

Universiteit
Leiden
The Netherlands

High-throughput mass spectrometric N-glycomics

Reiding, K.R.

Citation

Reiding, K. R. (2018, April 5). *High-throughput mass spectrometric N-glycomics*. Retrieved from https://hdl.handle.net/1887/61076

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/61076

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle http://hdl.handle.net/1887/61076 holds various files of this Leiden University dissertation.

Author: Reiding, K.R.
Title: High-throughput mass spectrometric N-glycomics
Issue Date: 2018-04-05

High-throughput mass spectrometric *N*-glycomics

Karli Robert Reiding

ISBN: 978-94-6295-891-3

© Karli Reiding. All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without permission of the author or the journals holding the copyrights of the published manuscripts. All published material was reprinted with permission.

The work presented in this thesis was performed at the Center for Proteomics and Metabolomics, Leiden University Medical Center, The Netherlands.

The work was supported by the European Union Seventh Framework Programme HighGlycan project, grant number: 278535.

Cover and chapter design: Boudien Reiding & Karli Reiding

Printing: ProefschriftMaken || www.proefschriftmaken.nl

High-throughput mass spectrometric *N*-glycomics

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 5 april 2018
klokke 10.00 uur

door

Karli Robert Reiding
geboren te Rotterdam
in 1985

Promotores:	Prof.dr. M. Wuhrer
	Prof.dr. A.M. Deelder
Leden promotiecommissie:	Prof.dr. C.H. Hokke
	Prof.dr. P.E. Slagboom
	Dr. R. Fahey (NIBRT Glycoscience Group, National Institute for Bioprocessing Research and Training, Dublin, Ireland)
	Dr. D.J. Lefeber (Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands; Translational Metabolic Laboratory, Radboud University Medical Center, Nijmegen, The Netherlands)

Table of contents

Chapter 1 – General introduction 9

1.1 Glycosylation 11

1.2 Monosaccharide structure 12

1.3 The biosynthesis of protein *N*-glycosylation 14

1.4 Glycosylation in health and disease 16

1.5 Glycan analysis 18

1.6 Matrix-assisted laser desorption/ionization mass spectrometry 19

1.7 Chemical derivatization for MALDI-MS glycan analysis 21

1.8 Scope 23

1.9 References 24

Chapter 2 – High-throughput analysis and automation for glycomics studies 31

2.1 Abstract 32

2.2 Introduction 33

2.3 High-throughput sample preparation 35

2.4 Use of internal standards 40

2.5 Integration of analysis workflows 42

2.6 Automation of sample preparation 43

2.7 Data analysis 45

2.8 Conclusion 46

2.9 Acknowledgments 49

2.10 References 49

Chapter 3 – High-throughput profiling of protein *N*-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification 55

3.1 Abstract 56

3.2 Introduction 57

3.3 Experimental section 58

3.4 Results 63

3.5 Discussion 71

3.6 Supporting information 72

3.7 Acknowledgments 73

3.8 References 73

Chapter 4 – Automation of high-throughput mass spectrometry-based plasma *N*-glycome analysis with linkage-specific sialic acid esterification 77

4.1 Abstract 78

4.2 Introduction 79

4.3 Experimental section 80

4.4 Results and discussion 84

4.5 Conclusion 87

4.6 Supporting information 87

4.7 Acknowledgments 88

4.8 References 88

Chapter 5 – Murine plasma *N*-glycosylation traits associated with sex and strain 91

5.1 Abstract 92

5.2 Introduction 93

5.3 Experimental section 94

5.4 Results 97

5.5 Discussion 103

5.6 Conclusions and perspectives 106

5.7 Acknowledgments 107

5.8 References 107

Chapter 6 – Human plasma *N*-glycosylation as analyzed by MALDI-FTICR-MS associates with markers of inflammation and metabolic health 111

6.1 Abstract 112

6.2 Introduction 113

6.3 Experimental procedures 114

6.4 Results 119

6.5 Discussion 128

6.6 Supporting information 132

6.7 Acknowledgments 132

6.8 References 132

Chapter 7 – Serum protein *N*-glycosylation changes with rheumatoid arthritis disease activity during and after pregnancy 139

7.1 Abstract 140

7.2 Introduction 141

7.3 Methods 142

7.4 Results 144

7.5 Discussion 149

7.6 Summary 151

7.7 Acknowledgments 151

7.8 Additional information 151

7.9 References 151

Chapter 8 – High-throughput serum *N*-glycomics: method comparison and application to study rheumatoid arthritis and pregnancy-associated changes 155

8.1 Abstract 156

8.2 Introduction 157

8.3 Results 158

8.4 Discussion 166

8.5 Summary 169

8.6 Methods 170

8.7 References 172

Chapter 9 – General discussion 177

9.1 Developments in linkage-specific sialic acid esterification 179

9.2 Persistent challenges in mass spectrometric *N*-glycomics 181

9.3 Developments human plasma *N*-glycosylation analysis 185

9.4 Future prospects of plasma *N*-glycosylation analysis by MALDI-MS 190

9.5 Concluding remarks 192

9.6 References 193

English summary 203

Nederlandse samenvatting 207

Acknowledgements 211

Curriculum Vitae 212

Publications and submitted manuscripts 213

Part I – Introduction

Chapter 1

General introduction

General introduction

The human body is an intricate and dazzling piece of machinery, with interlocking systems in place to allow us to do the things we want to do in daily life. In all of these systems we have the presence of countless large and small molecules, with only their combined action leading ultimately to a macrobiological phenotype such as fighting a disease, digesting lunch, or stretching your arm. Workhorses involved in this machinery are commonly proteins, which can cover a wide array of functions which depend on their precise makeup, from transport *via* replication to catalysis. The function, localization, and control of proteins is greatly enhanced by modifications they may undergo, prominent examples being phosphorylation and acetylation, but in the extracellular domain none of these is as abundant as the decoration of proteins with carbohydrate structures, coined glycosylation[1,2].

1.1 Glycosylation

Glycosylation, the reaction of a carbohydrate with another molecule, is an important way in which biomolecules can be modified. Enzymatic glycosylation may occur on lipids, proteins and on other carbohydrates, and respectively leads to the formation of glycolipids, glycoproteins and glycans (although "glycan" may also refer to the carbohydrate content of a glycoconjugate)[3]. Glycosylation is a wide-spread process in biology, and the presence or absence of a glycan (macroheterogeneity) or the actual size and structure thereof (microheterogeneity) are known to have major effects on a glycoconjugate. Glycans have for example been shown to affect glycoprotein folding and quality control[4-6], solubility[7], spatial organization[8], and receptor interaction[9,10], but many more functions have been described[11].

Human protein glycosylation occurs in two main forms, namely *N*-glycosylation and *O*-glycosylation, which differ by the type of linkage between the carbohydrate and the protein backbone, as well as by the amino acids on which they can occur[3,12]. *N*-Glycosylation, which yields *N*-glycans, is the enzymatic attachment of an oligosaccharide by *N*-glycosidic linkage to the nitrogen of the side chain amide of an asparagine. This process predominantly occurs within the amino acid consensus sequence Asp-X-Ser/Thr, where X can be any amino acid except proline[3,13]. On the other hand, *O*-glycosylation, yielding *O*-glycans, is the attachment of a monosaccharide by *O*-glycosidic linkage to the oxygen of the side-chain hydroxyl group, which may occur on serine or threonine in case of mucin-type *O*-glycosylation (although modification of hydroxylysine and tyrosine have also been described)[3,14,15]. After initial attachment, both *N*- and *O*-glycans can be further modified by glycosyltransferases and glycosidases, *i.e.*, enzymes that respectively add and remove monosaccharides, eventually resulting in a wildly divergent set of glycoforms that can be achieved for a given protein[3].

Both *N*- and *O*-glycosylation can be regarded as post-translational protein modification (PTM), sharing the nomenclature with, *e.g.*, phosphorylation and acetylation[2]. *N*-Glycosylation is sometimes also referred to as co-translational, for the reason that initial *N*-glycan attachment takes place during (and not after) translation of a protein into the rough endoplasmic reticulum (ER) lumen[3,12]. Among the known PTMs found in the extracellular space, *N*-glycosylation is estimated to be the most abundant, and the majority of plasma proteins are thought to be glycosylated with at least one *N*-glycan[1,2]. Nevertheless, it remains one of the most heavily under-studied biological topics to date, owing to the structural complexity involved and the resulting analytical challenges.

1.2 Monosaccharide structure

Glycans are polymers made up of monosaccharides just like deoxyribonucleic acid (DNA) and proteins are respective biopolymers of nucleotides and amino acids. However, whereas DNA and peptides are typically linear sequences, glycans tend to be arranged in complex branching structures, and in addition often show many kinds of stereoisomerism[3].

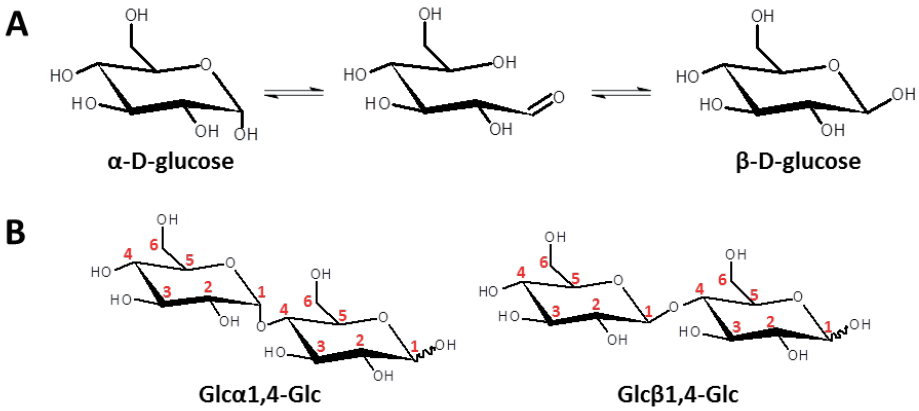


***Figure 1**. **A)** The α- and β-anomers of cyclic D-glucose and their open intermediate. **B)** Two glucose residues in either α1,4-linkage or β1,4-linkage, with numbered carbons. A perturbed line has been used to indicate the anomeric hydroxyl group involved in ring opening and closing.*

Taking as example the aldohexose, a group of monosaccharide isomers with six carbons and an aldehyde at the first carbon (chemical formula $C_6H_{12}O_6$), the presence of four chiral centers means that already 16 (2^4) stereoisomers are possible. The stereoisomerism at the fifth carbon is denoted by the letter substitution D or L, *e.g.*, D-aldohexose, while the stereoisomerism at the three remaining chiral carbons defines the overall name of the molecule, *e.g.*, glucose, mannose or galactose[3]. Free monosaccharides alternate in solution between an open-chain form and a cyclic form, but it is the cyclic form that is predominantly found in polysaccharides (**Figure 1A**)[3]. Cyclization of an aldohexose may either occur with the hydroxyl group at the fourth carbon, forming a furanose (*e.g.*, glucofuranose), or with

the hydroxyl group at the fifth carbon, forming a pyranose (*e.g.*, glucopyranose)[3]. In both cases this leaves the first carbon a new chiral center, of which the two configurations, named anomers, can be indicated with an α or β prefix (*e.g.*, β-glucopyranose)[3].

Polysaccharides, or glycans, are formed when monosaccharides are covalently joined by an *O*-glycosidic bond. While this could result from a condensation reaction between the anomeric carbon of a given monosaccharide with the hydroxyl group of another (such as the case with glycation), for glycosylation the bond is formed by the enzymatic transfer of a monosaccharide from a nucleotide-activated sugar. For example, beta-1,4-galactosyltransferase 1 may catalyze the transfer of the galactose of uridine diphosphate galactose (UDP-galactose) to glucose to form lactose (milk sugar)[16]. A resulting *O*-glycosidic linkage is typically described by the anomerism and carbons involved in the linkage, *e.g.*, α1,2 for a linkage between the first carbon of α-D-glucopyranose and the second carbon of D-fructofuranose to form sucrose (table sugar) (**Figure 1B**)[3]. An example underlining the functional differences between glycosidic linkage types can be found in glucose polymers, which when α1,4-linked forms amylose (a main component of starch), or when β1,4-linked instead forms cellulose (a main component of cotton and wood)[3].

Next to displaying stereoisomerism, monosaccharides in biology are often substituted at various hydroxyl groups with a wide range of other functional group. Examples of this include amination, amine acetylation, or the presence of a carboxylic acid group, which would for glucose respectively form glucosamine, *N*-acetylglucosamine, and glucuronic acid[3].

In all, given the total variety of monosaccharides available for polymerization, the number of potentially different glycan structures is truly staggering[17]. In reality, however, the types of monosaccharides involved in human protein glycosylation are rather constrained, and appear to be under tight control. The most common monosaccharides found in human *N*- and *O*-glycosylation are pyranoses, and principally comprise the deoxyaldohexose (dHex) L-fucose (Fuc), the aldohexoses (Hex) D-glucose (Glc), D-mannose (Man) and D-galactose (Gal), the *N*-acetylaldohexosamines (HexNAc) *N*-acetyl-D-glucosamine (GlcNAc) and *N*-acetyl-D-galactosamine (GalNAc), and the heavily substituted ketononose *N*-acetylneuraminic acid (Neu5Ac, belonging to the family of sialic acids) (**Figure 2**)[3,18]. The polymerization of these monosaccharide residues are what make up most of the human *N*- and *O*-glycans.

Whereas *N*- and *O*-glycosylation can be found within nearly all forms of life, different monosaccharides can be found within non-human glycan structures. For example, aldopentoses like D-xylose are commonly seen on the *N*-glycans of plants and parasites[3,20], and the mouse largely incorporates the sialic acid *N*-glycolylneuraminic acid (Neu5Gc) instead of *N*-acetylneuraminic acid[3,18].

1.3 The biosynthesis of protein *N*-glycosylation

The work presented in this thesis mainly involves *N*-glycosylation which, as prior indicated, can be regarded as both a post-translational and co-translational modification[3,12]. During protein synthesis into the lumen of the rough ER, *N*-glycosylation occurs when the oligosaccharide portion of a membrane-bound glycolipid, dolichol-$Glc_3Man_9GlcNAc_2$ (Glcα1,2-Glcα1,3-Glcα1,3-Manα1,2-Manα1,2-Manα1,3-[Manα1,2-Manα1,3-[Manα1,2-Manα1,6-]Manα1,6-]Manβ1,4-GlcNAcβ1,4-GlcNAcβ), is transferred in its entirety to the asparagine of a polypeptide containing the appropriate consensus sequence (Asn-X-Ser/Thr)[3,12]. The glucose residues of the then *N*-linked glycan are subsequently trimmed down by the action of glucosidase I and II, but the last glucose can repeatedly be removed and added (by glucosyltransferase activity) while the glycoprotein undergoes folding[3,12]. The presence of a monoglucosylated *N*-glycan facilitates interaction with the proteins calnexin and calreticulin, which form an important part of the protein quality control mechanism of the ER[3,12]. This glycan-mediated mechanism serves as checkpoint to decide between clearing proteins for further processing (requiring the persistent removal of the glucose) or alternatively sending them to the lysosomal pathway of protein degradation[12,21].

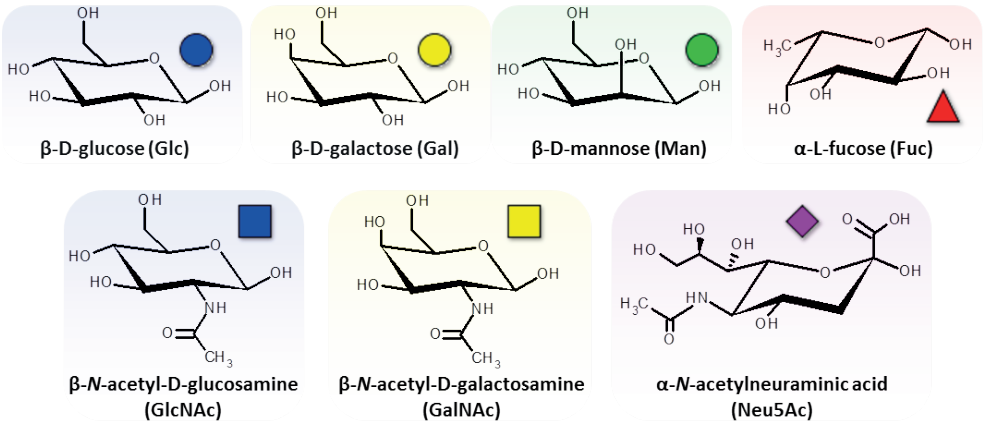


***Figure 2**. **A)** Monosaccharides commonly involved in human N- and O-glycosylation and their symbolical nomenclature as defined by the Consortium for Functional Glycomics (CFG)[19].*

From this point on, the type of *N*-glycosylation ($Man_9GlcNAc_2$) found on proteins may be conserved, eventually appearing as such in the extracellular space, or the glycan can be further processed in a variety of ways as the conjugate is guided through the ER and Golgi apparatus[3,12]. The processing of the glycan is not template-defined, and will ultimately depend on a multitude of factors, including the abundance of glycosidases and glycosyltransferases, the accessibility of the glycan for enzymatic alteration, the presence

of nucleotide sugars to serve as monosaccharide donor, the time spent in different compartments of the ER and Golgi, and so forth[3,12].

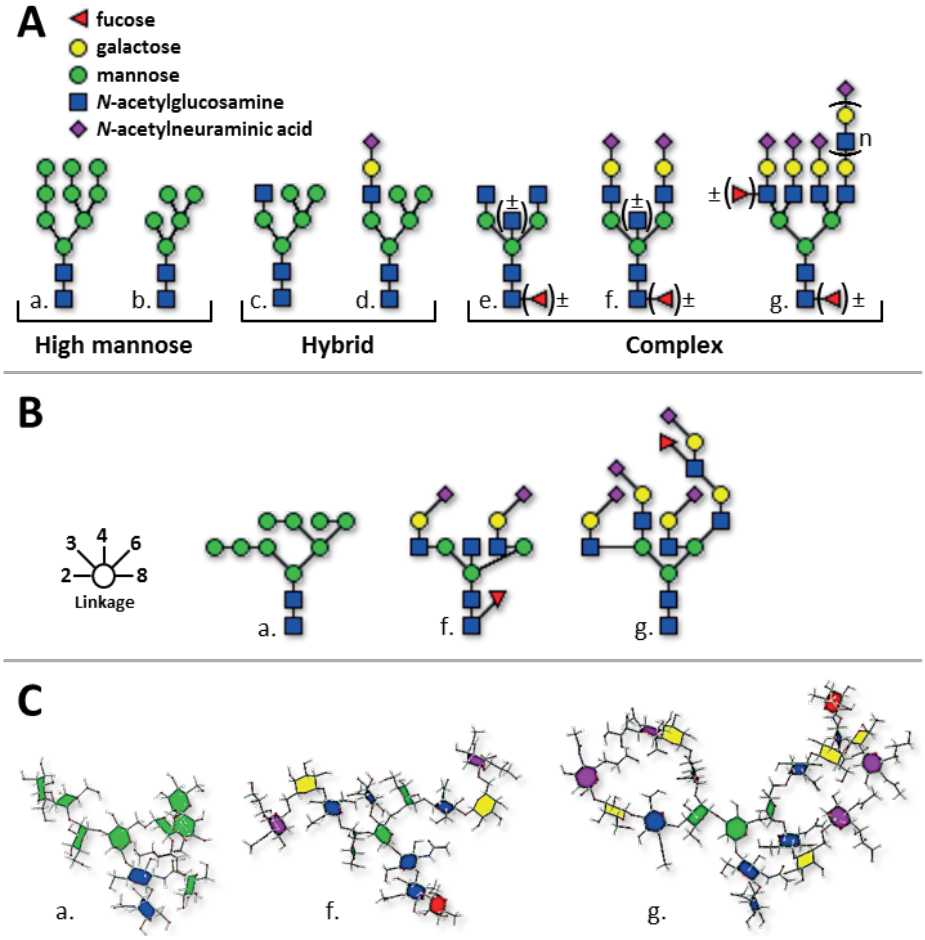


***Figure 3**. **A)** Typical N-glycan species as found on human plasma proteins. Distinction is made between high-mannose-type species (a. and b.), hybrid-type species (c. and d.) and complex-type species (e., f. and g.). This visualization style is simplified with regard to structure, and is commonly used to denote glycan compositions of which limited structural information is available. **B)** Species a., f. and g. with linkage information indicated by the angle. **C)** N-glycan structures for species a., f. and g, as edited from the Glycam structure prediction tool (glycam.org). Displayed here are the structures belonging to the lowest energy state, but glycans are highly mobile and will not maintain a static structure[3].*

In case of further processing, the glycan may be truncated to a $Man_5GlcNAc_2$ structure[3,12], after which an "antennary" GlcNAc residue can be appended in β1,2-linkage to the α1,3-linked mannose "branch" (referring to the branching mannoses on the β1,4-linked mannose) (**Figure 3**)[3,12]. The two mannoses remaining on the α1,6-linked branch may also be removed, and can be replaced by a β1,2-linked antennary GlcNAc[3,12]. An *N*-glycan is described as "high-mannose type" when both branches are occupied by mannose residues,

"hybrid type" when one of the branches has an antennary GlcNAc and the other contains mannoses, and "complex type" when both branches carry an antennary GlcNAc (**Figure 3**)[3,12]. Aside from having β1,2-linked antennae, human *N*-glycans may also have a β1,4-linked antenna on the α1,3-branch, and a β1,6-linked antenna on the α1,6-branch. These can form alternatively linked diantennary structures, or combine to yield tri- and tetraantennary species[3,12].

Antennary GlcNAcs can be extended in a variety of ways, the most common being the addition of a β1,4-linked galactose and subsequent addition of an *N*-acetylneuraminic acid (sialic acid) in either α2,6- or α2,3-linkage to the galactose[3,12]. Extension of structures may furthermore occur by α1,6-linked fucosylation of the innermost GlcNAc, by α1,2-, α1,3 or α1,4-fucosylation of antennae, by a "bisecting" GlcNAc that is β1,4-linked to the β1,4-linked core-mannose, by elongation of the β1,6-linked antenna to contain "LacNAc" repeats (Galβ1,4-GlcNAcβ1,3-), as well as by a variety of other modifications(**Figure 3**)[3,12]. Together this facilitates the highly diverse range of structures an *N*-glycan may have, potentially leading to specific oligosaccharide epitopes like sialyl-Lewis X (sLeX; Neu5Acα2,3-Galβ1,4-[Fucα1,3-]GlcNAc), a major ligand for E-selectins[3,22].

1.4 Glycosylation in health and disease

The process of glycosylation is not template-defined (such as DNA), but a considerable portion of the human genome is directly and indirectly involved in the process[23,24]. Genetic defects can lead to the absence of glycosylation on protein sequences where this is normally the case, a phenomenon named type I congenital disorder of glycosylation (CDG)[25]. Homozygous or compound heterozygous knock-outs are embryonically lethal, and individuals may otherwise suffer from severe symptoms including mental retardation and organ failure[25]. Genetic abnormalities leading to qualitative differences in glycosylation (*e.g.*, a decreased ability to form sialylated or fucosylated glycans) are grouped as type II CDG and may lead to a wide range of symptoms as well, including mental retardation, jaundice, and hypotonia[25]. Interestingly, some types of type II CGD appear treatable by dietary mannose, galactose or fucose supplementation, particularly when the genetic defect is in a nucleotide sugar pathway[25-28]. In both type I and type II CDGs, analysis of the glycan site occupancy and structures, either on specific proteins or within mixture of proteins such as in plasma, is highly useful for the detection and characterization of the defect.

Glycosylation may also be reflective of, or play a role in, tumor growth and metastasis. As glycosylation is ultimately produced by cells, aberrant regulation thereof, such as the case in cancer, may result in uncommonly glycosylated proteins being present in the circulation[29]. For example, while prostate cancer is diagnosed by the abundance of prostate-specific antigen (PSA) in the blood, it has also been shown that the type of glycosylation

present on the protein in circulation differs from that in the prostate, particularly by the degree of sialylation[30]. On the functional side, it is thought that the presence of glycan epitopes such as sLeX may be instrumental in allowing circulatory tumor cells to adhere to new positions, thereby facilitating tumor metastasis[29,31].

A setting in which glycosylation has shown diagnostic potential is in the discrimination of different types of diabetes. A particular genetic variant of the disease, *i.e.*, type 3 maturity-onset diabetes of the young (MODY 3), can often be distinguished from the other forms, *e.g.*, type I and II diabetes, by the decreased antennary fucosylation found on the *N*-glycosylation of plasma proteins[32]. This appears to be mainly driven by the *HNF1A* gene that is a risk factor for MODY, which is proposed to be a transcription factor for multiple genes responsible for antennary fucosylation (*e.g.*, *FUT3* and *FUT6*)[33]. The quantification of fucosylation is thought to be a simple and clinically interesting method for diabetes categorization, and may assist in the development of personalized treatment.

Glycosylation is also thought to play a large role in immunity and autoimmunity. Immunoglobulin G (IgG), the most abundant antibody in human plasma, serves as a major component in the adaptive immune system[34]. The protein generally carries one *N*-glycan on each of its two crystallizable fragment (Fc) domains, and displays *N*-glycosylation of the antibody-binding fragment (Fab) domains for around 20% of the circulatory IgGs[34,35]. The *N*-glycosylation of IgG-Fc has shown to heavily influence its Fc-gamma-receptor (FcγR) affinity, thereby modulating antibody-dependent cellular cytotoxicity (ADCC), with lack of a core-fucose leading to an up to 50-fold increased binding to FcγRIIIa and the degree of galactosylation and bisection having several-fold effect sizes[36-38]. Furthermore, various IgG-Fc glycosylation traits, notably galactosylation, have shown to associate with inflammation and autoimmune diseases such as rheumatoid arthritis (RA), in which lower galactosylation generally associates with higher degrees of inflammation[39-41]. Fab glycosylation is suspected to play a role in RA as well, as seen by the elevated degree of Fab-glycosylation present on disease-specific anti-citrullinated protein antibodies (ACPA) found in the subset of ACPA-positive RA patients[42]. The influence of glycosylation on antibody function extends to biopharmaceutical antibodies, meaning that production processes need to be under careful control and need to be monitored for glycan heterogeneity[43,44].

In addition to disease-associated glycosylation phenotypes, studies in healthy individuals have also shown numerous associations between glycosylation and phenotypes like age, sex, inflammation, body mass index, pregnancy and lipid levels, and so forth[45-48]. Interestingly, while the glycosylation within an individual is relatively stable, with changes usually being representative of a change in homeostasis, the variance between individuals is usually much greater[49,50]. This suggests a benefit for longitudinal study designs and implies that cross-sectional designs need to include considerable sample numbers.

1.5 Glycan analysis

While glycosylation is of clinical and biopharmaceutical importance, the analysis thereof is not easily performed. Since glycosylation does not have an equivalent to the polymerase-chain reaction found in genomics research[51], analyte concentrations are often low and detection methods need to be highly sensitive. In addition, analytical methods need to be of sufficient throughput to handle the samples sizes required to identify associations within the considerable population variance[49,50]. And, within these constraints, the methods need to have the actual capability to resolve the relevant glycosylation properties, which is complicated by glycan stereoisomerism and resulting numbers of chemically distinct structures that overlap in physical properties such as solubility, charge, and mass.

Glycoprotein analysis can in principle be performed at three levels, namely on intact glycoproteins, on glycopeptides, and on released glycans[44,52]. Intact glycoprotein analysis provides most information on the combination and interaction of glycans and other PTMs on a protein, approaching most closely the biological situation[53]. Proteolytic digestion of a glycoprotein into discrete peptides and glycopeptides, *e.g.*, by trypsin, GluC or proteinase K, may still retain the site-specific glycan information but reduces the analytical complexity to some degree[54]. Releasing glycans from the protein backbone, *e.g.*, by hydrazinolysis, β-elimination, or endoglycosidase digestion in case of *N*-glycosylation, can yield the most structural information on the glycan and has the highest tolerance to complex samples, but does not reveal the glycan attachment site or carrier protein[44,52,55,56].

Further distinction has to be made between the analysis of a single glycoprotein after targeted purification using, *e.g.*, affinity-capturing, and the analysis of a glycoprotein mixture. In-depth glycan analysis has been performed for a selection of single glycoproteins, examples including IgG, IgM, and alpha-1-antitrypsin[57-60], but these are limited by the availability of protein-specific purification protocols. For clinical biofluid glycosylation analysis, *e.g.*, serum or plasma, qualitative glycan information can be obtained as well, but this is usually restricted to the released glycan level due to the large array of glycoproteins involved[61,62]. These complex-sample released glycan analyses are generally referred to as "glycomics" experiments, and specific "glycomes" can be indicated by their sample source, *e.g.*, the total serum protein *N*-glycome (TSNG) or the total plasma protein *N*-glycome (TPNG). With the diversity of glycan structures involved in such samples, a combination of methods is typically required for complete structural characterization – each of the methods ideally providing orthogonal information[63].

