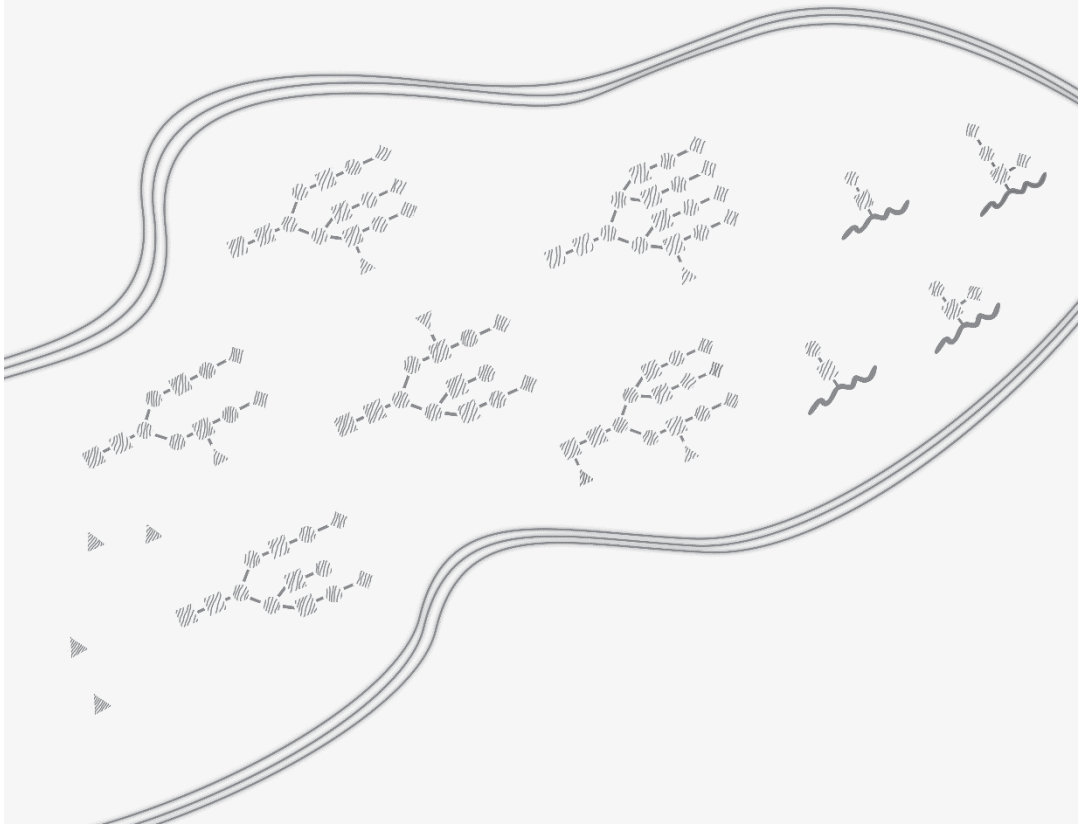
Biomarker discovery in diabetes mellitus and lipid metabolism: multi-platform glyco(proteo)mic approaches

Demus, D.A.

CHAPTER 1



1. Introduction

Glycosylation

Glycosylation is a co-/post-translational modification of proteins or lipids that occurs in the endoplasmic reticulum and Golgi apparatus of eukaryotic cells. It involves the enzymatic addition of sugar molecules (glycans) to specific sites of these molecules. Glycans are built from single monosaccharide units that combine through glycosidic bonds to form larger carbohydrates. The formation of glycans is mediated by specific enzymes – glycosyltransferases - which catalyse the synthesis of glycosidic linkages. These enzymes are classified into different families based on their amino acid sequence and the type of linkage they form between the donor and acceptor molecules. Some of the most important human glycosyltransferases belong to the families of *O*-linked glycosyltransferases (OGTs), *O*-linked *N*-acetylglucosamine (GlcNAc) and *O*-linked *N*-acetylgalactosamine (GalNAc) transferases, that catalyse the addition of a single GlcNAc and GalNAc, respectively, to the hydroxyl group of serine or threonine residues in proteins, forming *O*-linked glycosylation. Enzymes that belong to the families of *N*-linked glycosyltransferases (NGTs) form *N*-linked glycosylation by transferring an *N*-glycan via its reducing-end GlcNAc to the amide group of asparagine residues in proteins. These glycosyltransferases initiate the synthesis of two distinctive types of carbohydrate molecules: *N*-glycans and *O*-glycans[1].

