class is relative to the EIC measured for all glycans found in this class compared to all other glycans. SDs are shown (n=3).

Increased total glycans in receptive uterine surface compared to pre-receptive.

In order to measure the abundance of glycans on the surface of uterine scrapings, the Periodate acid-Schiff (PAS) carbohydrate staining method was used. In the above glycan structure analyses the initial protein amount added to each spot was normalised to total protein amount (5ug) via

152

Direct Blue staining. Saliva was used as a control with the same amount of protein to reflect the accuracy of measurements of this method seen from the standard deviation of n=2 (Figure 4.13, A). PAS stain of mucins scraped from uterine surface revealed an increase (around 300%) in the amount of carbohydrate displayed in the receptive phase compared to non-receptive phase. There was an increase in PAS stain relative to normalised mucin protein amount (Direct Blue stain) in receptive (3.5 C; 8.8±3.4) compared to pre-receptive (1.5 C; 1.4±0.6) mice uterus (Figure 4.13, B). Also, there was an increase (around 100%) in the amount of carbohydrate on the mucins of diabetic pre-receptive mice (1.5 D; 4.2±1.4) compared to control pre-receptive (1.5 C; 1.4±0.6); this increase was not evident in the comparison between diabetic receptive mice (3.5 D; 7.4±3.8) and control receptive mice (3.5 C). Overall, our results suggest an increase in carbohydrate on receptive uterine surface mucin.
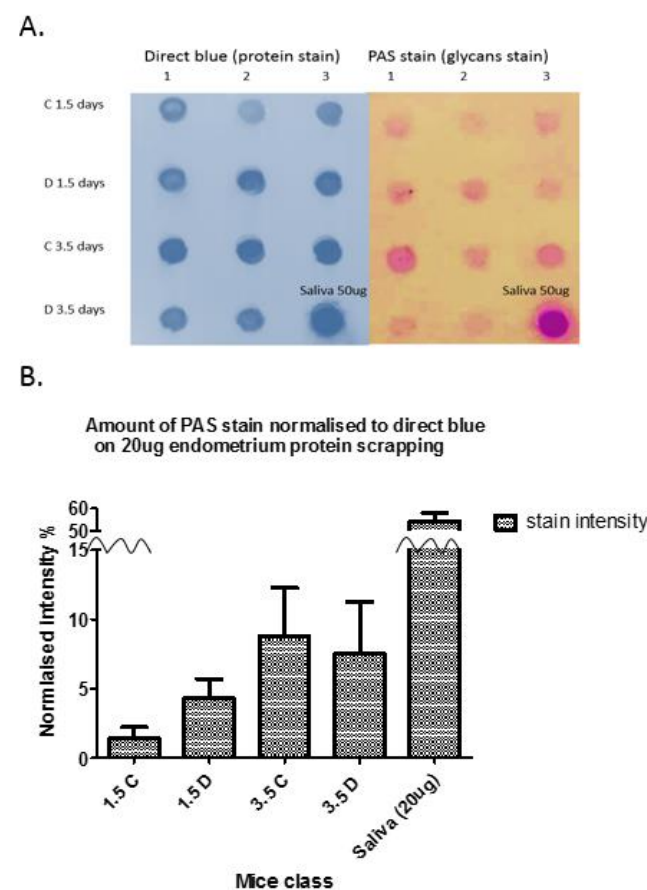

Figure 4.13: Analysis of MUC1 glycosylation in receptive and diabetic uterus tissues. A: PAS and direct blue staining of uterine scraping from control and diabetic of 1.5 and 3.5 days post coitus. B: normalised intensity obtained from part A using image G.

153

4.2.3 Discussion

Due to the importance of mucin proteins in uterine receptivity and the fact several proteins (PDGFR and EGFR) and processes involved in mucin secretion and maintenance showed significant changes in our proteomic analyses of oviduct and uterine tissues due to WOI and diabetes, the mucin component was analysed using carbohydrate staining as well as mass spectrometry to determine glycan amount as well as structures.

All mucin analysis protocols (PVDF (dot blotting), in-solution, CsCL density gradient, SMME and AgPAGE) provided a similar number *O*-glycans found and similar glycan type profile of, except for in-solution techniques which showed lower number of *O*-glycan found. Released *O*-glycans from PGM on a PVDF dot blot provided better signal-to-noise ratio compared to in-solution release. The other mucin analysis protocols had several disadvantages that made them inapplicable to the use of endometrial mucin analysis. These disadvantages include the requirement of large amount of sample, big loss of sample, irregularities with apparatus setup, hard to accurately reproduce and unsuitability for a large cohort of samples (Table 4.9). Endometrial mucins are made mainly of MUC1 [2][9] and produce low amount of sample. Also, our study has a large cohort of samples (total of 24 biological samples). As a result, PVDF blotting was chosen over other release protocols in our subsequent analysis of mucins in the oviduct and uterine scrapings.

Our study is the first report on the *O*-glycosylation of mice conductive oviduct and receptive uterine mucin using MS techniques. Our glycomic analysis of oviduct tissues might contain *O*-glycans expressed in the surface of other oviduct membrane proteins but mainly are MUC1 *O*-glycans [75]. Several glycan epitopes were present on both oviduct and receptive uterine tissues (Table 4.10).

Glycan epitopes found in both oviduct and uterine receptive surfaces	Importance	Previous observations	References
Sialylated epitopes	Sialic acid is proposed to inhibit blastocyst implantation	WGA binding (sialic acid) was shown to decrease in receptive uterus of rabbits, monkeys and human	[53, 54, 85]
Fucosylated epitopes	Fucose is proposed to form epitopes for blastocyst docking	Fucose is highly present in receptive endometrium	[49]
Expression of sialyl Lewis x, Lewis x and, Lewis a	These epitopes are known to bind to glycoproteins found in the surface of blastocyst such as L-selectin.	These epitopes were previously observed in endometrial cells	[61-63, 110]
Sulphated epitopes	N/A	Human endometrial scraping was previously reported to contain 8.4% sulphated mucins	[55]

Table 4.10: Summary table of glycan epitopes reported in oviduct and receptive uterine mucin surfaces.

To date, little is known about O-GalNAc linked glycan structures on proteins (membrane) in diabetes mellitus compared to studies on the single monosaccharide addition, O-GlcNAc (cytoplasmic) which has been extensively investigated due to its direct involvement in signalling and diabetes [111]. Using PGC-LC-MS/MS glycan analysis, no significant changes in the individual O-glycan structures were found on the surfaces from oviduct and uterine scraping between 1.5 and 3.5 days post coitus, as well as between control and diabetic states. Despite a common trend towards a lower relative abundance of core type 1 O-glycans and a trend to higher relative abundance of core type 2 and >6 subunits (large) structures exhibited by diabetic oviduct and uterine tissues seen in both time windows; this trend was not shown to be significant. There has been a reported increase in MUC1 expression in diabetic mice during receptive phase [23], and our carbohydrate staining results reflect this as an increase in total carbohydrate on the mucin proteins in pre-receptive uterine lining, as well as in diabetes (1.5 days) to a lesser extent.

Possible increase in the number of MUC1 O-glycosylation sites on the uterine lining during WOI
The comparison between pre-receptive and receptive stages showed more carbohydrate/protein staining of receptive uterine scrapings, whereas the MS glycomic analysis of the same samples

showed no significant difference in the relative abundance of observed structures. As a result, the increase in carbohydrate/protein staining and the absence of change in the relative abundance of the displayed structures suggests there was an increase in the number of O-glycosylation sites on the mucins which would account for the increase in overall glycan abundance. Alternatively, mucin proteins have been reported to decrease in expression in receptive uterus in the mouse (contrary to in humans) [86] [8] [4, 5], which implies that there may be less mucin proteins but with higher glycosylation present on the lining of the receptive uterus.

Using lectin binding, the uterine cell surface (Monkey) has been shown to exhibit changes in sialylation expression due to WOI [53, 58, 60]. As our MS analysis of the O-glycans on mucin scraping from oviduct and uterus have shown no structural differences due to WOI, any differences between our results and previously published reports could be a result of several reasons including differences in animal models or the fact our study measures relative abundances of each structures such that absolute glycan abundance may be different. Also, previous reports have used lectin binding, which will detect epitopes found on both lipids and proteins; our study only focuses on protein (mucin) glycosylation. Overall, there was an increase in carbohydrate staining of receptive uterine scaping (mucins) compared to pre-receptive uterine, which suggests an increase in the number of O-glycosylation sites on the expressed mucins.

4.3 Conclusions

Testing for uterine receptivity is likely to provide the next leap forward for infertility clinics. Several biological events are essential to allow for sperm-egg fertilisation in oviduct and subsequent blastocyst implantation in a receptive uterus during the window of implantation (WOI), including adhesion and proliferation. To gain insight into oviduct and receptive uterine surfaces, MS technologies were utilised to analyse changes in biological processes acquired from the membrane proteomes and the glycosylation of the mucin layer covering the surface of diabetic-mice uterus at 1.5 (conductive oviduct) and 3.5 days (receptive uterus) post coitus. The production of highly stringent protein IDs was facilitated by the use of a software package called

PeptideWitch. At the event of zygote formation (1.5 days post coitus), the membrane proteome of oviduct exhibited a change towards higher capacity to transport and deliver proteins to their proper locations, while uterus showed a decrease in oxidative phosphorylation; the latter oxidative phosphorylation process was shown to significantly increase in uterus in response to diabetes treatment. Diabetes resulted in two other changes in receptive uterus including increase in vesicle transport and a decrease in auto phagosome assembly. In addition, cell adhesion molecule (integrinB3) and growth factors were shown to be expressed significantly higher in control receptive uterine surface, but the growth factors group was significantly reduced due to diabetes. However, an important marker of a receptive uterus is the reduction in mucin surface proteins [2]. Diabetic mice have been shown to exhibit reduced embryo implantation rates and overexpression of the highly *O*-glycosylated MUC1 mucin protein in uterus at the receptive phase [23]. Using an optimised protocol of mucin analysis, no significant differences in individual *O*-glycan structures were found in oviduct and uterine surfaces due to WOI or diabetes. However, the carbohydrate/protein staining of receptive uterine lining increased in the WOI from 1.4±0.6 (1.5 C) to 8.8±3.4 (3.5 C), and in diabetes to a lesser extent from 1.4±0.6 (1.5 C) to 4.2±1.4 (1.5 D), implying that there is in an increase in the number of *O*-glycosylation sites on the expressed mucins in these conditions.

4.4 References:

1. Leung, C.Y. and M. Zernicka-Goetz, *Mapping the journey from totipotency to lineage specification in the mouse embryo.* Current opinion in genetics & development, 2015. **34**: p. 71-76.
2. Achache, H. and A. Revel, *Endometrial receptivity markers, the journey to successful embryo implantation.* Human reproduction update, 2006. **12**(6): p. 731-746.
3. Genbacev, O.D., et al., *Trophoblast L-selectin-mediated adhesion at the maternal-fetal interface.* Science, 2003. **299**(5605): p. 405-408.
4. Braga, V. and S.J. Gendler, *Modulation of Muc-1 mucin expression in the mouse uterus during the estrus cycle, early pregnancy and placentation.* Journal of cell science, 1993. **105**(2): p. 397-405.
5. Surveyor, G.A., et al., *Expression and steroid hormonal control of Muc-1 in the mouse uterus.* Endocrinology, 1995. **136**(8): p. 3639-3647.
6. Carson, D.D., M.M. Desouza, and E.G.C. Regisford, *Mucin and proteoglycan functions in embryo implantation.* Bioassays, 1998. **20**(7): p. 577-583.
7. Bowen, J.A., F.W. Bazer, and R.C. Burghardt, *Spatial and temporal analyses of integrin and Muc-1 expression in porcine uterine epithelium and trophectoderm in vivo.* Biology of reproduction, 1996. **55**(5): p. 1098-1106.
8. Hey, N.A., et al., *The polymorphic epithelial mucin MUC1 in human endometrium is regulated with maximal expression in the implantation phase.* The Journal of Clinical Endocrinology & Metabolism, 1994. **78**(2): p. 337-342.
9. Brayman, M., A. Thathiah, and D.D. Carson, *MUC1: a multifunctional cell surface component of reproductive tissue epithelia.* Reproductive Biology and Endocrinology, 2004. **2**(1): p. 4.
10. Garrido-Gomez, T., F. Dominguez, and C. Simon, *Proteomics of embryonic implantation*, in *Fertility Control.* 2010, Springer. p. 67-78.
11. Chen, J.I., et al., *Proteomic characterization of midproliferative and midsecretory human endometrium.* Journal of proteome research, 2009. **8**(4): p. 2032-2044.
12. Dominguez, F., et al., *Proteomic analysis of the human receptive versus non-receptive endometrium using differential in-gel electrophoresis and MALDI-MS unveils stathmin 1 and annexin A2 as differentially regulated.* Human reproduction, 2009. **24**(10): p. 2607-2617.
13. Maxwell, S., et al., *Antioxidant status in patients with uncomplicated insulin-dependent and non-insulin-dependent diabetes mellitus.* European journal of clinical investigation, 1997. **27**(6): p. 484-490.
14. Ročić, B., et al., *Total plasma antioxidants in first-degree relatives of patients with insulin-dependent diabetes.* Experimental and clinical endocrinology & diabetes, 1997. **105**(04): p. 213-217.
15. Newsholme, P., et al., *Diabetes associated cell stress and dysfunction: role of mitochondrial and non-mitochondrial ROS production and activity.* The Journal of physiology, 2007. **583**(1): p. 9-24.
16. Sandler, S. and I. Swenne, *Streptozotocin, but not alloxan, induces DNA repair synthesis in mouse pancreatic islets in vitro.* Diabetologia, 1983. **25**(5): p. 444-447.
17. Kröncke, K.-D., et al., *Nitric Oxide Generation during Cellular Metabolization of the Diabetogenic N-Mefhyl-N-Nitroso-Urea Streptozotozin Contributes to Islet Cell DNA Damage.* Biological Chemistry Hoppe-Seyler, 1995. **376**(3): p. 179-186.
18. Alexandrova, A., L. Petrov, and M. Kirkova, *Proteasome activity in experimental diabetes.* Open Life Sciences, 2006. **1**(2): p. 289-298.
19. Scheuner, D. and R.J. Kaufman, *The unfolded protein response: a pathway that links insulin demand with β-cell failure and diabetes.* Endocrine reviews, 2008. **29**(3): p. 317-333.

20. Roth, T., et al., *Integrin overexpression induced by high glucose and by human diabetes: potential pathway to cell dysfunction in diabetic microangiopathy.* Proceedings of the National Academy of Sciences, 1993. **90**(20): p. 9640-9644.

21. Platt, M., et al., *St Vincent's Declaration 10 years on: outcomes of diabetic pregnancies.* Diabetic Medicine, 2002. **19**(3): p. 216-220.

22. Garcia-Vargas, L., et al., *Gestational diabetes and the offspring: implications in the development of the cardiorenal metabolic syndrome in offspring.* Cardiorenal medicine, 2012. **2**(2): p. 134-142.

23. Albaghdadi, A.J. and F.W. Kan, *Endometrial receptivity defects and impaired implantation in diabetic NOD mice.* Biology of reproduction, 2012. **87**(2).

24. Li, C., et al., *Proteomic analysis of proteins differentially expressed in uterine lymphocytes obtained from wild-type and NOD mice.* Journal of cellular biochemistry, 2009. **108**(2): p. 447-457.

25. Neilson, K.A., et al., *Label-free quantitative shotgun proteomics using normalized spectral abundance factors*, in *Proteomics for Biomarker Discovery*. 2013, Springer. p. 205-222.

26. Dennis, G., et al., *DAVID: database for annotation, visualization, and integrated discovery.* Genome biology, 2003. **4**(9): p. R60.

27. Hibbert, S.A., et al., *Defining tissue proteomes by systematic literature review.* Scientific Reports, 2018. **8**(1): p. 546.

28. Wakabayashi, S. and H. Yoshida, *The essential biology of the endoplasmic reticulum stress response for structural and computational biologists.* Computational and structural biotechnology journal, 2013. **6**(7): p. e201303010.

29. Mansouri-Attia, N., et al., *Endometrium as an early sensor of in vitro embryo manipulation technologies.* Proceedings of the National Academy of Sciences, 2009. **106**(14): p. 5687-5692.

30. Riethmacher, D., V. Brinkmann, and C. Birchmeier, *A targeted mutation in the mouse E-cadherin gene results in defective preimplantation development.* Proceedings of the National Academy of Sciences, 1995. **92**(3): p. 855-859.

31. Fujimoto, J., et al., *Alteration of E-cadherin, α-and β-catenin mRNA expression in human uterine endometrium during the menstrual cycle.* Gynecological Endocrinology, 1996. **10**(3): p. 187-191.

32. van der Linden, P.J., et al., *Expression of cadherins and integrins in human endometrium throughout the menstrual cycle.* Fertility and sterility, 1995. **63**(6): p. 1210-1216.

33. Chen, G.T., et al., *Identification of the cadherin subtypes present in the human peritoneum and endometriotic lesions: Potential role for P-cadherin in the development of endometriosis.* Molecular Reproduction and Development: Incorporating Gamete Research, 2002. **62**(3): p. 289-294.

34. Poncelet, C., et al., *Expression of cadherins and CD44 isoforms in human endometrium and peritoneal endometriosis.* Acta obstetricia et gynecologica Scandinavica, 2002. **81**(3): p. 195-203.

35. Illera, M.J., et al., *A role for αvβ3 integrin during implantation in the rabbit model.* Biology of reproduction, 2003. **68**(3): p. 766-771.

36. Somkuti, S.G., et al., *Epidermal growth factor and sex steroids dynamically regulate a marker of endometrial receptivity in Ishikawa cells.* The Journal of Clinical Endocrinology & Metabolism, 1997. **82**(7): p. 2192-2197.

37. Hambartsoumian, E., *Endometrial leukemia inhibitory factor (LIF) as a possible cause of unexplained infertility and multiple failures of implantation.* American Journal of Reproductive Immunology, 1998. **39**(2): p. 137-143.

38. Lakhter, A.J. and E.K. Sims, *Minireview: emerging roles for extracellular vesicles in diabetes and related metabolic disorders.* Molecular Endocrinology, 2015. **29**(11): p. 1535-1548.

39. Møller, A.B., et al., *Altered gene expression and repressed markers of autophagy in skeletal muscle of insulin resistant patients with type 2 diabetes.* Scientific Reports, 2017. **7**: p. 43775.

40. Sivitz, W.I. and M.A. Yorek, *Mitochondrial dysfunction in diabetes: from molecular mechanisms to functional significance and therapeutic opportunities.* Antioxidants & redox signaling, 2010. **12**(4): p. 537-577.

41. Green, K., M.D. Brand, and M.P. Murphy, *Prevention of mitochondrial oxidative damage as a therapeutic strategy in diabetes.* Diabetes, 2004. **53**(suppl 1): p. S110-S118.

42. Skulachev, V.P., *Uncoupling: new approaches to an old problem of bioenergetics.* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1998. **1363**(2): p. 100-124.

43. Tei, C., et al., *Reduced expression of αvβ3 integrin in the endometrium of unexplained infertility patients with recurrent IVF-ET failures: improvement by danazol treatment.* Journal of assisted reproduction and genetics, 2003. **20**(1): p. 13-20.

44. Creus, M., et al., *The effect of different hormone therapies on integrin expression and pinopode formation in the human endometrium: a controlled study.* Human reproduction, 2003. **18**(4): p. 683-693.

45. Stewart, C.L., *The role of leukemia inhibitory factor (LIF) and other cytokines in regulating implantation in mammals.* Annals of the New York Academy of Sciences, 1994. **734**(1): p. 157-165.

46. Murphy, C., et al., *A freeze-fracture electron microscopic study of tight junctions of epithelial cells in the human uterus.* Anatomy and embryology, 1982. **163**(4): p. 367-370.

47. Marcobal, A., et al., *Consumption of human milk oligosaccharides by gut-related microbes.* Journal of agricultural and food chemistry, 2010. **58**(9): p. 5334-5340.

48. El-Hawiet, A., E.N. Kitova, and J.S. Klassen, *Recognition of human milk oligosaccharides by bacterial exotoxins.* Glycobiology, 2015. **25**(8): p. 845-854.

49. Lindenberg, S., et al., *The milk oligosaccharide, lacto-N-fucopentaose I, inhibits attachment of mouse blastocysts on endometrial monolayers.* Journal of reproduction and fertility, 1988. **83**(1): p. 149-158.

50. Whyte, A. and W. Allen, *Equine endometrium at pre-implantation stages of pregnancy has specific glycosylated regions.* Placenta, 1985. **6**(6): p. 537-542.

51. Whyte, A. and T. Robson, *Saccharides localized by fluorescent lectins on trophectoderm and endometrium prior to implantation in pigs, sheep and equids.* Placenta, 1984. **5**(6): p. 533-540.

52. Bychkov, V. and P.D. Toto, *Lectin binding to normal human endometrium.* Gynecologic and obstetric investigation, 1986. **22**(1): p. 29-33.

53. Niklaus, A.L., C.R. Murphy, and A. Lopata, *Ultrastructural studies of glycan changes in the apical surface of the uterine epithelium during pre-ovulatory and pre-implantation stages in the marmoset monkey.* The Anatomical Record, 1999. **255**(3): p. 241-251.

54. Anderson, T.L., G.E. Olson, and L.H. Hoffman, *Stage-specific alterations in the apical membrane glycoproteins of endometrial epithelial cells related to implantation in rabbits.* Biology of reproduction, 1986. **34**(4): p. 701-720.

55. Van Kooij, R., et al., *Synthesis of a mucous glycoprotein in the human uterus.* European Journal of Obstetrics & Gynecology and Reproductive Biology, 1982. **14**(3): p. 191-197.

56. Kupryjańczyk, J., *Cycle-and function-related changes in lectin binding to human endometrium: a histochemical study with pronase treatment.* Archives of gynecology and obstetrics, 1989. **246**(4): p. 211-221.

57. Jeschke, U., et al., *The human endometrium expresses the glycoprotein mucin-1 and shows positive correlation for Thomsen-Friedenreich epitope expression and galectin-1 binding.* Journal of histochemistry & cytochemistry, 2009. **57**(9): p. 871-881.

58. Hosie, M.J. and C.R. Murphy, *Unmasking of surface negativity on day 6 pregnant rat uterine epithelial cells by trypsin and pronase.* Acta histochemica, 1989. **86**(1): p. 33-38.

59. Murphy, C. and V. Turner, *Glycocalyx carbohydrates of uterine epithelial cells increase during early pregnancy in the rat.* Journal of anatomy, 1991. **177**: p. 109.

60. West, K. and J.L. Cope, *The binding of peroxidase-labelled lectins to human endometrium in normal cyclical endometrium and endometrial adenocarcinoma.* Journal of clinical pathology, 1989. **42**(2): p. 140-147.

61. Inoue, M., M. Nakayama, and O. Tanizawa, *Altered expression of Lewis blood group and related antigens in fetal, normal adult and malignant tissues of the uterine endometrium.* Virchows Archiv A, 1990. **416**(3): p. 221-228.

62. Inoue, M., et al., *Expression of blood group Antigens A, B, H, Lewis-a, and Lewis-b in fetal, normal, and malignant tissues of the uterine endometrium.* Cancer, 1987. **60**(12): p. 2985-2993.

63. Garin-Chesa, P. and W.J. Rettig, *Immunohistochemical analysis of LNT, NeuAc2----3LNT, and Lex carbohydrate antigens in human tumors and normal tissues.* The American journal of pathology, 1989. **134**(6): p. 1315.

64. Andersch-Björkman, Y., et al., *Large scale identification of proteins, mucins, and their O-glycosylation in the endocervical mucus during the menstrual cycle.* Molecular & Cellular Proteomics, 2007. **6**(4): p. 708-716.

65. Yuan, S., et al., *Oxidation increases mucin polymer cross-links to stiffen airway mucus gels.* Science translational medicine, 2015. **7**(276): p. 276ra27-276ra27.

66. Brum, A.M., et al., *Mucin 1 (Muc1) deficiency in female mice leads to temporal skeletal changes during aging.* JBMR Plus, 2018. 2(6).

67. McGuckin, M.A., et al., *Intestinal secretory cell ER stress and inflammation.* 2011, Portland Press Limited.

68. Birchenough, G.M., et al., *New developments in goblet cell mucus secretion and function.* Mucosal immunology, 2015. **8**(4): p. 712.

69. Singh, P.K., et al., *Platelet-derived growth factor receptor β–mediated phosphorylation of MUC1 enhances invasiveness in pancreatic adenocarcinoma cells.* Cancer research, 2007. **67**(11): p. 5201-5210.

70. Schroeder, J.A., et al., *Transgenic MUC1 interacts with epidermal growth factor receptor and correlates with mitogen-activated protein kinase activation in the mouse mammary gland.* Journal of Biological Chemistry, 2001. **276**(16): p. 13057-13064.

71. Gipson, I.K., et al., *The amount of MUC5B mucin in cervical mucus peaks at midcycle.* The Journal of Clinical Endocrinology & Metabolism, 2001. **86**(2): p. 594-600.

72. Thathiah, A., et al., *Tumor necrosis factor α stimulates MUC1 synthesis and ectodomain release in a human uterine epithelial cell line.* Endocrinology, 2004. **145**(9): p. 4192-4203.

73. Nikas, G. Uterine pinopods as markers of endoemtrial receptivity in clinical practice. Human Reproduction, 1999. **2:** p. 99-106.

74. Horne, A., et al., *The expression pattern of MUC1 glycoforms and other biomarkers of endometrial receptivity in fertile and infertile women.* Molecular reproduction and development, 2005. **72**(2): p. 216-229.

75. Gipson, I.K., et al., *Mucin genes expressed by human female reproductive tract epithelia.* Biology of reproduction, 1997. **56**(4): p. 999-1011.

76. Lagow, E., M.M. DeSouza, and D.D. Carson, *Mammalian reproductive tract mucins.* Human Reproduction Update, 1999. **5**(4): p. 280-292.

77. Gendler, S.J. and A. Spicer, *Epithelial mucin genes.* Annual review of physiology, 1995. **57**(1): p. 607-634.

78. Salamonsen, L.A., et al., *Proteomics of the human endometrium and uterine fluid: a pathway to biomarker discovery.* Fertility and sterility, 2013. **99**(4): p. 1086-1092.

79. Hoffman, L.H., et al., *Progesterone and implanting blastocysts regulate Muc1 expression in rabbit uterine epithelium.* Endocrinology, 1998. **139**(1): p. 266-271.

80. Horne, A.W., et al., *MUC 1: a genetic susceptibility to infertility?* The Lancet, 2001. **357**(9265): p. 1336-1337.

81. Meseguer, M., et al., *Human endometrial mucin MUC1 is up-regulated by progesterone and down-regulated in vitro by the human blastocyst.* Biology of reproduction, 2001. **64**(2): p. 590-601.