It is to no surprise then that a broad selection of analytical methods have been applied to the characterization of glycans and glycoconjugates. These include, among others, lectin or carbohydrate microarrays[64,65], nuclear magnetic resonance spectroscopy[60,66], liquid chromatography[59,67], capillary electrophoresis (CE)[68,69], and mass spectrometry (MS)[61,70], as

well as a variety of hyphenations among them (*e.g.*, CE-MS or LC-MS)[63,71,72]. Large-scale *N*-glycomics studies have so far been reported with hydrophilic-interaction (ultra-)high-performance liquid chromatography with fluorescence detection (HILIC-(U)HPLC-FLD)[33,46,73], capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF)[74], and matrix-assisted laser desorption/ionization (MALDI) MS[48,63,75].

1.6 Matrix-assisted laser desorption/ionization mass spectrometry

Mass spectrometry is an analytical technique to form gas-phase ions, separate them by mass per charge (*m/z*), and report on their relative abundances. Mass spectrometers usually consist of an ion source, a separation step, and a means of detecting the ions[76].

To generate gas-phase ions MS can make use of a large selection of ionization methods, but common variants for high-throughput glycan analysis are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI)[76]. ESI requires analytes to be dissolved in a liquid, which is dispersed into aerosols by the presence of a nebulizing gas and the application of a high voltage[76]. Ions may directly evaporate from these aerosols or solvent evaporation may leave behind ionized residual molecules[76]. MALDI, on the other hand, requires a sample of interest to be co-crystallized with a matrix substance. Gaseous molecules are generated by irradiating the crystals with a laser, leading to sublimation of the molecules. In the gas-phase the analytes of interest are ionized by deprotonation, protonation, or alkali metal adduction[76]. Unlike ESI, MALDI predominantly generates single-charged ions, which is unfavorable for the detection of very large masses, but on the other hand also leads to relatively clean mass spectra in the *m/z* range covering released human *N*-glycosylation[77]. Compared to ESI, MALDI is an extremely high-throughput method for the ionization of successive samples, as these can be pre-spotted on a target plate and indexed within seconds from each other[63,76].

After ionization, separation of the gas-phase ions can be achieved in a variety of ways as well. For time-of-flight (TOF) MS the separation occurs by application of a static electric field. This provides greater acceleration to molecules that are lower in mass (*m*) and carry more charges (*z*, hence the *m/z* separation dimension), which is then followed by a drift time in a field-free region towards an ion detector[76]. Alternatively, ions can be trapped in an electric or magnetic field and[76,78], and for example be detected by their ion cyclotron resonance as is the case for Fourier transform ion cyclotron resonance (FTICR) MS[79,80].

Whereas MALDI-FTICR-MS can achieve very high resolution *N*-glycan measurements, potentially separating isobaric chemicals that are milliDaltons apart[79,80], MALDI-TOF-MS requires additional equipment to achieve isotopic resolution for *N*-glycan analysis[81,82]. One issue with MALDI is that molecules form a three-dimensional cloud prior to acceleration,

leading to different exposure times to the electric field for different regions of the cloud. This results in different kinetic energies for equivalent molecules, and ultimately a broader time range in which they arrive at the detector[81,82]. One way to solve the resulting peak broadening is by using a mass reflectron (ion mirror), which reflects the molecules in such a way that it corrects the kinetic energy range (**Figure 4**)[81,82]. In way of resolution, reflectron-mounted MALDI-TOF-MS instruments clearly outperform their linear counterparts. However, when molecules break down by metastable decay in the field-free region, reflection can lead to inaccuracy in both resolution and the perceived *m/z* value of the analyte (**Figure 5**)[77,83].

Table 1. *Monosaccharides detectable by mass spectrometry.*

Monosaccharide	Short	CFG Symbol	Letter	Increment composition	Increment mass
Deoxyhexose	**dHex**	△	**F**	**$C_6H_{10}O_4$**	**146.0579**
Fucose	Fuc	▲	F	$C_6H_{10}O_4$	146.0579
Hexose	**Hex**	○	**H**	**$C_6H_{10}O_5$**	**162.0528**
Glucose	Glc	●	H	$C_6H_{10}O_5$	162.0528
Mannose	Man	●	H	$C_6H_{10}O_5$	162.0528
Galactose	Gal	●	H	$C_6H_{10}O_5$	162.0528
***N*-Acetylhexosamine**	**HexNAc**	□	**N**	**$C_8H_{13}NO_5$**	**203.0794**
N-Acetylglucosamine	GlcNAc	■	N	$C_8H_{13}NO_5$	203.0794
N-Acetylgalactosamine	GalNAc	■	N	$C_8H_{13}NO_5$	203.0794
***N*-Acetylneuraminic acid**	**Neu5Ac**	◆	**S**	**$C_{11}H_{17}NO_8$**	**291.0954**
***N*-Glycolylneuraminic acid**	**Neu5Gc**	◇	**Gs**	**$C_{11}H_{17}NO_9$**	**307.0903**

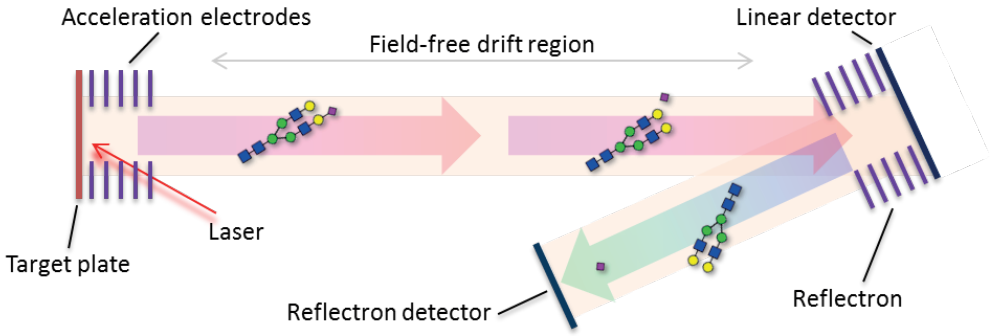


***Figure 4**. Scheme of a reflectron-mounted MALDI-TOF-MS system. Ions produced by MALDI are accelerated by a static electric field, further separated in the field-free drift region, and are either directly detected or after reflection to yield better resolution. In case an analyte breaks down by metastable decay reflection leads to inaccurate and imprecise signals.*

MALDI-MS has shown to be a powerful tool for the analysis of glycosylation, but several constraints apply as well[77]. As MS separation is primarily based on mass, different chemical compositions can usually be distinguished, but no information is obtained on the stereoisomeric aspects of a structure. As such, MS can define a glycan composition by the number of deoxyhexoses, hexoses and *N*-acetylhexosamines, but will not reveal whether these are derived from glucose, mannose or galactose, and in which linkages these are attached (**Table 1**)[77]. Sialic acids like *N*-acetylneuraminic acid and *N*-glycolylneuraminic acid can be distinguished by their unique masses as well, but analysis of sialylated glycans by MALDI-MS is challenging for other reasons. For one, while neutral glycans tend to ionize as cations together with sodium or potassium, sialylated glycans preferentially form anions by deprotonation of the carboxylic acids, which prevents their joint analysis when accelerated by a single polarity electric field[77]. Secondly, sialylated glycans are highly affected by in-source and metastable decay, meaning that mass spectra obtained by reflectron-mounted MALDI-TOF-MS are dominated by sialic acid loss and the appearance of unfocussed "metastable peaks" (**Figure 5**)[77,83].

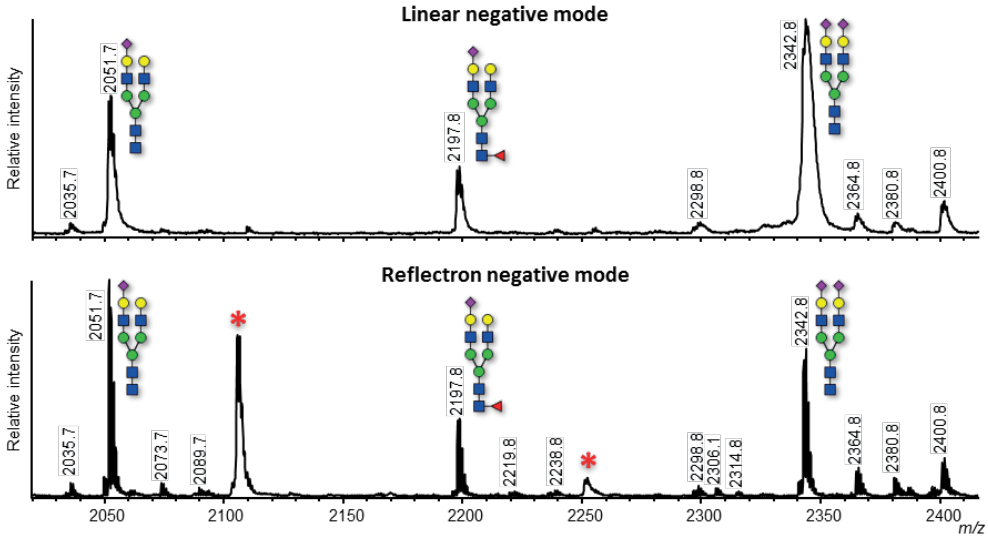


***Figure 5**. MALDI-TOF-MS profiles of 2-AA-labeled human plasma N-glycosylation in linear negative mode (**top**) and reflectron negative mode (**bottom**). Reflectron mode shows isotopic resolution whereas linear mode does not. However, reflectron mode shows the appearance of low-resolution "metastable peaks" (**asterisk**), at the expense of sialylated signals.*

1.7 Chemical derivatization for MALDI-MS glycan analysis

Without making use of derivatization, MALDI mass spectrometric experiments have tried to handle the analysis of sialylated glycan species in a variety of ways. Many reports have been made on the optimization of MALDI matrix conditions and constituents[77]. MALDI matrices such as 2,5-dihydroxybenzoic acid (2,5-DHB)[84], α-cyano-4-hydroxycinnamic acid (CHCA)[85],

or 5-chloro-α-cyanocinnamic acid (Cl-CCA)[86], can for instance be classified into "hot" and "cold" variants, which roughly describes the amount of energy they typically convey to their co-crystallized analytes of interest during the desorption / ionization process[77]. A matrix classified as "hot" allows easier ionization and fragmentation than "cold" matrices, but also leads to more in-source and metastable decay. However, no matrix has been described to date that fully prevents sialic acid decay.

Another way to improve the mass spectrometric detection of (sialylated) glycan species is the chemical alteration of problematic functional groups, and a variety of derivatization strategies have been reported over the years[56,75,77]. One popular method is permethylation, the conversion of a broad array of functional groups across a glycan into their methyl derivatives[55,87]. Permethylation amends the unfavorable ionization and stability situation of sialylated species and has proven to enhance positive ionization of carbohydrates in general[55,87]. A downside of permethylation is that even small reaction inefficiencies per functional group lead to significant peak splitting for larger glycan species. Another issue is the application of relatively harsh reaction conditions leading to sialic acid de-*O*-acetylation, an important modification which can occur on the hydroxyl groups of sialic acids[18,88,89]. Many studies have been published making use of glycan permethylation, both of *O*-glycans and *N*-glycans, and in general the derivatization is considered to be the baseline procedure for mass spectrometric glycan analysis[61,90-93].

Rather than chemical alteration of the wide variety of functional groups on a glycan, it also proved possible to specifically modify the carboxylic acids of the sialylated glycans, which tackled the ionization and stability issues as well. Workflows have been presented that rely on sialic acid esterification or amidation, encompassing a wide variety of acid activator reagents and nucleophiles to arrive at stabilized and neutralized *N*-glycans[56,75,77], several of which have proven viable for the chemical modification of glycopeptides as well[94-97].

Interestingly, the investigations into sialic acid derivatization have led to the observation that α2,3-linked *N*-acetylneuraminic acids – and not α2,6-linked *N*-acetylneuraminic acids – tend to lose water under harsh reaction conditions, understood to be lactonization with the subterminal galactose[98,99]. The first chemistry to exploit this linkage-specific modification for *N*-glycan analysis was with the use of the carboxylic acid activator DMT-MM, repeatedly added in crystalline form to a methanol solution with dissolved glycans[99]. As described, performing the reaction for 1 hour at 80 °C, or 6 hours at room temperature, led to the methyl esterification of α2,6-linked *N*-acetylneuraminic acids, and could also catalyze the lactonization of α2,3-linked sialic acids[99]. Next to neutralizing and stabilizing the sialylated glycans for MS analysis, the mass difference resulting from the linkage-specific reaction allowed mass-based distinction of their stereoisomerism[99].

Following a similar strategy, we showed that linkage-specific sialic acid esterification was as well achievable as a high-throughput one-pot reaction, with increased speed and at milder reaction conditions (1 hour at 37 °C proved sufficient in all cases), by making use of the carboxylic acid activator EDC and catalyst HOBt in an environment of ethanol[62,100]. For this reaction, the use of ethanol instead of methanol proved a requirement for the full linkage-specificity, consequentially leading to linkage-dependent ethyl esterification or lactonization[62,100].

1.8 Scope

The aim of this thesis is the development and application of new methodology for MALDI-MS *N*-glycomics, in order to improve our detection of glycosylation characteristics in a high-throughput manner, and to increase our understanding of the role of *N*-glycosylation in health and disease.

Part I introduces the field of research, and contains the general introduction (**Chapter 1**), as well as a literature review on high-throughput *N*-glycosylation analysis (**Chapter 2**).

Presented in **Part II** are the methodological advances that were achieved in glycosylation analysis by MALDI-TOF-MS. **Chapter 3** reports on a high-throughput workflow for the derivatization of sialylated *N*-glycans, allowing the concurrent detection of sialylated and non-sialylated glycan species in reflectron positive mode MALDI-TOF-MS, as well introducing a distinguishable mass difference between α2,3- and α2,6-linked sialylation isomers. This method was extended to an automated platform in **Chapter 4**.

Part III describes the application of the protocols in settings of clinical and biopharmaceutical interest. Much of current-day biopharmaceutical and fundamental research is performed on experimental animal models, of which the mouse is an abundant example. The plasma *N*-glycosylation of different mouse strains is reported in **Chapter 5**, as well as the differences between male and female individuals of those strains. In **Chapter 6** an alternative sialic acid protection method for MALDI-MS is described, namely the application of collisional cooling in an intermediate-pressure-sourced MALDI-FTICR-MS. This has led to the largest mass spectrometric TPNG study to date, and shows the association between *N*-glycan phenotypes and blood markers of inflammation and metabolic health. In **Chapter 7** the robotized MALDI-TOF-MS method is applied to study – in longitudinal fashion – the glycosylation parameters reflective of RA and the improvement of disease activity thereof during pregnancy. Finally, **Chapter 8** describes a collaboration with several leading high-throughput glycomics laboratories to compare the performance of the contemporary serum *N*-glycomics methodologies, *i.e.*, HILIC-UHPLC-FLD, multiplexed CGE-LIF, and MALDI-TOF-MS with ethyl esterification.

Part IV contains the general discussion (**Chapter 9**). This chapter puts the performed research in the literature context, and speculates on the requirements and challenges for the future of mass spectrometric *N*-glycosylation analysis.

1.9 References

1 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8 (1999).

2 Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* **1**, doi:10.1038/srep00090 (2011).

3 Varki, A. *et al. Essentials of Glycobiology*. 2nd edn, (2009).

4 Culyba, E. K. *et al.* Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science* **331**, 571-575, doi:10.1126/science.1198461 (2011).

5 Caramelo, J. J. & Parodi, A. J. A sweet code for glycoprotein folding. *FEBS Lett* **589**, 3379-3387, doi:10.1016/j.febslet.2015.07.021 (2015).

6 Roth, J. *et al.* Protein N-glycosylation, protein folding, and protein quality control. *Mol Cells* **30**, 497-506, doi:10.1007/s10059-010-0159-z (2010).

7 Duman, J. G. Animal ice-binding (antifreeze) proteins and glycolipids: an overview with emphasis on physiological function. *J Exp Biol* **218**, 1846-1855, doi:10.1242/jeb.116905 (2015).

8 Vaishnava, S. *et al.* The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255-258, doi:10.1126/science.1209791 (2011).

9 Sharon, N. & Lis, H. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* **14**, 53R-62R, doi:10.1093/glycob/cwh122 (2004).

10 Mimura, Y. *et al.* Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell*, doi:10.1007/s13238-017-0433-3 (2017).

11 Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3-49, doi:10.1093/glycob/cww086 (2017).

12 Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448-462, doi:10.1038/nrm3383 (2012).

13 Kaplan, H. A., Welply, J. K. & Lennarz, W. J. Oligosaccharyl transferase: the central enzyme in the pathway of glycoprotein assembly. *Biochim Biophys Acta* **906**, 161-173 (1987).

14 Van den Steen, P., Rudd, P. M., Dwek, R. A. & Opdenakker, G. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* **33**, 151-208, doi:10.1080/10409239891204198 (1998).

15 Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).

16 Amado, M., Almeida, R., Schwientek, T. & Clausen, H. Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. *Biochim Biophys Acta* **1473**, 35-53 (1999).

17 Laine, R. A. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **4**, 759-767 (1994).

18 Varki, A. Diversity in the sialic acids. *Glycobiology* **2**, 25-40 (1992).

19 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, doi:10.1093/glycob/cwv091 (2015).

20 Smit, C. H. *et al.* Surface expression patterns of defined glycan antigens change during Schistosoma mansoni cercarial transformation and development of schistosomula. *Glycobiology* **25**, 1465-1479, doi:10.1093/glycob/cwv066 (2015).
21 Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **4**, 181-191, doi:10.1038/nrm1052 (2003).
22 Tyrrell, D. *et al.* Structural requirements for the carbohydrate ligand of E-selectin. *Proc Natl Acad Sci U S A* **88**, 10372-10376 (1991).
23 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).
24 Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol Biosyst* **5**, 1087-1104, doi:10.1039/b907931a (2009).
25 Freeze, H. H. Genetic defects in the human glycome. *Nat Rev Genet* **7**, 537-551, doi:10.1038/nrg1894 (2006).
26 Niehues, R. *et al.* Carbohydrate-deficient glycoprotein syndrome type Ib. Phosphomannose isomerase deficiency and mannose therapy. *J Clin Invest* **101**, 1414-1420, doi:10.1172/JCI2350 (1998).
27 Marquardt, T. *et al.* Correction of leukocyte adhesion deficiency type II with oral fucose. *Blood* **94**, 3976-3985 (1999).
28 Tegtmeyer, L. C. *et al.* Multiple phenotypes in phosphoglucomutase 1 deficiency. *N Engl J Med* **370**, 533-542, doi:10.1056/NEJMoa1206605 (2014).
29 Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540-555, doi:10.1038/nrc3982 (2015).
30 Meany, D. L., Zhang, Z., Sokoll, L. J., Zhang, H. & Chan, D. W. Glycoproteomics for prostate cancer detection: changes in serum PSA glycosylation patterns. *J Proteome Res* **8**, 613-619, doi:10.1021/pr8007539 (2009).
31 Jorgensen, T. *et al.* Up-regulation of the oligosaccharide sialyl LewisX: a new prognostic parameter in metastatic prostate cancer. *Cancer Res* **55**, 1817-1819 (1995).
32 Thanabalasingham, G. *et al.* Mutations in HNF1A result in marked alterations of plasma glycan profile. *Diabetes* **62**, 1329-1337, doi:10.2337/db12-0880 (2013).
33 Lauc, G. *et al.* Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. *PLoS Genet* **6**, e1001256, doi:10.1371/journal.pgen.1001256 (2010).
34 Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol* **25**, 21-50, doi:10.1146/annurev.immunol.25.022106.141702 (2007).
35 Holland, M. *et al.* Differential glycosylation of polyclonal IgG, IgG-Fc and IgG-Fab isolated from the sera of patients with ANCA-associated systemic vasculitis. *Biochim Biophys Acta* **1760**, 669-677, doi:10.1016/j.bbagen.2005.11.021 (2006).
36 Shields, R. L. *et al.* Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J Biol Chem* **277**, 26733-26740, doi:10.1074/jbc.M202069200 (2002).
37 Thomann, M. *et al.* In vitro glycoengineering of IgG1 and its effect on Fc receptor binding and ADCC activity. *PLoS One* **10**, e0134949, doi:10.1371/journal.pone.0134949 (2015).
38 Thomann, M., Reckermann, K., Reusch, D., Prasser, J. & Tejada, M. L. Fc-galactosylation modulates antibody-dependent cellular cytotoxicity of therapeutic antibodies. *Mol Immunol* **73**, 69-75, doi:10.1016/j.molimm.2016.03.002 (2016).
39 Albrecht, S., Unwin, L., Muniyappa, M. & Rudd, P. M. Glycosylation as a marker for inflammatory arthritis. *Cancer Biomark* **14**, 17-28, doi:10.3233/CBM-130373 (2014).
40 van de Geijn, F. E. *et al.* Immunoglobulin G galactosylation and sialylation are associated with pregnancy-induced improvement of rheumatoid arthritis and the postpartum flare:

results from a large prospective cohort study. *Arthritis Res Ther* **11**, R193, doi:10.1186/ar2892 (2009).
41 Bondt, A. *et al.* Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res* **12**, 4522-4531, doi:10.1021/pr400589m (2013).
42 Hafkenscheid, L. *et al.* Structural Analysis of Variable Domain Glycosylation of Anti-Citrullinated Protein Antibodies in Rheumatoid Arthritis Reveals the Presence of Highly Sialylated Glycans. *Mol Cell Proteomics* **16**, 278-287, doi:10.1074/mcp.M116.062919 (2017).
43 Reusch, D. & Tejada, M. L. Fc glycans of therapeutic antibodies as critical quality attributes. *Glycobiology* **25**, 1325-1334, doi:10.1093/glycob/cwv065 (2015).
44 Zauner, G. *et al.* Glycoproteomic analysis of antibodies. *Mol Cell Proteomics* **12**, 856-865, doi:10.1074/mcp.R112.026005 (2013).
45 Knezevic, A. *et al.* Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology* **20**, 959-969, doi:10.1093/glycob/cwq051 (2010).
46 Ruhaak, L. R. *et al.* Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* **10**, 1667-1674, doi:10.1021/pr1009959 (2011).
47 Jansen, B. C. *et al.* Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep* **6**, 23296, doi:10.1038/srep23296 (2016).
48 Reiding, K. R. *et al.* Human Plasma N-glycosylation as Analyzed by Matrix-Assisted Laser Desorption/Ionization-Fourier Transform Ion Cyclotron Resonance-MS Associates with Markers of Inflammation and Metabolic Health. *Mol Cell Proteomics* **16**, 228-242, doi:10.1074/mcp.M116.065250 (2017).
49 Gornik, O. *et al.* Stability of N-glycan profiles in human plasma. *Glycobiology* **19**, 1547-1553, doi:10.1093/glycob/cwp134 (2009).
50 Hennig, R. *et al.* Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. *Biochim Biophys Acta* **1860**, 1728-1738, doi:10.1016/j.bbagen.2016.03.035 (2016).
51 Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491 (1988).
52 Geyer, H. & Geyer, R. Strategies for analysis of glycoprotein glycosylation. *Bba-Proteins Proteom* **1764**, 1853-1869, doi:10.1016/j.bbapap.2006.10.007 (2006).
53 Thompson, N. J., Rosati, S., Rose, R. J. & Heck, A. J. The impact of mass spectrometry on the study of intact antibodies: from post-translational modifications to structural analysis. *Chem Commun (Camb)* **49**, 538-548, doi:10.1039/c2cc36755f (2013).
54 Dalpathado, D. S. & Desaire, H. Glycopeptide analysis by mass spectrometry. *The Analyst* **133**, 731-738, doi:10.1039/b713816d (2008).
55 Morelle, W. & Michalski, J. C. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* **2**, 1585-1602, doi:10.1038/nprot.2007.227 (2007).
56 Gaunitz, S., Nagy, G., Pohl, N. L. & Novotny, M. V. Recent Advances in the Analysis of Complex Glycoproteins. *Anal Chem* **89**, 389-413, doi:10.1021/acs.analchem.6b04343 (2017).
57 Reusch, D. *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles--part 1: separation-based methods. *MAbs* **7**, 167-179, doi:10.4161/19420862.2014.986000 (2015).
58 Reusch, D. *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods. *MAbs* **7**, 732-742, doi:10.1080/19420862.2015.1045173 (2015).

59 Arnold, J. N. *et al.* Human serum IgM glycosylation - Identification of glycoforms that can bind to mannan-binding lectin. *Journal of Biological Chemistry* **280**, 29080-29087, doi:10.1074/jbc.M504528200 (2005).
60 Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. & Altmann, F. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics* **6**, 3369-3380, doi:10.1002/pmic.200500751 (2006).
61 Stumpo, K. A. & Reinhold, V. N. The N-glycome of human plasma. *J Proteome Res* **9**, 4823-4830, doi:10.1021/pr100528k (2010).
62 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).
63 Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).
64 Etxebarria, J., Calvo, J., Martin-Lomas, M. & Reichardt, N. C. Lectin-array blotting: profiling protein glycosylation in complex mixtures. *ACS Chem Biol* **7**, 1729-1737, doi:10.1021/cb300262x (2012).
65 de Boer, A. R., Hokke, C. H., Deelder, A. M. & Wuhrer, M. General microarray technique for immobilization and screening of natural glycans. *Anal Chem* **79**, 8107-8113, doi:10.1021/ac071187g (2007).
66 Stroop, C. J. *et al.* Characterization of the carbohydrate chains of the secreted form of the human epidermal growth factor receptor. *Glycobiology* **10**, 901-917 (2000).
67 Bones, J., Mittermayr, S., O'Donoghue, N., Guttman, A. & Rudd, P. M. Ultra performance liquid chromatographic profiling of serum N-glycans for fast and efficient identification of cancer associated alterations in glycosylation. *Anal Chem* **82**, 10208-10215, doi:10.1021/ac102860w (2010).
68 Callewaert, N., Geysens, S., Molemans, F. & Contreras, R. Ultrasensitive profiling and sequencing of N-linked oligosaccharides using standard DNA-sequencing equipment. *Glycobiology* **11**, 275-281 (2001).
69 Ruhaak, L. R. *et al.* Optimized workflow for preparation of APTS-labeled N-glycans allowing high-throughput analysis of human plasma glycomes using 48-channel multiplexed CGE-LIF. *J Proteome Res* **9**, 6655-6664, doi:10.1021/pr100802f (2010).
70 Goetz, J. A., Mechref, Y., Kang, P., Jeng, M. H. & Novotny, M. V. Glycomic profiling of invasive and non-invasive breast cancer cells. *Glycoconj J* **26**, 117-131, doi:10.1007/s10719-008-9170-4 (2009).
71 Khatri, K. *et al.* Microfluidic Capillary Electrophoresis-Mass Spectrometry for Analysis of Monosaccharides, Oligosaccharides, and Glycopeptides. *Anal Chem* **89**, 6645-6655, doi:10.1021/acs.analchem.7b00875 (2017).
72 Reusch, D. *et al.* High-throughput work flow for IgG Fc-glycosylation analysis of biotechnological samples. *Anal Biochem* **432**, 82-89, doi:10.1016/j.ab.2012.09.032 (2013).
73 Lauc, G. *et al.* Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet* **9**, e1003225, doi:10.1371/journal.pgen.1003225 (2013).
74 Ruhaak, L. R. *et al.* Targeted biomarker discovery by high throughput glycosylation profiling of human plasma alpha1-antitrypsin and immunoglobulin A. *PLoS One* **8**, e73082, doi:10.1371/journal.pone.0073082 (2013).
75 Shubhakar, A. *et al.* High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia* **78**, 321-333, doi:10.1007/s10337-014-2803-9 (2015).
76 de Hoffmann, E. & Stroobant, V. *Mass spectrometry: principles and applications*. 3rd edn, (2007).