N-glycans are complex, branched carbohydrate molecules that consist of a core structure of two GlcNAc and three mannose (Man) monosaccharide units. This core structure is typically further extended by the addition of various monosaccharides, including galactose, GlcNAc, *N*-acetylneuraminic acid, and fucose, resulting in a wide range of *N*-glycan structures with varying sizes and degrees of branching.

Unlike *N*-glycans, which have a conserved branched structure, *O*-glycans are highly diverse in terms of their size, composition, and branching. The biosynthesis of *O*-glycans begins with the addition of a single GalNAc residue, although other monosaccharides such as GlcNAc can also serve as initial substrates. The GalNAc

residue can then be further extended by the action of specific glycosyltransferases and glycosidases, resulting in a diverse array of core *O*-glycan structures.

The presence of *N*-glycan and *O*-glycan variants on proteins plays important roles in protein folding, secretion, stability and function. Glycoproteins on cell surface are involved in cell-cell recognition, signalling and adhesion, thus are involved in most biological processes[2]. Of note, adjusting a composition of glycan structures has found the application in biologics development to manipulate the stability and potency of therapeutic molecules[3].

Glycoproteins

Glycans are present on most human proteins and nearly all blood plasma proteins are known to be glycosylated[4]. Although there is a broad variety of glycoproteins circulating in the body, the most common classes are *N*-linked and *O*-linked glycoproteins, and glycoproteins containing multiple *N*- and *O*-glycosylation sites. The composition of glycans, which are attached to these proteins, generates proteins diversity and modulates their properties[5].

A common example of a glycoprotein whose functional characteristics are defined by *N*-glycan composition changes is a human immunoglobulin G (IgG). Afucosylated IgG is a glycoprotein that contains *N*-glycan structure in the IgG-Fc domain which lacks a fucose residue linked α1,6 on the first GlcNAc in the core. These afucosylated IgG molecules present the enhanced interaction of IgG with FcγRIII and thereby their activity. This characteristic has been broadly recognised in the biopharma industry and utilised to produce therapeutic antibodies with low core fucose content, which promotes the elevated killing activity through antibody-dependent cellular cytotoxicity mechanism of the immune system[6].

Acute phase proteins (APPs) are a major group of blood plasma proteins that undergo quantitative or qualitative changes during various pathophysiological conditions. Many of these APPs are glycosylated and it has been shown that glycosylation patterns can

be altered in pathophysiological conditions including acute inflammation, chronic inflammation (diabetes mellitus, inflammatory bowel disease), cancer (pancreatic, ovarian, lung, prostate)[7]. These glycosylation alterations correlate with disease severity in certain conditions and have been investigated as potential biomarkers. The most commonly observed APP glycosylation changes are alterations of the extent and type of fucosylation, sialylation and the number of antennae[8].

Apolipoproteins are protein components of lipoproteins - lipid-protein complexes that carry lipids in circulation, classified based on the lipid and apolipoprotein composition - that play a role in lipoprotein assembly, lipid transport, and receptor recognition[9]. Human plasma apolipoproteins, designated as A, B, (a), C, D, E, F, H, J, L, M, and O, can be classified into two categories: non-exchangeable, including ApoB-100, ApoB-48, and Apo(a), and exchangeable apolipoproteins having the ability to be transferred between different lipoprotein particle classes. Among them, apolipoproteins A, B, C, and L exhibit multiple subtypes. Glycosylation is a prevalent modification observed in nearly all human apolipoproteins and their aberrant glycosylation patterns have been linked to various diseases including metabolic syndrome[10], cancer[11], type 2 diabetes[12], autoimmune disease[13], and neurological disorders[14]. Glycosylation exerts diverse modulatory effects on apolipoprotein function, including ER-Golgi trafficking, lipoprotein assembly, stability, receptor binding and enzyme activity. For instance, the activity of lipoprotein lipase (LPL), an enzyme that degrades circulating triglycerides, is affected by the absence of *N*-glycans on apoD and the presence of sialylated *O*-glycans on apoC-III[15].