82. Thathiah, A. and D.D. Carson, *Mt1-mmp mediates muc1 shedding independent of tace/adam17.* Biochemical Journal, 2004. **382**(1): p. 363-373.

83. Jones, B.J. and C.R. Murphy, *A high resolution study of the glycocalyx of rat uterine epithelial cells during early pregnancy with the field emission gun scanning electron microscope.* Journal of anatomy, 1994. **185**(Pt 2): p. 443.

84. DeLoia, J., et al., *Regional specialization of the cell membrane-associated, polymorphic mucin (MUC1) in human uterine epithelia.* Human reproduction (Oxford, England), 1998. **13**(10): p. 2902-2909.

85. Aplin, J.D., N.A. Hey, and R.A. Graham, *Human endometrial MUC1 carries keratan sulfate: characteristic glycoforms in the luminal epithelium at receptivity.* Glycobiology, 1998. **8**(3): p. 269-276.

86. Carson, D.D., et al., *Embryo implantation.* Developmental biology, 2000. **223**(2): p. 217-237.

87. Brown, H.M., et al., *Periconception onset diabetes is associated with embryopathy and fetal growth retardation, reproductive tract hyperglycosylation and impaired immune adaptation to pregnancy.* Scientific reports, 2018. **8**(1): p. 2114.

88. Esposito, K., et al., *Inflammatory cytokine concentrations are acutely increased by hyperglycemia in humans: role of oxidative stress.* Circulation, 2002. **106**(16): p. 2067-2072.

89. Hu, H., et al., *AGEs and chronic subclinical inflammation in diabetes: disorders of immune system.* Diabetes/metabolism research and reviews, 2015. **31**(2): p. 127-137.

90. Singh, H., et al., *Early stages of implantation as revealed by an in vitro model.* Reproduction, 2010. **139**(5): p. 905-914.

91. Everest-Dass, A.V., et al., *Human disease glycomics: technology advances enabling protein glycosylation analysis–Part 2.* Expert review of proteomics, 2018. **15**(4): p. 341-352.

92. Peter-Katalinić, J., *Methods in enzymology: O-glycosylation of proteins.* Methods in enzymology, 2005. **405**: p. 139-171.

93. Baldus, S.E., K. Engelmann, and F.-G. Hanisch, *MUC1 and the MUCs: a family of human mucins with impact in cancer biology.* Critical reviews in clinical laboratory sciences, 2004. **41**(2): p. 189-231.

94. Jensen, P.H., D. Kolarich, and N.H. Packer, *Mucin-type O-glycosylation–putting the pieces together.* The FEBS journal, 2010. **277**(1): p. 81-94.

95. Dwek, R.A., et al., *Analysis of glycoprotein-associated oligosaccharides.* Annual review of biochemistry, 1993. **62**(1): p. 65-100.

96. Schumann, R.R., et al., *Structure and function of lipopolysaccharide binding protein.* Science, 1990. **249**(4975): p. 1429-1431.

97. Bäckström, M., et al., *Increased understanding of the biochemistry and biosynthesis of MUC2 and other gel-forming mucins through the recombinant expression of their protein domains.* Molecular biotechnology, 2013. **54**(2): p. 250-256.

98. Tian, E. and K.G. Ten Hagen, *Recent insights into the biological roles of mucin-type O-glycosylation.* Glycoconjugate journal, 2009. **26**(3): p. 325-334.

99. Davies, J.R., C. Wickström, and D.J. Thornton, *Gel-forming and cell-associated mucins: preparation for structural and functional studies*, in *Mucins*. 2012, Springer. p. 27-47.

100. Issa, S., et al., *Analysis of mucosal mucins separated by SDS-urea agarose polyacrylamide composite gel electrophoresis.* Electrophoresis, 2011. **32**(24): p. 3554-3563.

101. Schulz, B.L., et al., *Mucin glycosylation changes in cystic fibrosis lung disease are not manifest in submucosal gland secretions.* Biochemical Journal, 2005. **387**(3): p. 911-919.

102. Matsuno, Y.-k., et al., *Supported molecular matrix electrophoresis: a new tool for characterization of glycoproteins.* Analytical chemistry, 2009. **81**(10): p. 3816-3823.

103. Ramsey, K.A., Z.L. Rushton, and C. Ehre, *Mucin agarose gel electrophoresis: western blotting for high-molecular-weight glycoproteins.* Journal of visualized experiments: JoVE, 2016(112).

104. Matsuno, Y.K., et al., *Improved method for immunostaining of mucin separated by supported molecular matrix electrophoresis by optimizing the matrix composition and fixation procedure.* Electrophoresis, 2011. **32**(14): p. 1829-1836.

105. Jensen, P.H., et al., *Structural analysis of N- and O-glycans released from glycoproteins.* Nat. Protocols, 2012. **7**(7): p. 1299-1310.

106. Dong, W., Y.-k. Matsuno, and A. Kameyama, *A procedure for Alcian blue staining of mucins on polyvinylidene difluoride membranes.* Analytical chemistry, 2012. **84**(20): p. 8461-8466.

107. Hayes, C.A., et al., *Glycomic work-flow for analysis of mucin O-linked oligosaccharides*, in *Mucins*. 2012, Springer. p. 141-163.

108. Kameyama, A. and Y.-k. Matsuno, *Supported Molecular Matrix Electrophoresis.* Glycoscience: Biology and Medicine, 2014: p. 1-7.

109. Habte, H.H., et al., *The role of crude human saliva and purified salivary MUC5B and MUC7 mucins in the inhibition of Human Immunodeficiency Virus type 1 in an inhibition assay.* Virology journal, 2006. **3**(1): p. 99.

110. Fukuda, M.N. and K. Sugihara, *An integrated view of L-selectin and trophinin function in human embryo implantation.* Journal of Obstetrics and Gynaecology Research, 2008. **34**(2): p. 129-136.

111. Slawson, C., R.J. Copeland, and G.W. Hart, *O-GlcNAc Signaling: A Metabolic Link Between Diabetes and Cancer?* Trends in biochemical sciences, 2010. **35**(10): p. 547-555.

112. Nordman, H., et al., *Gastric MUC5AC and MUC6 are large oligomeric mucins that differ in size, glycosylation and tissue distribution.* Biochemical Journal, 2002. **364**(Pt 1): p. 191.

113. Gendler, S.J., *MUC1, the renaissance molecule.* Journal of mammary gland biology and neoplasia, 2001. **6**(3): p. 339-353.

114. Turner, B., et al., *Isolation and characterization of cDNA clones encoding pig gastric mucin.* Biochemical Journal, 1995. **308**(Pt 1): p. 89.

115. Devaraj, N., H. Devaraj, and V.P. Bhavanandan, *Purification of mucin glycoproteins by density gradientcentrifugation in cesium trifluoroacetatel.* Analytical biochemistry, 1992. **206**(1): p. 142-146.

4.5 Supplementary Data: Comparison between different mucin analysis protocols

In-solution vs PVDF release

As previously shown in figure 4.11, PVDF bound techniques has resulted in higher number of *O*-glycan structures found compared to in-solution, with in-solution technique given relatively higher signal to noise ratio (Figure 4.11).

CsCL density gradient centrifugation

On the other hand, the relatively more established CsCl protocol was tested for the purification of PGM mucins. Each of the overall 12 fractions was measured for its density giving a range between 1-1.5 g/ml. PGM fractions had a density of 1.3 g/ml as expected from mucin proteins (fraction 8 and 9). The same fractions stained positive to charged carbohydrate staining (Alcian blue). Fraction 12 is suspected to contain DNA and possible insoluble mucin because it stained positive for Alcian blue (Supplementary Figure 4.1, A). Both fractions (8-9) have also shown a very clean *O*-glycan spectrum but very different overall spectra suggesting a good isolation of different mucin proteins or possibly glycoforms. This method allowed more coverage of the PGM glycans with 29 glycan found in total (Supplementary Figure 4.1, B).
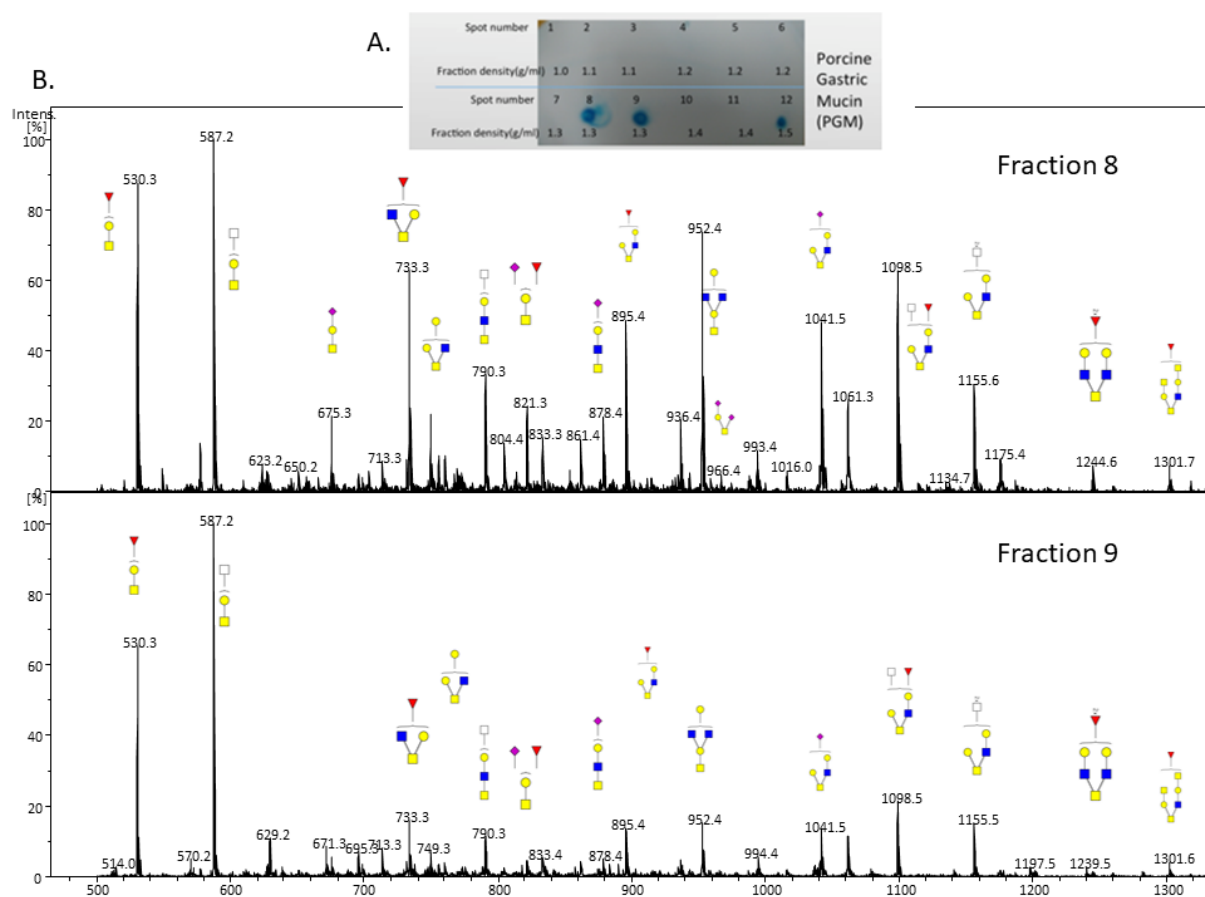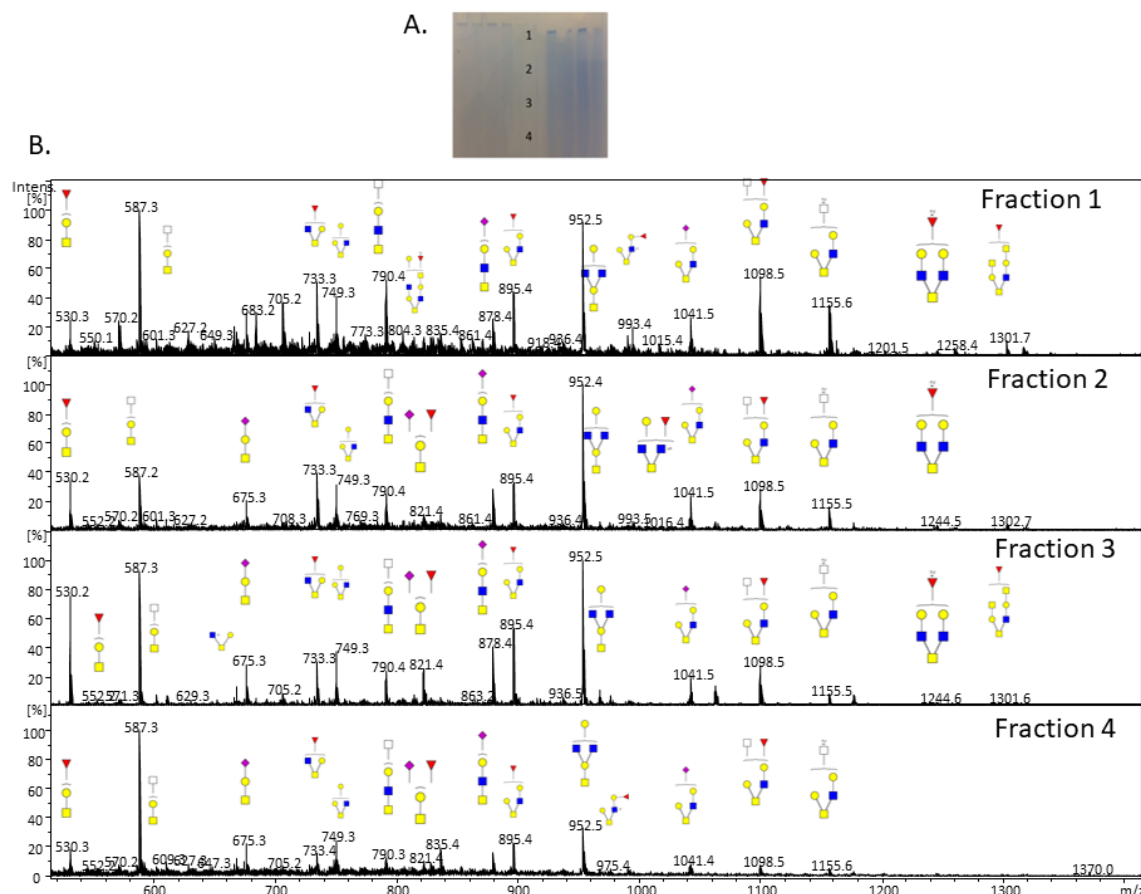
Supplementary figure 4.1: A. Alcian blue staining of 12 fractions generated via CsCL density gradient with their associated density g/ml; showing fraction 8 and 9 with 1.3 g/ml density and positive staining for sialylated and sulphated mucins. B. Summed MS of *O*-glycans released from two fractions (8 and 9) of PGM isolated via CsCl density gradient. The x-axis represents mass over charge with each signal characterized by its assigned *O*-glycan structure.

Agarose-Polyacrylamide (Ag-PAGE)

For Ag-PAGE analysis, the choice of 100ug PGM was made after testing several amounts of proteins in order to reach a successful membrane transfer and sufficient amount for subsequent glycan release. Also, several attempts were tried to reach a visual separation of the different PGM mucin proteins in the Ag-PAGE. The separation was seen irreproducible and hard to achieve. In this approach, PGM was allowed to enter the relatively large pore size gel resulting in a smear and for PGM proteins to separate according to different size and glycosylation profiles. The membrane was cut into four fractions, each used separately for subsequent *O*-glycan release (Supplementary figure 4.2, A). Due to the large amount of initial PGM used, all fractions gave a good signal-to-

noise ratio. The overall profile of the first three fractions seemed very similar, with only fraction 4 (the most travelled in the Ag-PAGE) containing lower amounts of large *O*-glycan structures ≥ 4 subunits. Overall, Ag-PAGE resulted in detection of a total of 28 *O*-glycans (Supplementary figure 4.2, B).
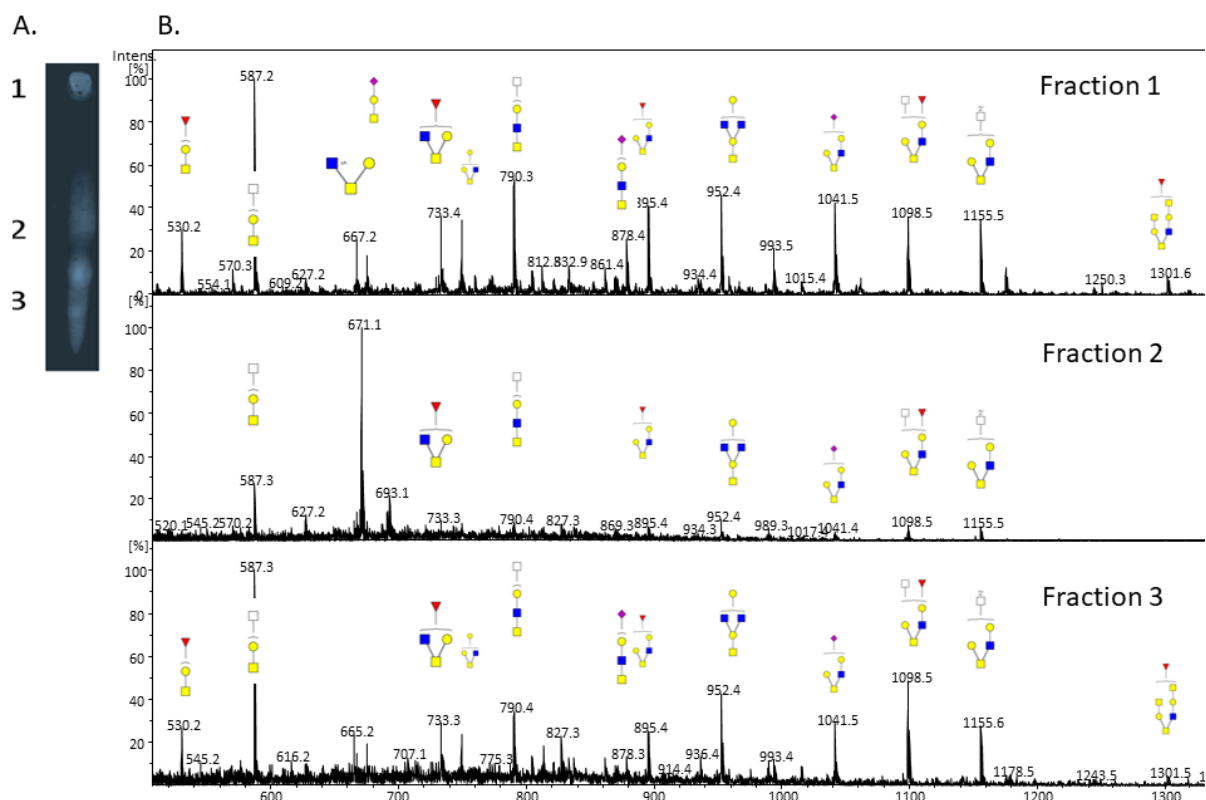

Supplementary figure 4.2: A. Commasie blue staining of gel before (right) and after (left) transfer. B. Summed MS of *O*-glycans released from four fractions of PGM isolated via AgPAGE. The x-axis represents mass over charge with each signal characterized by its assigned *O*-glycan structure.

Supported matrix membrane electrophoresis (SMME)

The investigation of the relatively new mucin purification technique SMME showed some promising results. Several spots were seen to travel through the electrophoresis membrane (Supplementary figure 4.3, A). From the MS glycan analysis, these spots were seen with the strongest differences between them compared to fractions obtained in Ag-PAGE and CsCL density gradient centrifugation, thus implying SMME provides the most resolution of PGM glycoforms. Fraction 1 (non-travelling from the original blotted spot) resembled the profile from the PVDF

released PGM experiment in Supplementary figure 4.3, inferring this spot contains non-travelling PGM spot (Supplementary figure 4.3, B). The other two fractions show a very good separation of two communities of *O*-glycan structures each with very different profiles. SMME enabled the characterisation of 29 *O*-glycan individual structures.

Supplementary figure 4.3: A. direct blue staining of PGM glycoforms migrated by SMME technique. B. Summed MS of *O*-glycans released from three fractions of PGM isolated via SMME. The x-axis represents mass over charge with each signal characterized by its assigned *O*-glycan structure.

Discussion

The porcine gastric mucin (PGM), although of different origin, provided a relatively cheap and commercially available representation of mucins, and was used for comparison between different protocols. Three mucin proteins are present in human stomach, namely: MUC 5AC (surface epithelium), MUC 6 (glandular) and MUC 1 (membrane epithelium) [112, 113]. In porcine, the gene for the PGM protein backbone is very similar to the human MUC5AC gene but the glycosylation of other mucin proteins are still not fully determined [114].

In CsCL gradient centrifugation, PGM was identified from its expected density at 1.3-1.4g/ml [115]. Staining with Alcian blue has been widely used to visualise acidic mucin and it is compatible with glycan analysis, unlike PAS staining of glycan structures, which is used for mucin characterisation but destroys the carbohydrates [106]. The spots seen in fraction 8 and 9 showed the presence of acidic mucins. CsCL density gradient centrifugation is essential to purify mucins from complex mixtures and tissues [99], but seems unessential for the use with relatively pure mucin samples in order to avoid sample loss. In fact, CsCL density gradient would require a lot of sample (100ug) to account for loss of samples during dialysis, mainly caused by mucins sticking to the dialysis bags.

Regarding Ag-PAGE, it was hard to accurately reproduce several steps in this protocol such as gel casting and membrane transfer. This technique has been previously used to separate different mucin proteins from each other [100, 101, 105], but it would be hard to reproducibly use this technique with a large cohort of samples. This technique has few disadvantages such as the need for large amount of mucin proteins and difficulties with transferring large mucin proteins. However, it would be a very useful technique in case of a large amount of mucins and small cohort studies.
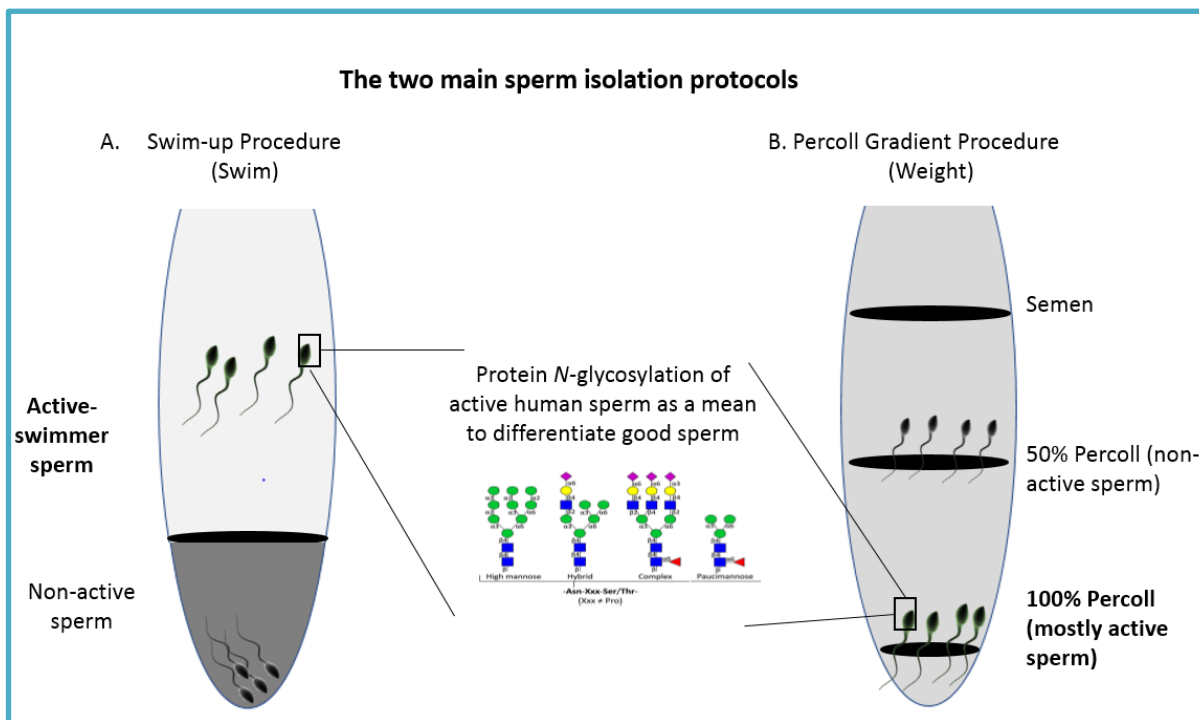
CsCL and SMME have shown a greater separation of different mucin glycoforms compared to Ag-PAGE. For SMME technique, there was an issue with reproducibility with mucin dots not moving occasionally. The issue seems to be concerned with unsuccessful loading of PGM into the matrix inside the PVDF. The very basic/hand-made apparatus might also be a disadvantage to this technique producing some irregularities. Finally, only few spots can be made per experiment which would prove difficult for a large cohort study.

Overall, when comparing between techniques, CsCl density gradient seems to provide a better glycoform separation compared to Ag-PAGE with both needing a very large amount of protein to work successfully (Supplementary Figure 1 and 2). On the other hand, SMME seems to show promising separation of different mucin glycoforms but it would require modifications to accommodate the PVDF apparatus and large cohorts (Supplementary Figure 3). As a result, each

mucin analysis protocol has its advantages and disadvantages and the choice among them is dependent on the case studied.

Chapter 5

Characterising the protein *N*-glycome of active and inactive sperm

The two main sperm isolation protocols

A. Swim-up Procedure (Swim)

Active-swimmer sperm

Non-active sperm

Protein *N*-glycosylation of active human sperm as a mean to differentiate good sperm

B. Percoll Gradient Procedure (Weight)

Semen

50% Percoll (non-active sperm)

100% Percoll (mostly active sperm)

The remaining two Chapters of this thesis will focus on the male contribution to successful reproduction, mainly the sperm capacity for fertilisation. The primary aim of the work in this Chapter was to profile and compare protein *N*-glycosylation of active human sperm as a means of differentiating active sperm from inactive sperm. Active sperm samples were generated via two main isolation protocols, namely Percoll density gradient separation and "swim-up" collection [1, 2]. Both techniques are regularly used for preparation of active sperm for *in vitro* fertilisation (IVF). Sperm samples from Percoll density gradient were separated into 50% v/v (inactive) and 100% v/v (active). Mostly active (100% v/v) sperm are able to fertilize an egg after the treatment with capacitating medium, and our active sperm were subjected to *in vitro* capacitation. All measure of sperm quality such as motility, concentration and morphology has largely been discredited for infertility diagnosis [3]. Using sperm glycosylation as a marker of sperm capacity to fertilise is promising, mainly for the relevance of glycosylation in sperm functionality such as sialic acid masking of sperm proteins involved in fertilization [4]. Global protein glycosylation of the human sperm proteins was first compared using three different protein preparation techniques, namely cell lysis and two membrane enrichment methods: differential detergent extraction and ultracentrifugation [5, 6]. The protein *N*-glycosylation of sperm isolated via Percoll density gradient and swim-up protocols was analysed using PGC-LC-MS/MS with the aim to utilise the glycome to differentiate between active and inactive sperm. The results of this work have been prepared as a manuscript to be submitted to Fertility and Sterility.