77 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev* **18**, 349-450, doi:10.1002/(SICI)1098-2787(1999)18:6<349::AID-MAS1>3.0.CO;2-H (1999).
78 Blaum, K. High-accuracy mass spectrometry with stored ions. *Phys Rep* **425**, 1-78, doi:10.1016/j.physrep.2005.10.011 (2006).
79 Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom Rev* **17**, 1-35, doi:Doi 10.1002/(Sici)1098-2787(1998)17:1<1::Aid-Mas1>3.0.Co;2-K (1998).
80 Zhang, L. K., Rempel, D., Pramanik, B. N. & Gross, M. L. Accurate mass measurements by Fourier transform mass spectrometry. *Mass Spectrom Rev* **24**, 286-309, doi:10.1002/mas.20013 (2005).
81 Berger, C. Compensator Role of the Electrostatic Mirror in Time of Flight Mass-Spectrometry. *International Journal of Mass Spectrometry and Ion Processes* **46**, 63-66, doi:Doi 10.1016/0020-7381(83)80053-9 (1983).
82 Mamyrin, B. A. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *Int J Mass Spectrom* **206**, 251-266, doi:Doi 10.1016/S1387-3806(00)00392-4 (2001).
83 Harvey, D. J., Hunter, A. P., Bateman, R. H., Brown, J. & Critchley, G. Relationship between in-source and post-source fragment ions in the matrix-assisted laser desorption (ionization) mass spectra of carbohydrates recorded with reflectron time-of-flight mass spectrometers. *Int J Mass Spectrom* **188**, 131-146, doi:Doi 10.1016/S1387-3806(99)00014-7 (1999).
84 Stahl, B., Steup, M., Karas, M. & Hillenkamp, F. Analysis of Neutral Oligosaccharides by Matrix-Assisted Laser Desorption - Ionization Mass-Spectrometry. *Anal Chem* **63**, 1463-1466, doi:Doi 10.1021/Ac00014a022 (1991).
85 Beavis, R. C., Chaudhary, T. & Chait, B. T. Alpha-Cyano-4-Hydroxycinnamic Acid as a Matrix for Matrix-Assisted Laser Desorption Mass-Spectrometry. *Org Mass Spectrom* **27**, 156-158, doi:DOI 10.1002/oms.1210270217 (1992).
86 Selman, M. H. *et al.* MALDI-TOF-MS analysis of sialylated glycans and glycopeptides using 4-chloro-alpha-cyanocinnamic acid matrix. *Proteomics* **12**, 1337-1348, doi:10.1002/pmic.201100498 (2012).
87 Ciucanu, I. & Kerek, F. A Simple and Rapid Method for the Permethylation of Carbohydrates. *Carbohyd Res* **131**, 209-217, doi:Doi 10.1016/0008-6215(84)85242-8 (1984).
88 Wu, Z. *et al.* Characterization of O-acetylation in sialoglycans by MALDI-MS using a combination of methylamidation and permethylation. *Sci Rep* **7**, 46206, doi:10.1038/srep46206 (2017).
89 Klein, A. & Roussel, P. O-acetylation of sialic acids. *Biochimie* **80**, 49-57 (1998).
90 Kang, P., Mechref, Y. & Novotny, M. V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun Mass Spectrom* **22**, 721-734, doi:10.1002/rcm.3395 (2008).
91 Shubhakar, A. *et al.* Automated High-Throughput Permethylation for Glycosylation Analysis of Biologics Using MALDI-TOF-MS. *Anal Chem* **88**, 8562-8569, doi:10.1021/acs.analchem.6b01639 (2016).
92 Parry, S. *et al.* The sperm agglutination antigen-1 (SAGA-1) glycoforms of CD52 are O-glycosylated. *Glycobiology* **17**, 1120-1126, doi:10.1093/glycob/cwm076 (2007).
93 Boztug, K. *et al.* JAGN1 deficiency causes aberrant myeloid cell homeostasis and congenital neutropenia. *Nat Genet* **46**, 1021-+, doi:10.1038/ng.3069 (2014).
94 Nishikaze, T., Nakamura, T., Jinmei, H. & Amano, J. Negative-ion MALDI-MS2 for discrimination of alpha2,3- and alpha2,6-sialylation on glycopeptides labeled with a pyrene derivative. *J Chromatogr B Analyt Technol Biomed Life Sci* **879**, 1419-1428, doi:10.1016/j.jchromb.2010.10.032 (2011).

95 Amano, J. *et al.* Derivatization with 1-pyrenyldiazomethane enhances ionization of glycopeptides but not peptides in matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* **82**, 8738-8743, doi:10.1021/ac101555a (2010).

96 de Haan, N. *et al.* Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. *Anal Chem* **87**, 8284-8291, doi:10.1021/acs.analchem.5b02426 (2015).

97 Gomes de Oliveira, A. G. *et al.* A systematic study of glycopeptide esterification for the semi-quantitative determination of sialylation in antibodies. *Rapid Commun Mass Spectrom* **29**, 1817-1826, doi:10.1002/rcm.7287 (2015).

98 Pudelko, M. *et al.* Formation of lactones from sialylated MUC1 glycopeptides. *Org Biomol Chem* **4**, 713-720, doi:10.1039/b514918e (2006).

99 Wheeler, S. F., Domann, P. & Harvey, D. J. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of alpha(2 --> 3)- and alpha(2 --> 6)-isomers. *Rapid Commun Mass Spectrom* **23**, 303-312, doi:10.1002/rcm.3867 (2009).

100 Reiding, K. R., Lonardi, E., Hipgrave Ederveen, A. L. & Wuhrer, M. Ethyl Esterification for MALDI-MS Analysis of Protein Glycosylation. *Methods Mol Biol* **1394**, 151-162, doi:10.1007/978-1-4939-3341-9_11 (2016).

Chapter 2

High-throughput analysis and automation for glycomics studies

Review article

Archana Shubhakar[1,2], Karli R. Reiding[3], Richard A. Gardner[1], Daniel I. R. Spencer[1], Daryl L. Fernandes[1], Manfred Wuhrer[1,2]

[1]Ludger Ltd, Culham Science Centre, Abingdon, Oxfordshire UK;

[2]Division of BioAnalytical Chemistry, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands;

[3]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands;

Reprinted and adapted with permission from Chromatographia, 2015; 78(5-6): 321–333.

Copyright © Springer-Verlag Berlin Heidelberg 2014

2.1 Abstract

This review covers advances in analytical technologies for high-throughput (HTP) glycomics. Our focus is on structural studies of glycoprotein glycosylation to support biopharmaceutical realization and the discovery of glycan biomarkers for human disease. For biopharmaceuticals, there is increasing use of glycomics in Quality by Design studies to help optimize glycan profiles of drugs with a view to improving their clinical performance. Glycomics is also used in comparability studies to ensure consistency of glycosylation both throughout product development and between biosimilars and innovator drugs. In clinical studies there is as well an expanding interest in the use of glycomics — for example in Genome Wide Association Studies — to follow changes in glycosylation patterns of biological tissues and fluids with the progress of certain diseases. These include cancers, neurodegenerative disorders and inflammatory conditions. Despite rising activity in this field, there are significant challenges in performing large scale glycomics studies. The requirement is accurate identification and quantitation of individual glycan structures. However, glycoconjugate samples are often very complex and heterogeneous and contain many diverse branched glycan structures. In this article we cover HTP sample preparation and derivatization methods, sample purification, robotization, optimized glycan profiling by UHPLC, MS and multiplexed CE, as well as hyphenated techniques and automated data analysis tools. Throughout, we summarize the advantages and challenges with each of these technologies. The issues considered include reliability of the methods for glycan identification and quantitation, sample throughput, labor intensity, and affordability for large sample numbers.

2.2 Introduction

This review focuses on advances in analytical technologies for high-throughput (HTP) glycomics (i.e. the large-scale study of glycoconjugate glycosylation patterns). Our emphasis will be on glycomics to support (1) realization of glycoprotein biopharmaceuticals and (2) development of glycan biomarkers of disease for clinical diagnostics.

Glycans play key roles in important biological processes such as protein folding, host-pathogen interaction and signal transduction[1]. They are present both in free form, or bound to proteins or lipids, and can modify the biological activities of the conjugate. The glycomics technologies covered in this review are mainly for analysis of protein glycosylation, the major types of which are *N*-linked glycans attached at Asn-X-Ser/Thr motifs on the peptide backbone and *O*-linked glycans attached to Ser/Thr[2]. Protein glycosylation is a co-translational and post-translational modification where glycans are attached to proteins during translation, assisting in the folding and quality control of the protein. Attached glycans are subsequently modified in the endoplasmic reticulum and Golgi to varying levels of complexity, causing glycosylation to be one of the most complex and diverse types of protein modification.

Glycomics is of interest to biopharma companies because glycans can greatly modify safety and efficacy profiles of the therapeutic protein to which they are attached. For example, IgG effector functions such as antibody-dependent cellular cytotoxicity and complement-dependent cytotoxicity of monoclonal antibodies are modulated by *N*-glycosylation of its fragment crystallizable (Fc) portion[3]. *N*-Glycosylation modifications are also involved in determining the plasma half-life of glycoproteins by modulating the interaction with various receptors[4,5]. Even small changes in glycosylation can lead to serious issues such as anaphylaxis in patients and the destruction of therapeutic activity. These effects have been found in drugs such as Cetuximab (a monoclonal antibody (mAb) for treatment of metastatic colorectal cancer), Herceptin (a mAb for treatment of breast cancer) and erythropoietin (EPO—a drug for increasing red blood cells in patients with renal failure and cancer patients undergoing chemotherapy). Given this, there are now increasing pressures from regulatory authorities for drug manufacturers to measure, optimize and control their drug's glycosylation[6]. This task is very challenging given the structural complexity of biopharmaceutical glycosylation and achieving it has required adoption of a new approach to drug design—namely Quality by Design (QbD). This methodology has been actively promoted by the FDA and other drug regulators and is being adopted by leading biopharma companies as the system of best practice for the design, development, and manufacture of biologic drugs. QbD has the potential to deal with the complexities of drug glycans and could help simplify difficult tasks such as demonstrating comparability of biopharmaceutical glycosylation[7]. However, the glycoprofiling effort required for QbD is considerable. Studies of the QbD approach by the CMC Biotech Working Group (a consortium of experts from

leading biopharma organizations including Genentech, Amgen, Eli Lilly, and Pfizer), indicate that definition of the QbD Design Space may require analysis of tens of thousands of drug samples[8]. Successful adoption of QbD for development of glycoprotein therapeutics requires glycoprofiling systems that can cope with this very high level of sample throughput—in effect it will need to be built on a glycomics framework[6].

In addition to its use in biopharmaceutical realization, there is also increasing interest in glycomics for development of new clinical diagnostics based on glycosylation markers of diseases. This follows from the key roles that glycans play in major biological processes from the fertilization of eggs by sperm to cell death and their involvement in a multitude of disease etiologies and progressions. The potential of glycans as sensitive disease biomarkers is particularly high, since they are synthesized by the concerted action of multiple proteins, and influenced by various disease-associated factors such as genetic variations, epigenetic signatures and metabolic distortions. Glycan patterns of secreted proteins often confer information on the pathophysiological status of the secreting cells, making glycosylation analysis of relevant glycoproteins a suitable diagnostic tool. Changes in glycosylation are associated with various diseases like cancer, neurodegenerative and inflammatory diseases[9-15]. These types of associated changes have resulted in an increased interest in studying the alterations in glycosylation patterns of biological fluids as disease biomarkers as well as for patient stratification and personalized medicine[16].

To date, two glycan biomarkers have been introduced for routine clinical diagnostics, namely carbohydrate-deficient transferrin for the detection of alcohol abuse[17] and fucosylated serum alpha-fetoprotein for the early diagnosis of hepatocellular carcinoma[18]. Importantly, glycomic changes are not limited to single conditions but are rather a hallmark of virtually any human disease, as stated recently in a whitepaper by Walt, *et al.*[16], and diverse promising glycomic biomarkers have been described for MODY-type diabetes and immunoglobulin G myeloma[19,20]. However, most of the glycan biomarkers discovered so far have been studied using glycoanalytical technologies that would not be suitable for use in routine clinical diagnostic labs—the shortfall being due mainly to limitations in sample throughput, resolution and affordability of the methods.

Given the above, it is clear that development of reliable, affordable, high-resolution HTP glycomics technologies would be essential in the biopharma industry for efficient optimization of the glycosylation patterns of glycoprotein therapeutics as well as advancement of practical clinical diagnostics based on glycosylation biomarkers of disease.

The rest of this review will cover a number of recent advances in glycomics technologies for these two applications. It highlights current challenges in analytical glycomics methods and presents some approaches of how the bottlenecks in protein glycosylation analysis may be overcome.

2.3 High-throughput sample preparation

Many reviews describe glycan sample preparation steps such as *N*- and *O*-glycan release, de-glycosylated sample desalting and clean-up, derivatization of released glycans and analysis using different analytical platforms[21-25]. Here we describe a number of recent sample preparation methods for HTP glycosylation analysis. Next to their application in biomedical research, these methods are often applied for the characterization of biopharmaceuticals.

Liquid chromatography (LC) is a widely used glycoanalytical technique with good quantification, reproducibility, and separation of glycan isomers. Royle, *et al.*[26], have demonstrated a HTP sample preparation for high performance LC (HPLC) using a polyvinylidene fluoride (PVDF) membrane and an in-gel block method for glycan release. The enzymatic release step is followed by 2-aminobenzamide (2-AB) labeling and purification using microplates, and subsequent HPLC analysis. Automation of the procedure was achieved to a large extent, as well as the automated analysis of the obtained results, but the resulting method is still labor intensive, particularly the in-gel block preparation and glycan release stage. Performing the whole process on 96 samples takes 3 days with an additional 2 days for the acquisition of the chromatographic data. Cook, *et al.*[27], used the rapid deglycosylation kit and instant 2-AB kit from Prozyme to demonstrate an approach that only took 3.5 h of sample preparation time. A downside here is that the required kits are a considerable cost factor, which may limit the number of samples that can be studied. Burnina, *et al.*[28], on the other hand, have described a cost effective HTP sample preparation method for the characterization of therapeutic antibody *N*-glycosylation taking approximately 90 min. The workflow included the following steps, denaturation, reduction, deglycosylation using a hydrophobic Immobilon-P PVDF membrane filter plate, fluorescent labeling and clean-up using a 96-well hydrophilic filter plate. They demonstrated the use of orthogonal assays and obtained LC and MS which were in agreement.

Callewaert, *et al.*[29-31], reported capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF) using a DNA sequencer for the analysis and profiling of total serum *N*-glycans labeled with the negatively charged, fluorescent label 8-aminopyrene-1,3,6-trisulfonic acid (APTS). Ruhaak, *et al.*[32], have demonstrated the HTP application of this by optimizing sample preparation and employing a system with multiplexed capillaries. After protein denaturation and *N*-glycan release, the glycans were labeled with APTS in a 96 well plate and subjected to hydrophilic interaction chromatography (HILIC) purification. Using the multiplexed CGE-LIF system, 48 samples can be analyzed in parallel (scalable to 96 samples) allowing the analysis of 96 samples with a hands-on time of 2.5 h. A recent large-scale application making use of the multiplexed CGE-LIF system is the analysis of alpha-1-antitrypsin and immunoglobulin A from over 2,400 human plasma samples[33]. Set up to

discover novel biomarkers, the method revealed protein glycosylation to be associated with age and sex, as well as with cardiovascular and metabolic diseases.

Varadi, *et al.*[34], described the use of magnetic beads for rapid, HTP sample preparation of *N*-glycans from therapeutic antibodies and indicate that this method is easily automatable. The carboxyl coated magnetic beads are capable of capturing released *N*-glycans due to the ionic interaction between the positively charged glycosylamines and the negatively charged carboxylated beads. The glycans were captured using the magnetic beads and were labeled with APTS for 2 h. Once again the magnetic beads were used for binding the labeled glycans and elution was accomplished by washing the beads with water. This eluent was suitable for direct analysis by CE-LIF.

Mass spectrometric analysis of glycans is difficult when dealing with isobaric structures and particularly difficult when dealing with sialylated species. Sialic acids are known to degrade under harsh sample preparation and ionization conditions, as well as showing a strong preference towards negative ionization. Sialylated glycans are often derivatized before mass spectrometric analysis, either by carboxylic acid-specific methods or permethylation[35].

Gil, *et al.*[36], reported a one pot sialic acid amidation reaction using acetohydrazide, allowing neutralization of sialylated *N*-glycans still attached to proteins. Accidental Schiff base formation with the acetohydrazide is prevented because the glycans are still attached to the protein. *N*-glycan release could then be performed afterwards, as well as labeling with the negatively charged fluorescent label 2-aminobenzoic acid (2-AA). Clean-up was performed by 96-well filter plate, and analysis by MALDI-time of flight (TOF)-MS. The authors used recombinant human glycoproteins to determine changes in fucosylation and changes in sialylation and compared the MALDI–TOF–MS data with HILIC HPLC data proving that they were in good agreement. Another approach used by Jeong, *et al.*[37], for HTP quantitative *N*-glycan analysis was solid phase permethylation. Sialylated *N*-glycans were protected with a methyl group at the carboxylic acid by the permethylation reaction, and glycan mixtures were subsequently analyzed by MALDI–TOF–MS in positive ion mode. Sample preparation steps were performed in a 96-well format, comprising glycoprotein binding on PVDF, deglycosylation, purification of *N*-glycans by porous graphitized carbon (PGC), and solid phase permethylation. By spiking a known concentration of internal standard mixture with ovalbumin and porcine thyroglobulin, the authors obtained absolute quantitation for these samples, as well as relative quantification of 49 glycans from human serum prostate specific antigen. The complete sample preparation utilizing this method takes 2 days, but is quite labor intensive due to multiple manual pipetting steps.

Miura, *et al.*[38], have developed a protocol for the methyl esterification of sialic acids using 3-methyl-1-p-tolyltriazene (MTT) as sole reactant and methyl group donor. This

derivatization was performed on both the free *N*-glycans, requiring a subsequent purification step, as well as after immobilization of the *N*-glycans on Affi-gel beads, allowing a much increased throughput. Glycans could be analyzed afterwards by MALDI–TOF–MS with significantly increased stability. Using this method, both the neutral glycans and the previously negatively charged sialylated glycans could be observed in the same positive mode spectrum. Liu, *et al.*[39], have shown a similar increase in sialic acid stability after methylamidation of the carboxylic acid residues with a combination of methylamine and tripyrrolidinophosphonium hexafluorophosphate (PyAOP). They have shown this approach to be suitable for the derivatization of *O*-acetylated sialic acids, and were able to perform both MALDI and electrospray ionization (ESI)-MS (/MS) experiments on the glycans. Both the MTT and the PyAOP stabilization methods only require a few hours of preparation time (purification included), making them suitable for HTP glycomics.

Another variant of methods for sialic acid modification can make use of the chemical nature of α2,3-linked sialic acids to lactonize with the subterminal galactose resulting in a loss of water that can be observed in mass spectrometry, whilst α2,6-linked sialic acids show no such behavior. Published by Wheeler, *et al.*[40], the use of 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium (DMT–MM) in combination with methanol showed good reaction selectivity for the sialic acid linkages, generating either lactones or methyl esters. Alley, *et al.*[41], have used DMT-MM to generate amides instead of esters, performing the reaction in the presence of ammonium chloride. Subsequent permethylation still showed linkage-specific mass differences, with the added benefit of more informative fragmentation spectra. Another extension of the DMT-MM deriviatization method was shown by Tousi, *et al.*[42], who performed nano-HILIC-Orbitrap-MS of the modified glycans, effectively increasing the number of individual resolvable peaks after HILIC separation, as well as allowing linkage-specific MS(/MS). Similar to MTT and PyAOP, DMT-MM-assisted modification can be performed in less than 3 h (with the reaction conditions and reaction time depending on the modification to be introduced), and is highly suitable for HTP sample preparation.

An alternative to the DMT-MM method for linkage-specific modification has recently been published by us using the combination of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 1-hydroxybenzotriazole (HOBt) as chemicals for derivatization[43]. Performed in ethanol, the reaction yields highly linkage-specific products when incubating at 37 °C for 1 h (lactones for α2,3-linked sialic acids and ethyl esters for α2,6-linked sialic acids), with more resistance to side reactivity (such as amidation) than with DMT-MM. As a result, the method could be performed directly on impure glycan-containing mixtures such as a peptide-*N*-glycosidase F (PNGase F)-released plasma *N*-glycome. The protocol has been validated for HTP application in 96-well plate format, and shows consistent results with ultra-performance liquid chromatography (UPLC)-based methods of analysis (**Figure 1**). An

application has recently been published by Bondt, *et al.*[44], who have studied glycosylation changes on IgG fragment antigen-binding (Fab) and Fc, at different time points during and after pregnancy. By analyzing a total of 576 glycan profile spectra, pregnancy-associated changes were revealed in IgG Fab glycan sialylation, galactosylation, bisection, fucosylation, and high mannose structures.

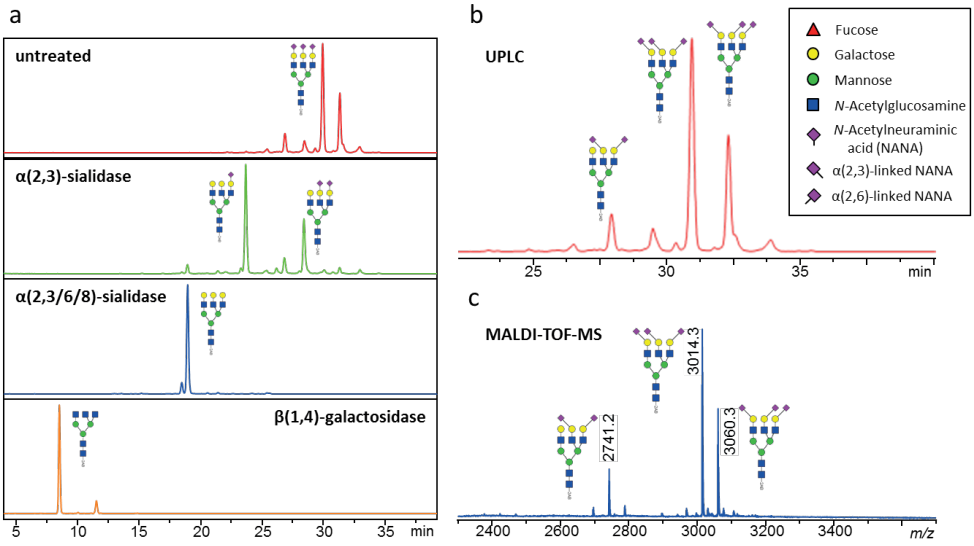


***Figure 1**. 2-AB-labeled A3 glycan standard from Ludger (CAB-A3-01) was analyzed by UPLC and MALDI–TOF–MS after ethyl esterification. **A)** Overnight incubation at 37 °C of Ludger 2-AB-labeled A3 standard with buffer (red), α(2,3)-sialidase (green), α(2,3/6/8)-sialidase (blue), α(2,3/6/8)-sialidase + β(1,4)-galactosidase (orange). Separation was performed on a Waters Acquity UPLC H-class system using an Acquity BEH 1.7 µm 2.1 × 150 mm glycan column. **B)** Linkage-specific assignment of the undigested UPLC data[101]. **C)** MALDI–TOF–MS spectrum of the 2-AB-labeled A3 standard ([M + Na]$^+$) after 1 h of ethyl esterification with EDC and HOBt at 37 °C and subsequent HILIC purification according to Reiding, et al.[43]. Profiles obtained from both methods are highly comparable and show similar ratios with regard to sialic acid occupancy and linkage. HILIC peak assignments are based on the exoglycosidase digests, as well as use of internal standards. Structural schemes of glycans are depicted following the CFG notation: N-acetylglucosamine (blue square), fucose (red triangle), mannose (green circle), galactose (yellow circle), N-acetylneuraminic acid (purple diamond). Known N-acetylneuraminic acid linkages are indicated by a left angle (α2,3) or right angle (α2,6), and otherwise unspecified.*

Modification and stabilization of glycoconjugates has recently been attempted by Nishikaze, *et al.*[45], showing a PyAOP-based methylamidation of all carboxylic acids present on sialylated glycopeptides. Not only does this derivatization protect the sialylated glycan species from breakdown, MALDI-quadrupole ion trap-TOF–MS fragmentation shows increased structural information as a consequence. Nishikaze and coworkers observed predominantly peptide fragmentation in positive mode collision-induced dissociation MS/MS, while negative mode led to more information on glycan structural features. Using

a different glycopeptide derivatization method, Amano, *et al.*[46], modified glycopeptides with 1-pyrenyldiazomethane, also specifically targeting carboxylic acids. Whilst this method leads to enhanced ionization for MALDI–MSn analysis, the subsequent exploration of the fragmentation has shown fragment ions reflecting predominantly sialic acid linkage cleavage[47].

For the HTP analysis of underivatized Fc IgG tryptic glycopeptides, Selman, *et al.*[48], have published a method employing MALDI-Fourier transform ion cyclotron resonance (FTICR)-MS. The advantage of this intermediate-pressure instrument is the tendency for labile substituents to remain intact without chemical stabilization, allowing directly for the analysis of sialylated glycopeptide species[49]. Using 96-well sample preparation, Selman, *et al.*, showed a highly repeatable analysis of 384 samples in less than 36 h, comprising all of the steps from capturing of the IgG to automatic processing of the spectra. Additionally, the same group established a HPLC–MS method for measuring IgG-Fc glycopeptides requiring 16 min analysis time per sample, employing a sheath-flow electrospray setup for enhanced sensitivity and robustness[50]. Large-scale applications of this method have been performed by Bondt, *et al.*[51], analyzing over 1,500 samples within 4 weeks with an internal standard coefficient of variation (CV) below 4%. The results indicated an association between IgG galactosylation and improvement of rheumatoid arthritis (RA) during pregnancy. Another application was performed by Rombouts, *et al.*[52], who have used the method to show changes in anti-citrullinated protein antibody Fc glycosylation even preceding the onset of RA. Both the FTICR and HPLC methods separate IgG glycopeptides on the basis of single amino acid differences (phenylalanine or tyrosine) allowing subclass-specific glycosylation analysis. An overview of various HTP methods for IgG glycosylation analysis has been published by Huffman, *et al.*[53], comparing UPLC with fluorescent detection, MALDI–TOF–MS, multiplexed CGE-LIF and LC–ESI–MS, as well as explaining the advantages and disadvantages of each method.

Another highly sensitive technique for analysis of protein glycosylation utilizes multiple reaction monitoring (MRM) which is an LC–MS/MS technique performed on a triple quadrupole (QqQ). MRM is known for its ability to quantify low abundant compounds in highly complex mixtures. For example, glycoproteins from biological fluids like serum or plasma are reduced, alkylated and digested with trypsin and analyzed without further fractionation[54]. In MRM glycans or glycopeptides are scanned in the first quadrupole (Q1) for user specified precursor ions and these precursors undergo CID in the second quadrupole (Q2) and fragments are produced. Finally the fragments of the precursor ions, known as transitions are scanned in the third quadrupole[55]. Song, *et al.*[55], used MRM for quantifying glycopeptides from both model glycoproteins and from depleted human blood serum. Hong, *et al.*[54,56], also reported the use of MRM for absolute quantitation of IgG glycopeptides and also published the use of MRM to study site specific glycosylation

alterations for identifying glycan biomarkers. The advantages of the technique are the very high sensitivity which together with the high selectivity allows for the direct analysis of the resulting glycopeptides on the QqQ without requiring prior protein enrichment or sample clean-up. A couple of drawbacks to this technique are that (1) MRM analyses are targeted, and analytes/glycopeptide species which are not included in the method set-up cannot be detected by data mining in hindsight. (2) A considerable amount of time and effort need to be invested for assay set-up and for establishing transitions for different analytes. These above mentioned MRM protocols were not adapted to HTP sample preparation, but can certainly be implemented for large sample numbers and MRM is envisaged to facilitate glycan biomarker discovery and protein characterization studies in the near future.

2.4 Use of internal standards

The analysis of glycosylation is made difficult by technical variation with respect to separation (retention time, *m/z* values) and intensity (fluorescent signal, ion detection). In particular, the analysis of complex samples is compromised by the presence of the biological matrix that can influence labeling as well as ionization efficacy and migration positions[57,58]. The use of standards has been found to assist tremendously in identification and quantification of glycans, and many interesting approaches have been shown in the literature to internalize standardization within HTP sample analysis. When using a DNA analyzer for CGE-LIF analysis of labeled glycans, the fact can be exploited that the system can analyze fluorescence on different wavelength channels. Using multiplexed CGE-LIF, co-injection of fluorescently labeled glycans and a base-pair nucleotide ladder fluorescing at a different wavelength was performed allowing for the correction of glycan migration times on the basis of the nucleotide ladder[32].