Glycosylation analysis

Studying and characterizing structures and compositions of glycans that are attached to protein or lipids is possible by applying several analytical methods. Several types of ultra-high performance liquid chromatography (UHPLC) separation techniques are used to separate and identify glycans and glycoproteins. Enzymatically or chemically released and fluorescent-labelled glycans are required for UHPLC-based analysis. The

most commonly used fluorescent tags used for glycan labelling are 2-aminobenzoic acid (2-AA), 2-aminobenzamide (2-AB) and procainamide that allow detection and relative quantification of chromatographically separated glycoforms[16, 17]. Due to the hydrophilic nature as well as size and charge variation, glycans can be separated using several molecule-stationary phase interaction mechanisms employed in UHPLC analytical columns. Hydrophilic interaction liquid chromatography (HILIC), which provides superior separation and reproducibility, has become the gold standard separation technique applied to released glycans and been applied in the majority of large-scale studies researching glycosylation changes by UHPLC methods[18, 19]. Nevertheless, reverse phase (RP)[20] and porous graphitised carbon (PGC)[21] have been proven applicable in UHPLC-based analysis of released glycans. Identification of fluorescent-labelled and chromatography separated glycans is possible by applying commercially-available glycan standards. However, UHPLC systems coupled with mass spectrometry (MS) or tandem mass spectrometry (MS/MS) detection - often supported by exoglycosidase digestions - enable the identification of rare glycan structures and overcome common hurdles of the released glycans analysis by UHPLC, such as co-eluting glycan species[22]. Similarly to the analysis of released glycans, UHPLC systems coupled with MS/MS detection can be applied to separate and identify glycoproteins. Glycoproteins can be analysed by RP, size-exclusion chromatography (SEC) and mixed-mode separation techniques.

Upon chromatographic separation, the composition of either glycan or glycoprotein samples can be analysed by electrospray ionization (ESI) MS that involves ionization of the analytes in a volatile solution. Electrospray occurs as the analyte solution is introduced into the mass spectrometer via a spray needle in the strong electric field by the application of a high voltage (2–6 kV), which causes the dispersion of the sample solution into highly charged electrospray (ES) droplets. As the solvent evaporates from the droplets in the presence of nitrogen (drying gas), multiple charged ions are formed which are transferred into the vacuum. Depending on the type of mass spectrometer used, ions with different mass-to-charge ratios (*m/z*) are then separated using a

combination of electric and magnetic fields and detected. ESI is a suitable technique for the MS analysis of glycans and glycoproteins as this “soft” ionization type does not cause significant fragmentation of these biomolecules. Identification of analytes is possible by studying their m/z values and ion fragmentation patterns, in cases when the parent ion fragmentation is applied. Different fragmentation types are available, with either collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD) and electron-transfer dissociation (ETD) combinations being the most suitable for the analysis of glycans and glycoproteins[23].

Matrix-assisted laser desorption/ionization (MALDI) MS is a commonly used mass spectrometry technique for the analysis of glycans and glycoproteins. In MALDI, a matrix, which is added to a sample, absorbs the laser energy and helps to desorb and ionize the analyte molecules. MALDI is considered a “soft” ionization technique, which in most cases, allows ionization of analytes without fragmenting or decomposing them[24]. However, the labile nature of sialic acids (SA) that are present on glycans makes MALDI analysis of glycans and glycoproteins challenging. Due to the loss of sialic acids upon MALDI, several SA derivatization (stabilization) methods have been developed for the analysis of released glycans[25, 26]. Although the application of SA derivatization methods compromise some of the benefits of MALDI analysis, such as the simplicity of sample preparation, it brings the advantage of distinguishing between SA linkages through which these residues are attached to glycans. This advantage has been used in several large sample cohort studies employing MALDI analysis of released *N*-glycans and enabled evaluation of distinctive α2,3- and α2,6-linkage specific derived sialylation traits[27, 28]. In some cases the use of "cold" MALDI matrices may minimize the loss of SA residues from glycans or glycoproteins, but it does not completely prevent it [29].