Acknowledgements

5.1 Analysis of the *N*-glycome of Human spermatozoa demonstrates that large differences exists between men (Manuscript 5, in-preparation for publication in Fertility and Sterility)

Abdulrahman M. Shathili[1,3], Robyn Peterson[1], Mark R. Baker[2], Nicolle H. Packer[1,3,4]

[1] *Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia*
[2] *School of Environmental and Life Sciences, University of Newcastle, Sydney, NSW, Australia*
[3] *ARC Centre of Nanoscale Biophotonics, Macquarie University, Sydney, NSW, Australia*
[4] *Institute for Glycomics, Griffith University, Gold Coast, Queensland, Australia*

Corresponding author: nicki.packer@mq.edu.au

Contribution: The writing of the manuscript, protein extractions and mass spectrometry analysis of sperm glycosylation were done by A. M. Shathili. R. Peterson edited and reviewed this manuscript. Associate Prof. M. R. Baker isolated and provided human sperm samples. Also, Associate Prof. M. R. Baker supervised and reviewed the manuscript. Prof. Nicolle H. Packer supervised, reviewed and edited the manuscript and Chapter.

Abstract

The protein *N*-glycosylation of sperm derived from five men, isolated via Percoll density gradient and swim-up protocols, was analysed using Porous Graphitized Carbon-Electrospray Ionization-Mass spectrometry (PGC-ESI-MS/MS). Protein *N*-glycosylation was compared using a cell lysis protocol and two membrane enrichment methods. All three methods produced similar protein glycomic profiles. Sperm samples were separated by Percoll density gradient centrifugation into poor quality (low Percoll density 50% v/v: 0.695 g/ml)) and high quality (high Percoll density 100% v/v: 1.39 g/ml) fractions. No significant correlations were found between active spermatozoa protein glycosylation isolated via Percoll density gradient and swim-up protocols when compared with the low quality Percoll density fraction. Higher quality spermatozoa showed trends toward less sialylation and an increase in paucimannose structures compared to low quality sperm. These trends might represent a link between the sperm *N*-glycome and its efficiency for fertilization. Active Percoll (100% v/v) spermatozoa were then subjected to *in vitro* capacitation. Contrary to expectation, no observed change was seen in sperm protein *N*-glycosylation compared to before capacitation. Above all, donors were seen with large individual variation in terms of the glycosylation of their sperm which shows the importance of considering personalised medicine as the appropriate treatment for infertility.

Introduction

Recently, Tecle and colleagues published a comprehensive review summarising the role of glycosylation in sperm survival and functionality [4]. This review elaborates that once spermatogenesis is complete in the seminiferous tubules, spermatozoa lose the ability to make new glycans yet new integral glycoproteins are incorporated and existing glycans are modified and elongated via the trans activity of glycosidases and glycosyltransferases in the lumen of the epididymis in order to form maturing spermatozoa [7-9].

The sperm glycocalyx has been implicated in several roles including capacitation, which is the physiological change that enables the cell to fertilise an egg [10-13]. As spermatozoa pass through the epididymis, heavy sialylation coats their surface which may allow them to survive as allogeneic cells in the female reproductive tract without rejection by the female immune system [14]. During *in vivo* capacitation, it has been reported in both human and mouse spermatozoa that a fraction of their glycosaminoglycans and their sialic acid moieties are lost. Loss of these negatively charged glycans is thought to promote sperm binding to oocytes [15, 16]. In confirmation of this notion, two sialidases (NEU1 and NEU3) have been shown to cleave sialic acid residues during capacitation *in vivo*, and furthermore, inhibition of the sialidases results in failure of sperm binding to zona pellucida [17].

There are two common sperm preparation techniques that are used to separate active sperm from seminal plasma for *in vitro* fertilisation (IVF) procedures, namely Percoll density gradient and 'swim-up' collection; each of which have reported advantages and disadvantages [2, 18]. Percoll is suitable for density gradient separation of live cells due to its low viscosity and toxicity [1]. Using isocratic gradients, sperm can be separated on the basis of their density. Whilst lower density cells have defective motility, poor morphology and poor chromatin compaction [18], the higher density cells are those used for the purpose of IVF [2]. Active-motile sperm (high-density) are able to be capacitated and eventually fertilize an oocyte whereas the inactive sperm show less capacity to fertilize oocytes. Capacitation involves shedding of the sperm glycocalyx by sperm glycosidases [4]. *In vitro* capacitated sperm was previously described by its reaction with

fluoresceinisocyanated conjugate-Pisum sativum agglutinin (FITC-PSA=specific to mannosylated structures) that stained only the anterior of the sperm head [36]. *In vitro* capacitation medium contains BSA which binds cholesterol to deplete sperm plasma membrane, and ions (HCO_3^- and Ca^{2+}) resulting in an increase in tyrosine phosphorylation which is essential for capacitation [36]. On the other hand, swim-up, which separates motile sperm from non-motile sperm and seminal plasma, is the most widely used technique in IVF [19].

In case of yak and bovine spermatozoa isolated via swim up technique, the cells showed higher motility and higher percentage of intact plasma membranes and acrosomes (sperm head), while Percoll gradient centrifugation produced higher recovery of spermatozoa [20, 21]. Swim-up preparations have been reported to produce higher rates of motility with less debris and percentage of sperm failing the hypo-osmotic swelling (HOS) test. This test determines membrane integrity by testing the ability of the sperm membrane to maintain equilibrium against the environment [22]. In contrast, density-gradient isolated spermatozoa achieve a much higher yield [23]. However, there is no difference when spermatozoa are isolated in terms of the number of successful blastocyst formations [21].

A key question in regard to the sperm glycome is whether it can be used to determine the fertility status of a spermatozoa. Despite intensive research effort, the use of sperm parameters such as motility, concentration and morphology for infertility diagnosis has largely been discredited [3], due to considerable variability, and the inability of a semen analysis to determine what is occurring at the molecular level [24]. Furthermore, a major reason for disparity in sperm parameters is within-subject biological variation [25]. Whether there is a correlation between sperm glycocalyx and the quality of sperm is yet to be fully addressed.

The *N*-glycome of sperm proteins has been shown to contain bi-, tri and tetra- antennary Lewis x and Lewis y sequences with varying degree of sialylation, as well as bisecting structures [26]. Individual sperm molecular composition and phenotypes are an expression of the male phenotype and not its inherited haploid genome [27, 28]. Despite the expected uniformity in sperm glycocalyx, some have reported sperm surface heterogeneity between different sperm from the

same donor with variations in glycosylation composition, such as polysialic acid (PolySia) [4, 29, 30]. The importance of sperm glycocalyx in sperm functionality suggests the potential of using sperm glycosylation as a means to select for active sperm for IVF use.

Little is known about the relationship between glycosylation variation and the fertility of individual spermatozoa, yet it is plausible to think that differences in the glycocalyx might influence the quality of individual spermatozoa [4]. The aim of this work was to compare the *N*-glycosylation of sperm samples prepared by different techniques namely, Percoll density gradient and swim-up, as well as the effect of *in vitro* capacitation, on sperm protein glycosylation. Analysing the *N*-glycome of active sperm may assist future IVF applications due to the reported relevance and importance of protein glycosylation in sperm functionality.

Results

All three protein extraction methods produced similar N-glycosylation profiles

Due to reported technical challenges associated with extraction, solubilisation and purification of glycoproteins [31], we compared three common methods for the isolation of sperm glycoproteins. Using active sperm isolated from high density Percoll gradient, the *N*-glycan profile of proteins extracted by three methods were compared; total cell lysis, membrane enrichment with differential detergent extraction and membrane enrichment by ultracentrifugation, to determine the optimal technique based on the number of observed *N*-glycans. All three methods generated similar *N*-glycan MS profiles. However, membrane enrichment by ultracentrifugation prior to PNGase F treatment generated the best signal-to-noise ratio *N*-glycan spectra. As a result, ultracentrifugation was used in the following analyses to purify sperm membrane proteins. Preliminary results showed expression of many types of *N*-glycans structures including bisecting GlcNAc, oligomannose, paucimannose, Lewis epitope containing, sialylated and GlcNAc/Gal capped (Figure 1).
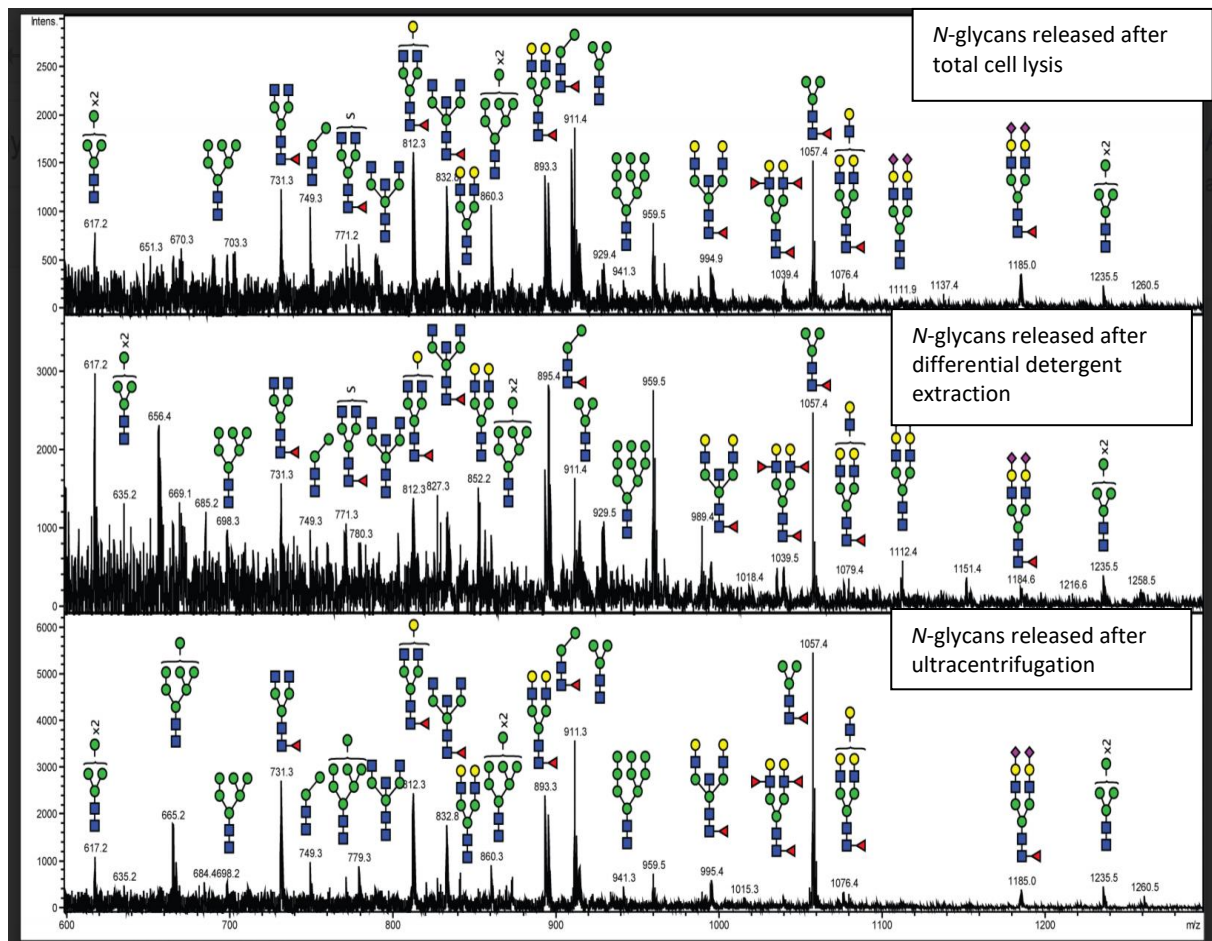
Figure 1: Comparison of *N*-glycans released from different protein extraction protocols of a capacitated human sperm sample using PGC-LC-MS/MS. A: whole cell lysis, B: membrane enrichment with differential detergent extraction, C: membrane enrichment with ultracentrifugation.

Sperm N-glycosylation profile is ineffective in differentiating spermatozoa isolated by Percoll density gradient or swim-up separation protocols.

Active spermatozoa prepared by swim-up were similar to the poor quality (low Percoll density) sperm cells in their relative abundance of oligomannose and sialylated structures classes. Similarly, the relative abundance of structures with Lewis epitopes was similar between active sperm prepared by swim-up and high Percoll density before capacitation (Figure 2). As a result, swim-up sperm share similar *N*-glycans to both high and low density Percoll separated sperm populations.

The comparison of the glycan classes between the two different Percoll density separated populations showed high quality sperm (high Percoll density) with a trend towards a lower relative

177

abundance of oligomannose, sialylated and Lewis epitope containing structures when compared to inactive low Percoll density sperm. Also, high quality spermatozoa exhibited a trend towards a higher relative abundance of paucimannose and GlcNAc/Gal-capped structures compared to low quality ones. This data suggests high quality sperm can possibly be differentiated by their protein *N*-glycan profile, exhibiting less sialylated structures and more paucimannosylated and GlcNAc/Gal-capped structures, than low quality sperm. Both swim-up and high density, quality spermatozoa have a lower relative abundance of sialylated and Lewis structures when compared to poor quality (low Percoll density) spermatozoa (Figure 2) which reinforces the possibility of using sperm glycosylation as a means to select for active sperm for IVF via measuring the level of sialylation/Lewis structure abundance. Little difference in the relative abundance of individual *N*-glycan structures was observed between swim-up and both populations isolated from Percoll density gradient.

In vitro capacitation resulted in no change to the relative abundance of the N-glycome.

In order to understand the effect that capacitation has on the *N*-glycome of human spermatozoa, we compared sugar structures before and after capacitation. In contrast to other reports, neither oligomannose, paucimannose, Lewis containing, sialylated and GlcNAc/Gal capped showed any significant difference after capacitation (Figure 2).
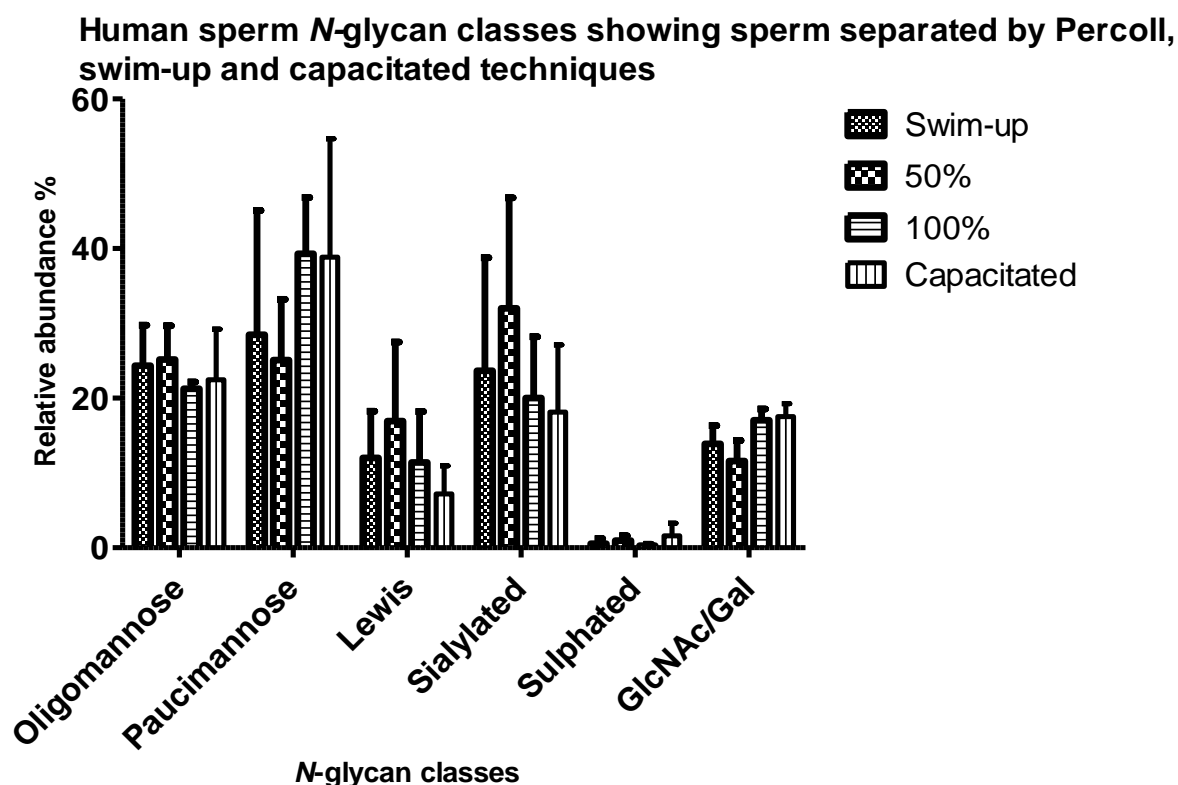
Figure 2: The relative abundance of *N*-glycan classes seen in sperm isolated by swim-up and Percoll density gradient tecqhniues. Percoll density gradient generated two populations of sperm: low quality (50% v/v) and high quality (100% v/v) sperm. High quality sperm were subjected to *in vitro* capacitation (capacitated). *N*-glycan classes includes: oligomannose, paucimannose, Lewis (Lewis x, Lewis y and sialyl Lewis x), total sialylated, sulphated and GlcNAc/Gal (structures capped with galactose or *N*-acetylglucosamine).

Large biological variations limited our statistical analysis

Interestingly, when we looked at individual variation across different sperm donors (swim-up preparation), we observed great differences, mainly in sialylated and paucimannosylated structures. To a lesser extent, individual variation was seen for oligomannose and GlcNAc-Gal capped classes (Figure 3). This large biological variation is also observed when comparing between sperm preparation techniques across different sperm donors (Figure 4).
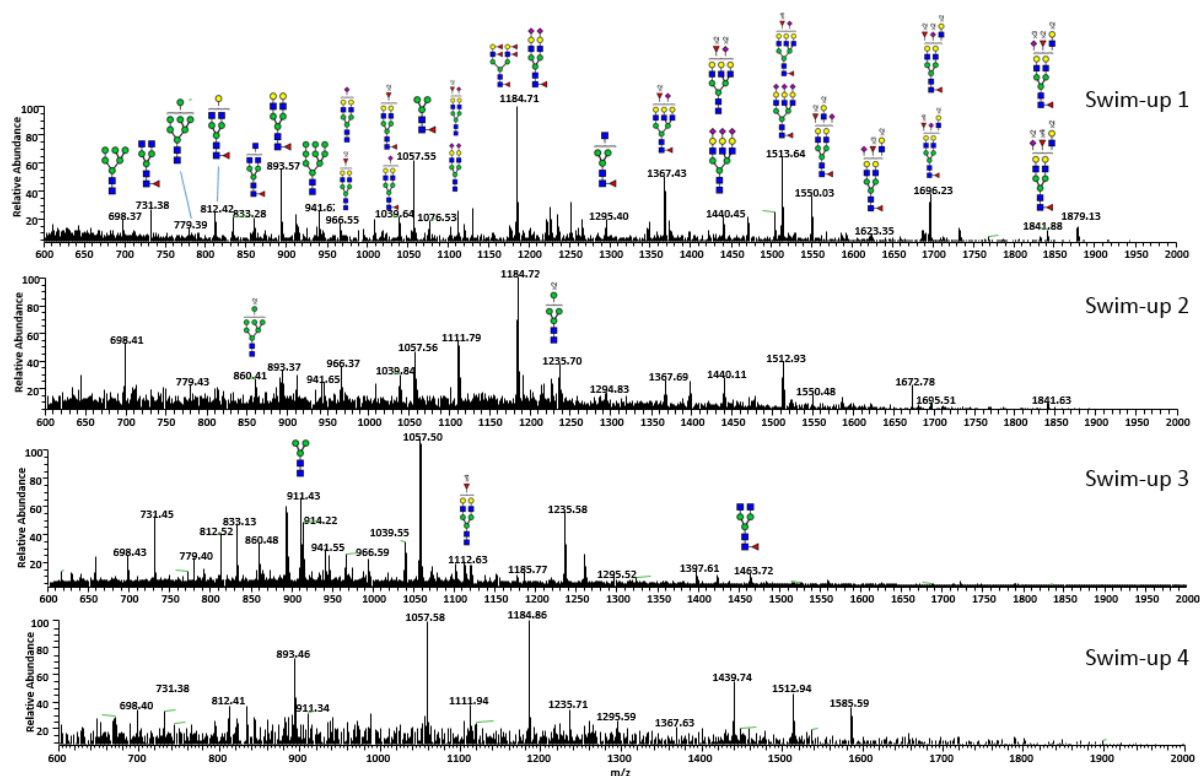
Figure 3: Comparison between sperm *N*-glycan structures from 4 men (using swim-up collection) analysed by PGC-LC-MS/MS. This graph shows the large individual variation exhibited in the individual glycan structures from 4 individuals.
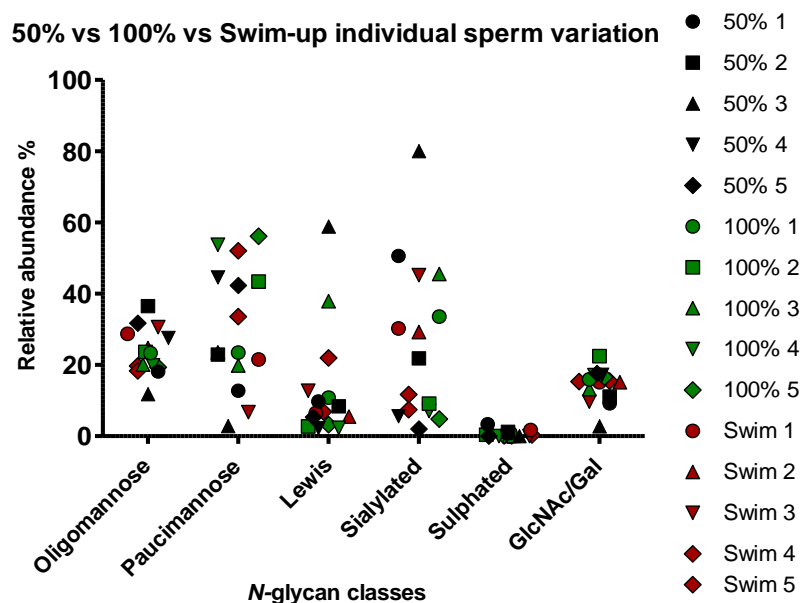

Figure 4: The relative abundance of *N*-glycan classes from 50% v/v, 100% v/v (non-capacitated) and swim-up sperm are displayed for each individual biological replicate (5 men). This graph shows the large individual variation exhibited across these samples. Each biological replicate was assigned a fixed shape to represent 50% v/v, 100% v/v (non-capacitated) and swim-up, black, white and red respectively.

Discrepancies between sperm N-glycome and published literature: the possible influence of seminal N-glycome

A previously published report [26] showed that swim-up sperm are heavily decorated with Lewis x and y structures and with a very low degree of sialylation (roughly 50% Lewis structures and only one low-abundant sialylated structure), whereas our work showed approximately 10% Lewis structures and 20% sialylated structures on sperm proteins. We also profiled human seminal plasma (semen without sperm) that also showed the presence of Lewis structures (50% Lewis structures) (Figure 5, B), which matched a previously published profile of human seminal plasma analysis [32]. Our human seminal plasma contained similar overall percentage of Lewis structures (50%) compared to the previously reported swim-up sperm profile, but showed more sialylated structures than the previously published swim-up sperm [26] (Figure 5, B),. The total seminal plasma protein samples that we analyzed contain mainly high-mannose and bi-antennary *N*-glycan structures with a large presence of tri-antennary structures with several Lewis antigens (Figure 5). Due to the small mass difference between sialic acid and two fucose units (0.5 Da), MS/MS fragments were needed to differentiate between the two types of terminal glycan structures. As an example, in the abundant bi-antennary core-fucosylated *N*-glycans, MS/MS analysis enabled the differentiation between Lewis and sialic acid containing structures in swim-up and semen samples (Figure 5).
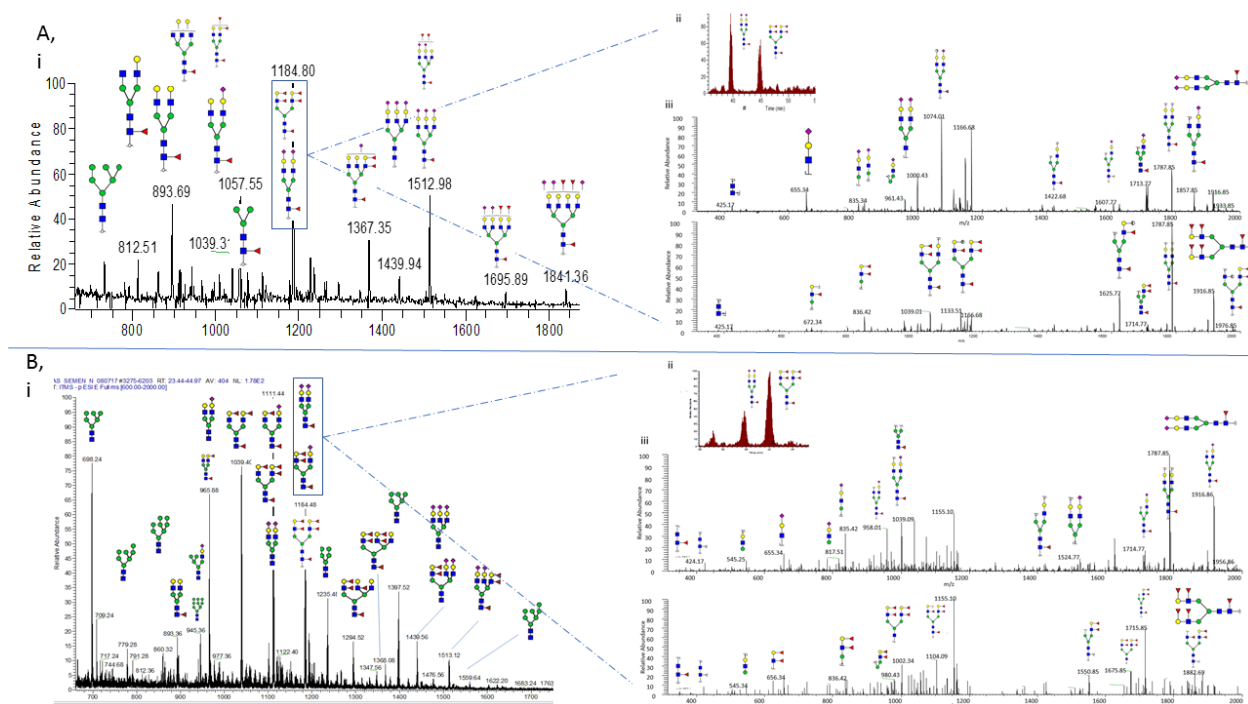
Figure 5: Comparison between semen and sperm (swim-up) *N*-glycans profile using PGC-LC-MS/MS. Part A (i) shows summed MS of *N*-glycans from swim-up sperm proteins. The figure displays the relative abundance of each individual glycan (intensity) against their mass/charge ratio. A (ii): EICs of two bi-antennary core-fucosylated *N*-glycan structures found in swim-up sperm: *m/z* 1184.5^{2-}(GlcNAc4Man3Gal2Fucose1NeuAc2) and *m/z* 1185.4^{2-} (GlcNAc4Man3Gal2Fucose3). A (iii): MS/MS fragment spectra from the doubly charged negative ions of two bi-antennary glycans: *m/z* 1184.5^{2-}(GlcNAc4Man3Gal2Fucose1NeuAc2) and *m/z* 1185.4^{2-} (GlcNAc4Man3Gal2Fucose3). Part B (i) shows summed MS of *N*-glycans from human seminal plasma isolated without the interference of sperm (2). B (ii): EICs of two Bi-antennary core-fucosylated *N*-glycan structures found in semen: *m/z* 1184.5^{2-}(GlcNAc4Man3Gal2Fucose1NeuAc2) and *m/z* 1185.4^{2-}(GlcNAc4Man3Gal2Fucose3). B (iii): MS/MS fragment spectra from the doubly charged negative ions of two bi-antennary glycans: *m/z* 1184.5^{2-} (GlcNAc4Man3Gal2Fucose1NeuAc2) and *m/z* 1185.4^{2-}(GlcNAc4Man3Gal2Fucose3). The MS/MS spectra are able to differentiate specific ions unique to each structure.