Glycan standards labeled with heavy isotopes facilitate quantitation during mass spectrometric analysis of derivatized glycans. For example, Prien, *et al.*[59], have quantitatively analyzed the *N*-glycosylation of several standard proteins by MALDI–TOF- and nanospray multistage-MS after labeling with normal 2-AA and a heavy ^{13}C version of 2-AA. The molecular weight difference between the two labels gives an offset of +6 Da. Mixing various ratios of the differently labeled samples revealed consistent quantitation by mass spectrometry. Using a different labeling chemistry, Walker, *et al.*[60], developed stable-isotope labeled hydrazide reagents for ESI–MS(/MS) analysis and showed quantitative detection after a sample preparation procedure that takes approximately 4 h. As another example, Zhang, *et al.*[61], have presented stable isotopically labeled phenyl-3-methyl-5-pyrazolone at the *N*-glycan reducing end with quantitation achievable within a 10-fold dynamic range.

Another labeling strategy for quantification lies in the use of isobaric tags. Variants of these labels have the same mass, but differ in the distribution of heavy isotopes on opposite sides

of a bond that readily undergoes fragmentation by CID. Reporter ions in MSn experiments can be used for relative quantification, making quantitation of highly complex samples intrinsically more robust than in MS mode quantitation, as the latter method can easily suffer from interfering, overlapping signals. An example of the use of isobaric tags in glycosylation analysis is provided by Ahn, *et al.*[62], who show LC–MS/MS data on tryptic glycopeptides derivatized with isobaric tags for relative and absolute quantitation (iTRAQ) at the peptide primary amines. Another version of isobaric tags are tandem mass tags (TMT), which have been used by Gong, *et al.*[63], to label released *N*-glycans still in their glycosylamine form. Glycosylamines are generated when PNGase F release is performed, but usually hydrolyze to the reducing sugar under most conditions. However, keeping the pH slightly alkaline preserved the glycosylamines and allowed for TMT labeling by *N*-hydroxysuccinimide ester chemistry[63]. Both iTRAQ and TMT not only facilitate internal standardization, but also allow the discrimination of multiple different samples during a single analysis, speeding up analysis time as a consequence.

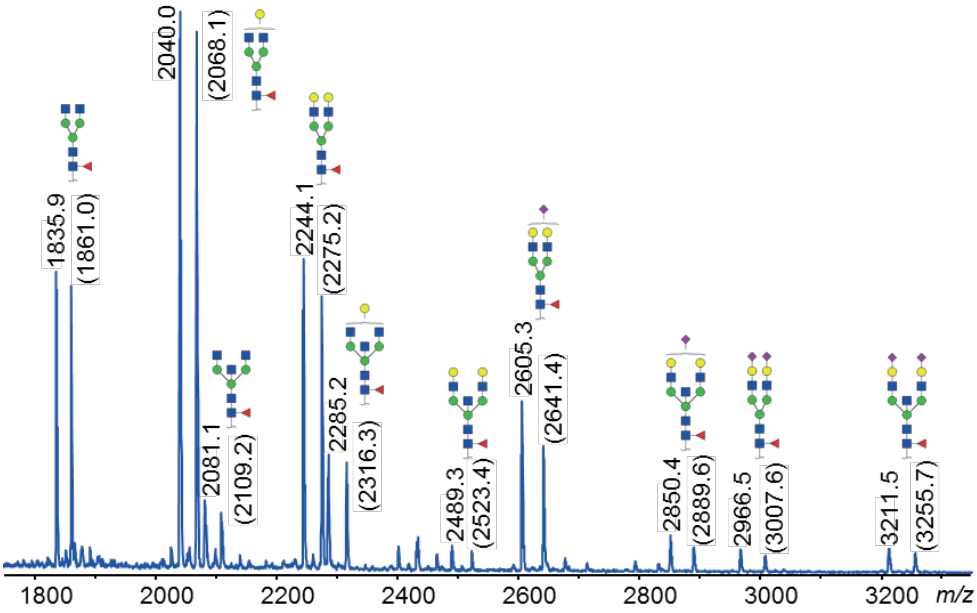


***Figure 2**. MALDI–TOF–MS spectrum of the ^{12}C and ^{13}C permethylated human-IgG N-glycan standards from Ludger (Cat# Cpm13C-IgG-01 and Cpm12C-IgG-01). ^{13}C was spiked with ^{12}C on the same sample spot to showcase the comparison of relative quantities of the major IgG N-glycans. The mass values shown in the spectra are $[M + Na]^+$ of permethylated glycans, with ^{13}C permethylated masses in parentheses.*

Permethylation provides another opportunity to generate internal standards. For example, when a ^{13}C permethylated glycan standard of known concentration is mixed with a ^{12}C permethylated glycan sample and analyzed by MS, relative glycan quantitation data can be

obtained. Manilla, *et al.*[64], have published the use of ^{13}C methyl iodide for permethylation, showing expected ratios at various concentrations when comparing to ^{12}C permethylated human milk oligosaccharides using MALDI–TOF–MS. ^{12}C and ^{13}C permethylated human IgG *N*-glycan standards developed at Ludger are an example of system suitability standards used for relative quantitation (**Figure 2**). When in possession of mass spectrometric instruments with high resolving power, Atwood, *et al.*[65], have shown isobaric variants can also be used for permethylation. Using $^{13}CH_3I$ and $^{12}CH_2DI$ as methyl donors, a minimal mass difference (0.0029 Da) is imparted upon the glycan permethylation site. When analyzed with high resolution instruments such as FTICR-MS the variants can be separated, achieving relative quantitation over two orders of magnitude, while a single peak can be generated for the two masses at lower resolution.

2.5 Integration of analysis workflows

Various groups have attempted to integrate different parts of the work-up procedure and simplified sample preparation steps. The integration of workflows has the potential to expedite analyses of a wide range of glycans, and the approaches can provide significant advantages in productivity, sensitivity, speed, and efficiency. Methods focusing on the integration of workflows are discussed below.

BlotGlyco kits produced by Sumitomo Bakelite in Japan integrate and support an "all-in-one" solution for automated and HTP analysis[66]. Glycoproteins are cleaved with trypsin, glycans are released by PNGase F, captured chemoselectively on hydrazide-functionalized beads coupled to 2-AB, and modified at the sialic acids by MTT as described previously. Subsequently, the *N*-glycans are released from the beads by reduction of sulfide bonds, but by that process retain 2-AB as a label. Using this integration method up to 96 samples can be purified and labeled in about 8 h. A newer method of Sumitomo Bakelite uses hydrazide beads to capture glycans in a process called glycoblotting, the captured glycans are purified, released from the beads and then labeled with 2AB on a 96 well filter plate[67].

Another integrated system is the Agilent mAb-glyco chip kit which has been designed for automated characterization of *N*-glycans from monoclonal antibodies (mAb)[68]. The system integrates on-chip deglycosylation of the mAb, HPLC chromatographic separation, quadrupole time-of-flight (QTOF)-MS analysis of the released glycans, as well as data analysis by Molecular Feature Extractor which is a part of Agilent's data analysis software MassHunter. A study comparing on-chip and in-solution deglycosylation methods showed that the analysis time could be reduced from 5 to 6 h to 12 min[69]. Another microchip-based integrated platform is the Caliper LabChip GXII Microchip-CE Platform, which uses the Caliper ProfilerPro glycan profiling kit to provide a HTP method for analyzing *N*-glycosylation patterns on mAbs and other glycosylated proteins[70]. The system contains plates for denaturation, deglycosylation, and glycan labeling. The analysis is performed on a LabChip

GXII CE instrument together with an internal standard. By this method, the relative abundance of the major *N*-glycans found on mAbs can be determined in a short span of time, requiring only 90 min for the preparation of 96 samples, although the glycan peak separation is compromised.

2.6 Automation of sample preparation

While the development of HTP methods for analysis of glycosylation facilitates the glycomic study of large clinical cohorts, performing manual sample preparation is still a time-consuming process. Additionally, while performing manual work-up is suitable for small sample sets, scaling up to larger sizes may lead to errors and inconsistency, resulting in poor repeatability. This set of circumstances has led to an increasing demand for automation of workflows, with laboratories requiring a system that is simple to operate, has scalable sample preparation and a process that is both repeatable and reliable. Currently, high quality glycomics data can be obtained using liquid handling robots for sample processing, clean-up and sample preparation[71]. A few approaches for automated and HTP glycan analysis will be reviewed.

Stöckmann, *et al.*, published an automated workflow for IgG *N*-glycan release, hydrazide bead capture of glycans through glycoblotting adapted from the protocol of Furukawa, *et al.*, mentioned above, 2-AB labeling, solid phase extraction (SPE), LC separation, and quantification. This workflow was developed for the efficient glycoprofiling of bio-therapeutics and clinical samples on the Hamilton Star liquid handling robot[72]. This approach was tested for the analysis of protein A captured IgG derived from CHO cells, as well as for the analysis of 100 samples of protein G-captured human IgG. The automated glycan preparation was verified with conventional and manual in-gel block release for the human serum IgG *N*-glycans, with CVs below 10% for all major glycan peaks, and with relative ratios in agreement with expectation.

A second integrated and automated technology developed specifically for biomolecule sample preparation is the Agilent AssayMAP platform, which is supported by the Bravo automated liquid-handling robot[73]. This workflow utilizes ProZyme GlykoPrep-plus chemistry kits, with an optional affinity purification step to capture glycoproteins from cell lysates, cell culture supernatants, or serum. Recoveries in glycoprotein purification approach 100% for most applications, with CVs of ≤5%. The purified glycoproteins are denatured and immobilized onto hydrophobic cartridges. Next, the immobilized glycoprotein samples are treated with PNGase F, the glycans fluorescently labeled, and cleaned on HILIC cartridges prior to analysis by HPLC, CE, or LC–MS[74]. The automation platform enables processing of complex samples, increases throughput, decreases variability and decreases assay time from a couple of days to 3–5 h. However, the cost of the kit could be a limiting factor when analyzing a large number of samples.

Another automated analytical workflow supported by the Hamilton STARlet liquid handling robot has been developed at Ludger for HTP glycomics for QbD studies. The automated workflow combines glycan release, 2-AB labeling, clean-up, and sample preparation for UPLC analysis. A validation study was performed, starting with 16 human IgG samples that were released by using PNGase F. Each sample was divided into four and the four subsets of samples were further divided equally into two 96 well PCR plates. Two operators used the liquid handling robot to label one plate each with 2-AB and cleaned-up using the liquid handling robot. The CVs for the major human IgG *N*-glycan peaks between the different operators were below 5%[75]. In addition, a fully robotized analysis for 48 samples of human IgG was performed, starting from the *N*-glycan release up to UHPLC sample preparation, showed CVs less than 5% for all peaks with average areas above 1% (**Figure 3**).

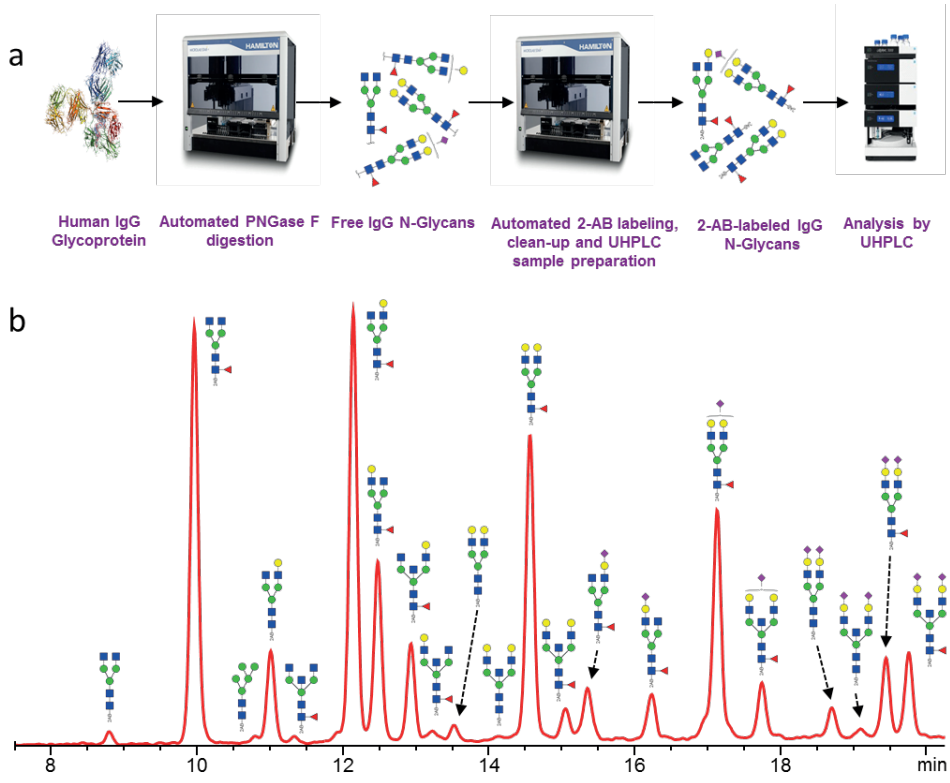


***Figure 3**. **A)** Workflow depicting validation study of semi-automated sample preparation of 48 replicates of human IgG samples, comprising digestion by PNGase F, 2-AB-labeling, HILIC SPE enrichment and preparation of samples for injection onto UHPLC performed with Hamilton Starlet liquid handling robot[75,102,103]. **B)** Typical UHPLC chromatogram of 2-AB-labeled human IgG glycans prepared using the robot. Analysis of the data from these 48 samples showed CVs less than 5% for all major peaks. The data shows that the automated, high-throughput method is repeatable. The liquid handling robot allows for fast, reliable and robust analyses of glycans. Glycan peaks in the chromatogram were assigned by exoglycosidase digestion, standard inclusion, and literature[104].*

For mass spectrometric analysis of glycopeptides, Reusch, *et al.*[76], have developed a fully automated HTP sample preparation method employing the Hamilton Microlab STAR to monitor mAb IgG Fc glycosylation by Orbitrap ESI–MS. They studied IgG from fermentation supernatant, automating the protein A capturing, tryptic digest, and HILIC-SPE, showing that the automation produces highly similar results when compared to HILIC-HPLC analysis of 2-AB labeled glycans, MALDI–TOF–MS and high-pH anion-exchange chromatography-pulsed amperometric detection of released glycans, as well as ESI-QTOF–MS of reduced mAb Fc. The glycopeptide sample preparation performed in this protocol facilitates discrimination between Fc and Fab portions, allows subclass-specific analysis, and is usable for a wide range of readout methods.

2.7 Data analysis

Innovations in sample preparation have increased the scope of glycomics considerably, allowing recently, for example, for large-scale GWAS[19,77]. However, increased data production leads to an accompanying need for robust and HTP data analysis procedures. The techniques focused on in this review allow the detection of large numbers of samples either by being HTP, or by featuring automation leading to decreased hands-on time. Such methods can be used for the discovery of analytes (glycans), but are most helpful for the repeated detection of similar glycosylation features across large cohorts. As a consequence, there is need for software capable of repeated and robust HTP feature extraction and quality control. Most analysis platforms and integrated methods have their software solutions, several having been mentioned above, but usually these are closed source, require licensing and may not cater to the functionality required by a user.

We recently reported an automated data extraction method for MALDI–TOF–MS spectra of *N*-glycans after sialic acid ethyl esterification[43]. This script requires Python, a programming language to be installed (freely available from www.python.org), which is capable of repeatedly integrating a list of glycan compositions from any number of mass spectra. The script calculates the expected glycan masses, allows a variety of modifications, takes the resulting isotopic distributions into account, and performs local background subtraction. As a downside, the program needs already calibrated data, and requires the user to still manually perform quality control.

Next to software for repeated analysis of data, several groups have published tools for the interpretation and assignment of glycomic and glycoproteomic data. These tools have been comprehensively reviewed by Woodin, *et al.*[78], and Dallas, *et al.*[79]. Notable and freely available examples of de novo interpretation of mass spectrometric data include GlycoPeakfinder[80], GlycoSpectrumScan[81], GlycoPep Grader[78], SysBioWare[82], GlycoMiner[83], and GlypID[84]. Interpretation by matching recorded data to a previously annotated database can be performed for mass spectrometric data by GlycoPeptide Search[85], GlycoSearchMS[86], and

GlycoPep DB[87], while HPLC derived glycan traits can be interpreted by GlycoExtractor[88] and autoGU[89].

Glycomic and glycoproteomic databases have been reviewed by Hizal, *et al.*[90], and Campbell, *et al.*[91], of note being the metadatabases GlycomeDB (containing entries from CFG, KEGG, GLYCOSCIENCES.de, BCSDB and Carbbank)[92], and UniCarbKB (containing entries from GlycosuiteDB, EUROCarbDB, UniCarb-DB and GlycoBase)[93]. GlycomeDB and UniCarbKB contain a wealth of theoretical and experimentally observed glycan structures, site-specific information on protein glycosylation, alongside various tools for retrieving and displaying the information.

Helpful with the graphical annotation of spectra are the web-based GlycanBuilder and DrawRings, allowing the facile construction of CFG or Oxford type glycan cartoons[94,95]. Particularly useful for glycan analysis is the desktop application GlycoWorkBench[96]. The program combines many aspects of glycan annotation and glycan spectral interpretation into a user-friendly package, incorporating GlycanBuilder for cartoon annotation, GlycoPeakfinder for theoretical structure prediction and interpretation, as well as several others for the automatic assignment of spectra, database searching, and many other functions.

All tools mentioned here are freely accessible web-based applets, or can be obtained without requiring payment.

2.8 Conclusion

Analysis of glycosylation is often a major challenge due to the vast macro-heterogeneity of glycoproteins, i.e., a single protein can have a varying number of occupied *N*- and *O*-glycosylation sites. Additionally, a single site can be occupied by a variety of glycans, which is known as micro-heterogeneity[2]. Furthermore, the analysis of glycans is complicated by variations in linkage and branching patterns which may lead to multiple isomers.

LC has long been the gold standard for analysis of fluorescently labeled glycans, providing good sensitivity, isomer separation, accurate glycan quantitation, repeatability, and in-depth characterization when supported by exoglycosidase sequencing. Downsides, however, are low sample throughput and relatively high cost per sample when comparing to, for example, MS-based methods of analysis[53]. MS on the other hand, provides good sensitivity, structural elucidation options via MSn experiments, good sensitivity, as well as very high throughput in case of MALDI ionization. Downsides here are the limitations in isomer separation and quantitation, and instability of labile groups when no derivatization is performed[97]. CGE-LIF supports high sensitivity, high throughput when multiplexed, and low running costs especially when comparing to LC. Unfortunately, databases for CGE-LIF annotation are small and could be a limiting factor.

Hyphenation of glycomic LC and CE methods with online MS detection, optionally in combination with fluorescence detection of labeled glycans, may add confidence in glycan structural assignment as both migration position and (tandem) mass spectrometric information can be combined[98,99]. However, it should be noted that certainly CE separation of glycans is often compromised by the restrictions faced with MS coupling. Further limitations in LC and CE hyphenation for glycomics are the relatively high costs due to medium sample throughput.

As demonstrated in this review, many of the analysis methods are moving towards increasingly HTP sample preparation, making use of multiwell formats, employing derivatizations with brief and manageable reaction times, integration of time-consuming steps and providing solutions for rapid analysis of the acquired data. Additionally, this progress towards simpler protocols provides the opportunity for liquid handling robots like the Hamilton and Bravo types to take over many of the time-consuming and labor- as well as cost-intensive steps. These automated platforms often show similar performance to expert manual labor, while even outperforming manual preparation when large cohorts are measured.

A common motif for HTP methodology is the inclusion of a solid support glycan binding or adsorption step. This can be used for purification, often in the form of covalent binding. When using chemical methods for capturing such as beads containing reactive hydrazide groups, this provides the opportunity for efficient labeling. This kind of integration is an excellent example of development towards shorter and easier protocols. Labeling itself is moving towards the use of heavy isotopes, to allow for reliable quantitation with mass spectrometry. The derivatization steps required for proper analysis of sialylated glycan species, notably permethylation and sialic acid-specific strategies, and several methods have been successful to date[100].

While mass spectrometry is one of the most HTP techniques possible for the analysis of glycans and glycoconjugates, it is hampered by its inability to provide quantitative data. This makes the various efforts undertaken for internal standardization, and thereby allowing quantitation, an exciting field to watch. The two main strategies for internal standardization are the use of heavy isotopic variants of glycans as well as the use of isobaric tags with unequal mass distribution. In both cases, the most convenient way to introduce mass differences is by employing derivatization steps already part of an analysis protocol, such as labeling or stabilization. A second strategy would be the spiking of a sample with a known concentration of a standard modified with a heavy isotope, which would also allow for the use of heavy isotopic variants of underivatized glycans. Convenient mass differences for internal standards are a few milliDalton showing parallel isotopic distributions on high resolution instruments, and onward from a few Dalton to prevent the overlapping of the natural istotopic distribution of glycans and glycopeptides.

With increasingly large studies being performed, the bottleneck of glycomics is shifting towards the data analysis. Current software for automation of analysis mainly focuses on interpretation rather than repeatable analysis of already interpreted and annotated spectra. There is a clear need for software allowing proper feature extraction, denoising, background subtraction, and most importantly of all, quality control both on individual glycans and complete spectra. Automated interpretation is mostly performed assisted by a predefined set of possibilities, or by matching experimental data to a database. While several programs exist to perform true unassisted de novo annotation, this approach has only shown successful on data that is of high quality. Particularly for hyphenated analysis methods such as LC–MS, manual annotation is still the way to go, and is a highly time consuming process.

Altogether, the methodological improvements discussed in this article show a clear trend towards more sensitive approaches, with faster analysis, lower costs and less hands-on time. These improvements allow for the increasing translation of glycomic approaches to clinical research, with envisioned applications in patient stratification and personalized medicine.

2.9 Acknowledgments

We would like to thank Louise Royle for expert assistance with the exoglycosidase digestion. This work was supported by the European Union (Seventh Framework Programme HighGlycan project, grant number 278535).

2.10 References

1 Varki, A. *et al. Essentials of Glycobiology*. 2nd edn, (2009).

2 Roth, Z., Yehezkel, G. & Khalaila, I. Identification and Quantification of Protein Glycosylation. *International Journal of Carbohydrate Chemistry* **2012**, 10, doi:10.1155/2012/640923 (2012).

3 Abes, R. & Teillaud, J. L. Impact of Glycosylation on Effector Functions of Therapeutic IgG. *Pharmaceuticals (Basel)* **3**, 146-157, doi:10.3390/ph3010146 (2010).

4 Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448-462, doi:10.1038/nrm3383 (2012).

5 Arey, B. J. in *Glycosylation* (ed Stefana Petrescu) Ch. 12 (InTech, 2012).

6 Fernandes, D. L. in *Biopharmaceutical Production Technology* 545-583 (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).

7 Fernandes, D. L. *A QbD approach to biopharmaceutical glycosylation.*, (Biopharm Knowledge Publishing, Hampshire, UK, 2008).

8 CMC Biotech Working Group. *A-Mab: a case study in bioprocess development.*, <http://www.ispe.org/pqli/a-mab-case-studyversion-2.1> (2009).

9 Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation [mdash] potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477-488 (2005).

10 Dennis, J. W., Granovsky, M. & Warren, C. E. Protein glycosylation in development and disease. *Bioessays* **21**, 412-421, doi:Doi 10.1002/(Sici)1521-1878(199905)21:5<412::Aid-Bies8>3.0.Co;2-5 (1999).

11 Theodoratou, E. *et al.* The role of glycosylation in IBD. *Nat Rev Gastro Hepat* **11**, 588-600, doi:10.1038/nrgastro.2014.78 (2014).
12 Kapur, R. *et al.* Low anti-RhD IgG-Fc-fucosylation in pregnancy: a new variable predicting severity in haemolytic disease of the fetus and newborn. *Brit J Haematol* **166**, 936-945, doi:10.1111/bjh.12965 (2014).
13 Albrecht, S., Unwin, L., Muniyappa, M. & Rudd, P. M. Glycosylation as a marker for inflammatory arthritis. *Cancer Biomarkers* **14**, 17-28, doi:10.3233/Cbm-130373 (2014).
14 Stockmann, H., Adamczyk, B., Hayes, J. & Rudd, P. M. Automated, High-Throughput IgG-Antibody Glycoprofiling Platform. *Anal Chem* **85**, 8841-8849, doi:10.1021/ac402068r (2013).
15 Adamczyk, B., Tharmalingam, T. & Rudd, P. M. Glycans as cancer biomarkers. *Bba-Gen Subjects* **1820**, 1347-1353, doi:10.1016/j.bbagen.2011.12.001 (2012).
16 Walt, D. Committee on assessing the importance and impact of glycomics and glycosciences board on chemical sciences and technology board on life sciences division on earth and life studies. (2012).
17 Bortolotti, F., De Paoli, G. & Tagliaro, F. Carbohydrate-deficient transferrin (CDT) as a marker of alcohol abuse: A critical review of the literature 2001-2005. *J Chromatogr B* **841**, 96-109, doi:10.1016/j.jchromb.2006.05.005 (2006).
18 Gish, R. G. Early Detection of Hepatocellular Carcinoma Through Surveillance Using Biomarkers. *Gastroenterology & Hepatology* **10**, 121-123 (2014).
19 Lauc, G. *et al.* Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. *PLoS Genet* **6**, e1001256, doi:10.1371/journal.pgen.1001256 (2010).
20 Novokmet, M. *et al.* Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* **4**, 4347, doi:10.1038/srep04347 (2014).
21 Carlson, D. M. Structures and Immunochemical Properties of Oligosaccharides Isolated from Pig Submaxillary Mucins. *Journal of Biological Chemistry* **243**, 616-+ (1968).
22 ONeill, R. A. Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis. *J Chromatogr A* **720**, 201-215, doi:Doi 10.1016/0021-9673(95)00502-1 (1996).
23 Geyer, H. & Geyer, R. Strategies for analysis of glycoprotein glycosylation. *Bba-Proteins Proteom* **1764**, 1853-1869, doi:10.1016/j.bbapap.2006.10.007 (2006).
24 Wada, Y. *et al.* Comparison of Methods for Profiling O-Glycosylation HUMAN PROTEOME ORGANISATION HUMAN DISEASE GLYCOMICS/PROTEOME INITIATIVE MULTI-INSTITUTIONAL STUDY OF IgA1. *Molecular & Cellular Proteomics* **9**, 719-727, doi:10.1074/mcp.M900450-MCP200 (2010).
25 Zauner, G. *et al.* Protein O-glycosylation analysis. *Biol Chem* **393**, 687-708, doi:10.1515/hsz-2012-0144 (2012).
26 Royle, L. *et al.* HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal Biochem* **376**, 1-12, doi:10.1016/j.ab.2007.12.012 (2008).
27 Cook, K. S., Bullock, K. & Sullivan, T. Development and qualification of an antibody rapid deglycosylation method. *Biologicals* **40**, 109-117, doi:10.1016/j.biologicals.2011.12.010 (2012).
28 Burnina, I., Hoyt, E., Lynaugh, H., Li, H. J. & Gong, B. A cost-effective plate-based sample preparation for antibody N-glycan analysis. *J Chromatogr A* **1307**, 201-206, doi:10.1016/j.chroma.2013.07.104 (2013).
29 Callewaert, N., Geysens, S., Molemans, P. & Contreras, R. Ultrasensitive profiling and sequencing of N-linked oligosaccharides using standard DNA-sequencing equipment. *Glycobiology* **11**, 275-281, doi:DOI 10.1093/glycob/11.4.275 (2001).
30 Callewaert, N. *et al.* Noninvasive diagnosis of liver cirrhosis using DNA sequencer-based total serum protein glycomics. *Nat Med* **10**, 429-434, doi:10.1038/nm1006 (2004).