Several types of plate-based assays, such as lectin microarrays and enzyme-linked lectin assays, have been developed and applied to study glycosylation. These assays typically employ specific antibodies or lectins for the capture and detection of specific glycan structures on a solid support to enable the characterization and quantification of

glycans present in biological samples[30]. A novel type of assays for the analysis of glycosylation features are enzymatic plate-based assays. These assays rely on the utilization of specific enzymes, including exoglycosidases, to catalyse reactions in microtiter plates and the subsequent quantification by detecting fluorescence signals[31].

Glycomics

Glycomics is a relatively young scientific field that is a subset of glycobiology, which aims to study complex sugars or glycans present in biological systems. The primary objective of glycomics is to gain understanding of the role that glycans play in biological processes, including molecular interactions involved in cellular communication and disease development. *N*-glycomics refers to the analysis of *N*-linked glycans, whereas *O*-glycomics specifically focuses on the study of the full repertoire of *O*-glycans, by identifying and quantifying these molecules in biological samples. In the past decade, clinical glycomics has been recognised for its contribution to understanding the biological and functional significance of glycans in various physiological and pathological processes. In the field of clinical glycomics, the analysis of a large number of samples is often required for cohort studies. Therefore, the high-throughput nature of both MALDI methods and plate-based assays makes them particularly suitable for such studies. Clinical glycomics frequently relies on the implementation of highly-automated analytical workflows supported by liquid handling platforms that employ a range of different analytical methods[32, 33]. Such workflows have been utilised to generate, as of now, blood plasma and IgG *N*-glycome datasets from large cohorts of clinical samples[28, 33]. By the identification of glycosylation aberrations in a significant number of patients, it is possible to propose new diagnostic and predictive biomarkers, and support the development of novel therapeutic approaches in various disease settings. Of note, with the advancements in analytical methods used in glycomics, particularly ultra-high resolution MS, there has been an increase in studies focusing on glycoproteomics, which involves analysing proteins that are glycosylated[34, 35].

Glycan biomarkers

Glycan biomarker refers to specific glycans or glycosylation patterns (often referred to as direct or derived traits, respectively) that are found to be associated with particular physiological or pathological conditions. Glycan biomarkers are identified through the combination of glycomics and association analysis by measuring and comparing the biomarker levels between disease and control groups of individuals. Glycan biomarkers have been shown to have the potential use in diagnosis, prognosis, and monitoring of various diseases, including cancer[36], autoimmune diseases[37], and infectious diseases[38]. It is noteworthy that *N*-glycomes of both human plasma and immunoglobulin G (IgG) remain stable over time under homeostatic conditions within an individual[39, 40]. The fact that these *N*-glycomes are highly sensitive to various pathological processes strengthens their diagnostic and prognostic potential[41, 42].

Diabetes

Diabetes mellitus is a group of chronic metabolic conditions, all sharing a common feature of heightened levels of glucose in the bloodstream. This elevation occurs due to the body's impaired capacity to produce insulin, resistance to insulin action, or both. Type 1, type 2, gestational diabetes and a group of other types of diabetes caused by genetic defects, pancreatic disease or other external diseases and influences are clinically distinct types of diabetes mellitus[43]. Type 1 diabetes refers to 5-10% of diabetes cases characterised by the autoimmune destruction of beta cells within the pancreas, which results in the complete absence of insulin production. Type 2 diabetes accounts for 90-95% of all diabetes cases and develops in cases of an abnormal increased resistance to the action of insulin, and when the body fails to produce enough insulin to overcome the resistance. If left untreated, diabetes can cause long term complications such as kidney damage, nerve damage, cardiovascular disease, and vision loss[44].