Discussion

In this work, two sperm preparation techniques were compared, namely Percoll density gradient and 'swim-up' collection of active sperm. Although the two techniques are commonly used for active sperm collection, they produce sperm with differences in motility, plasma membrane and acrosome components and sperm concentration [20, 21, 23]. Since all these measured parameters have failed to diagnose male infertility [3], the study of sperm *N*-glycome produced by these

common techniques might provide us a means to better differentiate active sperm for use in IVF. Furthermore, the swim-up protocol has been reported to produce active sperm with a higher percentage of plasma membrane (glycoproteins) when compared to the Percoll gradient technique [20, 21]. This was further evidenced by the lower MS signal to noise ratio presented in the MS spectrum of the *N*-glycans released from 100% v/v Percoll collected (active) sperm compared to 50% v/v (inactive) and swim-up sperm (Supplementary Figure 1).

In addition to comparing sperm preparation techniques, we also compared protein preparation techniques. Sperm *N*-glycome was released from the sperm glycoproteome using PNGase F and analyzed by the technique of PGC-LC-ESI-MS/MS. In order to isolate sperm glycoproteins from other cellular (nuclear and cytoplasmic) components, three protein isolation techniques were compared. All of the glycoprotein preparation techniques - cell lysis and membrane enrichment via ultracentrifugation and detergents - produced similar *N*-glycan profiles for high quality sperm isolated by Percoll density gradient (Supplementary Figure 1). As a result, the cell lysis protocol was chosen for subsequent analysis to maintain the same protocol as previously published literature, as well as to consider possible differences in membrane and cellular components arising from different sperm preparation techniques [20, 21].

In this work, we compared the *N*-glycome of sperm isolated from the two most common sperm preparation techniques. The comparison of protein *N*-glycosylation between active Percoll and swim-up sperm showed little difference in their *N*-glycome profiles. Percoll active (100% v/v) and inactive (50% v/v) human sperm (n=5) showed several differences in the relative abundance of *N*-glycan classes. These differences were not found to be statistically significant across the small number of donors, but a trend towards less sialylation and more paucimannosylation is evident in active compared to inactive sperm for all 5 human sperm samples (Figure 2). After capacitation, non-capacitated Percoll active (100% v/v fraction) and *in-vitro* capacitated Percoll active sperm (from the same donor) showed no significant differences in their *N*-glycosylation profiles for all 5 human sperm samples (Figure 2). Although capacitation has been previously correlated with a decrease in sperm sialylation [15, 16], the studies used lectin immunohistochemistry techniques

with sialic acid recognising lectins which detect the presence of glycan epitopes on both proteins and lipids and do not focus on protein glycosylation specifically. Also, the reported decrease of sialylated moieties on capacitated sperm could be associated with a total decrease in the glycoprotein content, which would also result in a decrease in total sialylation. Our data, and a current limitation in the analysis of glycans, show only the relative abundance of sialylation to other *N*-glycan classes and not the absolute abundance of *N*-glycans or glycoproteins. Overall, Percoll density gradient separates sperm into two fractions with different fertilisation capacity as well as appearing to produce different *N*-glycosylation profiles; this work requires more sample numbers but implicates a specific *N*-glycome with less sialylation and more paucimannosylation that may be useful in selecting for sperm efficiency for IVF.

Contrary to a previous publication [26] that described roughly 50% Lewis structures and only one low-abundant sialylated structure, the *N*-glycosylation profile of active swim-up, as well as of active Percoll purified, sperm contained 10% Lewis structures and 20% sialylated structures on sperm proteins (Figure 2) The under-representation of sialylated structures in the previous publication could be due to the use of the MALDI-TOF MS ionisation technique in the previous publication, which is not optimal for the ionisation of sialylated moieties and may result in the loss of sialic acid from structures [26, 32]. Our analysis showed seminal plasma *N*-glycome to be highly decorated with Lewis structures with a profile that correlates well with analysis from previous published reports (Figure 5, B) [32]. We suspect the difference in the abundance of Lewis structures between our sperm results and previous published data may be explained by the amount of seminal plasma contamination in the swim-up preparation procedure.

Above all, in our work, the most significant observation was the huge individual biological variation detected in the *N*-glycome of human sperm from five different male donors (Figure 3, 4) indicating the need for the analysis to be carried out on more samples to be meaningful in any trends observed. Heterogeneity in the individual sperm glycocalyx of the same donor can be due to differences in post-meiotic gene expression of glycosyltransferase and glycosidase in testes, variance in modification during epididymal transition and variation in addition of glycoconjugates

during maturation [4]. Others have reported significant biological variability of donor semen characteristics over time due to several factors such as diseases, use of medications, drug and alcohol, period of abstinence and others [33, 34]. As a result, large individual heterogeneity in glycocalyx of sperm within the same donor might result in a large variation in glycosylation among different donor's sperm. The large heterogeneity in sperm glycocalyx among individuals provide individuality and show future relevance of personalised medicine, where most appropriate treatments are tailored to individual patients or group of patients according to their specific sperm glycocalyx. In our analysis, the lack of data on individual donor fertility restricted our ability to correlate glycome with donor history of fertility. Furthermore, our work was not concerned with sperm *O*-glycosylation since only *N*-glycosylation was shown to be essential and *O*-glycans to be dispensable for spermatogenesis [35]. Yet to the best of our knowledge, this is the first report showing different sperm donors having large individual biological variation in their *N*-glycome and point to the need for a much larger study to determine if there are common changes associated with fertility.

Conclusions

In this work, sperm isolated by the two common active sperm preparation techniques were compared in their protein *N*-glycosylation profiles. There was a trend towards a decrease in sialylation and an increase in paucimannose structures in Percoll active (100% v/v) compared to inactive (50% v/v) sperm, which could possibly, in a larger study, associate a specific *N*-glycome characteristic of sperm IVF efficiency. *In vitro* capacitation resulted in no significant change in sperm protein *N*-glycosylation. Above all, the large individual biological variation in the human sperm protein *N*-glycome was the most significant observation in our work. This variation complicated our comparisons resulting in no observed statistical significance between glycan classes. Finally, our results will help future researchers in utilising the *N*-glycome for further research into sperm fertility while considering personalised medicine as an appropriate approach for infertility diagnosis.

Methods

Sperm preparation techniques (Percoll density gradient centrifugation and Swim-up)

Percoll density gradient centrifugation

Ejaculated semen samples were collected and stored in Beltsville thawing solution and processed as per [36] . Sperm were resuspended in non-capacitating medium (0.1 M NaCl, 0.36 mM NaH$_2$PO$_4$, 8.6 mM KCl, 0.5 mM MgCl$_2$, 11 mM glucose, 23 mM HEPES, pH 7.6). Percoll density gradients were made by mixing 50% v/v and 100% v/v Percoll, and motile (100% v/v) (active) sperm were collected from the bottom of the tube after centrifugation at 600g for 30min. In-active sperm were collected from 50% v/v Percoll density gradient, this layer was observed in the middle of the tube post centrifugation. Sperm were carefully collected using a pipette while preventing the disruption of the formed 50% v/v and 100% v/v density layers. Sperm were capacitated *in vitro* by treating active sperm with capacitating medium (0.1 M NaCl, 0.36 mM NaH$_2$PO$_4$, 8.6 mM KCl, 0.5 mM MgCl$_2$, 11 mM glucose, 23 mM HEPES, pH 7.6, 10 mM NaHCO$_3$, 2 mM CaCl$_2$, 5 mM pyruvate and 0.3% BSA) for 3 hours at 39°C with 5% CO$_2$.

Swim-up sperm collection

Sperm were incubated in Dulbecco's phosphate-buffered saline and overlaid with protein-free modified Ham's F-10 basal medium-HEPES as per [26]. Sperm were incubated at a 30° angle and were left to migrate for 1 hour at 37 °C before collection using plastic transfer pipette and washing with Dulbecco's phosphate-buffered saline.

Sperm protein preparation techniques

Experiments were performed on ice to avoid the activity of seminal glycosidases, and otherwise stored in -20°C.

Sperm cell lysis

Sperm were incubated in lysis buffer (25 mM Tris-HCL, 20 mM NaCL, 1% w/v CHAPS, 0.1% w/v protease inhibitor; pH 7.5) and probe sonicated (40A; 15sec on/10off; 2min), then centrifuged at 200g to collect unbroken cells. Protein concentration in the supernatant was measured by

bicinchoninic acid assay (BCA). After reduction (10 mM DTT 1 hour 37°C) and alkylation (25 mM IAA, overnight at room temperature), proteins were dot blotted onto a Polyvinylidene fluoride (PVDF) membrane (20 µg) in preparation for *N*-glycan release [38,39].

Differential detergent membrane extraction

Sperm membrane protein fractions were prepared as described previously [40]. Briefly, sperm cells were washed in PBS, re-suspended in 1 ml of cytosolic extraction buffer (0.01% w/v digitonin, 10mM PIPES (pH 6.8), 300 mM sucrose, 100 mM NaCl, 3 mM MgCl$_2$, 5 mM EDTA, 0.1% w/v protease inhibitor), and placed on ice for 30 min. The extract was centrifuged at 500 × g for 10 min, and then cytosolic extraction was repeated twice interspersed with 10 min incubation on ice. The final pellet was re-suspended in 500 µl membrane extraction buffer (M) (0.5 % v/v Triton X-100, 10mM PIPES (pH 7.4), 300mM sucrose, 100 mM NaCl, 3 mM MgCl$_2$, 3 mM EDTA, 0.1% w/v protease inhibitor), and placed on ice for 30 min. The extraction mix was centrifuged at 5000 × g for 10 min. The supernatant containing the cell membranes was collected and protein concentrations were determined using BCA assay with membrane extraction buffer as control. After reduction (10 mM DTT, 1 hour 37°C) and alkylation (25 mM IAA, overnight at room temperature), proteins were then dot blotted onto a Polyvinylidene fluoride (PVDF) membrane (20ug) in preparation for *N*-glycan release [38, 39].

Ultracentrifugation membrane enrichment

Sperms were lysed as above. After probe sonication and before reduction and alkylation, the sperm membrane fraction was collected by ultra-centrifugation (WX Ultra 80, Thermo Scientific, Australia) at 115,000 rpm for 1 h at 4°C. The membrane pellet was collected from the bottom of the ultra-centrifugation tubes with 500 µl solubilising solution (8M urea). Membrane proteins were measured via BCA assay, reduced, alkylated, then dot blotted onto a Polyvinylidene fluoride (PVDF) membrane (20ug) in preparation for *N*-glycan release [38, 39].

Semen N-glycan release

Proteins, isolated from the above techniques, were reduced and alkylated before dot blotting and release of *N*-glycans as described in section 2.2.

Glycan release, desalting, enrichment and mass spectrometry analysis

Samples were dot-blotted on a PVDF membrane prior to *N*-glycan release by an overnight incubation with 2.5 U *N*-glycosidase F (PNGase F, *Elizabethkingia miricola,* Roche) at 37°C. Released glycans were reduced, desalted and carbon cleaned in prepration for PGC-LC-ESI-MS/SM as described in section 2.2 [38].

Analysis of glycans was performed by using a porous graphitised carbon (PGC) column (5 μ particle size, 180 μ internal diameter x 10 cm; Hypercarb KAPPA Capillary Column, Thermo Fischer Scientific) as an online porous graphitised column-liquid chromatography-electrospray ionisation tandem mass spectrometry (PGC-LC-ESI-MS/MS). The column was equilibrated with 10 mM ammonium bicarbonate (Sigma Aldrich) and samples were separated on a 0-70% v/v acetonitrile in 10 mM ammonium bicarbonate gradient over 75 min, with a flow rate of 4 µl/min. Capillary voltage for ESI was set at 3.2 kV, full auto gain control of 80000, scanning for ion masses of m/z between 600-2000 and mass spectra were acquired in negative ion mode. For MS/MS experiments, CID fragmentation was carried out by 35 normalised-Collison Energy for the 5 top intense ions. Identified glycan masses were searched using GlycoMod (Expasy, http://web.expasy.org/glycomod/) for possible carbohydrate composition [41]. Analysis of MS/MS spectra was carried out using the Thermo Xcalibur Qual browser software. Possible glycan structures were identified based on diagnostic fragment ions and fragment ion mass differences (162Da for hexose, 203Da for N-acetylhexosamine (HexNac), 146Da for fucose, 291Da for N-acetylneuraminic acid (NeuAc), with a 0.2 Da allowance). Relative abundance was determined on MS signal acquisition only. MS/MS level of analysis allows further detailed characterisation of each structure [42]. Diagnostic ions include: 368 and 350 m/z for detecting core-fucosylation, and

others reported in previous literature [43] Relative abundance of each glycan structure was determined on MS signal acquisition data by calculating the area under the peak from the extracted ion chromatograms of the precursor m/z as a percentage of the total peak area.

References

1. Lessley, B.A. and D.L. Garner, *Isolation of motile spermatozoa by density gradient centrifugation in Percoll®.* Molecular reproduction and development, 1983. **7**(1): p. 49-61.

2. Gorus, F.K. and D.G. Pipeleers, *A rapid method for the fractionation of human spermatozoa according to their progressive motility.* Fertility and sterility, 1981. **35**(6): p. 662-665.

3. Guzick, D.S., et al., *Sperm morphology, motility, and concentration in fertile and infertile men.* New England Journal of Medicine, 2001. **345**(19): p. 1388-1393.

4. Tecle, E. and P. Gagneux, *Sugar-coated sperm: Unraveling the functions of the mammalian sperm glycocalyx.* Molecular reproduction and development, 2015. **82**(9): p. 635-650.

5. Ramsby, M.L. and G.S. Makowski, *Differential detergent fractionation of eukaryotic cells*, in *2-D Proteome Analysis Protocols*1999, Springer. p. 53-66.

6. Lai, X., *Reproducible method to enrich membrane proteins with high purity and high yield for an LC-MS/MS approach in quantitative membrane proteomics.* Electrophoresis, 2013. **34**(6): p. 809-817.

7. Bernal, A., et al., *Presence and regional distribution of sialyl transferase in the epididymis of the rat.* Biology of reproduction, 1980. **23**(2): p. 290-293.

8. Tulsiani, D.R., *Glycan-modifying enzymes in luminal fluid of the mammalian epididymis: an overview of their potential role in sperm maturation.* Molecular and cellular endocrinology, 2006. **250**(1-2): p. 58-65.

9. Sullivan, R., G. Frenette, and J. Girouard, *Epididymosomes are involved in the acquisition of new sperm proteins during epididymal transit.* Asian journal of andrology, 2007. **9**(4): p. 483-491.

10. Gilks, C.B., et al., *Histochemical changes in cervical mucus-secreting epithelium during the normal menstrual cycle.* Fertility and sterility, 1989. **51**(2): p. 286-291.

11. Pandya, I.J. and J. Cohen, *The leukocytic reaction of the human uterine cervix to spermatozoa.* Fertility and sterility, 1985. **43**(3): p. 417-421.

12. Thompson, L., et al., *The leukocytic reaction of the human uterine cervix.* American Journal of Reproductive Immunology, 1992. **28**(2): p. 85-89.

13. Tollner, T.L., C.L. Bevins, and G.N. Cherr, *Multifunctional glycoprotein DEFB126—a curious story of defensin-clad spermatozoa.* Nature Reviews Urology, 2012. **9**(7): p. 365.

14. Ma, X., et al., *Sialylation facilitates the maturation of mammalian sperm and affects its survival in female uterus.* Biology of reproduction, 2016. **94**(6).

15. Familiari, G. and P. Motta, *Morphological changes of mouse spermatozoa in uterus as revealed by scanning and transmission electron microscopy.* Acta biologica Academiae Scientiarum Hungaricae, 1980. **31**(1-3): p. 57-67.

16. Focarelli, R., F. Rosati, and B. Terrana, *Sialylglycoconjugates release during in vitro capacitation of human spermatozoa.* Journal of andrology, 1990. **11**(2): p. 97-104.

17. Ma, F., et al., *Sialidases on mammalian sperm mediate deciduous sialylation during capacitation.* Journal of Biological Chemistry, 2012. **287**(45): p. 38073-38079.

18. Netherton, J.K., et al., *Proteomic analysis of good-and poor-quality human sperm demonstrates that several proteins are routinely aberrantly regulated.* Biology of reproduction, 2017. **99**(2): p. 395-408.

19. Mahadevan, M. and G. Baker, *Assessment and preparation of semen for in vitro fertilization*, in *Clinical in vitro fertilization*1984, Springer. p. 83-97.

20. Ben, L., C. Yan, and S.-J. Yu, *Effect of swim-up and Percoll treatment on sperm quality and in vitro embryo development in yak.* Journal of Integrative Agriculture, 2013. **12**(12): p. 2235-2242.

21. Parrish, J., A. Krogenaes, and J. Susko-Parrish, *Effect of bovine sperm separation by either swim-up or Percoll method on success of in vitro fertilization and early embryonic development.* Theriogenology, 1995. **44**(6): p. 859-869.

22. Ramu, S. and R.S. Jeyendran, *The hypo-osmotic swelling test for evaluation of sperm membrane integrity*, in *Spermatogenesis*2013, Springer. p. 21-25.

23. Bollendorf, A., et al., *Comparison of direct swim-up, mini-Percoll, and Sephadex G10 separation procedures.* Archives of andrology, 1994. **32**(2): p. 157-162.

24. Poland, M.L., et al., *Variation of semen measures within normal men.* Fertility and sterility, 1985. **44**(3): p. 396-400.

25. Castilla, J., et al., *Influence of analytical and biological variation on the clinical interpretation of seminal parameters.* Human reproduction, 2005. **21**(4): p. 847-851.

26. Pang, P.-C., et al., *Expression of bisecting type and Lewisx/Lewisy terminated N-glycans on human sperm.* Journal of Biological Chemistry, 2007. **282**(50): p. 36593-36602.

27. Higginson, D.M. and S. Pitnick, *Evolution of intra-ejaculate sperm interactions: do sperm cooperate?* Biological Reviews, 2011. **86**(1): p. 249-270.

28. Dadoune, J.-P., J.-P. Siffroi, and M.-F. Alfonsi, *Transcription in haploid male germ cells.* International review of cytology, 2004. **237**: p. 1-56.

29. Cartwright, E.J., A. Cowin, and P.T. Sharpe, *Surface heterogeneity of bovine sperm revealed by aqueous two-phase partition.* Bioscience reports, 1991. **11**(5): p. 265-273.

30. Simon, P., et al., *Polysialic acid is present in mammalian semen as a post-translational modification of the neural cell adhesion molecule NCAM and the polysialyltransferase ST8SiaII.* Journal of Biological Chemistry, 2013. **288**(26): p. 18825-18833.

31. Smith, S.M., *Strategies for the purification of membrane proteins*, in *Protein Chromatography*2011, Springer. p. 485-496.

32. Kratz, E.M., et al., *The analysis of sialylation, N-glycan branching, and expression of O-glycans in seminal plasma of infertile men.* Disease markers, 2015. **2015**.

33. Schrader, S.M., *Individuality of DNA denaturation patterns in human sperm as measured by the sperm chromatin structure assay.* Re-prod Toxicol, 1991. **5**: p. 115125.

34. Zini, A., et al., *Biologic variability of sperm DNA denaturation in infertile men.* Urology, 2001. **58**(2): p. 258-261.

35. Batista, L., et al., *Complex N-Glycans Are Essential, but Core 1 and 2 Mucin O-Glycans, O-Fucose Glycans, and NOTCH1 Are Dispensable, for Mammalian Spermatogenesis.* Biological Reproduction, 2012. 86(6): p. 179

36. Khalil, M.B., et al., *Sperm capacitation induces an increase in lipid rafts having zona pellucida binding ability and containing sulfogalactosylglycerolipid.* Developmental biology, 2006. **290**(1): p. 220-235.

37. Álvarez-Guerrero, A., et al., *In vitro capacitation and acrosome reaction in sperm of the phyllostomid bat Artibeus jamaicensis.* In Vitro Cellular & Developmental Biology-Animal, 2016. **52**(4): p. 454-465.

38. Jensen, P.H., et al., *Structural analysis of N-and O-glycans released from glycoproteins.* Nature protocols, 2012. **7**(7): p. 1299.

39. Sutton-Smith, M., et al., *Analysis of protein-linked glycosylation in a sperm–somatic cell adhesion system.* Glycobiology, 2007. **17**(6): p. 553-567.

40. Ramsby, M.L. and G.S. Makowski, *Differential detergent fractionation of eukaryotic cells. Analysis by two-dimensional gel electrophoresis.* Methods Mol Biol, 1999. **112**: p. 53-66.

41. Cooper, C.A., E. Gasteiger, and N.H. Packer, *GlycoMod–a software tool for determining glycosylation compositions from mass spectrometric data.* Proteomics, 2001. **1**(2): p. 340-349.

42. Harvey, D.J., et al., *Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry.* Analytical biochemistry, 2008. **376**(1): p. 44-60.

43. Everest-Dass, A.V., et al., *Structural feature ions for distinguishing N-and O-linked glycan isomers by LC-ESI-IT MS/MS.* Journal of The American Society for Mass Spectrometry, 2013. **24**(6): p. 895-906.

Supplementary Data
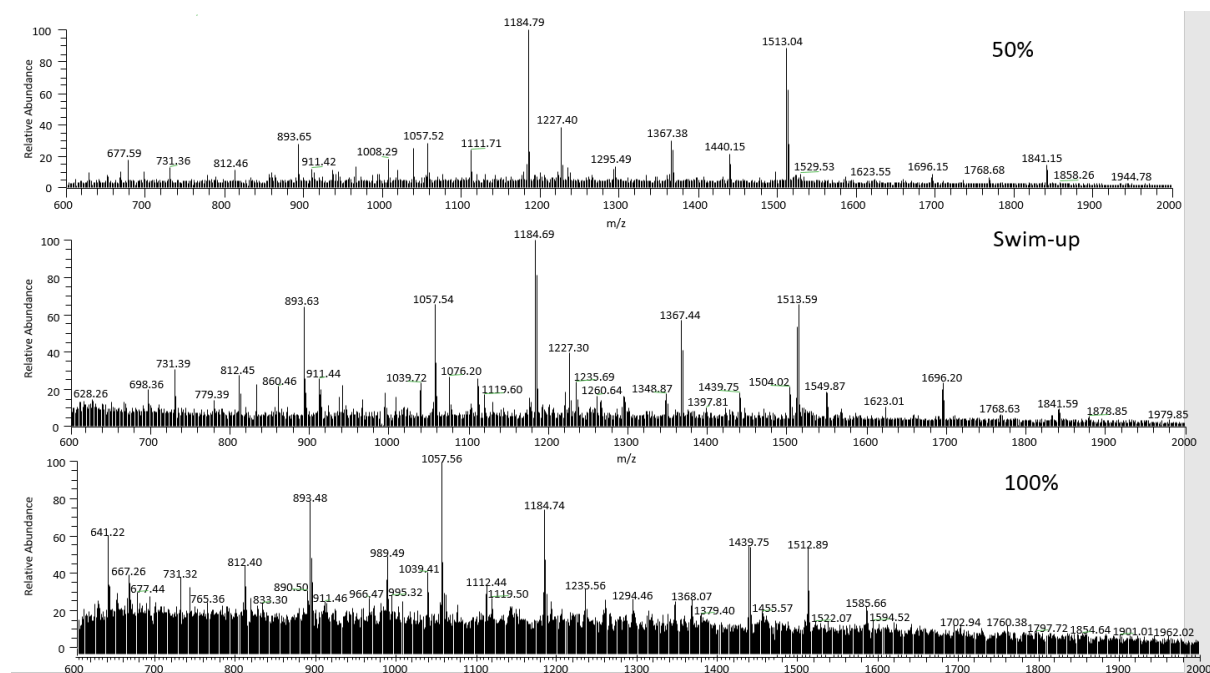
Supplementary Figure 1: Comparison between sperm *N*-glycan structures separated by swim-up and Percoll density gradient techqniues from 1 donor analysed by PGC-LC-MS/MS. This graph shows the lower signal to noise ratio found in high quality Percoll density gradient (100%) compared to low quality Percoll and swim-up sperm.