31 Laroy, W., Contreras, R. & Callewaert, N. Glycome mapping on DNA sequencing equipment. *Nat Protoc* **1**, 397-405, doi:10.1038/nprot.2006.60 (2006).
32 Ruhaak, L. R. *et al.* Optimized workflow for preparation of APTS-labeled N-glycans allowing high-throughput analysis of human plasma glycomes using 48-channel multiplexed CGE-LIF. *J Proteome Res* **9**, 6655-6664, doi:10.1021/pr100802f (2010).
33 Ruhaak, L. R. *et al.* Targeted biomarker discovery by high throughput glycosylation profiling of human plasma alpha1-antitrypsin and immunoglobulin A. *PLoS One* **8**, e73082, doi:10.1371/journal.pone.0073082 (2013).
34 Varadi, C., Lew, C. & Guttman, A. Rapid Magnetic Bead Based Sample Preparation for Automated and High Throughput N-Glycan Analysis of Therapeutic Antibodies. *Anal Chem* **86**, 5682-5687, doi:10.1021/ac501573g (2014).
35 Harvey, D. J. Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *J Chromatogr B* **879**, 1196-1225, doi:10.1016/j.jchromb.2010.11.010 (2011).
36 Gil, G. C., Iliff, B., Cerny, R., Velander, W. H. & Van Cottt, K. E. High Throughput Quantification of N-Glycans Using One-Pot Sialic Acid Modification and Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *Anal Chem* **82**, 6613-6620, doi:10.1021/ac1011377 (2010).
37 Jeong, H. J., Kim, Y. G., Yang, Y. H. & Kim, B. G. High-Throughput Quantitative Analysis of Total N-Glycans by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Anal Chem* **84**, 3453-3460, doi:10.1021/ac203440c (2012).
38 Miura, Y., Shinohara, Y., Furukawa, J. I., Nagahori, N. & Nishimurara, S. Rapid and simple solid-phase esterification of sialic acid residues for quantitative glycomics by mass spectrometry. *Chem-Eur J* **13**, 4797-4804, doi:10.1002/chem.200601872 (2007).
39 Liu, X., Qiu, H. Y., Lee, R. K., Chen, W. X. & Li, J. J. Methylamidation for Sialoglycomics by MALDI-MS: A Facile Derivatization Strategy for Both alpha 2,3-and alpha 2,6-Linked Sialic Acids. *Anal Chem* **82**, 8300-8306, doi:10.1021/ac101831t (2010).
40 Wheeler, S. F., Domann, P. & Harvey, D. J. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of alpha(2 --> 3)- and alpha(2 --> 6)-isomers. *Rapid Commun Mass Spectrom* **23**, 303-312, doi:10.1002/rcm.3867 (2009).
41 Alley, W. R. & Novotny, M. V. Glycomic Analysis of Sialic Acid Linkages in Glycans Derived from Blood Serum Glycoproteins. *J Proteome Res* **9**, 3062-3072, doi:10.1021/pr901210r (2010).
42 Tousi, F., Bones, J., Hancock, W. S. & Hincapie, M. Differential Chemical Derivatization Integrated with Chromatographic Separation for Analysis of Isomeric Sialylated N-Glycans: A Nano-Hydrophilic Interaction Liquid Chromatography-MS Platform. *Anal Chem* **85**, 8421-8428, doi:10.1021/ac4018007 (2013).
43 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).
44 Bondt, A. *et al.* Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics* **13**, 3029-3039, doi:10.1074/mcp.M114.039537 (2014).
45 Nishikaze, T., Kawabata, S. & Tanaka, K. In-Depth Structural Characterization of N-Linked Glycopeptides Using Complete Derivatization for Carboxyl Groups Followed by Positive- and Negative-Ion Tandem Mass Spectrometry. *Anal Chem* **86**, 5360-5369, doi:10.1021/ac500340t (2014).
46 Amano, J. *et al.* Derivatization with 1-Pyrenyldiazomethane Enhances Ionization of Glycopeptides but Not Peptides in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal Chem* **82**, 8738-8743, doi:10.1021/ac101555a (2010).

47 Nishikaze, T., Nakamura, T., Jinmei, H. & Amano, J. Negative-ion MALDI-MS2 for discrimination of α2,3- and α2,6-sialylation on glycopeptides labeled with a pyrene derivative. *Journal of Chromatography B* **879**, 1419-1428, doi:http://dx.doi.org/10.1016/j.jchromb.2010.10.032 (2011).

48 Selman, M. H. J. *et al.* Immunoglobulin G Glycopeptide Profiling by Matrix-Assisted Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* **82**, 1073-1081, doi:10.1021/ac9024413 (2010).

49 O'Connor, P. B. *et al.* A high pressure matrix-assisted laser desorption ion source for Fourier transform mass spectrometry designed to accommodate large targets with diverse surfaces. *J Am Soc Mass Spectr* **15**, 128-132, doi:10.1016/j.jasms.2003.10.005 (2004).

50 Selman, M. H. J. *et al.* Fc specific IgG glycosylation profiling by robust nano-reverse phase HPLC-MS using a sheath-flow ESI sprayer interface. *J Proteomics* **75**, 1318-1329, doi:10.1016/j.jprot.2011.11.003 (2012).

51 Bondt, A. *et al.* Association between Galactosylation of Immunoglobulin G and Improvement of Rheumatoid Arthritis during Pregnancy Is Independent of Sialylation. *J Proteome Res* **12**, 4522-4531, doi:10.1021/pr400589m (2013).

52 Rombouts, Y. *et al.* Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis. *Ann Rheum Dis* **74**, 234-241, doi:10.1136/annrheumdis-2013-203565 (2015).

53 Huffman, J. E. *et al.* Comparative Performance of Four Methods for High-throughput Glycosylation Analysis of Immunoglobulin G in Genetic and Epidemiological Research. *Molecular & Cellular Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).

54 Hong, Q. T., Lebrilla, C. B., Miyamoto, S. & Ruhaak, L. R. Absolute Quantitation of Immunoglobulin G and Its Glycoforms Using Multiple Reaction Monitoring. *Anal Chem* **85**, 8585-8593, doi:10.1021/ac4009995 (2013).

55 Song, E. W., Pyreddy, S. & Mechref, Y. Quantification of glycopeptides by multiple reaction monitoring liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Sp* **26**, 1941-1954, doi:10.1002/rcm.6290 (2012).

56 Hong, Q. T. *et al.* Label-Free Absolute Quantitation of Oligosaccharides Using Multiple Reaction Monitoring. *Anal Chem* **86**, 2640-2647, doi:10.1021/ac404006z (2014).

57 Mao, W. J., Thanawiroon, C. & Linhardt, R. J. Capillary electrophoresis for the analysis of glycosaminoglycans and glycosaminoglycan-derived oligosaccharides. *Biomed Chromatogr* **16**, 77-94, doi:10.1002/bmc.153 (2002).

58 Domann, P. J. *et al.* Separation-based Glycoprofiling Approaches using Fluorescent Labels. *Proteomics* **7**, 70-76, doi:10.1002/pmic.200700640 (2007).

59 Prien, J. M., Prater, B. D., Qin, Q. & Cockrill, S. L. Mass Spectrometric-Based Stable Isotopic 2-Aminobenzoic Acid Glycan Mapping for Rapid Glycan Screening of Biotherapeutics. *Anal Chem* **82**, 1498-1508, doi:10.1021/ac902617t (2010).

60 Walker, S. H., Budhathoki-Uprety, J., Novak, B. M. & Muddiman, D. C. Stable-Isotope Labeled Hydrophobic Hydrazide Reagents for the Relative Quantification of N-Linked Glycans by Electrospray Ionization Mass Spectrometry. *Anal Chem* **83**, 6738-6745, doi:10.1021/ac201376q (2011).

61 Zhang, P. *et al.* Relative quantitation of glycans using stable isotopic labels 1-(d0/d5) phenyl-3-methyl-5-pyrazolone by mass spectrometry. *Anal Biochem* **418**, 1-9, doi:10.1016/j.ab.2011.07.006 (2011).

62 Ahn, J. M. *et al.* Integrated Glycoproteomics Demonstrates Fucosylated Serum Paraoxonase 1 Alterations in Small Cell Lung Cancer. *Molecular & Cellular Proteomics* **13**, 30-48, doi:10.1074/mcp.M113.028621 (2014).

63 Gong, B. *et al.* N-Glycosylamine-mediated isotope labeling for mass spectrometry-based quantitative analysis of N-linked glycans. *Anal Bioanal Chem* **405**, 5825-5831, doi:10.1007/s00216-013-6988-9 (2013).

64 Alvarez-Manilla, G. *et al.* Tools for glycomics: relative quantitation of glycans by isotopic permethylation using (CH3I)-C-13. *Glycobiology* **17**, 677-687, doi:10.1093/glycob/cwm033 (2007).

65 Atwood, J. A., 3rd *et al.* Quantitation by isobaric labeling: applications to glycomics. *J Proteome Res* **7**, 367-374, doi:10.1021/pr070476i (2008).

66 Miura, Y. *et al.* BlotGIycoABC (TM), an integrated glycoblotting technique for rapid and large scale clinical glycomics. *Molecular & Cellular Proteomics* **7**, 370-377, doi:10.1074/mcp.M700377-MCP200 (2008).

67 Furukawa, J. I. *et al.* Comprehensive approach to structural and functional glycomics based on chemoselective glycoblotting and sequential tag conversion. *Anal Chem* **80**, 1094-1101, doi:10.1021/ac702124d (2008).

68 Bynum, M. A. *et al.* Characterization of IgG N-Glycans Employing a Microfluidic Chip that Integrates Glycan Cleavage, Sample Purification, LC Separation, and MS Detection. *Anal Chem* **81**, 8818-8825, doi:10.1021/ac901326u (2009).

69 Trojer, L., Gromadski, K. & Van Cottt, K. E. *The mAb-Glyco Chip Kit—a workflow solution for rapid and fully automated characterisation of N-linked glycans from monoclonal antibodies*, <http://www.chromatographytoday.com/article_read/1066/> (2011).

70 Caliper LifeSciences. *Rapid analysis of N-glycans on the LabChip GXII Microchip-CE platform. Application note 403.* , <http://www.level.com.tw/html/ezcatfiles/vipweb20/img/img/52857/(4)LCGX-Nglycan.pdf> (2010).

71 Doherty, M. *et al.* An automated robotic platform for rapid profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture. *Anal Biochem* **442**, 10-18, doi:10.1016/j.ab.2013.07.005 (2013).

72 Stockmann, H., O'Flaherty, R., Adamczyk, B., Saldova, R. & Rudd, P. M. Automated, high-throughput serum glycoprofiling platform. *Integr Biol (Camb)* **7**, 1026-1032, doi:10.1039/c5ib00130g (2015).

73 Russell, J. D., van den Heuvel, Z., Bovee, M. & Murphy, S. *Automation for LC/MS sample preparation: high throughput in-solution digestion and peptide cleanup enabled by the agilent AssayMAP Bravo Platform.*, <http://www.chem.agilent.com/Library/applications/5991-2957EN.pdf> (2013).

74 Murphy, S. *et al. Characterize N-glycans using a new quantitative, automated sample preparation platform. Agilent Technologies poster.* , <http://www.chem.agilent.com/Library/posters/Public/WCBP2013-GlykoPrep%20Poster%202013-01-26.pdf> (2013).

75 Hamilton. N-glycan characterisation of biopharmaceuticals: simplifying sample preparation with the Microlab STARlet.

76 Reusch, D. *et al.* High-throughput work flow for IgG Fc-glycosylation analysis of biotechnological samples. *Anal Biochem* **432**, 82-89, doi:10.1016/j.ab.2012.09.032 (2013).

77 Ruhaak, L. R. *et al.* Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* **10**, 1667-1674, doi:10.1021/pr1009959 (2011).

78 Woodin, C. L. *et al.* GlycoPep Grader: A Web-Based Utility for Assigning the Composition of N-Linked Glycopeptides. *Anal Chem* **84**, 4821-4829, doi:10.1021/ac300393t (2012).

79 Dallas, D. C., Martin, W. F., Hua, S. & German, J. B. Automated glycopeptide analysis-review of current state and future directions. *Brief Bioinform* **14**, 361-374, doi:10.1093/bib/bbs045 (2013).

80 Maass, K., Ranzingei, R., Geyer, H., von der Lieth, C. W. & Geyer, R. "Glyco-peakfinder" - de novo composition analysis of glycoconjugates. *Proteomics* **7**, 4435-4444, doi:10.1002/pmic.200700253 (2007).

81 Deshpande, N., Jensen, P. H., Packer, N. H. & Kolarich, D. GlycoSpectrumScan: Fishing Glycopeptides from MS Spectra of Protease Digests of Human Colostrum sIgA. *J Proteome Res* **9**, 1063-1075, doi:10.1021/pr900956x (2010).

82 Vakhrushev, S. Y., Dadimov, D. & Peter-Katalinic, J. Software Platform for High-Throughput Glycomics. *Anal Chem* **81**, 3252-3260, doi:10.1021/ac802408f (2009).

83 Ozohanics, O. *et al.* GlycoMiner: a new software tool to elucidate glycopeptide composition. *Rapid Commun Mass Sp* **22**, 3245-3254, doi:10.1002/rcm.3731 (2008).

84 Wu, Y. *et al.* Mapping site-specific protein N-glycosylations through liquid chromatography/mass spectrometry and targeted tandem mass spectrometry. *Rapid Commun Mass Sp* **24**, 965-972, doi:10.1002/rcm.4474 (2010).

85 Pompach, P., Chandler, K. B., Lan, R., Edwards, N. & Goldman, R. Semi-Automated Identification of N-Glycopeptides by Hydrophilic Interaction Chromatography, nano-Reverse-Phase LC-MS/MS, and Glycan Database Search. *J Proteome Res* **11**, 1728-1740, doi:10.1021/pr201183w (2012).

86 Lohmann, K. K. & von der Lieth, C. W. GlycoFragment and GlycoSearchMS: web tools to support the interpretation of mass spectra of complex carbohydrates. *Nucleic Acids Res* **32**, W261-W266, doi:10.1093/nar/gkh392 (2004).

87 Go, E. P. *et al.* GlycoPep DB: A tool for glycopeptide analysis using a "smart search". *Anal Chem* **79**, 1708-1713, doi:10.1021/ac061548c (2007).

88 Artemenko, N. V., Campbell, M. P. & Rudd, P. M. GlycoExtractor: A Web-Based Interface for High Throughput Processing of HPLC-Glycan Data. *J Proteome Res* **9**, 2037-2041, doi:10.1021/pr901213u (2010).

89 Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A. & Rudd, P. M. GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics* **24**, 1214-1216, doi:10.1093/bioinformatics/btn090 (2008).

90 Hizal, D. B. *et al.* Glycoproteomic and glycomic databases. *Clin Proteom* **11**, doi:Artn 1510.1186/1559-0275-11-15 (2014).

91 Campbell, M. P. *et al.* Toolboxes for a standardised and systematic study of glycans. *Bmc Bioinformatics* **15**, doi:Artn S910.1186/1471-2105-15-S1-S9 (2014).

92 Ranzinger, R., Herget, S., von der Lieth, C. W. & Frank, M. GlycomeDB-a unified database for carbohydrate structures. *Nucleic Acids Res* **39**, D373-D376, doi:10.1093/nar/gkq1014 (2011).

93 Campbell, M. P. *et al.* UniCarbKB: building a knowledge platform for glycoproteomics. *Nucleic Acids Res* **42**, D215-D221, doi:10.1093/nar/gkt1128 (2014).

94 Damerell, D. *et al.* The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol Chem* **393**, 1357-1362, doi:10.1515/hsz-2012-0135 (2012).

95 Akune, Y., Hosoda, M., Kaiya, S., Shinmachi, D. & Aoki-Kinoshita, K. F. The RINGS Resource for Glycome Informatics Analysis and Data Mining on the Web. *Omics* **14**, 475-486, doi:10.1089/omi.2009.0129 (2010).

96 Ceroni, A. *et al.* GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of Glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).

97 Alley, W. R., Mann, B. F. & Novotny, M. V. High-sensitivity Analytical Approaches for the Structural Characterization of Glycoproteins. *Chem Rev* **113**, 2668-2732, doi:10.1021/cr3003714 (2013).

98 Pabst, M., Bondili, J. S., Stadlmann, J., Mach, L. & Altmann, F. Mass plus 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* **79**, 5051-5057, doi:10.1021/ac070363i (2007).

99 Zhao, S. S., Zhong, X. F., Tie, C. & Chen, D. D. Y. Capillary electrophoresis-mass spectrometry for analysis of complex samples. *Proteomics* **12**, 2991-3012, doi:10.1002/pmic.201200221 (2012).

100 Alley, W. R. & Novotny, M. V. Structural Glycomic Analyses at High Sensitivity: A Decade of Progress. *Annu Rev Anal Chem* **6**, 237-265, doi:10.1146/annurev-anchem-062012-092609 (2013).

101 Ahn, J., Bones, J., Yu, Y. Q., Rudd, P. M. & Gilar, M. Separation of 2-aminobenzamide labeled glycans using hydrophilic interaction chromatography columns packed with 1.7 mu m sorbent. *J Chromatogr B* **878**, 403-408, doi:10.1016/j.jchromb.2009.12.013 (2010).

102 Fernandes, D. L. *Turning to colour*, <http://www.ludger.com/docs/articles/EBR-Glycan-Labelling-Summer-2009.pdf> (2009).

103 Tharmalingam, T., Adamczyk, B., Doherty, M. A., Royle, L. & Rudd, P. M. Strategies for the profiling, characterisation and detailed structural analysis of N-linked oligosaccharides. *Glycoconj J* **30**, 137-146, doi:10.1007/s10719-012-9443-9 (2013).

104 Lauc, G. *et al.* Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. *PLoS Genet* **9**, doi:ARTN e100322510.1371/journal.pgen.1003225 (2013).

Part II – Method development

Chapter 3

High-throughput profiling of protein *N*-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification

Research article

Karli R. Reiding[1], Dennis Blank[1], Dennis M. Kuijper[1,2], André M. Deelder[1], Manfred Wuhrer[1,2]

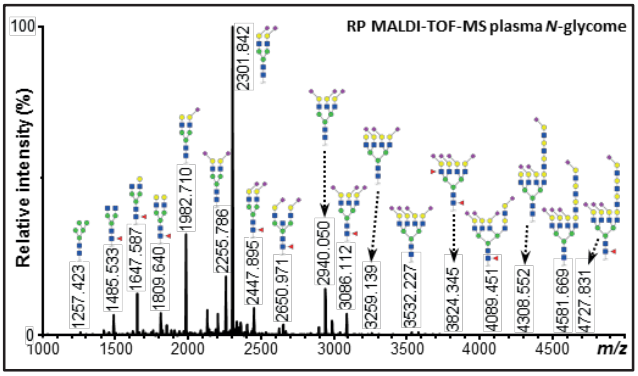
[1]Center for Proteomics and Metabolomics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

[2]Division of BioAnalytical Chemistry, VU University Amsterdam, 1081 HV Amsterdam, The Netherlands;

Reprinted and adapted with permission from Anal. Chem., 2014, 86 (12), pp 5784–5793; DOI: 10.1021/ac500335t; Publication Date (Web): May 15, 2014

Copyright © 2014 American Chemical Society

3.1 Abstract



Protein glycosylation is an important post-translational modification associated, among others, with diseases and the efficacy of biopharmaceuticals. Matrix-assisted laser desorption/ionization (MALDI) time-of-fight (TOF) mass spectrometry (MS) can be performed to study glycosylation in a high-throughput manner, but is hampered by the instability and ionization bias experienced by sialylated glycan species. Stabilization and neutralization of these sialic acids can be achieved by permethylation or by specific carboxyl group derivatization with the possibility of discrimination between α2,3- and α2,6-linked sialic acids. However, these methods typically require relatively pure glycan samples, show sensitivity to side reactions, and need harsh conditions or long reaction times. We established a rapid, robust and linkage-specific high-throughput method for sialic acid stabilization and MALDI-TOF-MS analysis, to allow direct modification of impure glycan-containing mixtures such as PNGase F-released human plasma *N*-glycome. Using a combination of carboxylic acid activators in ethanol achieved near-complete ethyl esterification of α2,6-linked sialic acids and lactonization of α2,3-linked variants, in short time using mild conditions. Glycans were recovered by hydrophilic interaction liquid chromatography solid phase extraction and analyzed by MALDI-TOF-MS in reflectron positive mode with 2,5-dihydroxybenzoic acid as the matrix substance. Analysis of the human plasma *N*-glycome allowed high-throughput detection and relative quantitation of more than 100 distinct *N*-glycan compositions with varying sialic acid linkages.

3.2 Introduction

Protein glycosylation is a common post-translational modification having effects on characteristics such as solubility, folding and receptor binding activity[1-3]. *N*-Glycans on mammalian proteins often exhibit terminal sialic acids such as *N*-acetylneuraminic acid, which is involved in binding to various human lectins, shows importance in cellular communication, and determines protein half-life[4-6]. Sialic acids are most frequently attached to a terminal galactose via α2,6 or α2,3 glycosidic linkage and show different functionality as a consequence. Examples include α2,3-linked *N*-acetylneuraminic acids being specifically required for formation of sialyl-Lewis structures, which have been found to be indicative of metastasis for several types of cancer[7-10], whereas α2,6-linked sialic acids have been linked to promoting cell survival by galectin inhibition[11-13]. With glycosylation and specific features such as sialylation, galactosylation and fucosylation becoming increasingly interesting as a clinical biomarker reservoir, and a necessity for the characterization of biopharmaceuticals, the development of methods for high-throughput glycan analysis is required.

A technique well-suited for high-throughput glycomics is matrix-assisted laser desorption/ionization (MALDI) time-of-fight (TOF) mass spectrometry (MS), as it can rapidly provide profiles with information on glycan composition[14]. MALDI-TOF-MS analysis of sialylated glycans is, however, compromised by the labile nature of the sialic acid glycosidic bond, resulting in the facile loss of the residue by in-source and metastable decay. In addition, sialylated glycan species tend to show different salt adducts, resulting in multiple signals for single glycan compositions. Moreover, the carboxyl group present on a sialic acid preferentially facilitates negative ionization by MALDI, generating a bias in signal intensity when comparing acidic and neutral oligosaccharides[15].

One way to improve MALDI-TOF-MS measurements of sialylated glycans is derivatization[16,17]. Commonly used is permethylation, which results in reproducible sialic acid stabilization as well as increased sensitivity of MS analysis but shows some variation in the number of methyl groups conveyed[18-21]. Another strategy leading to greatly enhanced stability involves the selective modification of the sialic acid carboxyl group[15,16,22]. Importantly, in contrast to permethylation, sialic acid-specific derivatization has been demonstrated in a manner dependent on the sialic acid linkage, with α2,6-linked sialic acids undergoing modifications such as esterification and amidation, whereas α2,3-linked sialic acids can instead form an intramolecular lactone, leading to water loss[23-25]. The resulting mass difference allows the direct distinction of sialic acid linkages by mass spectrometry. Methods described in literature to achieve this linkage-specific derivatization involve methyl esterification or (methyl)amidation, but these typically require highly purified glycan samples, harsh conditions or long reaction times, and often do not show complete derivatization and linkage specificity[23-25]. Although these procedures are informative for the

analysis of sialylated oligosaccharides, suitability for high-throughput analysis of complex samples appears limited.

Here we present a fast, easy and mild protocol for linkage-specific sialic acid derivatization, using 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC) and 1-hydroxybenzotriazole (HOBt) as activators in ethanolic solution. These chemicals are widely used in peptide synthesis, with EDC responsible for initial carboxylic acid activation and HOBt catalyzing the subsequent conversion to an ester or amide (**Figure S1**, Supporting Information)[26-28]. To arrive at the proposed conditions, several acid activators and alcohols were examined for reaction completeness and selectivity for sialic acid linkage. Repeatability of the final protocol was established within and across 3 days for PNGase F-released total plasma *N*-glycome and the suitability for high-throughput 96-well format analysis was demonstrated.

3.3 Experimental section

Samples

The plasma pooled from 20 human donors (Visucon-F frozen normal control plasma, citrated and 0.02 M HEPES buffered) was obtained from Affinity Biologicals (Ancaster, Canada). Oligosaccharide standards 3'-sialyllactose (Neu5Acα2,3-Galβ1,4-Glc; Neu5Ac = *N*-acetylneuraminic acid; Gal = galactose; Glc = glucose) sodium salt and 6'-sialyllactose (Neu5Acα2,6-Galβ1,4-Glc) sodium salt (both with purities higher than 98%) were purchased from Carbosynth (Compton, U.K.). Both 3'- and 6'-sialyllactose were dissolved to a concentration of 100 mg/mL in water. Fibrinogen from human plasma was acquired from Sigma-Aldrich (Steinheim, Germany), and was incubated for 4.5 h at 37 °C in 1× PBS, resulting in a fibrinogen solution of 24.12 mg/mL.

Chemicals, reagents and enzymes

Milli-Q water (MQ) used in this study was generated from a Q-Gard 2 system (Millipore, Amsterdam, Netherlands), maintained at ≥18 MΩ. Methanol, ethanol, 2-propanol, 1-butanol, trifluoroacetic acid (TFA), glacial acetic acid, sodium dodecyl sulfate (SDS), disodium hydrogen phosphate dihydrate ($Na_2HPO_4 \times 2H_2O$), potassium dihydrogen phosphate (KH_2PO_4) and sodium chloride (NaCl) were purchased from Merck (Darmstadt, Germany). *N*,*N*'-Dicyclohexylcarbodiimide (DCC), 1-hydroxybenzotriazole (HOBt) hydrate, 2-aminobenzoic acid (2-AA), 2-picoline borane (2-PB), dimethyl sulfoxide (DMSO), 50% sodium hydroxide, 98% formic acid, 25% ammonium hydroxide in water and Nonidet P-40 (NP-40) were obtained from Sigma-Aldrich, whereas 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC) hydrochloride and ethyl-2-cyano-2(hydroxyimino)acetate (trade name Oxyma Pure) originated from Fluorochem (Hadfield, U.K.). Additional components used for this study included recombinant peptide-*N*-glycosidase F (PNGase F) from Roche Diagnostics (Mannheim, Germany), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM) from Santa Cruz

Biotechnology (Santa Cruz, CA), 2,5-dihydroxybenzoic acid (2,5-DHB) from Bruker Daltonics (Bremen, Germany) and HPLC SupraGradient acetonitrile (ACN) from Biosolve (Valkenswaard, Netherlands).

N-Glycan release

N-Glycans were released from the human plasma and fibrinogen samples as described previously[29]. Ten microliters of sample was denatured by adding 20 µL of 2% SDS and incubating for 10 min at 60 °C. The release step was subsequently performed by adding 20 µL of release mixture containing 2% NP-40 and 0.5 mU PNGase F in 2.5× PBS (10× PBS containing 57 g/L $Na_2HPO_4 \times 2H_2O$, 5 g/L KH_2PO_4 and 85 g/L NaCl), and incubating overnight at 37 °C.

Cotton HILIC-SPE glycan enrichment

Glycan enrichment was performed by cotton hydrophilic interaction liquid chromatography (HILIC) solid phase extraction (SPE) as described previously[30], with a few modifications. Samples removed from −20 °C were allowed to return to room temperature before proceeding. Twenty microliter pipet tips (Rainin Instrument, Oakland, CA) were packed with 3 mm cotton thread (180 µg, Pipoos, Utrecht, Netherlands), which was then conditioned and equilibrated by pipetting three times 20 µL of MQ, followed by three times 20 µL of 85% ACN. The sample was loaded by pipetting 20 times into the reaction mixture, making sure not to include the sedimented precipitate, in order to prevent clogging of the tips. The tips were washed three times with 20 µL of 85% ACN 1% TFA, and three times with 20 µL of 85% ACN. Subsequent elution was performed in 10 µL of MQ. A 12-channel pipet was used for all steps.

96-well format Sepharose HILIC-SPE

A 96-well format glycan enrichment was performed by Sepharose HILIC-SPE, using previously reported conditions with some modifications[31]. Ten microliters of Cl-4B Sepharose beads (45–165 µm, GE Healthcare, Uppsala, Sweden) were added in 20% ethanol to a 96-well filter plate (0.7 mL/well, PE frit, Orochem, Naperville, IL), and solvent was removed by vacuum using a vacuum manifold (Millipore, Billerica, MA). Beads were washed three times with 180 µL of MQ and three times with 180 µL of 85% ACN, after which the samples were transferred to the plate. To facilitate binding, the plate was put on a multiwell plate shaker (1.5 mm orbit, VWR, Amsterdam, Netherlands) for 5 min at 1000 rpm, after which the remaining solvent was removed by vacuum. The plate was subsequently washed three times with 85% ACN 0.1% TFA and 85% ACN before applying 30 µL of MQ for elution. Again the plate was incubated for 5 min on a shaking platform at 1000 rpm, and elution performed by centrifuging 1 min at 200 *g*.

MALDI-TOF-MS

For MALDI-TOF-MS analysis, 1 µL of glycan sample purified by HILIC-SPE was spotted on a MTP AnchorChip 800/384 TF MALDI target (Bruker Daltonics, Bremen, Germany), mixed on plate with 1 µL 2,5-DHB (5 mg/mL) 1 mM NaOH in 50% ACN and left to dry by air. The matrix crystals were made uniform by addition of 0.2 µL ethanol, causing rapid recrystallization.

Analyses were performed RP-mode on an UltraFlextreme MALDI-TOF-MS equipped with a Smartbeam-II laser and Flexcontrol 3.4 software Build 119 (Bruker Daltonics). The instrument was calibrated using a peptide calibration standard (Bruker Daltonics, Bremen, Germany). A 25 kV acceleration voltage was applied after a 140 ns extraction delay. A mass window of *m/z* 100 to 1500 was used for α2,3- and α2,6-sialyllactose analysis, and a mass window of *m/z* 1000 to 5000 with suppression up to *m/z* 900 for fibrinogen and plasma *N*-glycan samples. For each spectrum, 20 000 laser shots were accumulated at a laser frequency of 2000 Hz, using a complete sample random walk with 200 shots per raster spot. High laser intensity was used for sample profiling to allow for ionization of larger glycan species, while the monoisotopic peak was still clearly defined for all detectable glycan masses.