The group of diabetes types, which are distinct from type 1, type 2 and gestational diabetes, includes monogenic diabetes syndromes. Maturity-onset diabetes of the

young (MODY), a rarer type of diabetes consisting of less than 5% diabetes cases, is one of the main forms of monogenic diabetes caused by a single gene mutation. There are 14 known subtypes of MODY, each caused by mutations in different genes that leads to distinct clinical features[45].

Disease complications are defined as medical conditions that arise during a disease or are a result of unfavorable evolution of an existing disease, a health condition, or a medical treatment. Long term complications of untreated or inadequately regulated diabetes, classified as microvascular or macrovascular, pose a significant burden on public health systems. Microvascular complications affect nervous system (neuropathy), renal system (nephropathy) and eye health (retinopathy) whereas macrovascular complications are the cause of stroke, cardiovascular disease and peripheral vascular disease[46].

Glycan biomarkers in diabetes

Several studies suggest that protein *N*-glycosylation may play a role in the development of type 2 diabetes[47, 48]. *N*-glycosylation changes have been observed in various diseases such as type 1 diabetes, type 2 diabetes and HNF1A-MODY[49]. These *N*-glycosylation changes are considered as potential biomarkers, which can differentiate between different types of diabetes and predict the risk of developing diabetes in the future.

A large cohort study identified that individuals at a higher risk of developing type 2 diabetes can be identified based on variations in glycosylation. An increased complexity of *N*-glycan structures, specifically highly branched, galactosylated, and sialylated structures in total plasma proteins, was found to be associated with an elevated risk of developing type 2 diabetes[47]. Another study has found that a proinflammatory state, which is characterised by a decrease in galactosylation and sialylation of IgG, is observed in individuals with type 2 diabetes. Moreover, an increase in pro-inflammatory IgG fucosylated structures with bisecting GlcNac was observed in these individuals[48, 50].

Several studies have shown that by measuring antennary fucosylation (α1-3,4-linked fucose residues) of *N*-glycans, it is possible to discriminate HNF1A-MODY and other types of diabetes[51, 52]. The *N*-glycan antennary fucosylation levels are significantly decreased in individuals with HNF1A-MODY as the consequence of the presence of genetic variants in the *HNF1A* gene[53]. Recently, fucosylated alpha-1-acid glycoprotein (AGP) has been evaluated for its diagnostic potential in HNF1A-MODY[54]. The best performing AGP variant was found to provide a very good discriminatory potential with an AUC of 0.94, which is in line with the results obtained in the previous studies that considered released *N*-glycans and absolute α1-3,4 fucosylation levels of total plasma proteins[51, 55].

Glycan biomarkers in complications of diabetes

There have been several reports on the associations between either total plasma *N*-glycome or IgG *N*-glycome and type 2 diabetes, however, there is a limited number of studies on the involvement of *N*-glycosylation in type 2 diabetes complications. In a recent study on two large Dutch cohorts using mass spectrometry to assess total plasma *N*-glycome, several glycosylation features, including fucosylation, galactosylation and sialylation, were found to be associated with prevalent and incident complications of type 2 diabetes. High levels of *N*-glycan bisection was strongly associated with prevalent cardiovascular disease, while the increase in 2,6-sialylation on tri-antennary glycans was observed in cases of prevalent nephropathy[28].

Dyslipidemia, a cardiovascular risk factor prevalent in approximately 50% of individuals with type 2 diabetes, is characterised by a preponderance of small, dense, low-density lipoprotein particles, elevated triglyceride and low high-density lipoprotein cholesterol levels[56]. Alterations in apolipoprotein glycosylation have been linked to various diseases associated with dyslipidemia, including diabetes[57]. Apolipoprotein C-III (ApoC-III) is one of the main four forms of ApoC found in chylomicrons, VLDL, and HDL, and ApoC-III *O*-glycosylated variants have been found to be associated with fasting plasma triglyceride levels[58]. The complete molecular mechanisms underlying these

associations have not been fully elucidated. However, it is noteworthy that sialylated apoC-III glycoforms play a role in the differential clearance of triglyceride-rich lipoproteins[59]. These findings suggest that apoC-III has the potential to serve as a biomarker and be targeted for the treatment of lipoprotein metabolism disorders.