Chapter 6

Immunosuppression by CD52 specific sialoforms
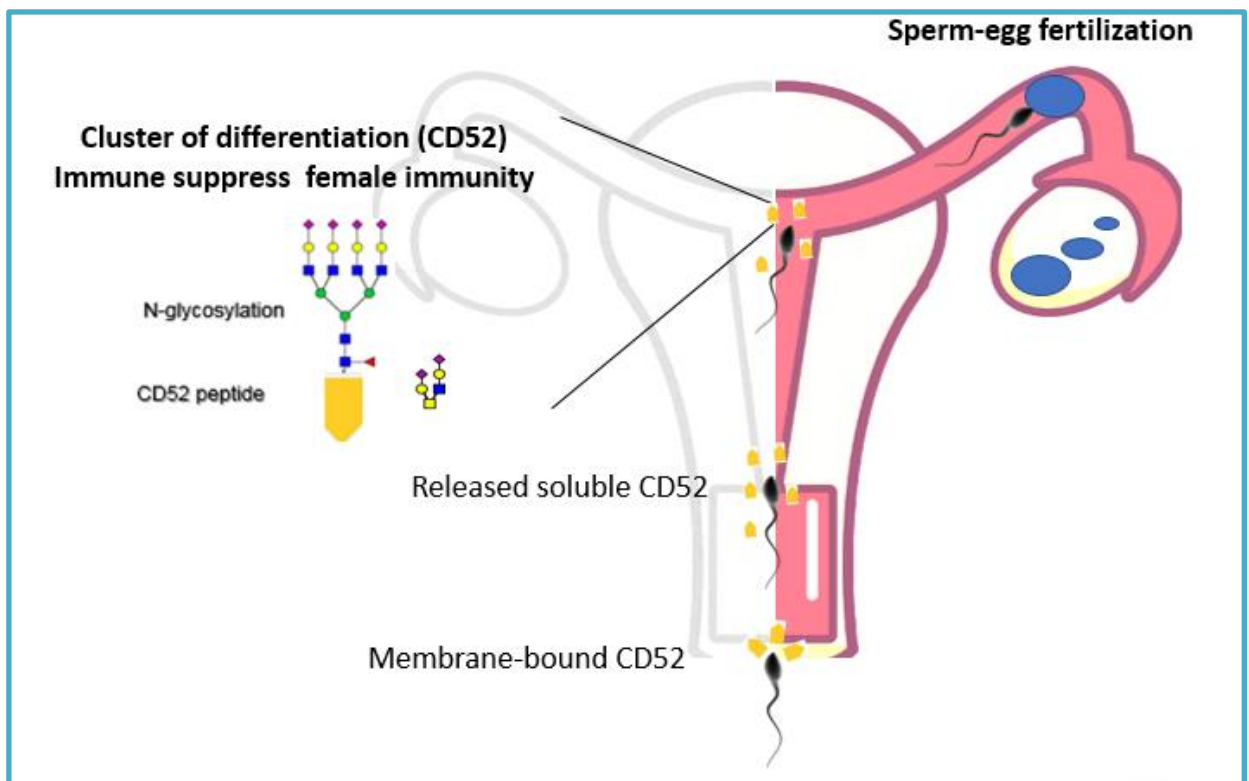
The work described in Chapter 5 reported little difference in the macro glycosylation between active and inactive sperm isolated with different techniques. The aim of the work in this Chapter was to investigate protein-specific glycosylation at the micro level, namely on the immunosuppressive protein, Cluster of differentiation (CD) 52. In semen, CD52 has a proposed role in the ability of human sperm to evade the female immunity. CD52 is immune-suppressive in semen, but only lymphocyte CD52 has been fully characterised from the activity point of view [1] and not in terms of its glycoforms. As a result, lymphocyte CD52 was analysed in the work reported in this Chapter due to its relevance and availability. We initially analysed human CD52 purified from spleen, then recombinant CD52 was expressed in various host cell lines with differential immune suppression activity on T cells in order to relate function to glycosylation status. The active recombinant CD52 was seen to contain significantly higher tri- and tetra-antennary sialylated structures than a similarly expressed product with less immunosuppresive activity. The glycoforms were further resolved via anion exchange chromatography producing specific fractions with increased immunosuppressive activity. The relative abundance of α-2,3 sialic acid linkage correlated with the higher immunosuppressive activity. *O*-glycan core type-2 di-sialylated structures were most abundant in the active fractions, located on Ser12. The results of this work have been prepared as a manuscript to be submitted to *Frontiers of Immunology*, but this manuscript is currently under publication embargo for the protection of intellectual property (IP).

Acknowledgements

6.1 Introduction

One of the main roles of sperm glycosylation is the protection from female immunity in the uterus including antibodies and immune cells such as macrophages and neutrophils [2, 3]. The inhibition of the female immune response has been attributed to the abundant sialylated structures decorating sperm, that function to modulate innate immune receptors such as the sialic acid binding immunoglobulins (Ig)-like lectins (Siglecs) [4]. Siglecs are inhibitory signalling molecules expressed on most immune cells and differ in their sialic acid binding specificity [4].

Furthermore, seminal plasma contains several glycoproteins that are added into the sperm membrane during maturation along the male reproductive tract and some of these glycoproteins function to interact with the female reproductive tract to prime the immune response [5]. Cluster of Differentiation (CD)52, the focus of this Chapter, is secreted in the epididymis (male reproductive tract) and inserted into the sperm membrane with a proposed role in immunosuppression [6, 7]. CD52 is a low molecular weight GPI-anchored glycoprotein (12 amino acid residues) with one *N*-linked glycosylation site and six possible *O*-glycosylation sites [8]. CD52 is also found in lymphocytes but they seem to exhibit a different glycosylation profile than male reproductive tract (mrt) CD52 since a CD52 antibody isolated from an infertile woman (termed MAb H6-3C4) was shown to recognize the *N*-glycan structures of mrtCD52, but not that of the lymphocyte CD52 using Western blot before and after the enzymatic removal (PNGase F) of the *N*-linked carbohydrate moiety [9]. Lymphocyte CD52 immunosuppresses T-cells via binding to Siglec-10 and inhibiting T-cell activation (Figure 6.1) and as little is known about the mechanism of seminal CD52 biological activity, lymphocyte CD52 was used in the first instance in this Chapter to relate glycosylation to immunosuppressive activity.
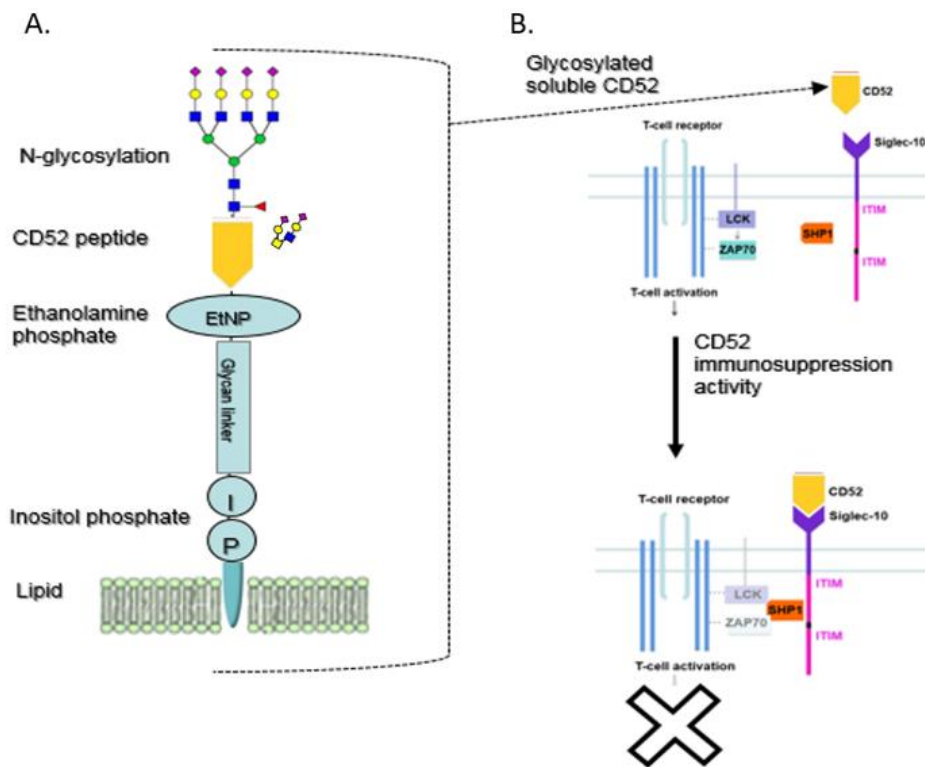
Figure 6.1: A. Molecular features of CD52 GPI-anchored protein. B. CD52 immunosuppression pathway of T-cell activity via SIGLEC-10 ITIM receptors.

Despite our understanding of the activity of lymphocyte CD52, a few key areas remain undefined: the CD52 glycans structure is yet to be fully understood; whether *O*-glycans are also presented on CD52 has not been fully characterised; and the most bioactive glycoform(s) remain unknown. The overall aims of the work reported in this Chapter were:

i. To characterise the *N*- and *O*-glycosylation of human CD52 purified from spleen

ii. To compare the glycosylation of two CD52 recombinant protein constructs that show differential immune suppressive activities

iii. To characterise recombinant CD52 fractionated via monoQ anion exchange chromatography.

6.2 Specific Sialoforms Required for the Immunosuppressive Activity of Human Soluble CD52 (Manuscript 6 under publication embargo for the protection of intellectual property, in-preparation for publication in Frontiers in Immunology)

Shathili A.M.[1,2], Bandala-Sanchez E.[3,4], John A.[3,4], Goddard-Borger E.D.[3,4], Thaysen-Andersen M.[1], Everest-Dass A.[5], Harrison L.C*.[3,4] and Packer N.H*.[1,2,5]

1 Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia

2 ARC Centre of Nanoscale Biophotonics, Macquarie University, Sydney, NSW, Australia

3 The Walter and Eliza Hall institute of Medical Research, Parkville. VIC, Australia

4 Department of Medical Biology, University of Melbourne, Parkville, VIC, Australia

5 Institute for Glycomics, Griffith University, Gold Coast campus, QLD, Australia

A.M.S. and E.B-S. contributed equally to this work.

Corresponding authors: nicki.packer@mq.edu.au and harrison@wehi.edu.au

Contribution: A. M. Shathili performed all the glycan analysis and E. Bandala-Sanchez all the cellular studies. A. John performed MonoQ fractionation. A. M. Shathili, E. Bandala-Sanchez, N. H. Packer and L. C. Harrison analysed data and drafted the manuscript. E. D. Goddard-Borger provided advice and technical support. All authors discussed and commented on the manuscript.

Abstract

Human CD52 is a small glycopeptide (12 amino acid residues) with one *N*-linked glycosylation site at Asn3 and several possible *O*-glycosylation serine/threonine sites. Soluble CD52 is released from the surface of activated T cells and initiates immunosuppression by first sequestering pro-inflammatory HMGB1 followed by binding to the inhibitory sialic acid-binding immunoglobulin-like lectin-10 (Siglec-10) receptor on activated T cells. Our analysis of native CD52 purified from human spleen confirmed extensive heterogeneity in *N*-glycosylation and the presence of multi-antennary sialylated *N*-glycans with abundant polyLacNAc extensions, together with mainly di-sialylated *O*-glycosylation type structures. We aimed to define the glycan structures on recombinant soluble human CD52- immunoglobulin Fc fusion protein that correlate with its immune suppressive activity. Glycomics (porous graphitised carbon-ESI-MS/MS) and glycopeptide (C8-LC-ESI-MS) analysis after factor Xa-based Fc removal revealed that bioactive CD52 was characterised by a higher abundance of tetra-antennary α-2,3/6 sialylated *N*-glycans compared to less bioactive CD52. Moreover, the relative abundance of the α-2,3 sialic acid linkage correlated with higher bioactivity. Removal of α-2,3 sialylation abolished bioactivity, which was restored by re-sialylation with α-2,3 sialyltransferases. Bioactive glycoforms of CD52-Fc were isolated by fractionation on an anion exchange MonoQ-GL column. The CD52 component of fractionated bioactive CD52-Fc displayed mainly tetra-antennary α-2,3 sialylated *N*-glycan structures and decreased relative abundance of bisecting GlcNAc structures compared to non-bioactive adjacent fractions. In addition, *O*-glycan core type-2 di-sialylated structures at Ser12 were more abundant in bioactive CD52 fractions. These findings describe the active glycoforms of CD52 that confer immunosuppressive activity of soluble CD52.

Key words: CD52; immunosuppression; glycosylation; tetra-antennary; α-2,3 sialic acid

Introduction

CD52 is a glycoprotein composed of only 12 amino acid extensively modified by both *N*-linked and possible *O*-linked glycosylation, anchored by glycosylphosphatidylinositol (GPI) to the surface of lymphoid and male reproductive cells. The conserved CD52 peptide backbone probably functions only as a scaffold for presentation of the large *N*-linked glycan, which masks the small GPI-anchored peptide and acts as the prime feature of the CD52 antigen with respect to cell-cell contacts [8, 10]. This notion is supported by the recent discovery of the immune suppressive role of CD52 [1]. Human T cells with high expression of CD52 were found to exhibit immune suppressive activity via phospholipase C-mediated release of soluble CD52, which was shown to bind to the inhibitory sialic acid-binding immunoglobulin (Ig)-like lectin-10 (Siglec-10) receptor on neighbouring T cell populations [1]. This sialic acid interaction was subsequently shown to require initial binding of soluble CD52 glycan with the damage-associated molecular pattern (DAMP) protein, high-mobility group box 1 (HMGB1). Complexing of soluble CD52 with HMGB1 promoted binding of CD52 *N*-glycan, preferentially in α-2,3 sialic acid linkage, to Siglec-10 [11].

In the only previous detailed MS analysis, the *N*-glycans on human leukocyte CD52 exhibited extensive heterogeneity with multi-antennary complexes containing core α-1,6 fucosylation, abundant polyLacNAc extensions and variable sialylation [8]. With recent insights into the function of soluble CD52, and its potential as an immunotherapeutic agent, the glycan structure-function determinants of CD52 require more detailed investigation. In particular, although the CD52 *N*-glycan is known to be required for bioactivity [1], its structure is not fully elucidated and the glycoforms required for bioactivity have not been identified. In addition, even with a total of six potential amino acid sites, *O*-glycosylation of CD52 has not been analysed. We aimed therefore to address these questions using both purified native CD52 from human spleen and recombinant soluble CD52 expressed as a fusion protein with immunoglobulin Fc.

Materials and Methods

Human blood and spleen donors

Cells were isolated from human blood buffy coats (Australian Red Cross Blood Service, Melbourne, VIC, Australia) or blood of de-identified healthy volunteers with informed consent through the Volunteer Blood Donor Registry of The Walter and Eliza Hall Institute of Medical Research (WEHI), following approval by WEHI and Melbourne Health Human Ethics Committees. Peripheral blood mononuclear cells (PBMCs) were isolated on Ficoll/Hypaque (Amersham Pharmacia, Uppsala, Sweden), washed in phosphate-buffered saline (PBS) and re-suspended in IMDM medium containing 5% pooled, heat-inactivated human serum (PHS; Australian Red Cross, Melbourne, Australia), 100 mM non-essential amino acids, 2 mM glutamine and 50 μM 2-mercaptoethanol (IP5 medium).

Healthy human spleen from cadaveric organ donors were obtained from Australian Islet Transplant Consortium and trained coordinators of Donate Life from heart-beating, brain dead donors with informed written consent of next of kin. All studies were approved by WEHI Human Research Ethics Committee (Project 05/12).

Purification of native CD52 from human spleen

Frozen human spleen tissue (10 mg) was homogenized with three volumes of water as per described in [12]. In brief, homogenate was mixed with methanol and chloroform 11:5.4 volumes, respectively. Samples were left to stir for 30 min and allowed to stand for one hour. The upper (aqueous) phase was collected, evaporated, dialyzed and freeze dried. NHS-activated Sepharose 4 Fast Flow resin was incubated with 1 mg of purified anti-CD52 antibody in 0.5 mL of PBS for 3 hr at RT. The mixture was incubated overnight at 4°C and quenched with 1 M ethanolamine. A Bio-Rad 10-mL Poly-Prep column was used for packing and resins were washed with sequential treatment of 5 mL of PBS, 5 mL of pH 11.5 diethylamine, and 5 mL of PBS /0.02% sodium azide. The column was stored at 4°C in 5 mL of PBS/0.02% sodium azide before use. Spleen extracts

were solubilized with 2mL of 2% sodium deoxycholate in PBS, and then added to the packed column and washed with 5 mL of PBS containing 0.5% sodium deoxycholate. The sample was eluted with six times 500 µl of elution buffer (50 mM diethylamine, 500 mM NaCl, pH 11.5) containing 0.5 % sodium deoxycholate. The eluate was collected, neutralized with 50 µl of HCl (0.1M) and dialyzed against PBS and water.

CD52 recombinant proteins

Human CD52-Fc recombinant proteins; CD52-Fc I (Expi293), CD5-Fc II (HEK293) and CD52-Fc III (Expi293) were produced as described [1]. Expi293 and HEK293 are human embryonic kidney cells. Expi293 cells generate superior protein yield compared to HEK293 cells. The signal peptide sequences joined to human IgG1 Fc were constructed with polymerase chain reaction (PCR) then digested and ligated into a FTGW lentivirus vector or pCAGGS vector for the transfection of FreeStyle HEK293F and Expi293 cells. The construct included a flexible GGSGG linker, a strep-tag II sequence for purification [13], and a cleavage sites for Factor Xa protease between the signal peptide and Fc molecule. The recombinant proteins were purified from the medium by affinity chromatography on Streptactin resin and eluted with 2.5 mM desthiobiotin [1].

^{3}H-thymidine incorporation assay

PBMCs (2×10^5 cells/well) in IP5 medium were incubated for up to 3 d at 37 °C in 5% (X/X) CO_2 in 96-well round-bottomed plates with or without tetanus toxoid (10 Lyons flocculating units per ml), and various concentrations of CD52-Fc or control Fc protein, in a total volume of 200 µL. In some wells, ^{3}H-thymidine (1 µCi) was added and, after 16 h, cells were collected and radioactivity in DNA measured by scintillation counting.

ELISpot assay

PBMCs (2×10^5 cells/well) were cultured in 200 µL of IP5 medium in triplicate wells of a 96-well ELISpot plate (PVDF MultiScreen) from Merck Millipore (Bayswater, Australia) containing anti-IFN-γ monoclonal antibody pre-bound (1 µg/mL) at 4°C. PBMCs were incubated with tetanus

toxoid (10 Lfu/mL) added to the wells together with CD52-Fc I, CD5-Fc II and CD52-Fc III (5, 25 and 50 µg/mL). After 24 h, cells were removed by washing and IFN-γ spots were developed by incubation with biotinylated anti-IFN-γ antibody (1 µg/mL) followed by streptavidin-alkaline phosphatase and BCIP/NBT colour reagent (Resolving Images, Melbourne, Australia).

Lectin ELISA

A 96-well flat-bottom plate was coated with 20 µg/mL of *Maackia amurensis* lectin (MAL-I) overnight at 4°C and subsequently blocked with 200 µl of 1 % BSA for 1 h. After washing with PBS, CD52-Fc (20 µg/mL) was added and incubated at RT for 1 hr and washed twice with PBS. After washing with PBS, 50 µl of a 1:1000 dilution of HRP-conjugated antibody to CD52 (Campath H1; 1 µg/mL) was added and incubated at RT for 1 h. 50 µl of TMB substrate was added and colour development stopped by addition of 50 µl of 0.5 M H_2SO_4. Absorbance was measured at 450 nm in a Multiskan Ascent 354 microplate photometer (Thermo Labsystems, San Francisco, Calif.).

De-sialylation and re-sialylation of recombinant CD52-Fc protein

De-sialylation and re-sialylation of recombinant CD52-Fc proteins were performed by a modification of the method of Paulson and Rogers [14]. Briefly, CD52-Fc (500 µg/each) was incubated with *Clostridium perfringens* type V sialidase (50 mU/mL) for 3 hr at 37 ºC to remove all types of sialic acids. Samples were then passed through a Protein G-Sepharose column, which was washed twice with PBS before the bound protein was eluted with 0.1 M glycine-HCl, pH 2.8 into 1 M Tris-HCl, pH 8.0, followed by dialysis against PBS. Binding to MAL-I lectin was performed to confirm removal of sialic acids. CD52-Fc from Expi293 cells was then incubated with either of two sialyltransferases, PdST6GalI which restores sialic acid residues in α-2,6 linkage with underlying galactose or CstII which restores sialic acid residues in α-2,3 linkage with galactose, in the presence of 0.46 mM-0.90 mM CMP-N-acetylneuraminic acid sodium salt (Carbosynth, Compton Berkshire, United Kingdom) for 3 h at 37°C. The different CD52-Fc (III) proteins with different linkages (α-2,3 or α-2,6) were passed through Protein G-Sepharose

columns, washed twice with PBS and eluted with 0.1 M glycine-HCl, pH 2.8, into 1 M Tris-HCl, pH 8.0, followed by dialysis against PBS. Samples were freeze-dried, re-suspended in PBS at 200 µg/mL and stored at -20 °C.

Fc fragment removal

CD52-Fc recombinant protein fractions (50-200 µg) were incubated with 4 µL of Factor Xa protease (purified from bovine plasma, New England Biolabs, United States) in a total volume of 1 mL of cleavage buffer (20 mM Tris-Hcl, pH 8, 100 mM NaCl, 2 mM $CaCl_2$). Samples were incubated overnight at RT. Samples were mixed three times with Protein G-Sepharose beads for 1 hr at RT and centrifuged at 10,000 rpm for 15 min. Fc fragment removal was confirmed by Western blot using anti-human IgG (Fc specific produced in goat; Sigma, United States) and anti-CD52 (rabbit) antibodies (Santa Cruz Biotechnology, United States).

N- and O- linked glycan release for mass spectrometry analysis

Mono Q fractionated and whole (non-fractionated) recombinant CD52-Fc were dot-blotted on a PVDF membrane. Soluble CD52 with the Fc removed was kept in-solution prior to *N*-glycan release by an overnight incubation with 2.5 U *N*-glycosidase F (PNGase F, *Elizabethkingia miricola,* Roche) at 37°C followed by a NaBH4 reduction (1 M NaBH4, 50mM KOH) for 3 hr at 50°C. *O*-glycan were subsequently released by overnight reductive β-elimination using 0.5 M NaBH4, 50mM KOH at 50°C. The released and reduced *N*- and *O*-glycans were thoroughly desalted prior to the LC-MS/MS as described previously [15].

Mass spectrometry and data analysis of released glycans

The separation of glycans was performed by using a porous graphitised carbon (PGC) column (5 µ particle size, 180 µ internal diameter x 10 cm column length; Hypercarb KAPPA Capillary Column, Thermo Scientific) operated at a constants flow rate of 4 µl/min using a Dionex Ultimate 3000 LC (Thermo Scientific). The separated glycans were detected online using liquid chromatography-electrospray ionisation tandem mass spectrometry (LC-ESI-MS/MS) using an

LTQ Velos Pro mass spectrometer (Thermo Scientific). The PGC column was equilibrated with 10 mM ammonium bicarbonate (Sigma Aldrich) and samples were separated on a 0-70% (v/v) acetonitrile in 10 mM ammonium bicarbonate gradient over 75 min. The ESI capillary voltage was set at 3.2 kV. The full auto gain control was set to 80,000. MS1 full scans were made between *m/z* 600-2000. All glycan mass spectra were acquired in negative ion mode. The LTQ mass spectrometer was calibrated with a tune mix (Pierce[TM] ESI negative ions, Thermo Scientific) for mass accuracy of 0.2 Da. The CID-MS/MS was carried out on the five most abundant precursor ions in each full scan by using 35 normalized collision energy. Possible monosaccharide compositions were provided by GlycoMod (Expasy, http://web.expasy.org/glycomod/) based on the molecular mass of glycan precursor ions [16]. Analysis of MS/MS spectra was carried out using the Thermo Xcalibur Qual browser software. Possible glycan structures were identified based on diagnostic fragment ions 368 for core fucosylation and others as reported in previous literature [17], and B/Y- and C/Z-glycan fragments in the CID-MS/MS spectra. A mass tolerance of 0.2 Da was allowed for both the precursor and product ions. The relative abundances of the identified glycans were determined as a percentage of the total peak area from the MS signal strength using area under the curve (AUC) of extracted ion chromatograms of glycan precursor ion [18].

Profiling the N- and O- glycans on the CD52 peptide

MonoQ fractionated and unfractionated CD52 glycoforms without the Fc were desalted on C18 micro-SPE stage tips (Merck-Millipore). Elution was performed with 90% Acetonitrile (ACN) and samples were dried and redissolved in 0.1% Formic acid (FA). The desalted CD52 glycopeptides were analysed by ESI-LC-MS in positive ion polarity mode using a Quadrupole-Time-of-flight (Q-TOF) 6538 mass spectrometer (Agilent technologies)-HPLC (Agilent 1260 infinity). In parallel experiments, *N*-glycosidase F was used to remove *N*-glycans from some samples of CD52 (with a resulting Asn->Asp conversion i.e. +1 Da) to enable better ionization of the highly heterogeneous and anionic CD52 glycopeptides. The *N*- and *O*-glycan occupancy was determined by comparing

the AUC of the deamidated and *O*-glycosylated CD52 glycoforms. Samples (roughly 500 ng) were injected onto C8 column (Protecol C8, 3 µm particle size, 300 A pore size, 300 nm inner diameter 10 cm length, SGE analytical science). The HPLC gradient was made starting with 0.1% FA with a linear rise to 60% (v/v) ACN 0.1% FA over 30 min. The column was then washed with 99% ACN (v/v) for 10 min before re-equilibration with 0.1% FA for another 10 min. The flow rate was set to 4 µL/min with an optimised fragmentor positive potential of 200 V with the following MS setting: m/z range 400-2500, nitrogen drying gas flow rate 8 L/min at 300°C, nebulizer pressure was 10 psi, capillary positive potential was 4.3 kV, skimmer potential was 65 V. The mass spectrometer was calibrated with a tune mix (Agilent technologies) to reach a mass accuracy typically better than 0.2 ppm. MassHunter workstation vB.06 (Agilent technologies) was used for analysis and deconvolution of the resulting spectra. The previously determined glycans from the PGC-LC-MS/MS analysis were used to guide the assignment of glycoforms to deconvoluted CD52 peptides based on the accurate molecular mass.

Mono Q fractionation

CD52-Fc III was diluted into 5 mL 50 mM Tris-HCl, pH 8.3 and applied to a Mono Q column (Mono Q 5/50 GL, GE Lifesciences). The column was washed with 10 column volumes of 50 mM Tris-HCl, pH 8.3 and then eluted with 50 column volumes of 50 mM Tris-HCl, 500 mM NaCl, pH 8.3 in 0.5 mL fractions. Fractions were then collected and analyzed by isoelectric focusing (IEF).

IEF

Novex pH 3-10 IEF gels were used for pI determination. CD52-Fc fractions were loaded with sample buffer and run at 100 V for 2 h, then at 250 V for 1 h and, finally, the voltage was increased to 500 V for 30 min. After electrophoresis, the gel was carefully transferred to a clean container, washed and fixed with 20% trichloroacetic acid (TCA) for 1 h at RT, rinsed with distilled water, stained with colloidal Coomasie blue for 2 h at RT, and thoroughly destained with distilled water.