Tandem mass spectrometry (MALDI-TOF/TOF-MS/MS) was performed on the most abundant peaks of human plasma *N*-glycome via laser-induced disassociation.

Analysis of MALDI-TOF-MS data

Using flexAnalysis v3.3 build 65 (Bruker Daltonics), the acquired MALDI-TOF-MS spectra were internally recalibrated using a set of calibration masses (**Table S1**, Supporting Information). More specifically, masses were calculated as $[M + Na]^+$, taking into account the different masses for the sialic acids, depending on their lactonization or alkylation. Glycan compositions used in the calibration of plasma *N*-glycome were H3N4F1, H4N4F1, H5N4F1, H5N4S1, H5N4F1S1, H5N4S2, H6N5S3, H6N5F1S3 and H7N6S4, (H = hexose, N = *N*-acetylhexosamine, F= fucose and S = *N*-acetylneuraminic acid with unspecified linkage, the number of residues given after the letter). For fibrinogen samples, only H5N4S1 and H5N4S2 were used for calibration, whereas for sialyllactose, the H2S1 values were calculated for alkylated and lactonized sialic acid variants. Masses were picked in the spectra using a centroid algorithm with a mass window of 0.5 Da, followed by quadratic calibration. Recalibrated spectra were exported as a text format and further analyzed using an in-house developed Python script (see the Supporting Information). In short, the analysis was performed as a targeted data extraction using a determined list of glycan compositions. For each composition, the isotopic distribution and its masses were calculated. For each isotopic peak within 95% of the cumulative isotopic peak pattern, the intensities were summed within a 1 Da mass window. Noise intensities were then summed for a 1 Da window located 1 Da lower than the isotope cluster and subtracted from each isotopic peak (resulting in the background-corrected signal values). Observed isotopic ratios were

compared to the calculated ones to prevent errors due to overlap, after which the intensities for the individual isotopic peaks were summed per glycan composition. Percentages were calculated with standardization to the overall glycan signal intensity of each spectrum. Averages and standard deviations were calculated for repeated experiments. Figures were assigned with glycan structures using GlycoWorkbench 2.1 build 146[32,33].

Activator comparison

A selection of carboxylic acid activator reagents and their combinations was tested for suitability with sialic acid esterification of unpurified plasma *N*-glycome samples obtained from PNGase F treatment. DMT-MM, HOBt, Oxyma Pure, DCC and EDC were dissolved at 0.5 M in methanol, with and without 0.2% TFA. One microliter of released plasma *N*-glycome (containing approximating 14 µg of plasma proteins and the corresponding released *N*-glycans) was added to 20 µL of each reagent. In addition, 1 µL of released plasma *N*-glycome was added to 20 µL of methanol, containing the following reagent combinations: DCC with HOBt, DCC with Oxyma Pure, EDC with HOBt and EDC with Oxyma Pure, each component at a concentration of 0.25 M, and each condition with and without 0.2% TFA. All samples were incubated for 1 h at 60 °C. After the reaction, the samples were brought to an ACN concentration of 50% (*v/v*) by addition of 20 µL ACN, followed by 15 min incubation at −20 °C to stimulate additional precipitate formation. Cotton HILIC-SPE was performed to achieve enrichment and purification of the derivatized glycans from the reaction mixture.

Comparison of different alcohols for derivatization

0.25 M EDC with 0.25 M HOBt was selected as the most promising reagent combination for sialic acid esterification, and additional alcohols besides methanol were tested for use as both solvent and alkyl donor. Solutions of 0.25 M EDC together with 0.25 M HOBt were prepared using four different alcohols: methanol, ethanol, 2-propanol and 1-butanol. One microliter of PNGase F-released plasma *N*-glycome was added to 20 µL of each of the four solutions, and the samples were incubated for 1 h at 60 °C. After this, 20 µL of ACN was added and the samples were incubated at −20 °C for 15 min before glycan enrichment by cotton HILIC and MALDI-TOF-MS analysis.

Linkage specificity on α2,3- and α2,6-sialyllactose

The linkage specificities of the 0.25 M EDC and 0.25 M HOBt solutions in both methanol and ethanol were tested using oligosaccharide standards with a known sialic acid linkage position. One microliter of either 100 mg/mL (100 µg) α2,3- or α2,6-sialyllactose was added to 20 µL of methanol or ethanol containing 0.25 M EDC and 0.25 M HOBt. Samples were incubated for 1 h at 60, 50, 37, 21 or 4 °C. After the addition of 20 µL of ACN, all samples were incubated for 15 min at −20 °C before proceeding with glycan enrichment and analysis.

Repeatability testing

Repeatability of the established ethyl esterification conditions was ascertained by multiple analyses of the same sample across several days, performing all steps in a 96-well polypropylene PCR plate (Greiner Bio-One), sealed with adhesive tape (Nutacon, Leimuiden, Netherlands) during incubation steps. For each day, 24 independent samples were taken from the same pooled plasma and deglycosylated with PNGase F as described before. One microliter of each released sample was ethyl esterified by adding it to a new PCR plate containing 20 µL of 0.25 M HOBt 0.25 M EDC in ethanol and incubating the solution for 1 h at 37 °C. Twenty microliters of ACN was then added and the plate incubated at −20 °C for 30 min. Samples were purified by cotton HILIC-SPE and measured by MALDI-TOF-MS. The entire procedure was repeated twice on consecutive days with freshly prepared reagents to establish day-to-day variability.

Fibrinogen glycan labeling

Aliquots of PNGase F-released *N*-glycans from fibrinogen were modified by either ethyl esterification or fluorescent labeling of the reducing end with 2-AA to allow for a comparison between MALDI-TOF mass spectrometric data and profiles obtained by HILIC ultra high-performance liquid chromatography (UHPLC) with fluorescent detection. For ethyl esterification, the optimal conditions were established in previous experiments. One microliter of fibrinogen release mix (corresponding to 24.12 µg protein) was added to 20 µL of 0.25 M EDC and 0.25 M HOBt in ethanol, and incubated for 1 h at 37 °C. Subsequently, 20 µL of ACN was added and the mixture was stored at −20 °C for 15 min before glycan enrichment and MALDI-TOF-MS analysis.

For HILIC-UHPLC, the released *N*-glycans were labeled with 2-AA as described previously[34]. In short, 20 µL of PNGase F-released fibrinogen glycans was mixed with 10 µL of 2-AA (48 mg/mL) in DMSO with 15% glacial acetic acid and 10 µL of 2-PB (107 mg/mL) in DMSO. The mixture was incubated for 2 h at 65 °C and diluted to 75% ACN before HILIC-UHPLC analysis.

HILIC-UHPLC measurement

For the separation and analysis of 2-AA labeled *N*-glycans, HILIC-UHPLC with fluorescent detection was performed on a Dionex Ultimate 3000 (Thermo Fisher Scientific, Breda, Netherlands) equipped with a 1.7 µm × 2.1 × 100 mm Acquity UPLC BEH Glycan column (Waters, Etten-Leur, Netherlands). Separation was performed at 60 °C with a flow rate of 0.6 mL/min. Two solutions were used for gradient generation, ACN as solution A and 100 mM ammonium formate pH 4.4 (prepared as formic acid buffered to pH 4.4 by NH_4OH) as solution B. The column was equilibrated by 85% solution A for 0.5 min. After this, the samples were loaded in 75% A, and the excess fluorescent reagent was eluted from the column by washing with 85% A for 10 min. The separation gradient started at 75% A and decreased linearly to 57% A in 45 min. The column was then flushed again with 40% A for 10 min followed by 10 min of 85% A for re-equilibration. For fluorescent detection, 330 nm

was used for excitation and the emission recorded at 420 nm. The resulting chromatograms were analyzed using Chromeleon version 7.1.2.1713 (Thermo Fisher Scientific).

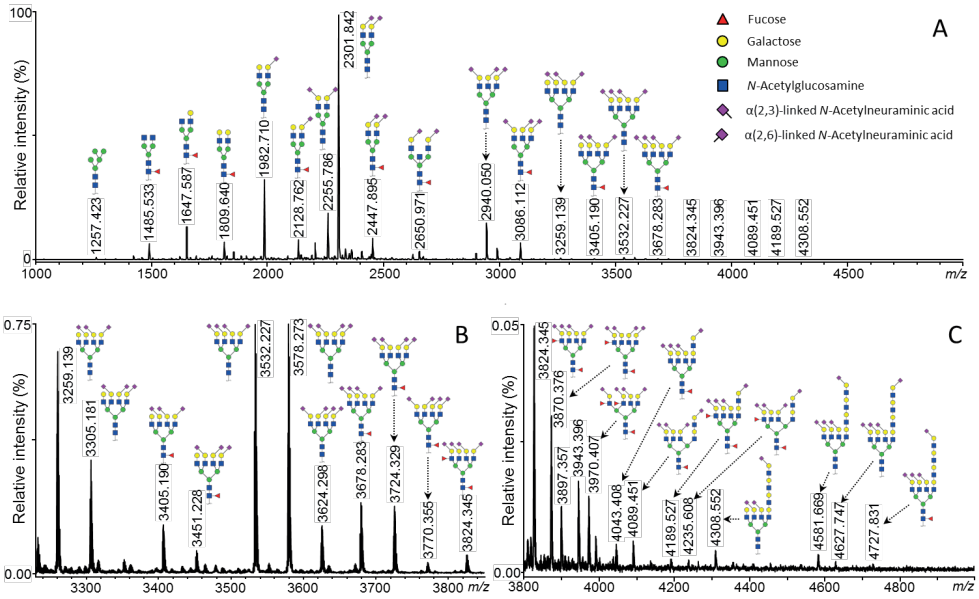


***Figure 1**. RP MALDI-TOF-MS spectrum after applying the 1 h at 37 °C EDC + HOBt ethyl esterification protocol for plasma N-glycome analysis. **A)** Overview spectrum. **B)** Intermediate m/z range spectrum (m/z 3200 to 3800). **C)** High m/z range spectrum (m/z 3800 to 5000) showing glycan masses of up to 4727.640 Da $[M + Na]^+$. Proposed glycan structures are based on literature[36,41-43], as well as MS/MS data (**Table S2**, Supporting Information). Structures are depicted following the CFG notation. Linkage positions of sialic acid residues are indicated by differing angles.*

3.4 Results

A robust, high-throughput MALDI-TOF-MS method was developed for profiling human plasma *N*-glycans with linkage-specific derivatization of sialic acid residues. Reaction conditions were optimized to achieve ethylation of α2,6-linked sialic acid residues with parallel lactone formation of α2,3-linked sialic acids. Commercially available pooled plasma (10 µL) was subjected to PNGase F treatment to obtain a complex sample with free *N*-glycans. The resulting plasma *N*-glycome contains a large set of *N*-glycans of different compositions, including neutral as well as highly sialylated species with varying linkages[35,36], making it a demanding sample to study the overall effects of sialic acid modification methods. Ethylation using carboxylic acid activators EDC and HOBt was chosen due to its high catalytic efficacy and specificity. Glycan profiles were obtained, allowing the differentiation of 77 glycan compositions, which resulted in 108 glycan species due to variation in sialic acid linkage (**Figure 1**). Method development and validation are described in the following sections.

Comparison of different activators

A number of coupling reagents as well as combinations thereof were compared for linkage-specific esterification of sialic acid residues in impure mixtures containing native *N*-glycans. All reactions were performed directly on the unpurified PNGase F released *N*-glycan mixture, using methanol as both methyl donor and solvent. Samples were allowed to react for 1 h at 60 °C in methanol containing coupling/activation reagents DMT-MM, DCC, EDC, HOBt or Oxyma Pure, each tested with and without 0.2% TFA. Additionally, reagent combinations DCC + HOBt, DCC + Oxyma Pure, EDC + HOBt and EDC + Oxyma Pure were tested, each with and without 0.2% TFA. During the 1 h incubation step, a white precipitate formed at the bottom of the samples, causing separation between the methanol soluble components within the mixture (including glycans and the coupling reagents), and the methanol insoluble fraction containing proteins.

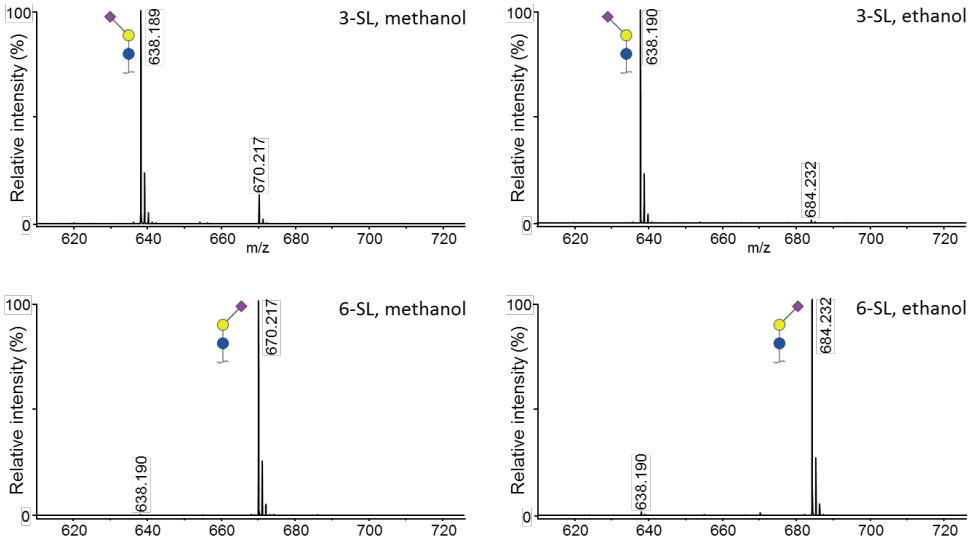


***Figure 2**. RP MALDI-TOF-MS spectra of 3'- and 6'-sialyllactose when alkyl esterified for 1 h at 37 °C with EDC + HOBt in methanol or ethanol. The lactonized reaction product is visible at 638.190 Da [M + Na]⁺, with the esterified masses being 670.217 and 684.232 Da for methanol and ethanol, respectively. In no case was the mass of 678.183 Da [M – H + 2Na]⁺ visible, which would be indicative for incomplete reaction. Reaction in ethanol shows the highest specificity for sialic acid linkage with undesired reaction products only appearing in minimal amounts.*

Samples were purified by cotton HILIC-SPE. To this purpose, various volumes of ACN were added, and concentrations between 25 and 75% were found to result in excellent enrichment without noticeable bias for low- or high-mass species. The intermediate value of 50% ACN was chosen for purification. One microliter of the purified samples was spotted on an AnchorChip MALDI plate and cocrystallized with 1 µL of 2,5-DHB (5 mg/mL) as the matrix. To suppress potassium adduct formation, 1 mM NaOH was added to the matrix

solution, yielding almost exclusively $[M + Na]^+$ species in reflectron positive mode. Recrystallization was performed with 0.2 µL ethanol to decrease shot-to-shot variation and further decrease the variation in salt adduct formation.

Efficacy of sialic acid derivatization was determined for the various reaction conditions by evaluating the signals of the sialylated *N*-glycans (**Figure S2**, Supporting Information). The mass shift induced by the desired methyl esterification is +14.016 Da per sialic acid, resulting in a mass of 2273.823 Da $[M + Na]^+$ for the fully esterified major plasma *N*-glycan H5N4S2. Glycans with unmodified sialic acids were observed in RP mode with full exchange of proton for sodium on the carboxylic acid groups (+21.982 Da per sialic acid), resulting in signals at *m/z* 2281.763 and 2289.740, indicative of incomplete reactions. In addition, lactonization of one of the sialic acids could be observed, yielding an 18.011 Da mass loss, resulting in a signal at *m/z* 2241.767 $[M + Na]^+$.

Single reagents DCC, EDC, HOBt and Oxyma Pure proved unable to induce modification of sialic acids, regardless of whether acid was used during the reaction or not. DMT-MM did show methyl esterification of sialic acid residues, but also generated a prominent peak at *m/z* 2240.811, as well as a signal at *m/z* 2258.817 (amidation of carboxylic acid groups, resulting in a −0.984 Da mass difference per acid). These side products were largely prevented by adding 0.2% TFA to the reaction, but strong acidic conditions may also result in partial desialylation and loss of other labile substituents. Reagent combinations DCC + HOBt, DCC + Oxyma Pure and EDC + Oxyma pure resulted in prominent products with methyl esterification of the sialic acid residues, but the presence of sodium adducts also indicated the conversion to be incomplete. This was largely, but not completely, corrected by using acidified reaction mixtures. EDC + HOBt appeared to be the most promising reagent combination, showing complete conversion regardless of the use of acid, and this conditions was therefore selected for further experimentation.

Testing of different alcohols

EDC + HOBt was selected as the most promising reagent mixture for sialic acid esterification, and various alcohols were tested as combined protein precipitation agent and alkyl donor. After a solution of 0.25 M EDC and 0.25 M HOBt in 20 µL of methanol, ethanol, 2-propanol or 1-butanol was added to 1 µL of plasma *N*-glycome solution, the reaction was performed for 1 h at 60 °C. EDC and HOBt dissolved only partially in 2-propanol and 1-butanol at room temperature; therefore, a well-homogenized suspension was used instead. Purification of the reaction products and analysis by RP MALDI-TOF-MS were performed as described above.

Analysis of sialylated glycan compositions showed all alcohols to be an alkyl group donor for sialic acid esterification (**Figure S3**, Supporting Information). Trisialylated glycan compositions H6N5S3 and H6N5F1S3 were studied as examples. Expected masses after

derivatization with methanol, ethanol, 2-propanol and 1-butanol were 2944.047, 2986.094, 3028.141 and 3070.188 Da for composition H6N5S3 and 3090.105, 3132.152, 3174.199 and 3216.246 Da for H6N5F1S3, all observable in the respective spectra. In addition, lactonization products were present in all spectra, with mass differences of 32.026, 46.042, 60.058 and 74.073 Da as compared to the alkyl esterification products, but relative ratios of the lactonized and alkyl esterified species differed per alcohol. 2-Propanol showed a large variety of reaction products in addition to the expected species, and both 2-propanol and 1-butanol yielded relatively low signals for the sialylated glycans. Methanol and ethanol were both selected as the most promising solvents and alcohol donors for linkage-specific sialic acid modification and were therefore used for further method development.

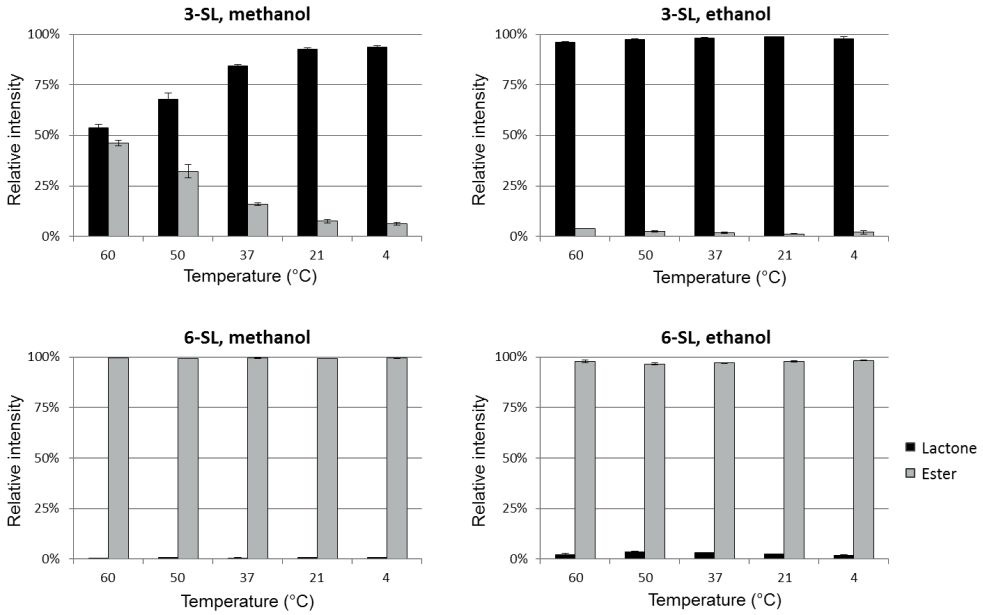


***Figure 3**. Triplicate analysis of 3'- and 6'-sialyllactose after EDC + HOBt alkylation in either methanol or ethanol, at varying temperatures. Shown in the graphs are the average relative intensities of the lactonized and esterified reaction products, with the error bars indicating standard deviation. α2,6-Linked sialic acids show to preferably form an ester with both methanol and ethanol at all temperatures. Reacting α2,3-linked sialic acids in methanol, however, shows a temperature dependent effect, at best 93.6% (SD ± 0.9%) of the desired lactonized product being formed at 4 °C. Using ethanol as the alkyl donor, on the other hand, shows on average 97.7% (SD ± 0.9%) lactonization across all temperatures and is therefore most suitable for separating α2,3- and α2,6-linked sialic acids.*

Achieving linkage-specificity

The linkage specificity of EDC + HOBt in combination with methanol and ethanol was studied using oligosaccharide standards with known sialic acid linkage, namely 3'-sialyllactose (Neu5Acα2,3-Galβ1,4-Glc) and 6'-sialyllactose (Neu5Acα2,6-Galβ1,4-Glc), both with purities

higher than 98%. To study the temperature effect on linkage-specificity, the samples were reacted not only for 1 h at 60 °C but also at 50, 37, 21 and 4 °C. Purification of reaction products and RP MALDI-TOF-MS analysis were performed as before (**Figure 2**).

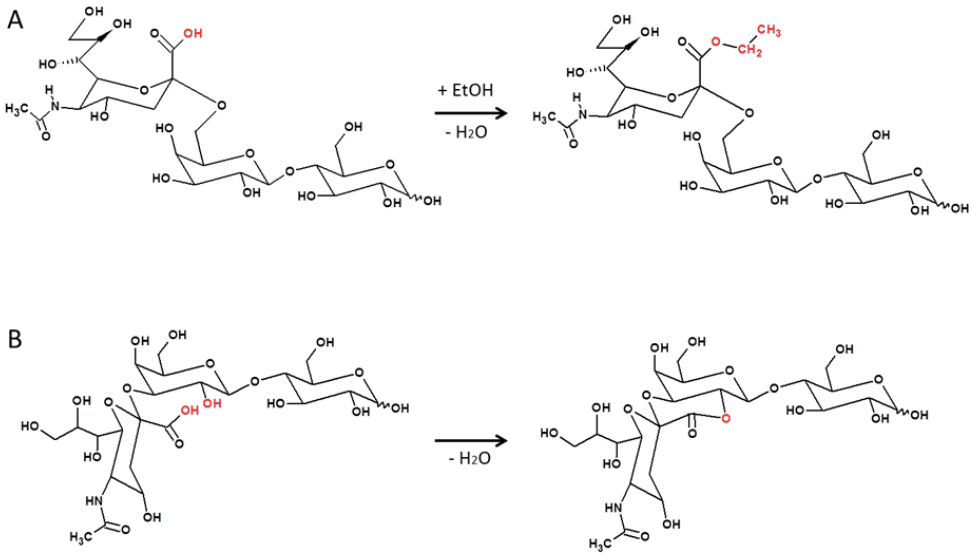


Figure 4*. Structures of* ***A)*** *6'-sialyllactose and* ***B)*** *3'-sialyllactose before and after ethyl esterification. Whereas α2,6-linked sialic acids show ethyl esterification of the carboxylic acid, α2,3-linked sialic acids instead form a lactone with the neighboring galactose.*

For each experiment, relative quantitation of the lactonized reaction products (638.190 Da in all conditions) and the methyl- and ethyl esterified products (670.217 and 684.232 Da, respectively) was performed (**Figure 3**). The unmodified glycan was not found (expected signal for $[M - H + 2Na]^+$ at 678.183 Da), indicating that all sialic acids were either esterified or lactonized. α2,6-Linked sialic acids were shown to be highly susceptible to alkylation, as both methanol and ethanol led to a near complete esterification. However, lactonization of α2,3-linked sialic acids in methanol displayed a temperature dependent conversion, with a lower amount of side reaction (methanol esterification) formed at lower temperatures. Using methanol, correct lactonization ranged from only 53.9% (standard deviation (SD) ± 1.5%) at 60 °C to 93.6% (SD ± 0.6%) at 4 °C. The use of ethanol, however, resulted in a much higher preference for α2,3-linked sialic acid lactonization, showing, on average, only 2.3% (SD ± 0.9%) of side reaction (ethanol esterification) across all temperatures. Because of the nearly complete selectivity between 3'- and 6'-sialyllactose reaction products (**Figure 4**), ethanol was selected as the solvent and donor for linkage-specific sialic acid modification by EDC + HOBt. Reaction conditions were set at 1 h at 37 °C for further experiments, to allow for a controlled temperature condition that is readily available in most laboratories.

Sialic acid stability

The ratios of H5N4S1 and H5N4S2 glycans released from fibrinogen were determined by MALDI-TOF-MS after ethyl esterification and compared with ratios determined after 2-AA labeling by HILIC-UHPLC and fluorescence detection. Triplicate analysis and relative quantification of the monosialylated and disialylated glycans showed signals averaging on 65.3% and 34.7% (SD ± 1.5%) for MALDI-TOF-MS and 66.4% and 33.6% (SD ± 0.2%) for the UHPLC measurements (**Figure 5**). The standard deviation of the MALDI method was noticeably higher than that for the UHPLC method, with a coefficient of variation (CV) of 4.3% for the smaller peak and 2.3% for the larger peak. Importantly, the proportions of the major monosialylated and disialylated diantennary glycans proved to be highly similar between mass spectrometry and UHPLC, and correspond with the literature[37-40].

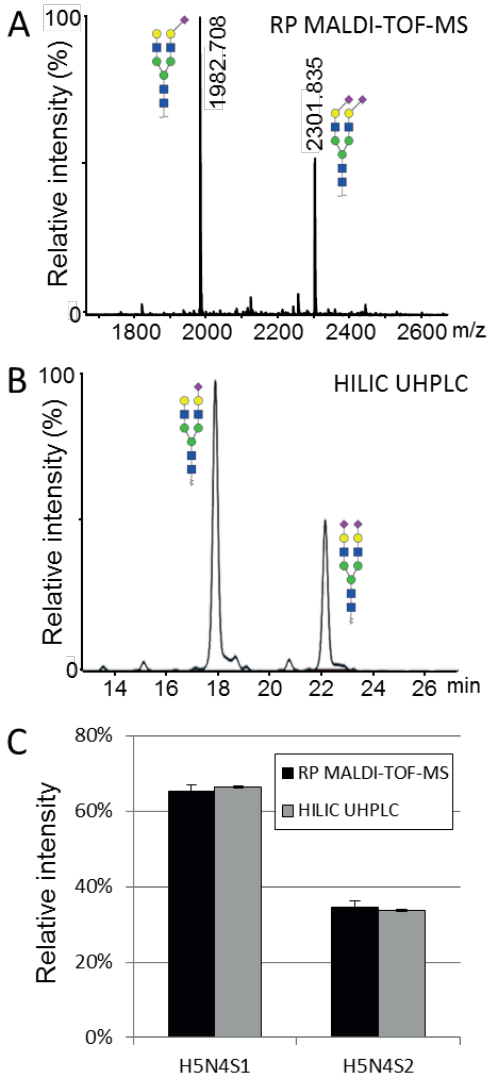


***Figure 5**. Released fibrinogen N-glycans as studied by **A)** RP MALDI-TOF-MS after 1 h 37 °C EDC + HOBt ethyl esterification, and **B)** 2-AA labeling and HILIC UHPLC with fluorescent detection. **C)** Triplicate analysis and relative quantification of the mono- and disialylated structures shows highly comparable relative signal intensities between MALDI-TOF-MS and UHPLC analysis. Abbreviations used are hexose (H), N-acetylhexosamine (N), fucose (F), and N-acetylneuraminic acid with unspecified linkage (S). The accompanying number indicates the number of residues. The error bars show standard deviations.*

Repeatability

Multiple repeat analyses were performed to demonstrate the repeatability of EDC + HOBt ethyl esterification (1 h at 37 °C) for human plasma *N*-glycome profiling. The control plasma was divided into 24 separate samples, glycans were released with PNGase F, ethyl esterified, purified by cotton HILIC-SPE and analyzed by RP MALDI-TOF-MS. Glycan signals were integrated, normalized to the sum of intensities, and the relative signals and standard

deviations were calculated. This protocol was performed for two additional days to account for the day-to-day variation.