Scope of the thesis

The aim of the research presented in this thesis is development and optimization of analytical methods to study glycosylation changes as potential stratification biomarkers in large clinical sample cohorts of patients with two types of diabetes and diabetes-related complications. The analytical workflows proposed in this research were designed to address and overcome challenges in areas of high-throughput sample preparation, the analysis of these samples, subsequent data processing and statistical analysis. The developed methods are applicable for the analysis of released *N*-glycans, *O*-glycosylated proteins and absolute fucosylation levels of proteins derived from blood plasma.

Since the first study in 2013 which proposed that blood plasma protein *N*-glycan antennary fucosylation is altered in patients with a monogenic type of diabetes, HNF1A-MODY, there have been several studies evaluating the potential of full plasma protein antennary fucosylated *N*-glycans as stratification biomarkers to differentiate patients with HNF1A-MODY from other diabetes cases. In **chapter 2**, a new LC-MS/MS method was developed to assess *N*-glycan antennary fucosylation levels in a large number of patients and perform a first inter-laboratory evaluation of this glycan biomarker by comparing the biomarker's diagnostic performance and the consistency of *N*-glycan antennary fucosylation level measurements that were performed in two independent research centres.

The application of LC methods is poorly recognised in public diagnostic centres, which sets a major bottleneck in translating glycan biomarkers into clinical practice. In **chapter 3**, an enzymatic plate-based assay, which employs high-throughput sample preparation using a liquid-handling platform and simplified data processing, was developed to assess absolute α1-3,4 fucosylation levels of blood plasma proteins in a large cohort of patients with HNF1A-MODY and type 2 diabetes. The diagnostic performance of α1-3,4 fucosylation levels was evaluated against the diagnostic performance of antennary fucosylated *N*-glycans obtained in the previous studies applying LC methods.

Elevated triglyceride levels are observed in diabetic dyslipidemia. Altered profiles of apoC-III *O*-glycosylated sialylated variants have been linked to increased plasma triglyceride levels based on results from several studies applying various analytical methods. Small sample cohorts that were applied in the current studies might be a significant concern in consideration of the validity of their research discoveries. In **chapter 4**, a previously established highly-automated ultra-high resolution MALDI-FTICR MS method was adjusted to facilitate the analysis of apoC-III *O*-glycosylation profiles in large sample cohorts. The workflow that was applied to collect, process and curate MALDI MS-derived data involves the use of in-house developed high-throughput processing software, MassyTools. The association analysis was performed for apoC-III *O*-glycosylation profiles and a panel of lipid biomarkers.

The method for the analysis of apoC-III *O*-glycosylation, which was optimised and validated in **chapter 4**, was applied for a well characterised large cohort of patients with type 2 diabetes. ApoC-III *O*-glycosylation profiles were obtained for these patients, including a control group of patients without diabetes, and undergone genome-wide association studies (GWASs) to identify associations between genetic variants and apoC-III glycoform levels in **chapter 5**. Further, apo-CIII *O*-glycosylation profiles derived from groups of patients with type 2 diabetes, which were categorised based on the presence of micro- and macrovascular complications, were subjected to the association analysis to uncover associations between apo-CIII *O*-glycoform levels and either the prevalence or the incidence of these complications in **chapter 6**. The research findings were strengthened by performing the meta-analysis of genetic variants linked to apoC-III glycosylation with lipid biomarkers and micro-/macro-vascular complications of type 2 diabetes in **chapter 5 and chapter 6**, respectively.

In the final chapter, **chapter 7**, research studies presented in this thesis are discussed and the research findings critically evaluated in the context of their significance for clinical use. Aspects that pose a challenge in translating glycan biomarkers into clinical practice are highlighted. Moreover, the potential of analytical methods and statistical

approaches employing large omics data, which are proposed in this thesis, is evaluated for its use in the biopharma industry.