Sequential sialidase treatment

N-glycans released from cleaved CD52 (roughly 2 µg) were treated with α-2-3-specific sialidase

(1 mU, Merck) and broad (α-2-3,6,8 sialidase-reactive) sialidase *V. cholera* (1 mU, Sigma Aldrich). Both reactions were carried out in 50mM sodium phosphate reaction buffer at 37°C for 3 h. Desialylated CD52 *N*-glycans were dried and solubilised in water for downstream MS analysis. Fetuin was used as positive control for successful sialic acid removal since, like cleaved CD52, this model glycoprotein carries multi-antennary sialylated *N*-glycans.

EThcD fragmentation for O-glycans site localisation on the CD52 peptide

Fractionated CD52 glycoforms were treated with PNGase F prior to *O*-glycan site localisation analysis. CD52 peptides were analysed using a Dionex 3500RS nanoUHPLC coupled to an Orbitrap Fusion™ Tribrid™ Mass Spectrometer in positive mode with the same LC gradient mentioned in 'Profiling the *N*- and *O*- glycans on intact CD52' but while using a nano-flow (250 nL/min). The following MS settings were used: spray voltage 2.3 kV, 120k orbitrap resolution, scan range *m/z* 550-1,500, AGC target 400,000 with one microscan. The HCD-MS/MS used 40% nCE. Precursors that resulted in fragment spectra containing diagnostic oxonium ions for glycopeptides i.e. *m/z* 204.08671, 138.05451 and 366.13961, were selected for a second EThcD (nCE 15%) fragmentation. The analysis of all fragment spectra was carried out using Thermo Xcalibur Qual browser software with the aid of Byonic (v2.16.11, Protein Metrics Inc) using the following parameters: precursor mass tolerance 6 ppm, fragment mass tolerance 1 Da and 10 ppm to respectively account for possible proton transfer during ETD fragment formation and the MS/MS resolution, deamidated (variable) and two core type 2 *O*-glycans, previously seen in intact mass analysis.

Statistical analyses

Data are expressed as mean ± standard deviation (SD). The significance of differences between groups was determined using two-sided t-test, with Prism software (GraphPad Software). $p < 0.05$ was used throughout as the significance threshold.

Results

Human spleen-derived CD52 exhibits extensive N- and O-glycosylation heterogeneity

To characterize the natural glycosylation of human CD52, we purified CD52 from human spleen and performed a comprehensive analysis of released *N*- and *O*-glycans by porous-graphitised carbon (PGC)-ESI-MS/MS (Figure 1A, B). We confirmed high *N*-glycosylation heterogeneity, expressed as multi-antennary sialylated *N*-glycans with abundant polyLacNAc extensions (Figure 1A). Similar *N*- glycans have been previously reported for natural occurring human CD52 [8]. The *O*-glycosylation profile was characterized as core type 1 and core type 2 sialylated structures with mainly (66%) di-sialylated core type 2 *O*-glycans (Figure 1B). This glycan heterogeneity raises the question whether particular bioactive glycoforms of CD52 exist and whether such heterogeneity is reflected in the recombinant form of human CD52.
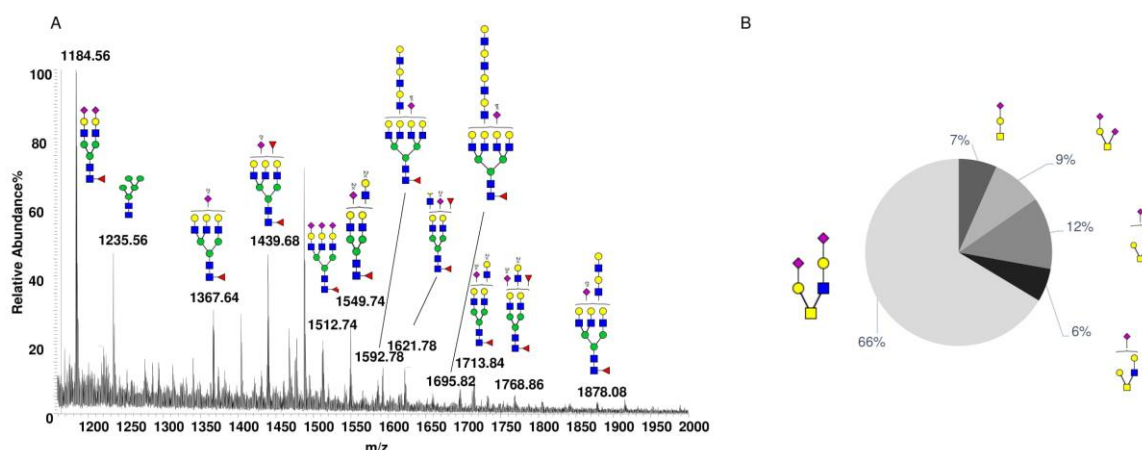
Figure 1

Figure 1: Glycosylation analysis of human spleen CD52. A, Summed MS profile of released *N*-glycans from CD52 purified from human spleen tissue. B, Distribution of *O*-linked glycans released from human spleen CD52.

Recombinant CD52 shares common N-glycan structures with Fc carrier protein

The yield of native soluble CD52 was insufficient to enable us to pinpoint the bioactive glycoforms on the naturally occurring glycoprotein. Therefore, we engineered human CD52 as a recombinant

fusion protein conjugated with an IgG1 Fc fragment as described [1]. Previously, we demonstrated the ability of recombinant CD52-Fc, but not its Fc component, to suppress a range of immune functions [1, 11]. The two recombinant human CD52-Fc batches we generated for this study recapitulated this previously observed immuno-suppressive bioactivity (Figure 2A). However, the Fc has a single *N*-linked glycosylated site at N297 (Figure 2Bi), which had to be considered in characterizing and assessing the impact of the *N*-glycosylation of recombinant CD52-Fc. This was addressed in two ways: i) by analysing a recombinant form of human CD52-Fc in which Fc contained an N297A mutation, allowing analysis of CD52 *N*-glycosylation profile at the released glycan level without interference from the Fc *N*-glycan (Figure 2Bii), and 2) by removal of the Fc component from CD52-Fc via Factor Xa proteolysis of a cleavage site appropriately incorporated in the CD52-Fc construct, as shown by a Western blot using a specific antibody for CD52 (Figure 2C).
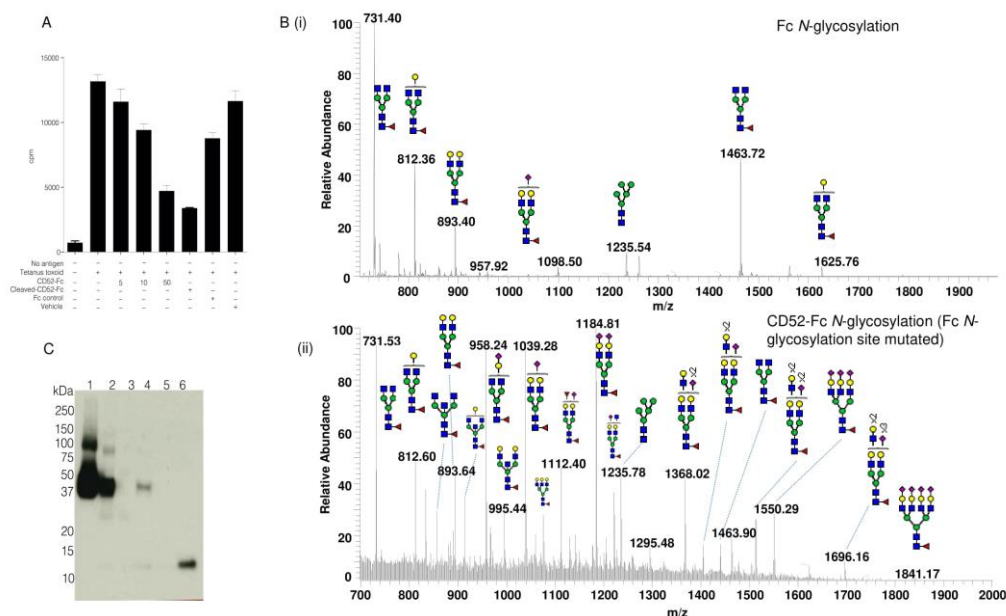

Figure 2: Comparative *N*-glyco-profiling of recombinant human IgG Fc and CD52. A, Proliferation of human PBMCs (^{3}H thymidine uptake) followed 5 days incubation with tetanus toxoid. Histograms show mean ± SD of within-assay triplicates, in the presence of different concentration of proteins. The Fc component was cleaved from CD52-Fc with Factor Xa. B, Summed MS profile of *N*-glycans released from the Fc (I) and CD52 (II); the latter variant was generated by introducing a point mutation (A297N) into the conventional Fc *N*-glycosylation site. C, Factor Xa treated-CD52 was analyzed by Western blotting with Campath-H1-HRP antibody.

Bioactive recombinant CD52 glycoforms displays more abundant tri- and tetra-antennary sialylated N-glycans

We had noted that the specific bioactivity of recombinant CD52-Fc varied from batch to batch. Therefore, we compared the impact of sialylation between two CD52-Fc variants made from different host cells that were shown display higher and lower immunosuppressive activity, here referred to respectively as CD52-Fc I (from Expi 293 cells) and CD52-Fc II (from HEK 293 cells) (Figure 3A).

N-glycans were released *via* in-solution treatment with PNGase F and subsequently analysed by PGC-ESI-MS/MS [15]. *N*-glycans on cleaved CD52 I had greater relative abundances of bi-, tri- and tetra- antennary sialylated glycans compared to CD52 II (Figure 3B). Also, CD52 I displayed a significantly higher relative abundance of sialylated structures possibly containing LacNAc moieties (Figure 3B). Not only the numbers of antennae, but also their degree of sialylation differed between the two recombinant CD52 glycoforms: tetra-sialylated *N*-glycans were significantly more abundant in CD52 I ($6.9\pm0.9\%$) compared to CD52 II (4.9 ± 1.3) ($p < 0.05$). In contrast, CD52 II displayed significantly greater abundance of non-sialylated bi-antennary and bisecting structures (30% and 5% compared to 19% and 2%, respectively) (Figure 3B), which could exclude the involvement of such structures in the bioactivity of CD52.
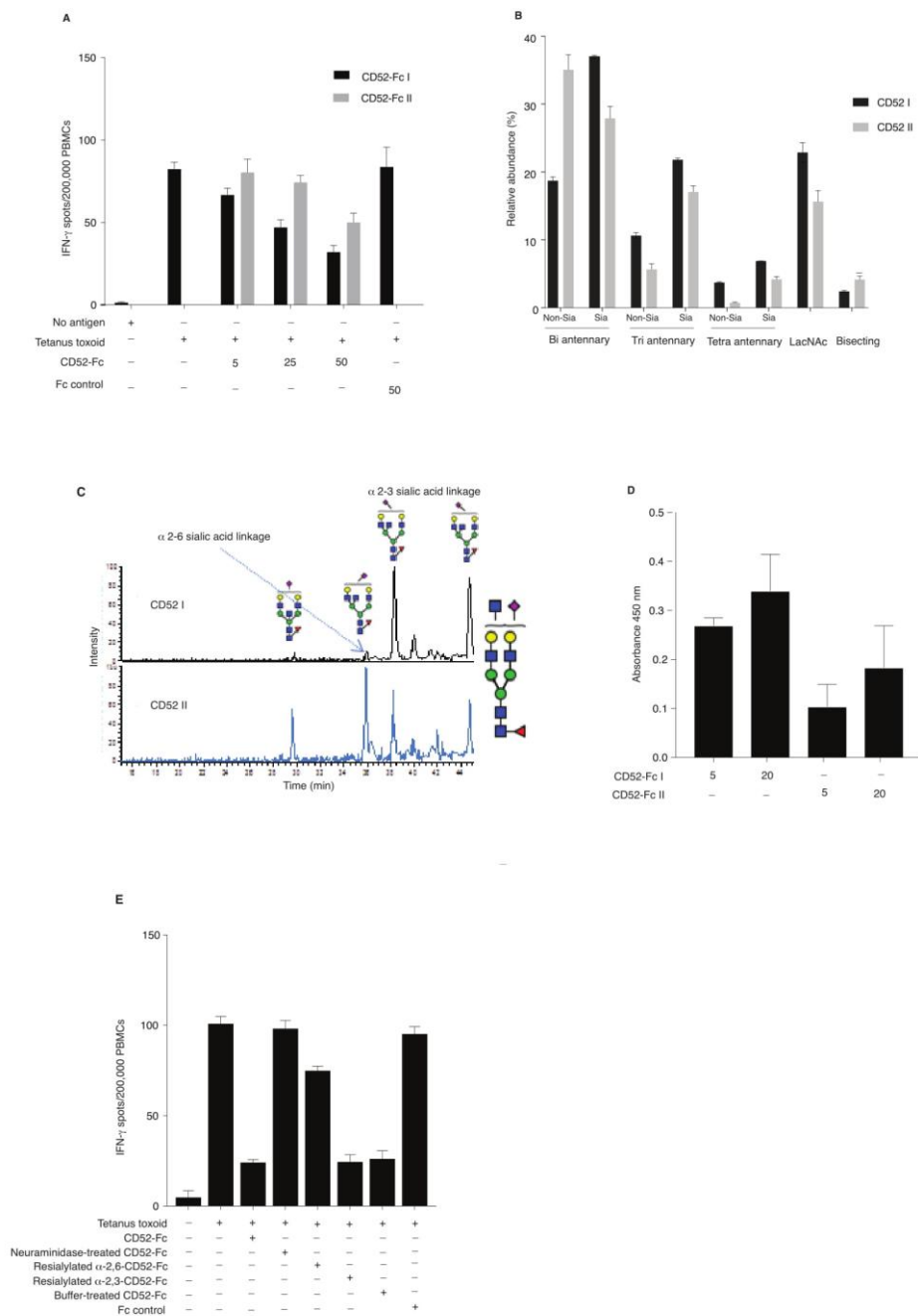
Figure 3: Comparison of two recombinant human CD52-Fc variants (I and II) with different immunosuppressive activities. A, IFN-γ production measured by ELISpot assay from human PBMCs (2 x 10^6) in 200 µL/well. Samples were incubated with no antigen or tetanus toxoid in the presence of two different preparations of CD52-Fc (CD52 I or CD52 II; 5, 25, and 50 µg/ml). B, N-linked glycans released from cleaved CD52 I and CD52 II. The abundance of each N-glycan class is the sum of all EICs measured for all glycans in that class relative to the total of all EICs observed for all N-glycans. C, EIC of m/z 1140.4^{2-} (GlcNAc$_5$Man$_3$Gal$_2$NeuAc$_1$) demonstrating the PGC-based separation of sialo-glycan isomers observed in CD52 I and CD52 II. D, Binding of CD52-Fc I and CD52-Fc II (5 and 20 µg/ml) to the α-2,3 sialic acid recognizing lectin MAL-1. E, ELISpot assay showing activity of CD52-Fc I reconstituted with

210

sialic acid in α2-6, α2-3 and α2-8 linkages with galactose. The data points in panel A, D and E are plotted as mean ±
SEM of three independent replicate experiments. Data in B and C are mean ± SDs (n = 3). Student t test was used to
test for significant difference between group means.

After the removal of Fc, recombinant CD52 I and CD52 II were then subjected to high-resolution
intact peptide analysis using C8-LC-ESI-MS. Both proteins showed *N*-glycosylation profiles
similar to those of released glycans. The high resolution of the Q-TOF instrumentation used even
in the high m/z range enabled the identification of very elongated sialylated antennary structures
including searching for *N*-glycans carrying Lewis-type structures (antenna-type fucosylation). The
experimental isotopic distribution of both variants of recombinant CD52 matched the theoretical
isotopic distribution of the 90% tri-sialylated (non-Lewis fucosylated) CD52 glycoforms, which
indicate that the main glycoforms of recombinant CD52 do not carry Lewis-type fucosylation
(Supplementary Figure 1A). The more bioactive CD52 I displayed a higher level of multi-
antennary sialylated and possible LacNAc elongated structures (Supplementary Figure 1B).
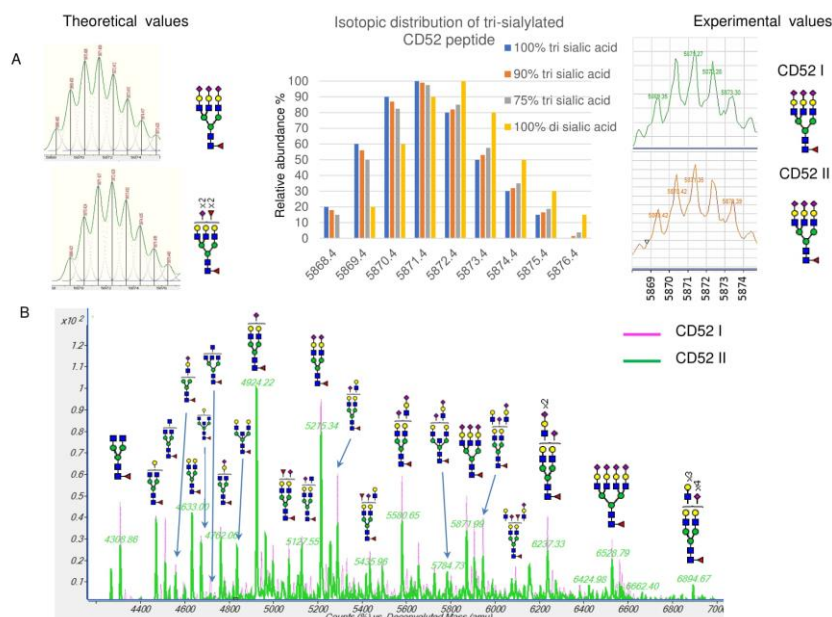
Supplementary Figure 1: Analysis of CD52 I and CD52 II at the intact peptide level. A, The theoretical isotopic distribution of deconvoluted 5871.99 (amu) CD52 glycoform as as tri-sialylated (GlcNAc5Man3Gal3NeuAc3Fucose1) or di-sialylated with two outer fucoses (GlcNAc5Man3Gal3NeuAc2Fucose3). The bar graph shows the theoretical isotopic envelopes generated when different amount of these two glycans are present. Experimental isotopic distribution values suggest a population of 90-100% tri-sialylated structures. B, High-resolution intact mass analysis of CD52 I (pink) and CD52 II (green).

α-2,3 sialylated N-glycans are indispensable for CD52 activity

CD52 N-glycans displaying α-2,3 sialylation preferentially bind to Siglec-10 [11]. PGC-LC-MS/MS glycan analysis and *Maackia amurensis-1* (MAL-I) lectin blotting were used to identify any differences in sialic acid linkage between the two variants of recombinant CD52-Fc (CD52-Fc I and CD52-Fc II). MAL-I has a known preferential recognition of glycoconjugates displaying α-2,3 sialylation. Despite the high separation power of PGC for sialoglycans, this technique has difficulty resolving very large multi-antennary sialylated glycans, but can easily discriminate between α-2,3 and α-2,6-sialylation on the more common bi- and tri-antennary N-glycans. Several abundant bi-antennary α-2,3 sialoglycans were observed on CD52 I. For one sialylated glycan, *m/z* 1140.4^{2-} (GlcNAc$_5$Man$_3$Gal$_2$NeuAc$_1$), only the α-2,3 sialic acid glycan isomer was observed on CD52 I. On the other hand, the less bioactive CD52 II carried both α-2,3 and α-2,6 sialo-N-glycans

212

(Figure 3C). This differential sialyl linkage presentation between the two recombinant CD52 variants was supported by MAL-I lectin binding, which was higher for the more bioactive CD52-Fc I (Figure 3D). MAL-I lectin favours binding to α-2,3 sialic acid linked tri- and tetra-sialylated *N*-glycans [19]. The importance of the α-2,3 sialylation for bioactivity of CD52-Fc was confirmed in a parallel experiment where the immuno-suppressive activity of sialidase-treated and re-sialylated CD52-Fc was determined relative to the original recombinant variant. Treatment of CD52-Fc with sialidase completely abolished its immunosuppressive activity, which was fully restored upon re-sialylation with α-2,3, but not α-2,6 sialylation (Figure 3E). Overall, these findings indicate that the bioactivity of CD52-Fc is associated with the presence of α-2,3-linked tetra-sialylated *N*-glycans found on CD52 and not the Fc.

Active CD52 glycoforms resolved by anion exchange chromatography

We then performed anion exchange chromatography on a MonoQ column in order to separate recombinant CD52-Fc variants based on their degree of sialylation, with the aim of identifying the most bioactive forms (Figure 4A).
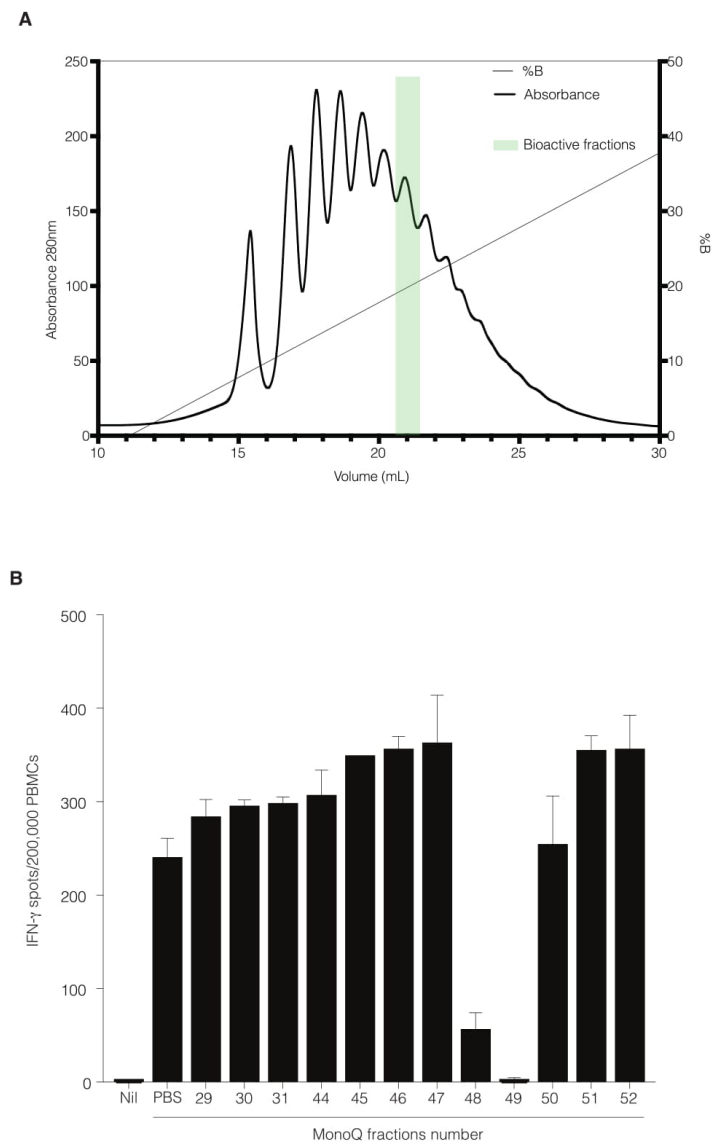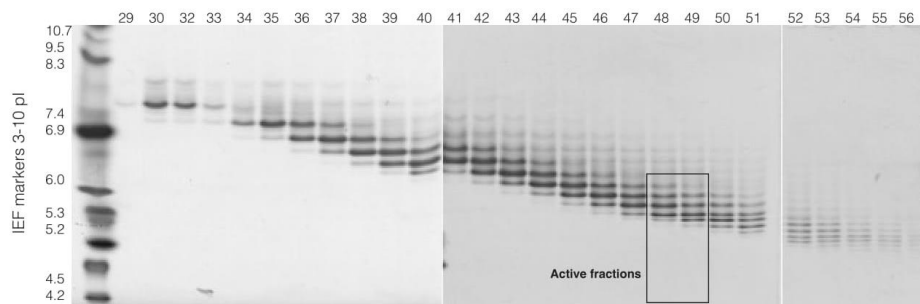
Figure 4

A

B

Figure 4: CD52-Fc after fractionation by anion-exchange chromatography. A, Anion exchange chromatography on a MonoQ-GL column fractionated the recombinant human CD52-Fc into a gradient of anionic glycoforms displaying a spectrum of pI (see Supplementary Figure 2). B, IFN-γ ELISpot assay with 2 x 10^6 PBMCs in 200 µL/well incubated with no antigen or with anti-CD3/CD28 antibody Dynabeads in the presence of recombinant human CD52-Fc fractions (F29-F53; 5 µg/ml).

The increasing degree of sialylation (decreasing isoelectric point [pI]) of CD52-Fc in the collected fractions was validated by isoelectric focusing (IEF) (Supplementary Figure 2) and mass spectrometry.
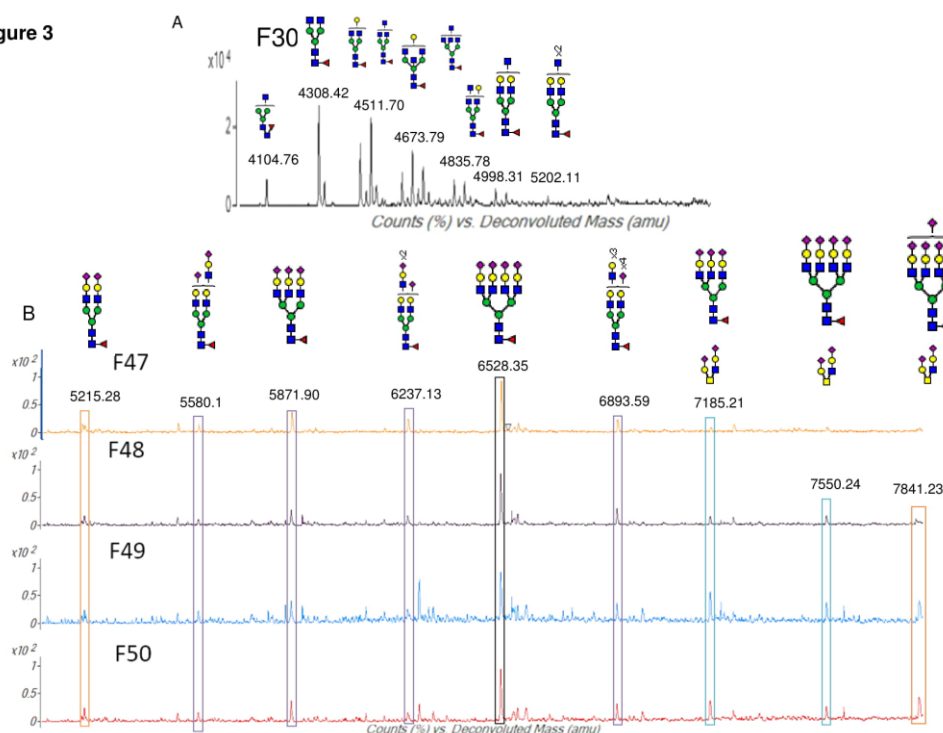
Supplementary Figure 2: CD52-Fc fractions resolved in isoelectric focusing (IEF) gel. Colloidal Coomassie Blue gel showing protein in MonoQ fractions (F29-54). Fractions showed a gradual decrease in isoelectric point (pI) values.