Repeatability analysis for the 20 most abundant glycans (responsible for 94% of the cumulative glycan distribution of a spectrum) across 24 independent samples originating from the same plasma pool revealed high repeatability within one batch, as well as between three batches prepared and measured on different days. Average relative intensity values for the highest peak (H5N4E2, the two esterified sialic acids (E) indicating α2,6-linkage) were around 54.7% (SD ± 2.3%) across all measurements, with the CV averaging to 3.8% (**Figure 6**). Analysis of all assigned glycan masses shows low variability even for values below 0.1% of the cumulative distribution, including those in very high mass ranges (**Figure S4**, Supporting Information).

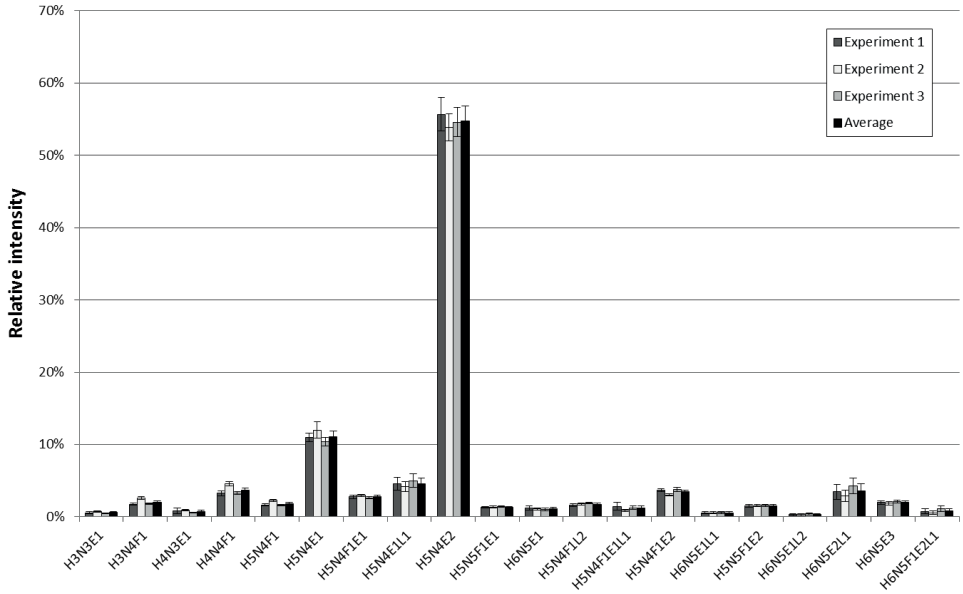


***Figure 6**. Repeatability of plasma N-glycan profiling by MALDI-TOF-MS after 1 h at 37 °C EDC + HOBt ethyl esterification. Twenty-four samples originating from a stock of control plasma were independently released by PNGase F, ethyl esterified, purified by HILIC-SPE, crystallized with matrix, recrystallized with ethanol and analyzed by RP MALDI-TOF-MS. The experiment was performed three times on separate days, indicated as experiments 1–3. The graph shows the average relative intensities observed for the 20 major glycan species (normalized to the overall sum of intensities), with error bars for standard deviation. Abbreviations used are hexose (H), N-acetylhexosamine (N), fucose (F) and N-acetylneuraminic acid with either unspecified linkage (S), α2,3-linkage as indicated by lactonization (L) or α2,6-linkage as indicated by esterification (E).*

96-well plate protocol

To validate the ethyl esterification protocol for applicability in high-throughput format, all steps were performed on a 96-well plate, employing Sepharose HILIC-SPE rather than

cotton purification. MALDI-TOF-MS spectra obtained after Sepharose were similar to those obtained from cotton and high quality spectra could be obtained across the plate (**Figure S5**, Supporting Information). Relative quantification revealed only slight variations between the two HILIC modes, noticeable in the H5N4E1, H5N4E1L1 and H5N4E2 compositions.

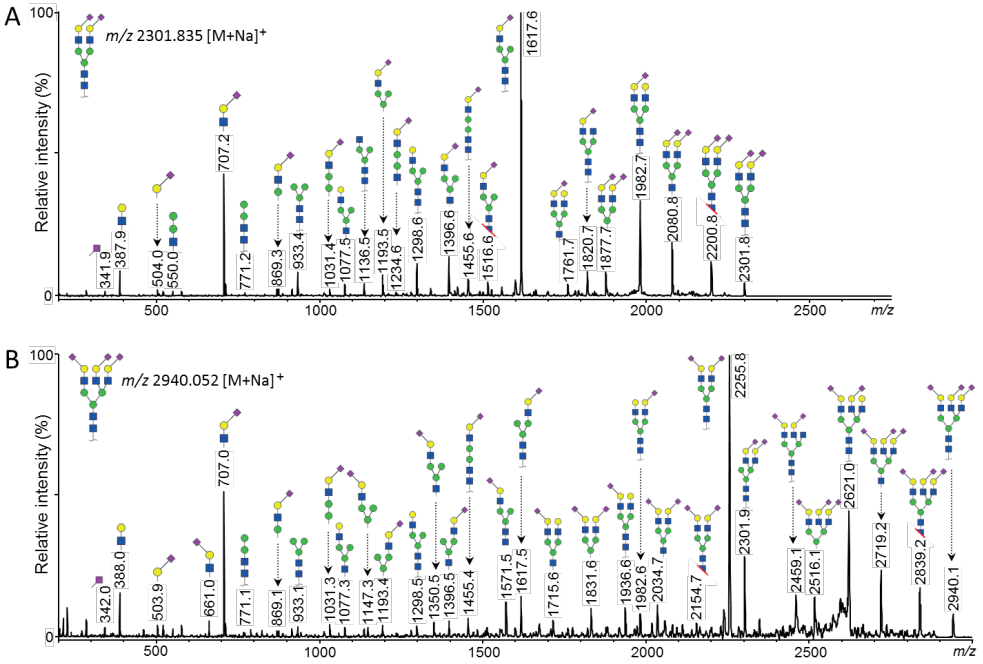


***Figure 7**. MALDI-TOF/TOF-MS/MS spectra of two major glycans of the ethyl esterified plasma N-glycome. **A)** Diantennary species carrying two α2,6-linked sialic acids. **B)** Triantennary species carrying two α2,6-linked sialic acids and one α2,3-linked sialic acid. Prominent losses include the α2,6-sialylated arms (684.3 Da), which can be observed in both spectra (resulting in signals at m/z 1617.6 and 2255.8, respectively), as well as the unique loss of an α2,3-sialylated arm (638.2 Da) in the triantennary spectrum (m/z 2301.9). Additional common losses include the loss of the reducing end N-acetylglucosamine (221.1 Da, visible at m/z 2080.8 and 2719.2, respectively) and an $^{0,2}A$ crossring fragment thereof (101.1 Da, visible at m/z 2200.8 and 2839.2, respectively) (indicated by a red line).*

Structural assignment

Plasma profiles studied using the ethyl esterification protocol showed glycans $[M + Na]^+$ ranging in mass from 1257.423 Da (H5N2) to 4727.640 Da (H10N9F1 with four lactonized (L) sialic acids indicating α2,3-linkage). In total, 217 isotope clusters could be detected in the spectra, 108 of which could be attributed to specific glycan compositions within a low mass deviation (average mass deviation 0.015 Da), accounting for approximately 90% of the total spectrum intensity (**Table S2**, Supporting Information). The most commonly found signals that could not be assigned to a native glycan structure showed masses 101.051 and 221.090

Da lower than the associated major glycan peaks, respectively indicating an ^{0,2}A cross-ring fragmentation or complete loss of the reducing end *N*-acetylglucosamine. The relative abundances of these fragments increase with increasing laser power (respectively, around 4.3% and 0.3% of the major peak for low laser power settings and 9.6% and 0.7% for high laser power settings) but do not seem to discriminate between specific glycan types. Another side-reaction, which could arise at +357.184 Da from major peaks, may be attributable to a yet unidentified reducing end modification. Altogether, these unwanted products are minor and do not overlap or change the relative intensities of glycans in the spectrum, allowing analysis of glycan profiles without interference.

Glycans were structurally assigned on the basis of literature knowledge on the human plasma *N*-glycome and established biosynthetic pathways[36,41-43], together with analytical data acquired in this study. MALDI-TOF/TOF-MS/MS was performed on the 38 highest intensity glycan peaks in the plasma *N*-glycome spectrum in order to confirm compositions and derive key structural features (**Table S2**, Supporting Information). The observed fragmentation spectra proved informative with regard to antenna composition and fucosylation (**Figure 7**). Particulary high intensity fragments included the loss of the reducing end *N*-acetylglucosamine (221.090 Da), which was observed in all spectra, except for the spectra of core-fucosylated precursors which instead showed the loss of the reducing end *N*-acetylglucosamine with fucose (367.148 Da). Furthermore, sialylated glycans showed high intensity signals at losses of 638.217 Da (antenna carrying an α2,3-linked sialic acid), 684.259 Da (antenna carrying an α2,6-linked sialic acid) or 784.275 Da (antenna carrying an α2,3-linked sialic acid with additional fucose). The latter was only assigned as a fucosylated sialylated antenna if an accompanying loss of 221.090 Da would indicate lack of core fucosylation, and a 638.217 Da fragment could not be seen. No evidence was obtained for fucosylation of α2,6-sialylated antennae.

3.5 Discussion

In this study, we present a high-throughput glycan derivatization method for MALDI-TOF-MS analysis, which allows the stabilization of sialylated species in a linkage-specific manner. Although the differential reactivity of α2,3- and α2,6-linked sialic acids has previously been reported, as well as derivatization methods relying on this feature[23,24], the ethanol esterification described in this article shows marked advantages over current literature. First, we set up a one-pot approach resulting in a rather simple protocol with minimized sample handling steps. This allowed us to establish a robust high-throughput method in 96-well plate format. Second, we succeeded in increasing the specificity of the reaction by choosing efficient activators and mild reaction conditions: therefore, the presence of detergents, salts and amines, at the levels found in plasma glycan release samples, did not noticeably impede the specific derivatization of sialic acids by ethyl ester formation and lactonization. Third, the chemicals needed in the procedure are inexpensive and of only low

toxicity, increasing the attractiveness of the method for routine, high-throughput applications.

We succeeded in performing ethyl esterification directly on impure samples like PNGase F-released plasma *N*-glycome, which may be attributed to several factors. Using ethanol as a solvent induces precipitation of proteins and peptides, reducing their interference with the reaction, while the glycans remain soluble and available for modification[44]. Not surprisingly, when the ethylation reaction was performed with shaking instead of at a stable position, we witnessed a significant decrease in reaction completeness (not shown). Another reason could be the excess of ethanol in the reactive environment. Esterification of α2,6-linked sialic acids with ethanol hereby becomes a high probability event, while steric hindrance still allows full lactonization of α2,3-linked sialic acids. Glycan samples modified by the ethyl esterification show markedly improved spectra over native ones in reflectron positive MALDI-TOF-MS. The resulting spectrum of total human plasma *N*-glycans is largely in line with patterns determined by more quantitative chromatographic methods[35,45].

The mass changes induced by the sialic acid derivatization method allow the discrimination of α2,3- and α2,6-linked sialic acids directly from the overview spectrum and, contrary to underivatized samples, separate the isotopic peak patterns of sialylated and difucosylated glycan species (respectively, 291.095 and 292.116 Da mass increment without derivatization). However, while the current protocol is optimized for the modification of α2,3- and α2,6-linked *N*-acetylneuraminic acids, the efficacy of related derivatization schemes for other acidic residues present on glycans remains unknown. Notable examples include glucuronic acid, which has recently been found on human plasma *N*-glycans[45], as well as monosaccharide modifications like phosphorylation and sulfation. In addition, whereas Neu5Acα2,3-Gal and Neu5Acα2,6-Gal can be discriminated by the proposed method, the reactivities of *N*-glycolylneuraminic acids, α2,8-linked *N*-acetylneuraminic acids, *O*-acetylated sialic acids and sialic acids linked to *N*-acetylhexosamines remain to be investigated.

To conclude, we have introduced a facile method for linkage-specific derivatization of sialic acid residues, using the combination of EDC and HOBt in ethanol, to allow for enhanced glycan detection by MALDI-TOF-MS. The method is characterized by high linkage specificity, overall mild and short reaction conditions, resistance to impurities, low cost and toxicity and repeatability within and between batches of samples. Altogether, we expect the proposed ethyl esterification to be an excellent method for the high-throughput analysis of sialylated glycans derived from biological sources and biopharmaceuticals.

3.6 Supporting information

Figures containing the reaction mechanism of ethyl esterification, RP MALDI-TOF-MS analysis of PNGase F-released plasma *N*-glycome after methyl esterification, RP MALDI-TOF-

MS of the plasma *N*-glycome after esterification with methanol, ethanol, 2-propanol and 1-butanol as solvents and alkyl donors using EDC + HOBt activators, repeatability of plasma *N*-glycan profiling by MALDI-TOF-MS after 1 h 37 °C EDC + HOBt ethyl esterification, RP MALDI-TOF-MS example spectrum of Sepharose purified plasma *N*-glycans after ethyl esterification, comparison of relative intensities obtained from cotton HILIC and Sepharose HILIC plate purification procedures, and an Excel workbook containing glycan signal assignments. This material is available free of charge via the Internet at http://pubs.acs.org.

3.7 Acknowledgments

This work was supported by the European Union (Seventh Framework Programme HighGlycan project, grant number 278535). The authors thank Emanuela Lonardi for critically reading the paper, and David Falck for his assistance with the displayed reaction mechanism.

3.8 References

1 Wormald, M. R. & Dwek, R. A. Glycoproteins: glycan presentation and protein-fold stability. *Struct Fold Des* **7**, R155-R160, doi:Doi 10.1016/S0969-2126(99)80095-1 (1999).

2 Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and their roles in the immune system. *Nat Rev Immunol* **7**, 255-266, doi:10.1038/nri2056 (2007).

3 Muramatsu, T. Essential roles of carbohydrate signals in development, immune response and tissue functions, as revealed by gene targeting. *J Biochem-Tokyo* **127**, 171-176 (2000).

4 Dall'Olio, F. & Chiricolo, M. Sialyltransferases in cancer. *Glycoconj J* **18**, 841-850, doi:Doi 10.1023/A:1022288022969 (2001).

5 Schauer, R. Sialic acids as regulators of molecular and cellular interactions. *Curr Opin Struc Biol* **19**, 507-514, doi:10.1016/j.sbi.2009.06.003 (2009).

6 Sperandio, M., Gleissner, C. A. & Ley, K. Glycosylation in immune cell trafficking. *Immunol Rev* **230**, 97-113 (2009).

7 Isozaki, H., Ohyama, T. & Mabuchi, H. Expression of cell adhesion molecule CD44 and sialyl Lewis A in gastric carcinoma and colorectal carcinoma in association with hepatic metastasis. *Int J Oncol* **13**, 935-942 (1998).

8 Nakayama, T., Watanabe, M., Katsumata, T., Teramoto, T. & Kitajima, M. Expression of Sialyl Lewis(a) as a New Prognostic Factor for Patients with Advanced Colorectal-Carcinoma. *Cancer* **75**, 2051-2056, doi:Doi 10.1002/1097-0142(19950415)75:8<2051::Aid-Cncr2820750804>3.0.Co;2-4 (1995).

9 Jeschke, U. *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, invasive carcinomas and their lymph node metastasis. *Anticancer Res* **25**, 1615-1622 (2005).

10 Jorgensen, T. *et al.* Up-Regulation of the Oligosaccharide Sialyl Lewis(X) - a New Prognostic Parameter in Metastatic Prostate-Cancer. *Cancer Res* **55**, 1817-1819 (1995).

11 Schultz, M. J., Swindall, A. F. & Bellis, S. L. Regulation of the metastatic cell phenotype by sialylated glycans. *Cancer Metast Rev* **31**, 501-518, doi:10.1007/s10555-012-9359-7 (2012).

12 Zhuo, Y. & Bellis, S. L. Emerging Role of alpha 2,6-Sialic Acid as a Negative Regulator of Galectin Binding and Function. *Journal of Biological Chemistry* **286**, 5935-5941, doi:10.1074/jbc.R110.191429 (2011).

13 Hsu, D. K., Yang, R. Y. & Liu, F. T. Galectins in apoptosis. *Method Enzymol* **417**, 256-273, doi:10.1016/S0076-6879(06)17018-4 (2006).

14 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev* **18**, 349-450, doi:10.1002/(SICI)1098-2787(1999)18:6<349::AID-MAS1>3.0.CO;2-H (1999).

15 Powell, A. K. & Harvey, D. J. Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption ionization mass spectrometry. *Rapid Commun Mass Sp* **10**, 1027-1032 (1996).

16 Harvey, D. J. Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *J Chromatogr B* **879**, 1196-1225, doi:10.1016/j.jchromb.2010.11.010 (2011).

17 Ruhaak, L. R. *et al.* Glycan labeling strategies and their use in identification and quantification. *Anal Bioanal Chem* **397**, 3457-3481, doi:10.1007/s00216-010-3532-z (2010).

18 Ciucanu, I. & Kerek, F. A Simple and Rapid Method for the Permethylation of Carbohydrates. *Carbohyd Res* **131**, 209-217, doi:Doi 10.1016/0008-6215(84)85242-8 (1984).

19 Morelle, W. & Michalski, J. C. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* **2**, 1585-1602, doi:10.1038/nprot.2007.227 (2007).

20 Kang, P., Mechref, Y., Klouckova, I. & Novotny, M. V. Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun Mass Sp* **19**, 3421-3428, doi:10.1002/rcm.2210 (2005).

21 Kang, P., Mechref, Y. & Novotny, M. V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun Mass Spectrom* **22**, 721-734, doi:10.1002/rcm.3395 (2008).

22 Miura, Y., Shinohara, Y., Furukawa, J. I., Nagahori, N. & Nishimurara, S. Rapid and simple solid-phase esterification of sialic acid residues for quantitative glycomics by mass spectrometry. *Chem-Eur J* **13**, 4797-4804, doi:10.1002/chem.200601872 (2007).

23 Wheeler, S. F., Domann, P. & Harvey, D. J. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of alpha(2 --> 3)- and alpha(2 --> 6)-isomers. *Rapid Commun Mass Spectrom* **23**, 303-312, doi:10.1002/rcm.3867 (2009).

24 Alley, W. R. & Novotny, M. V. Glycomic Analysis of Sialic Acid Linkages in Glycans Derived from Blood Serum Glycoproteins. *J Proteome Res* **9**, 3062-3072, doi:10.1021/pr901210r (2010).

25 Liu, X., Qiu, H. Y., Lee, R. K., Chen, W. X. & Li, J. J. Methylamidation for Sialoglycomics by MALDI-MS: A Facile Derivatization Strategy for Both alpha 2,3-and alpha 2,6-Linked Sialic Acids. *Anal Chem* **82**, 8300-8306, doi:10.1021/ac101831t (2010).

26 Neises, B. & Steglich, W. Simple Method for the Esterification of Carboxylic Acids. *Angewandte Chemie International Edition in English* **17**, 522-524, doi:10.1002/anie.197805221 (1978).

27 König, W. & Geiger, R. Eine neue Methode zur Synthese von Peptiden: Aktivierung der Carboxylgruppe mit Dicyclohexylcarbodiimid unter Zusatz von 1-Hydroxy-benzotriazolen. *Chemische Berichte* **103**, 788-798, doi:10.1002/cber.19701030319 (1970).

28 Chan, L. C. & Cox, B. G. Kinetics of Amide Formation through Carbodiimide/N-Hydroxybenzotriazole (HOBt) Couplings. *The Journal of Organic Chemistry* **72**, 8863-8869, doi:10.1021/jo701558y (2007).

29 Ruhaak, L. R. *et al.* Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem* **80**, 6119-6126, doi:10.1021/ac800630x (2008).

30 Selman, M. H. J., Hemayatkar, M., Deelder, A. M. & Wuhrer, M. Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal Chem* **83**, 2492-2499, doi:10.1021/ac1027116 (2011).

31 Selman, M. H. J. *et al.* Immunoglobulin G Glycopeptide Profiling by Matrix-Assisted Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* **82**, 1073-1081, doi:10.1021/ac9024413 (2010).
32 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).
33 Damerell, D. *et al.* The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol Chem* **393**, 1357-1362, doi:10.1515/hsz-2012-0135 (2012).
34 Ruhaak, L. R., Steenvoorden, E., Koeleman, C. A. M., Deelder, A. M. & Wuhrer, M. 2-Picoline-borane: A non-toxic reducing agent for oligosaccharide labeling by reductive amination. *Proteomics* **10**, 2330-2336, doi:10.1002/pmic.200900804 (2010).
35 Knezevic, A. *et al.* High throughput plasma N-glycome profiling using multiplexed labelling and UPLC with fluorescence detection. *Analyst* **136**, 4670-4673, doi:10.1039/c1an15684e (2011).
36 Stumpo, K. A. & Reinhold, V. N. The N-glycome of human plasma. *J Proteome Res* **9**, 4823-4830, doi:10.1021/pr100528k (2010).
37 Adamczyk, B., Struwe, W. B., Ercan, A., Nigrovic, P. A. & Rudd, P. M. Characterization of Fibrinogen Glycosylation and Its Importance for Serum/Plasma N-Glycome Analysis. *J Proteome Res* **12**, 444-454, doi:10.1021/pr300813h (2013).
38 Townsend, R. R. *et al.* Carbohydrate Structure of Human-Fibrinogen - Use of 300-Mhz H-1-Nmr to Characterize Glycosidase-Treated Glycopeptides. *Journal of Biological Chemistry* **257**, 9704-9710 (1982).
39 Townsend, R. R., Heller, D. N., Fenselau, C. C. & Lee, Y. C. Determination of the Sialylation Pattern of Human-Fibrinogen Glycopeptides with Fast Atom Bombardment. *Biochemistry-Us* **23**, 6389-6392, doi:Doi 10.1021/Bi00321a016 (1984).
40 Maghzal, G. J., Brennan, S. O. & George, P. M. The sialic acid content of fibrinogen decreases during pregnancy and increases in response to fibrate therapy. *Thromb Res* **115**, 293-299, doi:10.1016/j.thromres.2004.08.013 (2005).
41 Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu Rev Biochem* **54**, 631-664, doi:DOI 10.1146/annurev.biochem.54.1.631 (1985).
42 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).
43 Royle, L. *et al.* HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal Biochem* **376**, 1-12, doi:10.1016/j.ab.2007.12.012 (2008).
44 Yoshikawa, H., Hirano, A., Arakawa, T. & Shiraki, K. Mechanistic insights into protein precipitation by alcohol. *Int J Biol Macromol* **50**, 865-871, doi:10.1016/j.ijbiomac.2011.11.005 (2012).
45 Huffman, J. E. *et al.* Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum Mol Genet* **20**, 5000-5011, doi:10.1093/hmg/ddr414 (2011).

Chapter 4

Automation of high-throughput mass spectrometry-based plasma *N*-glycome analysis

Technical note

Karli R. Reiding[1†], Marco R. Bladergroen[1†], Agnes L. Hipgrave Ederveen[1], Gerda C. M. Vreeker[1], Florent Clerc[1], Stephanie Holst[1], Albert Bondt[1,2], Manfred Wuhrer[1,3], Yuri E. M. van der Burgt[1]

†These authors contributed equally

[1]Center for Proteomics and Metabolomics, Leiden University Medical Center (LUMC), PO Box 9600, 2300 RC Leiden, The Netherlands;

[2]Department of Rheumatology, Erasmus University Medical Center, 3000 CA Rotterdam, The Netherlands;

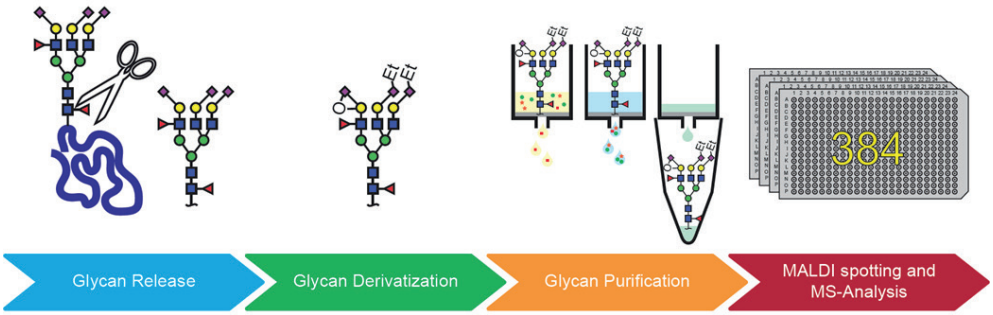
[3]Division of BioAnalytical Chemistry, VU University Amsterdam, 1081 HV Amsterdam, The Netherlands;

Reprinted and adapted with permission from J. Proteome Res., 2015, 14 (9), pp 4080–4086; DOI: 10.1021/acs.jproteome.5b00538; Publication Date (Web): July 16, 2015

Copyright © 2015 American Chemical Society

4.1 Abstract

Glycosylation is a post-translational modification of key importance with heterogeneous structural characteristics. Previously, we have developed a robust, high-throughput MALDI-TOF–MS method for the comprehensive profiling of human plasma *N*-glycans. In this approach, sialic acid residues are derivatized with linkage-specificity, namely the ethylation of α2,6-linked sialic acid residues with parallel lactone formation of α2,3-linked sialic acids. In the current study, this procedure was used as a starting point for the automation of all steps on a liquid-handling robot system. This resulted in a time-efficient and fully standardized procedure with throughput times of 2.5 h for a first set of 96 samples and approximately 1 h extra for each additional sample plate. The mass analysis of the thus-obtained glycans was highly reproducible in terms of relative quantification, with improved interday repeatability as compared to that of manual processing.



4.2 Introduction

Glycosylation is an important post-translational protein modification with significant effects on, for example, conjugate stability, folding, solubility, and receptor interaction[1-3]. These effects also apply to biopharmaceuticals, in which glycans have been shown to influence glycoconjugate half-life and biological activity[4,5]. In addition, glycosylation has proven to be reflective of (and in some cases causal to) disease etiology. Notable examples include tumor growth and metastasis, inflammation, and congenital disorders of glycosylation[6-8]. To study glycosylation for the development of biopharmaceuticals and clinical biomarkers, large numbers of samples are required. Sample processing times, however, are often a limiting factor in high-throughput screening studies. This has prompted efforts toward integrated methodology and automated workflows on robotic liquid-handling platforms[9-11].

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF–MS) is an attractive and high-throughput method for compositional glycan analysis[12,13]. However, such an analysis can be severely hampered by the instability and ionization bias of sialylated glycan species. To overcome these limitations, we recently established a rapid, robust, and linkage-specific high-throughput method for sialic acid stabilization and applied MALDI-TOF–MS analysis to map PNGase F-released human plasma *N*-glycan mixtures[14]. The stabilization and neutralization of these sialic acids was achieved by specific and near-complete carboxyl group derivatization with discrimination between α2,3- and α2,6-linked sialic acids, namely through the ethyl esterification of α2,6-linked sialic acids and lactonization of α2,3-linked variants, leading to the mass-based discrimination of sialylation linkage isomers. Derivatized glycans were then recovered by hydrophilic interaction liquid chromatography (HILIC) solid-phase extraction (SPE) followed by MS analysis. The discrimination of the sialic acid linkages is an important aspect of glycosylation analysis because the isomers may differ in biological functionality. It was, for example, shown that α2,6-sialylation (and not α2,3-sialylation) has an effect on the anti-inflammatory activity of immunoglobulin G, whereas the up-regulation of α2,3-sialylation (and the corresponding increase of sialyl-Lewis X) has been associated with the metastasis of various cancer types[15-17]. The ethyl esterification method has thus far been applied to the in-depth analysis of plasma *N*-glycosylation, as well as to explore the differences between glycans on the IgG variable and constant region and their changes during pregnancy[14-18]. Although the derivatization, SPE, and MALDI measurement can be performed within a few hours for 96 samples, the processing of large cohorts is still time-consuming and tends to result in considerable batch-to-batch variation.

In this paper, we present a fully automated protocol for the high-throughput analysis of *N*-glycans released from plasma proteins with linkage-specific derivatization of sialic acid residues. The protocol makes use of a 96-well filter plate with GH Polypro (GHP) membrane, which has previously been reported to function as the HILIC stationary phase in SPE[19]. Using

a robotic liquid-handling system, 384 samples can be prepared without user intervention within a running time of 5.5 h, followed by MALDI-TOF–MS analysis within an additional 1.5 h. In addition to the beneficial aspect with regard to sample throughput, the automation of this protocol for *N*-glycan release combined with the linkage-specific derivatization of sialic acid residues ensures robustness and standardization.