The released *N*-glycans from fractions 46 to 51 (F46-F51) revealed a gradual increase in sialic acid content, and structures containing a higher number of antennae (Supplementary Table I), as shown also from intact glycopeptide analysis (Supplementary Figure 3). Released and intact glycan analysis from fraction 30 revealed various GlcNAc and Gal capped structures and a complete absence of sialic acid moieties (Supplementary Table I and Supplementary Figure 3). Remarkably, only two fractions, F48-F49, with pI in the 5-6 range displayed significant immunosuppressive activity (Figure 4B). The adjacent fractions were not bioactive, even at higher concentrations of protein (Supplementary Figure 4A and B). These late-eluting, uniquely bioactive fractions (F48-49) were highly enriched (60-70%) in tri- and tetra-sialylated glycans, further confirming that sialylation strongly impacts the bioactivity of CD52. However, the adjacent non-bioactive fractions also contained similar glycoforms with small differences in overall relative abundance, indicating that the highly sialylated glycoforms, while necessary, are not sufficient for bioactivity (Supplementary Table I and Supplementary Figure 3).
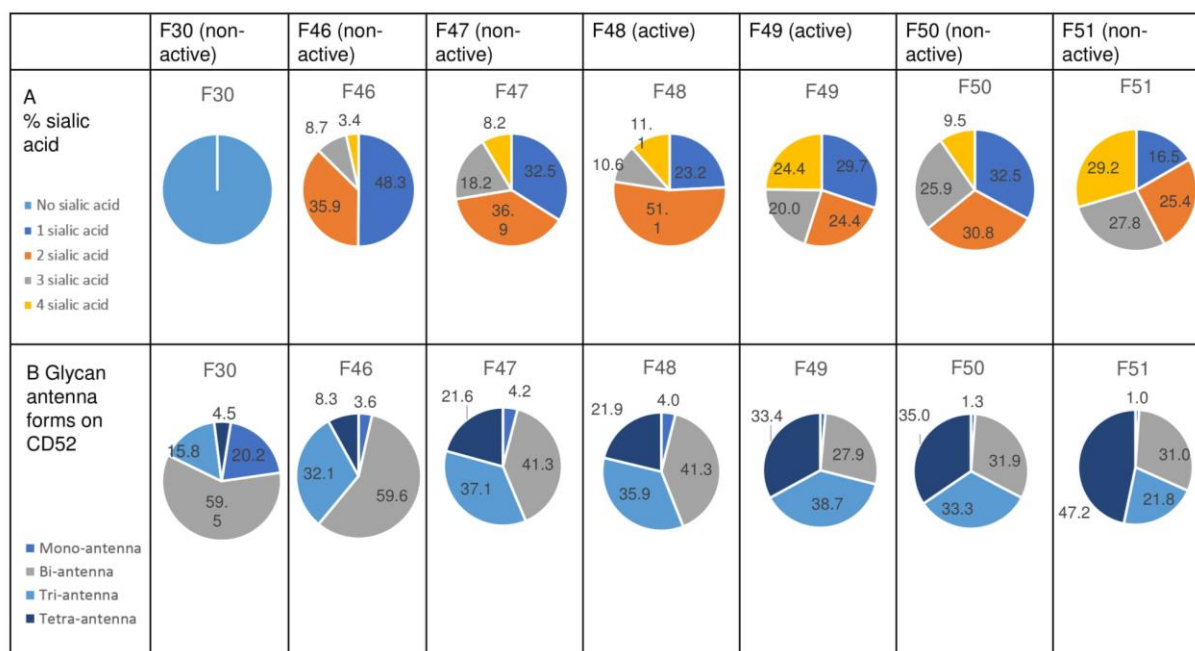
Supplementary Table I: Sialic acid content and antennae distribution of recombinant human CD52 fractions separated by anion chromatography. A, (upper panel) The total number of sialic acid residues and B, (lower panel) The antennae distribution identified on CD52 fractions (F30, F46, F47, F49, F50 and F51) using PGC-LC-MS/MS.


Supplementary Figure 3: High-resolution intact mass analysis of MonoQ fractions (F30 and F47-50). A, F30 intact mass analysis of the CD52 part showed absence of sialic acid molecules. B, MonoQ fractionation was able to separate CD52 sialylated structures according to their amount of sialic acid as well as number of antennae. Among fractions F47-50, F49 and F50 contained more of the bigger sialylated structures.

A

B
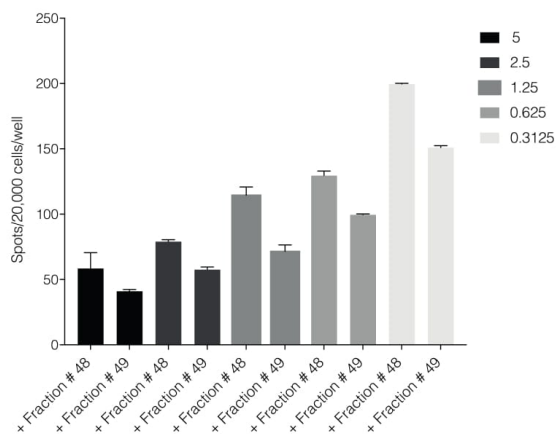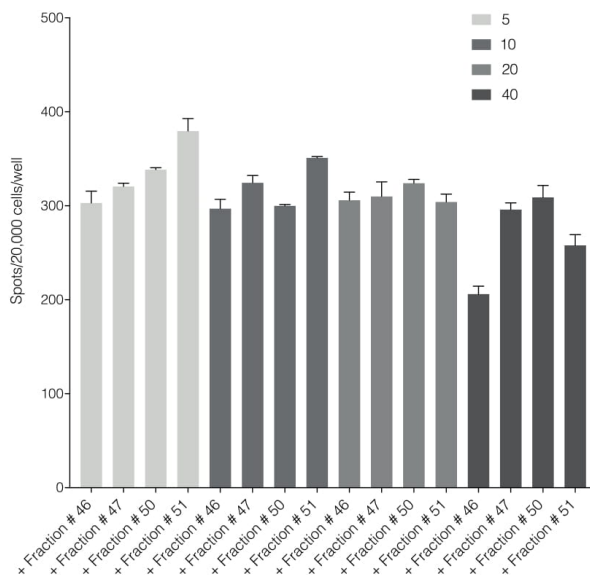
Supplementary Figure 4. Active MonoQ fractions suppress in a dose-dependent manner. A and B, IFN-γ production measured by ELISpot assay from human PBMCs (2×10^5) incubated in IP5 medium with no antigen or anti-CD3/CD28 antibody Dynabeads. A) Active Mono-Q fractions (F48-49) suppressed in a dose-dependent manner (0.3125, 0.625, 1.25, 2.5, and 5 µg/ml). B) Adjacent fractions (inactive; F46, F47, F50 and F51) do not suppress despite the increase of protein added (5, 10, 20 and 40 µg/ml). The data points in panels A and B are plotted as mean ± SEM of three independent replicates.

Active CD52 MonoQ fractions are enriched with α-2,3 sialylated structures

It is challenging to determine the sialylation linkages of large multi-sialylated *N*-glycans using mass spectrometry. Therefore, differences in sialic acid linkage of active and adjacent non-active MonoQ fractions were probed by α-2,3-specific sialidase treatment. The linkage-specific activity of α-2,3 sialidase was confirmed on bovine fetuin as a control protein, which demonstrated specific removal of α-2,3-linked sialic acid residues from the known bi-antennary sialylated glycan *m/z* 1111.5[2-] (Figure 5A). The glycan products resulting from α-2,3 sialidase treatment of the active fractions of CD52 were determined via PGC-LC-MS/MS (Figure 5B i and ii). The active MonoQ fractions (F48/F49) had a higher proportion of α-2,3 sialic acid (58%) compared to adjacent earlier (F46/F47) and later (F50/F51) eluting fractions (51% and 25%, respectively) and less bisecting structures than the adjacent non-active fractions (1% compared to 4% and 5%, respectively) (Figure 5C). Finally, the profile of the most active CD52 fractions at the intact peptide level supported a predominance of tri- and tetra-antennary sialylated structures (Figure 5D).
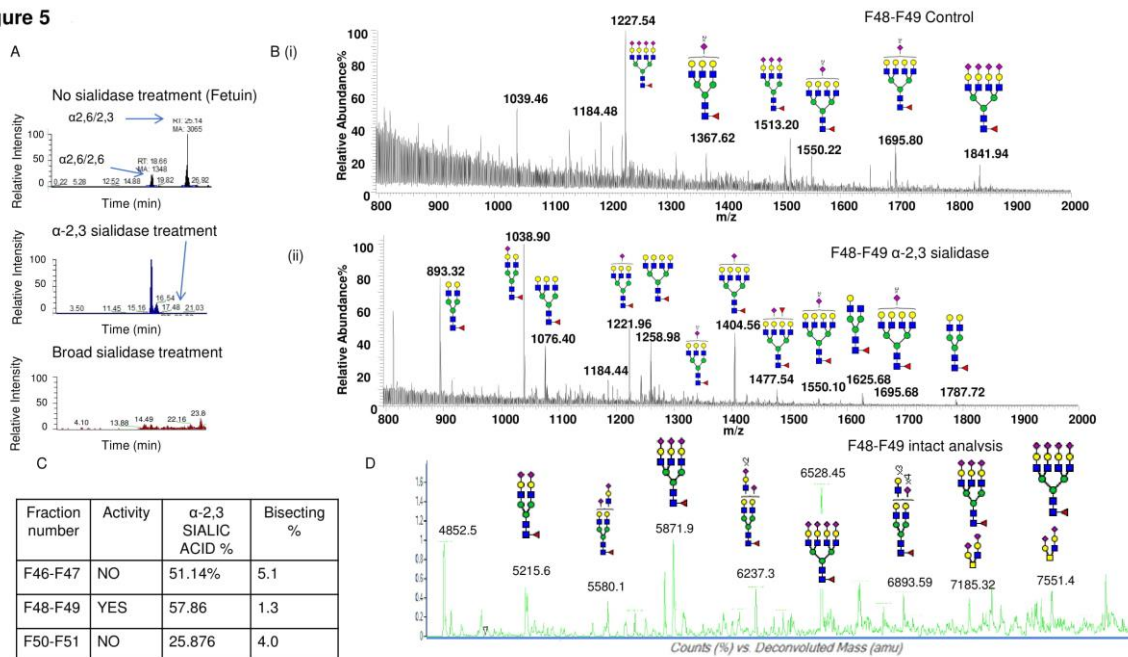

Figure 5: Sialic linkage analysis of active monoQ active fractions. A, EICs of the di-sialylated *N*-glycan *m/z* 1111.4[2-] after sequential α-2,3 sialidase treatment of bovine fetuin, known to carry tri-antennary α-2,3-sialylated *N*-glycans. The EICs assess the removal of each of the sialic acid residues. B, Summed MS of all *N*-glycans observed for the active CD52 fractions F48 and F49 before and after treatment with α-2,3-specific sialidase. C, Summary of the immunosuppressive bioactivity, degree of α-2,3 sialylation and bisecting GlcNAcylation of late-eluting MonoQ

218

fractions of particular interest. D, High-resolution intact mass analysis of the immune suppressive CD52 fractions (F48/F49).

The highly anionic MonoQ fractions are enriched in O-sialylated glycans

Initially, *O*-glycosylation analysis of de-*N*-glycosylated CD52 at the intact peptide level revealed that both variants of recombinant CD52 (CD52 I and CD52 II) had very low (4%) *O*-glycan occupancy (Figure 6A), casting doubt on the relevance of *O*-glycosylation for CD52 activity. Non-deamidated signatures were absent in the spectra for both CD52 I and II, indicating that the CD52 peptides were fully *N*-glycosylated (black data points). Like human spleen CD52, the recombinant CD52 proteins were found to contain mainly core type 2 *O*-glycans with one or two sialic acid residues, shown in grey and orange, respectively. Sialylated core type 1 *O*-glycans were also identified albeit at very low abundance (<0.5%) (Data not shown). Interestingly, the most anionic MonoQ fractions of CD52 (F46-F51) showed a considerably higher *O*-glycan occupancy (around 15-20%). Extracted ion chromatograms (EIC) of the bioactive fractions (F48/F49) showed an absence of sialo-isomers for the most abundant *O*-glycan structure m/z 665.2^{2-} ($GalNAc_1GlcNAc_1Gal_2NeuAc_2$), but not for m/z 1040.4^{1-} ($GalNAc_1GlcNAc_1Gal_2NeuAc$) (Figure 6B). Finally, *O*-glycan site localisation was determined by electron transfer/higher-energy collision dissociation (EThcD), which provided c and z ions to conclude that di-sialylated *O*-glycans were conjugated to Ser12 and possibly Ser10 due to lack of diagnostic ions, whereas the mono-sialylated *O*-glycans were only found on Thr8 (Figure 6C). Hence further experiments are desirable to elucidate the role of *O*-glycans for bioactivity of CD52.
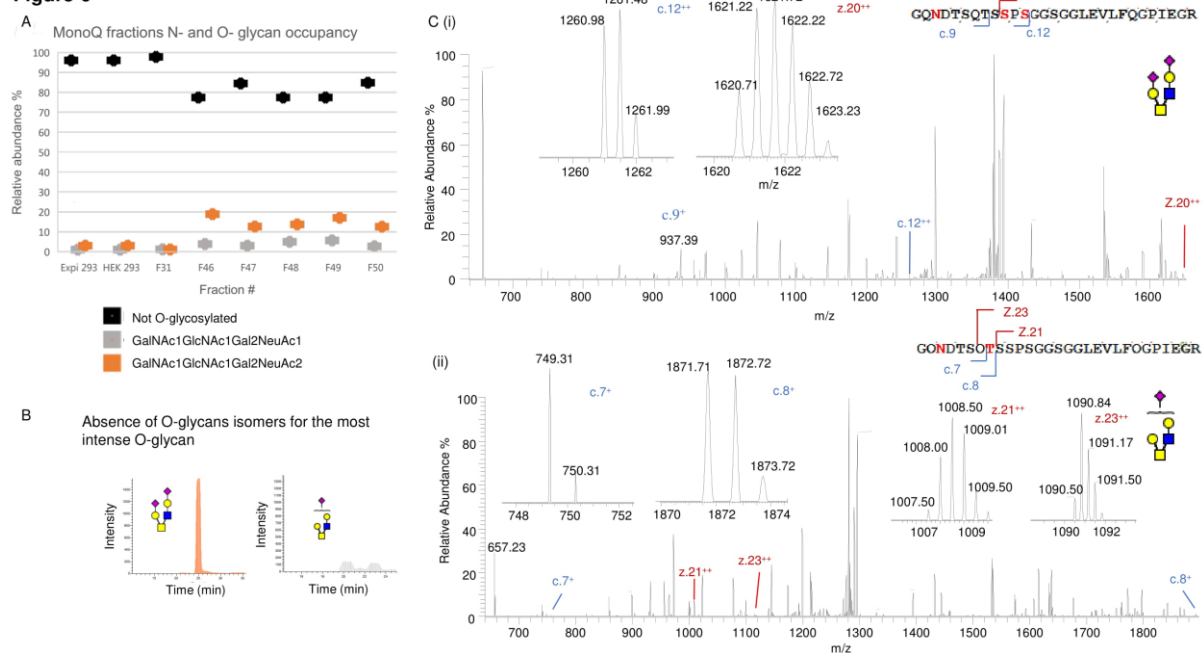
Figure 6: Mapping the *O*-glycosylation of recombinant human CD52. A, *N*- and *O*-glycan occupancy of CD52 I, CD52 II and selected MonoQ fractions (F31 and F46-F51) measured at the protein level after de-*N*-glycosylation. B, PGC resolution of *O*-glycosylated isomers from active fractions *m/z* 665.2^{2-} (GalNAc$_1$GlcNAc$_1$Gal$_2$NeuAc$_2$) and *m/z* 1040.4^{1-} (GalNAc$_1$GlcNAc$_1$Gal$_2$NeuAc$_1$). C, EThcD-MS/MS based site localisation analysis showing the peptide backbone fragments and the ions diagnostic of the amino acid site for both aforementioned *O*-glycans.

Discussion

In this study, we identified that CD52 from human spleen tissue and recombinant forms of human CD52 carry *N*-glycans that display complex type core fucosylation, abundant sialylation and LacNAc extensions. These features corroborate a previous report on the *N*-glycans on human spleen CD52 [8]. However, we found that there were differences in the specific bioactivity of two recombinant CD52-Fc variants made from different host cells. Only CD52 in the more bioactive variant of the CD52-Fc fusion protein was significantly abundant in tetra-sialylated *N*-glycan structures with α-2,3 sialic acid linkage. On the contrary, CD52 in the more bioactive CD52-Fc exhibited lower bisecting structures. Using Mono Q anion exchange chromatography, CD52-Fc was separated into a gradient of anionic glycoforms, some of which exhibited distinctly different immunosuppressive activities. The bioactive glycoforms uniquely displayed an abundance of tri- and tetra-sialylated glycans (60-70%), high levels of α-2,3 sialylation (58%) and an absence of

220

bisecting GlcNAcylation. Moreover, the most anionic tri- and tetra-sialylated glycans had a unique abundance in core type 2 di-sialylated *O*-glycan, at Ser 12, 20% compared to 4% in non-active fractions of CD52-Fc.

Both glycan and glycopeptide based analytical approaches were used to correlate the CD52 glycosylation with its bioactivity. The glycan approach depended on the high resolving power of PGC columns to separate glycan isomers and isobaric structures. It was used in conjunction with negative mode ionisation to provide fragment ions of certain glycan structural features [15, 18]. The glycopeptide-based approach allowed analysis of CD52 glycans directly bound to the peptide backbone with the assurance of no interference by Fc glycan. The two approaches largely corroborated each other adding confidence in the reported structures. Indeed, we found the same results after CD52-Fc purification by anion exchange chromatography, as described above. Anion exchange chromatography to fractionate sialylated glycoforms was previously employed to purify the soluble and sperm-associated form of CD52 in the mouse reproductive tract [20] but glycan structure was not analysed.

We confirmed the importance of the α-2,3 sialic acid linkage for CD52-Fc bioactivity. Previously, we showed that soluble CD52 mediates T-cell suppression by binding to Siglec-10 [1]. The diverse family of mostly inhibitory Siglec receptors has evolved to recognise linkage-specific sialic acid residues on host cells and pathogens [21]. Siglec-10 is highly expressed on leukocytes [22, 23] and plays significant roles in regulating the innate and adaptive immune response to tissue injury, sepsis and viral invasion [24]. Previously, Siglec-10 was found to have no binding preference for α-2,3 or α-2,6-sialylation [25]. However, we recently reported that human CD52-Fc bound to Siglec-10 preferentially through the α-2,3 sialic acid linkage [11]. In the present study, the bioactive CD52-Fc was characterized by a high abundance of the α-2,3 sialic acid linkage, and re-sialylation with α-2,3 restored the bioactivity of sialidase-treated CD52-Fc.

Regarding CD52 *O*-glycosylation, Ermini et al [26] deduced the presence of *O*-glycosylation of CD52 by antibody binding, but did not determine the type, occupancy or localisation of *O*-glycans. We characterised for the first time the *O*-glycans on human spleen CD52. In addition, bioactive recombinant CD52 was found to contain a low abundance (4%) of mainly core type 2 *O*-glycans with one or two sialic acid residues, on Ser 12 and Thr 8. Due to the proximity of the *N*- and *O*-glycosylation sites of CD52 peptide, the low degree of *O*-glycosylation could arise from steric hindrance by the bulky *N*-glycans that occupy most *N*-glycosylation sites. Determination of the *O*-glycan sites and occupancies on human spleen CD52 was challenging due to the availability of very limited biological sample amount. However, with continuing developments in highly sensitive and informative glycoprotemics [27], it is likely that we will soon be able to identify the site-specific *O*-glycosylation of CD52 directly from tissues and bodily fluids without prior purification. Our results also indicate that recombinant human CD52 does not require fucosylated *O*-glycans for bioactivity, as found for CD52 of the male reproductive tract (21). The polypeptide of recombinant human CD52 is identical to spleen CD52 and shares the core type 2 and core type 2 sialylated *O*-glycans with mrtCD52 [7]. However, we identified a dramatic enrichment of *O*-glycosylation in the bioactive CD52-Fc glycoforms strongly implying a role for both *N*- and *O*-glycosylation in the bioactivity of CD52.

Another striking observation was the association between CD52 bioactivity and bisecting GlcNAcylation. Previously, *N*-glycans displaying bisecting GlcNAc were found to correlate with a decrease in tri- and tetra-sialylated structures, since bisecting GlcNAc residues inhibit the activity of GlcNAc-transferases required to generate multi-antennary sialoglycans [28]. Furthermore, an increase in bisecting GlcNAcylation has been linked with a decrease in α-2,3 sialylation [29], which we have here shown is important for CD52 bioactivity. The functions of bisecting GlcNAc are not fully understood, but they have been associated with a decrease in target-cell susceptibility for NK cell-induced lysis [30]. However, the relatively small reduction in bisecting GlcNAc-containing CD52-Fc in active Mono Q fractions casts doubt on any direct role (e.g. Siglec or

HMGB1 binding) of this glycoepitope in the function of CD52. Interestingly, CD52 in recombinant human CD52-Fc resembled naturally-occurring CD52 purified from human spleen with respect to the *N*- and *O*-glycosylation, except in the degree of polyLacNAc elongation, which was greater in the native form. Although bioactive CD52 was characterized by higher abundance of sialylated structures with possible polyLacNAc, the contribution of polyLacNAc units to CD52 activity is yet to be determined.

In conclusion, the comparison of native and recombinant human CD52-Fc, and CD52-Fc variants differing in bioactivity, enabled us to identify a CD52 glycoform-specific for bioactivity. The characteristics of the bioactive CD52-Fc glycoforms can be summarised as an abundance of tri- and tetra-antennary α-2,3-sialylated *N*-glycans, a concomitant absence of bisecting GlcNAcylation and the presence of the di-sialylated type 2 *O*-glycosylation. However, further glycomic analysis of CD52 will be required to detail for example the length of polyLacNAc extensions and the degree of polyLacNAc branching. The present study extends our knowledge of the glycan structure underlying the immunosuppressive activity of soluble CD52 and may refine the design and production of CD52-Fc as an immunotherapeutic agent.

Acknowledgments

Conflict of interest statement

The authors declare they have no competing interests or other interests that might be perceived to influence the results and discussion reported in this paper.

6.3 References

1. Bandala-Sanchez, E., et al., *T cell regulation mediated by interaction of soluble CD52 with the inhibitory receptor Siglec-10.* Nature immunology, 2013. **14**(7): p. 741.
2. Pandya, I.J. and J. Cohen, *The leukocytic reaction of the human uterine cervix to spermatozoa.* Fertility and sterility, 1985. **43**(3): p. 417-421.
3. Thompson, L., et al., *The leukocytic reaction of the human uterine cervix.* American Journal of Reproductive Immunology, 1992. **28**(2): p. 85-89.
4. Ghaderi, D., et al., *Sexual selection by female immunity against paternal antigens can fix loss of function alleles.* Proceedings of the National Academy of Sciences, 2011: p. 201102302.
5. Robertson, S.A. and D.J. Sharkey, *Seminal fluid and fertility in women.* Fertility and sterility, 2016. **106**(3): p. 511-519.
6. Diekman, A.B., et al., *Evidence for a unique N-linked glycan associated with human infertility on sperm CD52: a candidate contraceptive vaccinogen.* Immunological reviews, 1999. **171**(1): p. 203-211.
7. Parry, S., et al., *The sperm agglutination antigen-1 (SAGA-1) glycoforms of CD52 are O-glycosylated.* Glycobiology, 2007. **17**(10): p. 1120-1126.
8. Treumann, A., et al., *Primary structure of CD52.* Journal of Biological Chemistry, 1995. **270**(11): p. 6088-6099.
9. Hasegawa, A., et al., *Epitope analysis for human sperm-immobilizing monoclonal antibodies, MAb H6-3C4, 1G12 and campath-1.* Molecular human reproduction, 2003. **9**(6): p. 337-343.
10. Schröter, S., et al., *Male-specific modification of human CD52.* Journal of Biological Chemistry, 1999. **274**(42): p. 29862-29873.
11. Bandala-Sanchez, E., et al., *CD52 glycan binds the proinflammatory B box of HMGB1 to engage the Siglec-10 receptor and suppress human T cell function.* Proceedings of the National Academy of Sciences, 2018: p. 201722056.
12. Xia, M.Q., et al., *Characterization of the CAMPATH-1 (CDw52) antigen: biochemical analysis and cDNA cloning reveal an unusually small peptide backbone.* European journal of immunology, 1991. **21**(7): p. 1677-1684.
13. Schmidt, T.G. and A. Skerra, *The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins.* Nature protocols, 2007. **2**(6): p. 1528.
14. Paulson, J.C. and G.N. Rogers, *[11] Resialylated erythrocytes for assessment of the specificity of sialyloligosaccharide binding proteins*, in *Methods in enzymology*1987, Elsevier. p. 162-168.
15. Jensen, P.H., et al., *Structural analysis of N-and O-glycans released from glycoproteins.* Nature protocols, 2012. **7**(7): p. 1299.
16. Cooper, C.A., E. Gasteiger, and N.H. Packer, *GlycoMod–a software tool for determining glycosylation compositions from mass spectrometric data.* Proteomics, 2001. **1**(2): p. 340-349.
17. Everest-Dass, A.V., et al., *Structural feature ions for distinguishing N-and O-linked glycan isomers by LC-ESI-IT MS/MS.* Journal of The American Society for Mass Spectrometry, 2013. **24**(6): p. 895-906.
18. Harvey, D.J., et al., *Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry.* Analytical biochemistry, 2008. **376**(1): p. 44-60.
19. Wang, W.-C. and R. Cummings, *The immobilized leukoagglutinin from the seeds of Maackia amurensis binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked alpha-2, 3 to penultimate galactose residues.* Journal of Biological Chemistry, 1988. **263**(10): p. 4576-4585.

20. Giovampaola, C.D., et al., *Surface of human sperm bears three differently charged CD52 forms, two of which remain stably bound to sperm after capacitation.* Molecular reproduction and development, 2001. **60**(1): p. 89-96.

21. Khatua, B., S. Roy, and C. Mandal, *Sialic acids siglec interaction: A unique strategy to circumvent innate immune response by pathogens.* The Indian journal of medical research, 2013. **138**(5): p. 648.

22. Li, N., et al., *Cloning and characterization of Siglec-10, a novel sialic acid binding member of the Ig superfamily, from human dendritic cells.* Journal of Biological Chemistry, 2001. **276**(30): p. 28106-28112.

23. Whitney, G., et al., *A new siglec family member, siglec-10, is expressed in cells of the immune system and has signaling properties similar to CD33.* The FEBS Journal, 2001. **268**(23): p. 6083-6096.

24. Chen, G.-Y., et al., *Siglec-G/10 in self–nonself discrimination of innate and adaptive immunity.* Glycobiology, 2014. **24**(9): p. 800-806.

25. MUNDAY, J., et al., *Identification, characterization and leucocyte expression of Siglec-10, a novel human sialic acid-binding receptor.* Biochemical Journal, 2001. **355**(2): p. 489-497.

26. Ermini, L., et al., *Different glycoforms of the human GPI-anchored antigen CD52 associate differently with lipid microdomains in leukocytes and sperm membranes.* Biochemical and biophysical research communications, 2005. **338**(2): p. 1275-1283.

27. Lee, L.Y., et al., *Toward automated N-glycopeptide identification in glycoproteomics.* Journal of proteome research, 2016. **15**(10): p. 3904-3915.

28. Schachter, H., *The joys of HexNAc. The synthesis and function of N-andO-glycan branches.* Glycoconjugate journal, 2000. **17**(7-9): p. 465-483.

29. Lu, J., et al., *Expression of N-acetylglucosaminyltransferase III suppresses α2, 3 sialylation and its distinctive functions in cell migration are attributed to α2, 6 sialylation levels.* Journal of Biological Chemistry, 2016: p. jbc. M115. 712836.

30. Yoshimura, M., et al., *Bisecting N-acetylglucosamine on K562 cells suppresses natural killer cytotoxicity and promotes spleen colonization.* Cancer research, 1996. **56**(2): p. 412-418.

Chapter 7

Summary and future directions
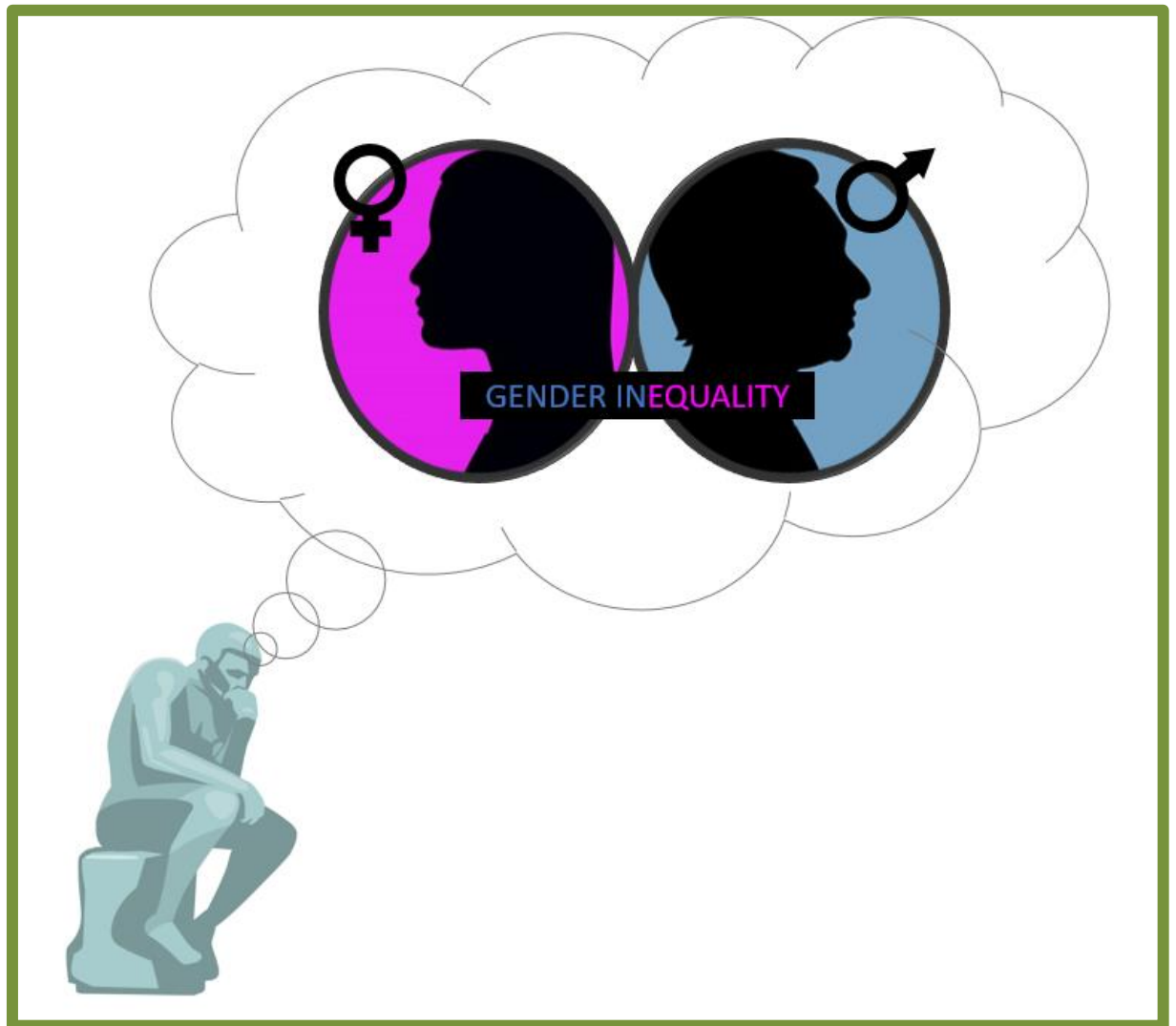
7.1 Infertility and gender inequality

Any disease of the reproductive process that results in inability to conceive and establish a viable pregnancy is referred to as infertility. Infertility is a global health issue with a significant overall burden that has not displayed any decrease over the last 20 years [1]. According to available data, this burden traumatizes at least 50 million couples, 186 million people, worldwide [2, 3]. Male, female or combinations of both are usually the causative agents for infertility. Female fertility can be affected by ovulation disorders, blocked oviduct, endometriosis and others [4]. Although most research in this area is concerned with women, infertility is also prevalent in men. Male infertility provides more than half of all cases, yet infertility remains a women's social burden [2]. In fact, from 1973 to 2011, there was a sharp decline of more than 50% in total sperm count and concentration among men from Western countries, with reports showing similar decline for data 2011- 2017 [5].

"When I talk to people about infertility, they tend to automatically assume that it is a woman's problem and are genuinely surprised to find out that it's pretty much 50-50" Clare Brown, (the chief executive of the Infertility Network, United Kingdom).

*****

The global data on male infertility is still lacking worldwide – both in terms of men's experience across cultures and prevalence estimates - contributing gender-related stigma. As a result, infertility requires gender unbiased research [6].

"We want to end gender inequality- and to do that we need everyone to be involved" Emma Watson, (UN Women Goodwill Ambassador).

*****

We did our part by designating equal number of Chapters for each gender! From the glycosylation perspectives, our thesis investigated the challenges diabetic women face regarding their reproductive organs in Chapters 3 and 4, while sperm viability and resistance to female immunity was investigated in Chapters 5 and 6 respectively.

7.2 Thesis overview

The importance of protein glycosylation in several biological processes is highlighted by its involvement in cellular activity, cell-cell interaction, and protein folding and interaction [3]. As a result, any aberrant glycosylation, in abundance or type, can influence the overall physiology of tissues and the functions of proteins [7]. The two main types of mammalian protein glycosylation, N-linked and O-GalNac linked, are decorated with functional determinants usually consisting of sialic acid, fucose, galactose, N-acetylglucosamine and N-acetylgalactosamine building blocks, and are often altered in disease states [8, 9]. Overall, the physiological and the interactive status of the cells are influenced by the global glycosylation observed at the macro-cellular level [10, 11], whereas protein properties such as function, size, solubility, hydrophobicity and stability are affected by aberrant glycosylation at the micro, single-protein level [7, 12].

In this thesis, for the relevance of infertility and possibility of aiding assisted reproductive technology (ART), we wanted to investigate the role of protein glycosylation at the macro (tissue) and micro (single protein) levels and in both female (with effect of diabetes) and male (with effects on immunity of sperm) reproductive tracts.

7.2.1 Protein glycosylation in the female reproductive tract and the effect of diabetes in the ovary, oviduct and uterus

In Chapter 3, the macro-glycosylation of ovaries was investigated in diabetic mice as a relevant model. Diabetes was chosen as an investigation model for its unfavourable effect(s) on fertility as well as the known effect of diabetes on glycosylation [13-15]. Our comprehensive protein N- and O-glycosylation analysis of two different diabetic models (STZ -induced diabetic = type 1 diabetes, and high-glucosamine injected = type 2 diabetes) showed similarities in their ovary glycosylation profiles, with some differences attributed to the different types of diabetes models,

described in Chapter 3. The two diabetic models shared a relative decrease in oligomannose, and a decrease in sialylated, structures compared to control. The differences in protein glycosylation were observed as a decrease in core-fucosylated structures in the STZ-induced diabetic ovary which could be attributed to the confirmed inflamed state of the STZ treated mice [16], the latter proven by qPCR analysis of known inflammatory gene expression (TNF and IL1B). Our results suggest changes occur in the ovarian glycan biosynthesis pathway due to diabetes. Overall, we are the first to report aberrant protein glycosylation in any diabetic female reproductive organs, where our study showed global glycome changes with major possible effects on ovary functions (macro glycosylation). Also, to the best of our knowledge, our study provides the first evidence for differential protein glycosylation profiles generated in different diabetes models in the same organ; highlighting the complexity of each diabetes type and their effect on protein glycosylation.

The data presented in Chapter 4 aimed to provide a comprehensive membrane-proteome and glycome view of the oviduct and uterine surfaces during the uterine window of implantation (WOI), as well as highlighting changes due to STZ- induced diabetes in mice. Our interest in STZ-induced diabetes was due to the fact that non-obese mice type 1 mice have significantly higher loss of embryo compared to normal with the cause thought to be due mainly to uterine receptivity [17]. In addition, at the time of WOI, there are changes in membrane protein biological processes as well as glycan expression in both oviduct and uterine tissues. The receptive uterine membrane-proteome showed a decrease in oxidative phosphorylation that was shown to be significantly increased in response to diabetes. Furthermore, one cell adhesion molecule (IntegrinB3, responsible for temporal adhesiveness between endometrium and embryo [18]) and two growth factors (EGFR and PDGFR that are regulators of protein expression in WOI in human [19]) exhibited a significant increase in the receptive non-diabetic uterine surface, with the growth factors significantly reduced in diabetic receptive uterus. These growth factors (especially PDGFR) have been linked with differential expression of cytokines required for a normal blastocyst implantation process such as leukemia inhibitory factor (LIF).

Since glycoconjugates play an important role in blastocyst uterine implantation, especially the highly glycosylated mucin layer found on the surface of oviduct and uterine tissues (MUC1), an optimized protocol of mucin analysis (PVDF blotting), was used to profile the surface *O*-glycoprofile of diabetic oviduct and uterine surfaces. These exhibited no difference in individual *O*-glycan structures due to WOI or diabetes but the carbohydrate/protein abundance increased on MUC1 on the uterine lining during the receptive window, associated with an increase in the number of *O*-glycosylated sites on the expressed mucins. This is the first account of the protein glycosylation of the embryo implantation process as well as the influence of diabetes on the blastocyst journey using MS based technologies.

Overall, the data presented in Chapters 3 and 4 of this thesis highlighted changes in glycosylation that are seen in the female reproductive tract of diabetic mice models.

7.2.2 Protein glycosylation in male reproductive tract: governance of sperm individuality and immunity, the latter via CD52.

In the Chapters 5 and 6, for equality purposes and for the fact that 50% of infertility causes are due to male reproductive issues [2], we tested the possibility of utilising sperm protein glycosylation to aid in IVF selection for their capacity for fertilization. Glycosylation measure is very promising as a measure of sperm quality due to its relevance in sperm functionality [20] and the fact that all current measures of sperm quality have failed [21-23].

In Chapter 5, detailed *N*-glycosylation analysis of the sperm surface was tested as a means to compare and differentiate active and inactive sperm generated via two main isolation protocols, namely Percoll density gradient centrifugation and "swim-up" collection. In our analyses, active sperm capable of fertilization exhibited a trend towards a decrease in sialylation and an increase in paucimannose structures compared to inactive sperm, which could possibly associate a specific *N*-glycome with sperm efficiency for IVF application. Above all, the large individual biological variation in human sperm *N*-glycome was the most significant observation in our work, which complicated our comparisons between active and inactive sperm. Finally, our results suggest the

need for a larger sample cohort needed to exploit the opportunity of utilising the *N*-glycome as a measure of individual male fertility.

The data presented in Chapter 6 of this thesis suggests that glycosylation of a protein has a significant impact on its function. Specifically, Cluster of differentiation 52(CD52) protein was subjected to comprehensive *N*- and *O*-glycan analysis. CD52 has a proposed role in sperm evading female immunity inside the female reproductive tract [20] but only lymphocyte CD52 has been fully characterised from the activity point of view [24]. To correlate glycosylation with immunosuppressive activity, lymphocyte CD52 was analysed in the work reported in this Chapter. In our analysis, we successfully described the most active immunosuppressive glycoforms of lymphocyte CD52. These particular findings will feed into the CD52 drug discovery pipeline developments.

Overall, the data presented in Chapters 5 and 6 of this thesis identifies that protein glycosylation has the potential to be used as a means to differentiate active sperm, and to identify the glycoforms associated with suppression of the female immune system.

7.3 Summary

The aim of this thesis was to enhance our understanding of the role of glycosylation in the reproductive process in both male and female tracts from both a glycomics and glycoproteomics perspective. However, at this stage it is unknown whether the reported glycan changes in the diabetic state are the cause or the result. Also, the direct relationship between the observed aberrant diabetic ovarian glycosylation and ovarian function is yet to be determined. The same level of analysis (macro-glycosylation) was investigated as a means of differentiating active sperm from inactive sperm. The most significant result was the large individual biological variation of each donor's sperm. Further work is needed to correlate the glycosylation profile of individual sperm donors with their fertility status On the other hand, the micro-glycosylation (at the single glycoprotein level) focused on two functionally relevant reproductive proteins, namely: endometrial MUC1 and seminal Cluster of Differentiation 52 (CD52). MUC1 showed no

significant difference in *O*-glycan structures at the time of window of implantation (WOI) or in diabetes, although the carbohydrate/protein abundance increased on MUC1 on the uterine lining in the WOI. CD52 has a proposed immune-suppressive function in semen and as a prelude to understanding the role of CD52 glycosylation in reproduction we showed sialylation was essential for the immunosuppression and identified the most immune suppressive CD52 glycoforms. Overall, this thesis provides support for the inherent contribution of protein glycosylation, on a macro and micro protein level, in the process of reproduction in both males and females.

7.4 Future directions and potential applications

This thesis has provided a detailed representation of the involvement of protein glycosylation in both males and female reproduction at the macro and micro levels. We would like to provide some ideas for future investigations of the role of glycans in female and male reproductive tracts, Regarding the significant correlation between diabetes and aberrant glycosylation in ovary, future studies can focus on glycoproteomic analysis for the identification of the proteins that carry the aberrant glycans detected in our study. This kind of approach would provide invaluable information on the glycoproteome of diabetic ovary and help in the identification of glycome changes at the single glycoprotein level, and may help in the discovery of new prognostic markers for better disease prediction and diabetic patient treatment strategies. In particular, probing the inflamed state of the ovary of the high-glucosamine injected mice (type 2 diabetes) will help correlate the decrease in core fucosylation as a direct response from STZ- induced inflamed state (type 1 diabetes). This link between decrease in core-fucosylation and type 1 diabetes can be further investigated by treating ovarian cell lines with TNF-α and IL1B to relate changes in glycosylation (possibly a decrease in core fucosylation) to the expression of these specific pro-inflammatory cytokines. Potentially, the significant correlation between diabetes and aberrant protein glycosylation in the two diabetic mice ovarian models as well as the role of protein glycosylation in the overall tissue/ovary functions suggest that the observed glyco-signatures of relative decrease in oligomannose and sialylated structures, as tools to evaluate the ovary's overall

functional status. These glyco- signatures have the potential to be used to predict diabetic status in ovarian tissues, which can be proven supportive of various techniques used in assisted reproductive technology (ART) such as ovulation and oocyte picking.

Regarding the rest of the reproductive organs, proteomic analysis of membranes of oviduct and uterine can be further validated by the use of more biological replicates. Another suggestion is fine tuning of how oxidative phosphorylation and vesicle mediated transport are involved and differentially expressed in the WOI and under the diabetic state in the blastocyst journey, in order to determine the fine scale protein effects using targeted mass spectrometry (MRM) or western-blot approach. Furthermore, our analysis on uterine scrapings shows a significant increase in carbohydrate/protein abundance which has direct implications on the uterine hydrophilicity, embryo implantation and overall functions. Hence, our results show the importance of studying the mucin glycosylation in addition to mucin protein expression as previous studies have addressed. In fact, the *O*-glycans displayed on the mucin surface present a large variety of carbohydrate epitopes that can act as binding ligands on the surface of oviduct and uterus [2][25]. Potentially, analysis on the type of mucin *O*-glycans presented on MUC1 of infertile women may enable the use of these molecules as potential contraceptive agents.

In regard to male reproduction, the notion to correlate sperm activity with specific glycosylation through the female reproductive tract has been studied previously [21-23]. In our study the use of glycosylation as a means to differentiate active sperm made it clear that knowing the history of individual donors may enable the correlation between donor fertility and their glycosylation profiles. The use of a larger sample cohort from individuals with normalised age and health status would further elucidate glycosylation trends found in our comparison between active and inactive sperm including the decrease in sialylation and increase in paucimannose structures expressed on the sperm membrane. Once validated, lectins, by their ability to bind specific glycan epitopes such as sialic acid (Wheat germ agglutinin binds-WGA binds to sialic acid) and mannose (Mannose-binding lectin-MBL binds to mannose), could be used to enrich active sperm from inactive ones. Furthermore, collecting sperm samples from the same donor over multiple time points would

emphasise any biological variation exhibited in the same man. Potentially, our results show the importance of considering individual variation in future fertility experiments, especially those involved in sperm activity.

In vivo experimentation with recombinant CD52-Fc, especially using the most active CD52-Fc MonoQ fractions found in this study, will further our quest towards human trials for testing CD52 as an immunosuppressive drug that can be prescribed to patients with autoimmune diseases such as sepsis and type 1 diabetes [26][24]. Our finding that there are active CD52 glycoforms has enhanced our understanding of the requirement for specific glycan structures for bioactivity and will help to improve the recombinant protein production of a specific CD52 glycoform that may be a better immunotherapeutic agent against autoimmune diseases such as type 1 diabetes and sepsis [24]. Within the reproductive scope, this analysis of the importance of glycosylation for lymphocyte CD52 activity has paved the way to future investigations on the activity of glycosylated seminal CD52 in supressing female immunity. Lymphocyte CD52 has been found to initiate its immunosuppressive activity via binding to SIGLEC-10 [24]t little is known about seminal CD52 from the point view of activity. Researching seminal CD52 would require further understanding of its binding partner inside the female reproductive tract and an assay to test for the most active CD52 glycoforms.

In general, the approach we used here; the generation of recombinant glycoproteins from different host cells, correlation between activity and glycosylation status using PGC-LC-MS/MS and C8-LC-MS analyses and the purification of the active components by the use of different types of liquid chromatography such as ion exchange or size exclusion chromatography together with an assay to measure biological activity, are steps that can be used for the identification of the most active glycoforms of any glycoprotein protein of interest.

7.5 Conclusions

In the female reproductive tract, diabetes is associated with many detrimental outcomes including: abnormal ovarian function, irregularity in the expression of glycosylated proteins found in the

uterine lining, and overall lower chances of fertility. In this thesis, we show the established aberrant glycosylation in several diabetic organs is also extended to reproductive organs which points to the need to investigate the glycan portion of proteins in studies involving diabetic reproductive organs. In fact, many glycoproteins are involved in the process of ovulation and embryo implantation, all of which could be functionally altered in diabetes. However, the extent of diabetes alteration on protein glycosylation structures does not necessarily affect all glycoproteins expressed in the female reproductive tract including endometrial mucin (MUC1) although the abundance of glycosylation was seen to increase. However, as previously stated in Chapter 4, human and mice exhibit a differential MUC1 response to the window of implantation. Unlike human, receptive mice uterus display a decrease in MUC1 expression [27-31]. As a result, uterine receptivity remains an elusive process and further analyses are required to decipher this challenging phenomenon.

We also correlated sperm activity with trends in protein *N*-glycosylation confirming the involvement of protein glycosylation in the male reproductive tract and sperm activity. Above that, sialylation was found to be essential for the innate and adaptive immunosuppressive activity of lymphocyte CD52.Seminal CD52 is also decorated with heavy sialylation inside the female reproductive tract which could potentially enable the sperm survival against the female immune system.

The work in this thesis points out to the need to shift the typical focus from genome and proteome into the glycome in order to answer relevant biological questions in the scope of reproduction.

7.6   References

1. Inhorn, M. and F. Van Balen, *Infertility around the globe: New thinking on childlessness, gender, and reproductive technologies*2002: Univ of California Press.

2. Inhorn, M.C. and P. Patrizio, *Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century.* Human reproduction update, 2015. **21**(4): p. 411-426.

3. Daar, A.S. and Z. Merali, *Infertility and social suffering: the case of ART in developing countries.* Current practices and controversies in assisted reproduction, 2002: p. 15-21.

4. Healy, D.L., A.O. Trounson, and A.N. Andersen, *Female infertility: causes and treatment.* The Lancet, 1994. **343**(8912): p. 1539-1544.

5. Levine, H., et al., *Temporal trends in sperm count: a systematic review and meta-regression analysis.* Human reproduction update, 2017. **23**(6): p. 646-659.

6. Rouchou, B., *Consequences of infertility in developing countries.* Perspectives in public health, 2013. **133**(3): p. 174-179.

7. Dwek, R.A., *Glycobiology: toward understanding the function of sugars.* Chemical reviews, 1996. **96**(2): p. 683-720.

8. Varki, A. and J.B. Lowe, *Biological roles of glycans, in Essentials of Glycobiology, A.* Varki, et al., Editors. 2017: Cold spring Harbor (NY).

9. Rini, J.M. and J.D. Esko, *Glycosyltransferases and glycan-processing enzymes, IN Essenrial of Glycobiology,* A. Varki, et al, Editors. 2017: Cold Spring Harbor (NY).

10. Arnold, J.N., et al., *The impact of glycosylation on the biological function and structure of human immunoglobulins.* Annu. Rev. Immunol., 2007. **25**: p. 21-50.

11. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum.* Annual review of biochemistry, 2004. **73**(1): p. 1019-1049.

12. Bieberich, E., *Synthesis, processing, and function of N-glycans in N-glycoproteins*, in *Glycobiology of the Nervous System*2014, Springer. p. 47-70.

13. Platt, M., et al., *St Vincent's Declaration 10 years on: outcomes of diabetic pregnancies.* Diabetic Medicine, 2002. **19**(3): p. 216-220.

14. Garcia-Vargas, L., et al., *Gestational diabetes and the offspring: implications in the development of the cardiorenal metabolic syndrome in offspring.* Cardiorenal medicine, 2012. **2**(2): p. 134-142.

15. Everest-Dass, A.V., et al., *Human disease glycomics: technology advances enabling protein glycosylation analysis–Part 2.* Expert review of proteomics, 2018. **15**(4): p. 341-352.

16. Okada, M., et al., *Blockage of core fucosylation reduces cell-surface expression of PD-1 and promotes anti-tumor immune responses of T cells.* Cell reports, 2017. **20**(5): p. 1017-1028.

17. Albaghdadi, A.J. and F.W. Kan, *Endometrial receptivity defects and impaired implantation in diabetic NOD mice.* Biology of reproduction, 2012. **87**(2).

18. Dominguez, F., et al., *Proteomic analysis of the human receptive versus non-receptive endometrium using differential in-gel electrophoresis and MALDI-MS unveils stathmin 1 and annexin A2 as differentially regulated.* Human reproduction, 2009. **24**(10): p. 2607-2617.

19. Chen, J.I., et al., *Proteomic characterization of midproliferative and midsecretory human endometrium.* Journal of proteome research, 2009. **8**(4): p. 2032-2044.

20. Tecle, E. and P. Gagneux, *Sugar-coated sperm: Unraveling the functions of the mammalian sperm glycocalyx.* Molecular reproduction and development, 2015. **82**(9): p. 635-650.

21. Guzick, D.S., et al., *Sperm morphology, motility, and concentration in fertile and infertile men.* New England Journal of Medicine, 2001. **345**(19): p. 1388-1393.

22. Poland, M.L., et al., *Variation of semen measures within normal men.* Fertility and sterility, 1985. **44**(3): p. 396-400.

23. Castilla, J., et al., *Influence of analytical and biological variation on the clinical interpretation of seminal parameters.* Human reproduction, 2005. **21**(4): p. 847-851.

24. Bandala-Sanchez, E., et al., *T cell regulation mediated by interaction of soluble CD52 with the inhibitory receptor Siglec-10.* Nature immunology, 2013. **14**(7): p. 741.

25. Naughton, J., et al., *Interaction of microbes with mucus and mucins: recent developments.* Gut Microbes, 2014. **5**(1): p. 48-52.

26. Rashidi, M., et al., *CD52 inhibits Toll-like receptor activation of NF-κB and triggers apoptosis to suppress inflammation.* Cell death and differentiation, 2018. **25**(2): p. 392.

27. Braga, V. and S.J. Gendler, *Modulation of Muc-1 mucin expression in the mouse uterus during the estrus cycle, early pregnancy and placentation.* Journal of cell science, 1993. **105**(2): p. 397-405.

28. Surveyor, G.A., et al., *Expression and steroid hormonal control of Muc-1 in the mouse uterus.* Endocrinology, 1995. **136**(8): p. 3639-3647.

29. Carson, D.D., M.M. Desouza, and E.G.C. Regisford, *Mucin and proteoglycan functions in embryo implantation.* Bioessays, 1998. **20**(7): p. 577-583.

30. Bowen, J.A., F.W. Bazer, and R.C. Burghardt, *Spatial and temporal analyses of integrin and Muc-1 expression in porcine uterine epithelium and trophectoderm in vivo.* Biology of reproduction, 1996. **55**(5): p. 1098-1106.

31. Hey, N.A., et al., *The polymorphic epithelial mucin MUC1 in human endometrium is regulated with maximal expression in the implantation phase.* The Journal of Clinical Endocrinology & Metabolism, 1994. **78**(2): p. 337-342.