4.3 Experimental section

Chemicals and reagents

Throughout this study, ultrapure water (MQ) was used (≥18.2 MΩ at 25 °C). Nonidet P-40 substitute (NP-40), 1-hydroxybenzotriazole monohydrate (HOBt), and super-DHB were purchased from Sigma-Aldrich (Steinheim, Germany). Peptide-*N*-glycosidase F (PNGase F) was obtained from Roche Diagnostics (Mannheim, Germany), HPLC-grade acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands), analytical grade ethanol was purchased from Merck (Darmstadt, Germany), and 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide hydrochloride (EDC) was purchased from Fluorochem (Hadfield, United Kingdom).

Samples

Plasma standard (Visucon-F frozen normal control plasma, pooled from 20 human donors, citrated, and buffered with 0.02 M HEPES) was obtained from Affinity Biologicals (Ancaster, ON) and used for all repeatability experiments.

Glycan release

N-Glycans were enzymatically released from plasma proteins using PNGase F as described previously, with minor modifications[20]. In short, a 6 µL plasma sample was denatured by the addition of 12 µL of 2% SDS and incubation for 10 min at 60 °C. For the glycan release, 12 µL of 2.5× PBS containing 2% NP-40 and 0.4 mU PNGase F were added followed by incubation for 16 h at 37 °C. After glycan release, the 96-well plate was either directly transferred to the robot to prepare samples for MALDI-TOF–MS analysis or stored at −20 °C until further use.

Robotic platform

The automated liquid handling platform used in this study is identical to the robotic system that has been described in detail[21,22]. The platform is depicted in **Figure 1** and consists of three units, namely (1) an 8-channel Hamilton MicrolabSTAR; (2) a Hamilton MicrolabSWAP, plate storage, and automated vacuum centrifuge; and (3) a Hamilton MicrolabSTARplus, hereafter referred to as Unit1, Unit2, and Unit3, respectively. The two pipetting platforms (*i.e.*, Unit1 and Unit3) are controlled by two independent computers that are “mastered” through Hamilton scheduler software on a third computer that also controls Unit2.

A modular design was followed to program such an automated protocol. In each procedure, either one or more units or parts of these units (“resources”, *e.g.*, STAR, STARplus, RVC, SWAP, or a combination of these) are allocated to perform an “activity” as controlled through the scheduler software. An overview of all activities that are applied in a procedure is given in **Figure 2A**. All activities are performed via a programmed method. The first method in a procedure initially creates a unique identifier. User input is then required to set up the procedure. This information included, among others, the activities to be performed (*e.g.*, selecting whether MALDI spotting is required after derivatization and cleanup), the method of MALDI spotting, and sample and matrix volumes in case of MALDI spotting.

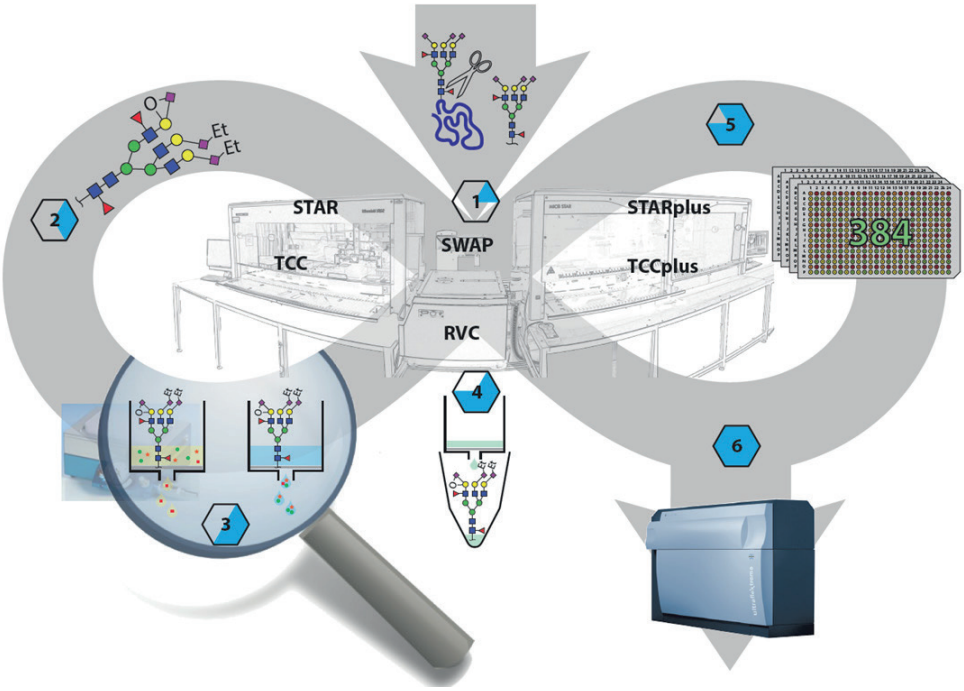


***Figure 1**. Graphical representation of the system setup and the consecutive processing steps locally represented therein: (**1**) PNGase F release, (**2**) ethyl esterification, (**3**) glycan binding and enrichment by GHP HILIC–SPE, (**4**) elution of the glycans from the GHP membrane, (**5**) MALDI target spotting, and (**6**) MALDI-TOF–MS analysis.*

Automated ethyl esterification of released glycans

Before starting a procedure on the robot, freshly prepared chemicals and solutions are supplied to the designated reagent containers. The robotics procedure comprised the following steps: a microtitration plate (96-well PCR plate, polypropylene, Greiner Bio-One, Alphen a/d Rijn, The Netherlands) was filled with 60 µL of ethylation reagent (250 mM EDC and 250 mM HOBt in ethanol) per well. A total of 3 µL of released glycan sample was added, and the plate was covered to prevent solvent evaporation and incubated for 75 min at room

temperature. Before proceeding to purification, 120 µL of ACN was added to the derivatized glycans to prepare the sample for cleanup.

Purification of esterified glycans

A 96-well filter plate with a GHP membrane (GHP plate, Pall AcroPrep Advance 96 Filter plate, Pall Corporation, Ann Arbor, MI) was prewetted, activated, and equilibrated with 100 µL of 70% ethanol, 100 µL of water, and 100 µL of ACN, respectively. The complete volume of the esterification reaction (183 µL) was applied onto the GHP plate, followed by 5 min of incubation, after which the lowest possible vacuum was applied (about 3 kPa below ambient pressure) to ensure a slow flow. The GHP plate was washed three times with 100 µL of 96% ACN. Subsequently, the GHP plate was gently tapped on a disposable hand towel (*e.g.*, Kleenex) to remove residual solvents and transferred onto the top of a PCR plate. For elution, 30 µL of water was added to the filter plate, followed by 5 min of incubation. Recovery was performed by centrifugation at approximately 300 *g* for 5 min.

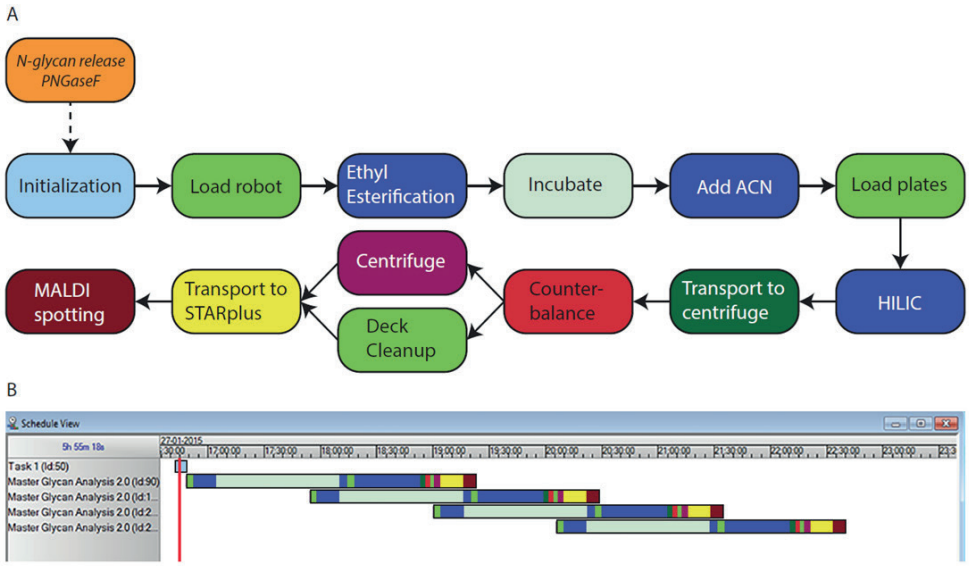


***Figure 2**. Workflow showing the modular setup **A)** and a graphical representation of the scheduled processes for four plates **B)**. Each colored box represents an activity that launches individual programmed methods on the respective resources. The colors in panel A match those in panel B. The resources used are STAR and SWAP (light green); STAR and TCC (dark blue); none (light blue), STAR, RVC, and SWAP (dark green); SWAP and RVC (light red); RVC (purple); SWAP, RVC, STARplus, and TCCplus (yellow); and STARplus and TCCplus (dark red).*

Target plate spotting and MALDI-TOF–MS

Samples were spotted onto a MALDI target plate using the same liquid handling instrument and procedure as described previously[21]. In short, 10 µL of sample was premixed with 10 µL of matrix (5 mg/mL super-DHB in 99% ACN with 1 mM NaOH) in a 384-well plate. A sample

of 2 µL of this mixture was spotted onto a MALDI target plate (800/384 MTP AnchorChip, Bruker Daltonics, Bremen, Germany) and allowed to dry. The MALDI-TOF–MS experiments were performed in reflectron positive mode on an UltrafleXtreme mass spectrometer with a Smartbeam-II laser (Bruker Daltonics) controlled by flexControl 3.4 (Build 135). Within a window of *m/z* 1000 to 5000, spectra were recorded with 10 000 laser shots at a frequency of 1000 Hz, employing a full sample random walking pattern of 100 shots per raster spot.

Method repeatability

To test the repeatability of the automated protocol for the ethyl esterification, purification, MALDI target spotting, and MALDI-TOF–MS analysis of plasma *N*-glycosylation, we prepared a master plate containing 88 released glycan samples (100 µL) originating from a stock solution of mixed releases, as well as 8 PBS blanks. These samples were processed with the robotized method on three different days (2 and 8 days apart) to establish the interday variability (day 1–3), as well as three times within those days to establish the intraday variability (run 1–3). To additionally evaluate the variability introduced by the PNGase F glycan release step, we processed a set of 16 independently released standards on the robotic system as well (day 4).

Analysis

Before the annotation of the obtained mass spectra, an internal calibration was performed by flexAnalysis 3.3 (Build 65, Bruker Daltonics) using a set of *N*-glycan calibration masses (**Table S1** in the Supporting Information). A full list of the detected glycan compositions is provided as **Table S2** in the Supporting Information. Figures were constructed using flexAnalysis, and annotation was performed according to CFG annotation using GlycoWorkbench[23].

For the analysis of the method repeatability, flexAnalysis was used to export all obtained mass spectra into the text format (x,y). The calibration of these exported files and the targeted extraction of the detected glycan compositions was performed by an in-house developed program “MassyTools”. Only spectra that showed a signal-to-noise ratio (S/N) of 9 or above (root-mean-square) for at least five calibration masses were included for further analysis. This method excluded all blanks (72) as well as 9 additional spectra. Furthermore, 20 spectra showing less than 70% of the analyte area above S/N values of 9 were excluded from analysis. For the remaining 779 spectra, the observed areas were normalized to the sum of all areas for each spectrum. For these normalized values, the averages were calculated per plate and day as well as the standard deviations (SDs) thereof. Derived traits were calculated using SPSS Statistics version 20 (IBM) (**Table S3** in the Supporting Information).

4.4 Results and discussion

Protocol implementation on an automated liquid handling platform

We here describe a new fully automated protocol that was implemented on a three-unit Hamilton robotic platform. The protocol is suitable for high-throughput studies on protein *N*-glycosylation by MALDI-TOF–MS involving sequential *N*-glycan release from plasma proteins with PNGase F, ethyl esterification, HILIC-SPE cleanup, and MALDI spotting[14]. The *N*-glycan release from plasma proteins with PNGase F is performed offline because this activity includes an overnight digestion step and does not necessarily benefit from an automation strategy. Further manual steps preceding automation include solution preparation and the placement of plates, pipetting tips, a tissue, and the aforementioned solutions on the robot deck. In addition, the scheduler software requires input on the amount of plates to process and tip numbers and locations, as well as the MALDI target spotting locations and volumes.

Compared to the manual version, the following adjustments were made: (1) chemical derivatization was carried out at room temperature; (2) instead of the previous Sepharose- or cotton-based approach, a GHP plate was used for the purification of esterified glycans; and (3) super-DHB was used as a MALDI matrix.

The sample handling on the robot starts with Unit2 loading the sample plates and processing plates onto Unit1 (step 1 in Figure 1), which in turn places the plates on the required deck positions. During transfer, the plate barcodes are read and saved in the database with the corresponding unique identifier. Through these barcodes, sample data can be imported from a sample list in either the laboratory information management system (LIMS), an Excel spreadsheet, or a text file. This sample list provides not only sample identification but also information on the number and location of samples in the plate to also allow the processing of partially filled plates.

After setup and loading, the first part of the ethyl esterification reaction (until the incubation step) is performed (step 2 in **Figure 1**). Using this setup, Unit1 is liberated for parallel actions during incubation. Following incubation, the addition of ACN and the subsequent purification by HILIC-SPE are performed as separate activities (step 3 in **Figure 1**), as well as the transfer of the resulting filter plate–PCR plate stack by Unit1 and Unit2 into the vacuum centrifuge (step 4 in **Figure 1**). After centrifugation, the stack is transferred to Unit3, the filter plate is discarded, and the eluted samples are spotted onto a MALDI target plate (step 5 in **Figure 1**). Finally, the sample and elution plates are stored in a plate stacker to make the positions available for the next experiment.

Elution of the glycans from the filter plate could alternatively be achieved by applying a vacuum to the GHP plate and collecting the flow-through. We experienced, however, that

droplets may stick to the nozzles of the filter plate, resulting in inconsistent elution volumes and sample loss. Therefore, we prefer a centrifugation procedure and indeed eliminated this problem. For similar reasons, the filter plate is pressed onto a piece of filter paper prior to building the PCR plate–filter plate stack to remove residual liquid from the bottom of the filter plate. A normal centrifuge is not incorporated in our system, but the centrifugal force of the vacuum centrifuge (approximately 300 *g* at a maximum speed of 1350 rpm for 5 min) proved sufficient to achieve complete elution and collection of the eluates in the PCR plate.

Eluates were mixed 1:1 with matrix solution on a separate 384-well plate and spotted on an AnchorChip MALDI target. We found that a high matrix ACN percentage was required to prevent the formation of air bubbles during mixing (leading to spotting inconsistencies). Eluates of a sample plate were directly spotted after the processing of that plate (1 h intervals) and left to dry by air. The resulting spots were analyzed by MALDI-TOF–MS in reflectron positive mode.

The processing of the sample plates on the robot was scheduled to include an overlap of procedures (see **Figure 2B**), allowing the parallel handling of multiple plates in a shorter time frame. This setup of the experiments is managed through an in-house developed Web site which stores the data in a mySQL database. The unique identifier is used here as a reference to the experiment (the sample plate). The current throughput allows the processing of 384 samples (four plates) using a total time of 7 h, including robot preparation and MALDI-TOF–MS measurement. Small changes in the deck layout of the STAR are expected to increase the uninterrupted sample throughput to 768 samples (eight plates) in 9.5 h.

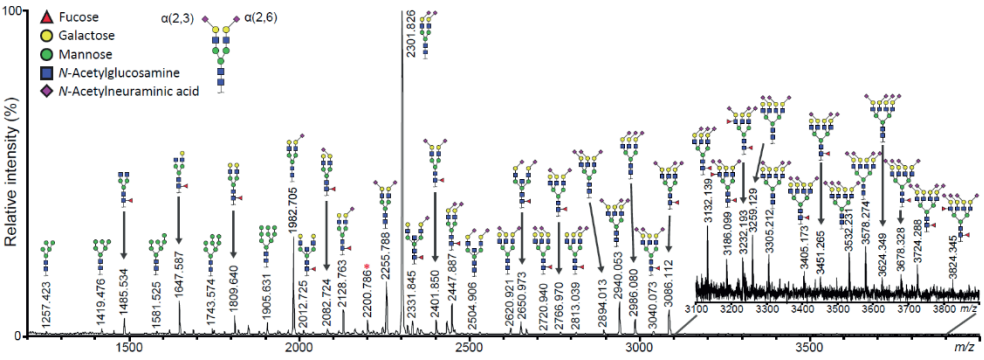


***Figure 3**. Reflectron positive-ion-mode MALDI-TOF–MS spectrum of released plasma N-glycome after robotic ethyl esterification, HILIC–SPE, and MALDI target spotting. Signals are assigned as [M + Na]⁺ glycan compositions on the basis of the observed mass and literature[24-27]. The N-acetylneuraminic acid linkage type could be determined by the mass spectrometric method, but monosaccharide positions and other displayed linkages are presumed on the basis of literature. The asterisk indicates an $^{0,2}A$ fragmentation of the main peak-reducing end N-acetylglucosamine.*

In summary, in this work, the implementation of a fully automated protocol was achieved on a three-unit Hamilton robotic system. Such application of multiple units provides time-efficient throughput; however, it is difficult to transfer to a different platform. Nevertheless, in principle, each part of the automated protocol can be implemented on hardware systems ("resources") from different manufacturers due to the modular design of the workflow (acknowledging that this requires specific adjustments and optimization strategies).

Repeatability testing

To prove the applicability and the advantages of the system, we performed experiments on three different days (day 1–3) with three 96-well sample plates per day (run 1–3). The sample plate used for the repeatability analysis contained 88 wells with a released plasma *N*-glycome standard from a single stock of mixed releases as well as 8 PBS blanks. In addition, separate releases were performed on 16 plasma samples to evaluate the added variation of the PNGase F release step (day 4). Obtained reflectron positive mode MALDI-TOF–MS spectra proved to be informative for the detection of 80 glycan compositions (average mass error 0.01 Da SD ± 0.01 Da), including various sialylated glycans separated by their linkage isomers (**Figure 3** and **Table S2** in the Supporting Information). In addition, no glycan signals were observed in any of the 72 blanks, indicating an absence of sample cross-contamination throughout the automated workflow.

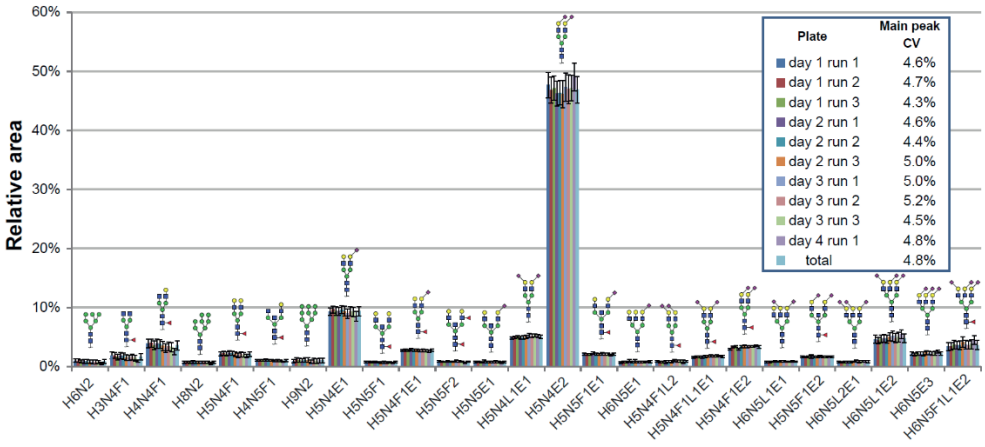


***Figure 4**. Repeatability of the robotized esterification workflow, represented by the relative profile of the 25 most abundant glycan compositions. A single mixture of released plasma N-glycome was distributed across a 96 well plate and subjected to the robotized sample preparation method for 3 days, 3 times per day, with subsequent MALDI-TOF–MS measurement. The fourth day was composed of 16 independent releases showing minimal contributions of the release step to the variation of the profiles. Error bars indicate standard deviation. H = hexose, N = N-acetylhexosamine, F = deoxyhexose (fucose), L = lactonized N-acetylneuraminic acid (α2,3-linked), E = ethyl esterified N-acetylneuraminic acid (α2,6-linked), CV = coefficient of variation.*

The consecutive analyses of the standard samples showed to be highly repeatable, both intraday (main peak averages per day being 47.2% (SD ± 2.2%), 46.2% (SD ± 2.2%), and 47.2% (SD ± 2.3%) of the total area of all assigned peaks) and interday (the main peak average across all days being 46.8% (SD ± 2.3%)) (**Figure 4** and **Figure S1** in the Supporting Information). The main peak coefficient of variation (CV) was, on average, 4.7% (SD ± 0.2%) but was generally higher for peaks that comprise a smaller portion of the total profile. For day 4, the main peak average was 49.0% (SD ± 2.4%, CV 4.8%), indicating a minor contribution of the PNGase F release to the profile repeatability. Method repeatability is at the level reported for manual preparation by a well-trained technician[14].

More clinically interesting than the individual relative glycan levels are the derived traits calculated from a subset of glycans. Examples of these include the percentage of high-mannose-type glycans (M), fucosylation of diantennary glycans (A2F), and sialylation of triantennary fucosylated glycans (A3FS) (Table S3 in the Supporting Information). Advantageously, these derived traits showed a higher repeatability than the individual traits from which they were calculated (Figure S2 in the Supporting Information). When we calculated 47 derived traits over the 779 spectra, the SDs were on average 1.6% (SD ± 0.8%). The average CV for the 43 derived traits with intensity values above 5% was 4.2% (SD ± 3.3%). The higher repeatability of the derived traits as compared to that of single glycan values might be due to the chemical similarity of the glycans contributing to the derived traits. This similarity may cause them to be similarly affected by small biases during the workflow (e.g., during HILIC-SPE, crystallization, or MS ionization), making ratios within them more robust than those obtained across a chemically more diverse spectrum of analytes (as is the case with individual glycan values relative to a total area or reference peak).

4.5 Conclusion

The objective of this study was to speed up and standardize glycan derivatization, purification, and analysis by MALDI-TOF–MS through automation using a robotic liquid-handling system. We were able to implement a time-efficient procedure on an automated system that showed to be highly repeatable with performance similar to or better than that of manual processing. Throughput times are 2.5 h for the first set of 96 samples and approximately 1 h extra for each additional sample plate. Automation of the protocol resulted in high intraday and interday repeatability, with standard deviations of the individual glycan compositions being below 3% on average and still lower for traits derived from those.

4.6 Supporting information

Figure S1: repeatability of the robotized esterification workflow for 80 detected glycan compositions. **Figure S2**: repeatability of derived glycosylation traits across all analyzed

spectra. **Table S1**: calibration masses. **Table S2**: compositions detected by reflectron positive mode MALDI-TOF–MS after robotized ethyl esterification. **Table S3**: derived glycan trait properties and calculations. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00538.

4.7 Acknowledgments

This work was supported by the Seventh Framework Programme projects HighGlycan (grant no. 278535) and IBD-BIOM (305479) of the European Union.

4.8 References

1 Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and their roles in the immune system. *Nat Rev Immunol* **7**, 255-266, doi:10.1038/nri2056 (2007).

2 Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370-2376, doi:DOI 10.1126/science.291.5512.2370 (2001).

3 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130 (1993).

4 Alessandri, L. *et al.* Increased serum clearance of oligomannose species present on a human IgG1 molecule. *MAbs* **4**, 509-520, doi:10.4161/mabs.20450 (2012).

5 Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery* **8**, 226-234, doi:10.1038/nrd2804 (2009).

6 Adamczyk, B., Tharmalingam, T. & Rudd, P. M. Glycans as cancer biomarkers. *Bba-Gen Subjects* **1820**, 1347-1353, doi:10.1016/j.bbagen.2011.12.001 (2012).

7 Albrecht, S., Unwin, L., Muniyappa, M. & Rudd, P. M. Glycosylation as a marker for inflammatory arthritis. *Cancer Biomark* **14**, 17-28, doi:10.3233/CBM-130373 (2014).

8 Guillard, M. *et al.* Plasma N-glycan profiling by mass spectrometry for congenital disorders of glycosylation type II. *Clin Chem* **57**, 593-602, doi:10.1373/clinchem.2010.153635 (2011).

9 Shubhakar, A. *et al.* High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia* **78**, 321-333, doi:10.1007/s10337-014-2803-9 (2015).

10 Doherty, M. *et al.* An automated robotic platform for rapid profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture. *Anal Biochem* **442**, 10-18, doi:10.1016/j.ab.2013.07.005 (2013).

11 Stockmann, H., Adamczyk, B., Hayes, J. & Rudd, P. M. Automated, High-Throughput IgG-Antibody Glycoprofiling Platform. *Anal Chem* **85**, 8841-8849, doi:10.1021/ac402068r (2013).

12 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev* **18**, 349-450, doi:10.1002/(SICI)1098-2787(1999)18:6<349::AID-MAS1>3.0.CO;2-H (1999).

13 Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).

14 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).

15 Anthony, R. M. *et al.* Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* **320**, 373-376, doi:10.1126/science.1154315 (2008).

16 Kannagi, R. Molecular mechanism for cancer-associated induction of sialyl Lewis X and sialyl Lewis A expression - The Warburg effect revisited. *Glycoconj J* **20**, 353-364, doi:Doi 10.1023/B:Glyc.0000033631.35357.41 (2003).

17 Takada, A. *et al.* Contribution of Carbohydrate Antigens Sialyl Lewis-a and Sialyl Lewis-X to Adhesion of Human Cancer-Cells to Vascular Endothelium. *Cancer Res* **53**, 354-361 (1993).

18 Bondt, A. *et al.* Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics* **13**, 3029-3039, doi:10.1074/mcp.M114.039537 (2014).

19 Burnina, I., Hoyt, E., Lynaugh, H., Li, H. J. & Gong, B. A cost-effective plate-based sample preparation for antibody N-glycan analysis. *J Chromatogr A* **1307**, 201-206, doi:10.1016/j.chroma.2013.07.104 (2013).

20 Ruhaak, L. R. *et al.* Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem* **80**, 6119-6126, doi:10.1021/ac800630x (2008).

21 Bladergroen, M. R. *et al.* Standardized and automated solid-phase extraction procedures for high-throughput proteomics of body fluids. *J Proteomics* **77**, 144-153, doi:10.1016/j.jprot.2012.07.023 (2012).

22 van den Broek, I. *et al.* Quantification of serum apolipoproteins A-I and B-100 in clinical samples using an automated SISCAPA-MALDI-TOF-MS workflow. *Methods* **81**, 74-85, doi:10.1016/j.ymeth.2015.03.001 (2015).

23 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).

24 Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu Rev Biochem* **54**, 631-664, doi:DOI 10.1146/annurev.biochem.54.1.631 (1985).

25 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).

26 Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **13**, 2314-2327, doi:10.1021/pr401092y (2014).

27 Stumpo, K. A. & Reinhold, V. N. The N-glycome of human plasma. *J Proteome Res* **9**, 4823-4830, doi:10.1021/pr100528k (2010).

Part III – Method application

Chapter 5

Murine plasma *N*-glycosylation traits associated with sex and strain

Research article

Karli R. Reiding[1], Agnes L. Hipgrave Ederveen[1], Yoann Rombouts[1,2], Manfred Wuhrer[1]

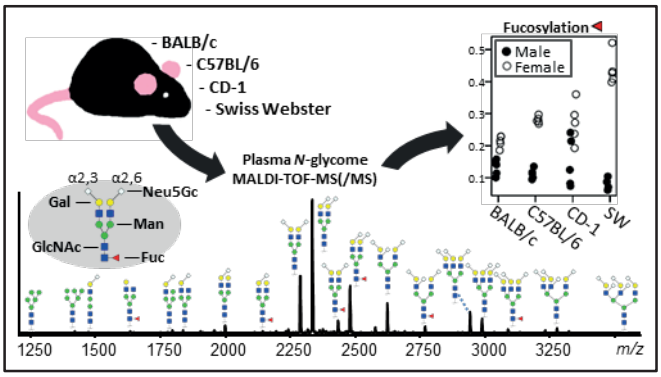
[1]Leiden University Medical Center, Center for Proteomics and Metabolomics, Leiden 2333ZA, The Netherlands;

[2]Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, Toulouse 31077, France;

Reprinted and adapted with permission from J. Proteome Res., 2016, 15 (10), pp 3489–3499; DOI: 10.1021/acs.jproteome.6b00071; Publication Date (Web): August 22, 2016

Copyright © 2016 American Chemical Society

5.1 Abstract



Glycosylation is an abundant and important protein modification with large influence on the properties and interactions of glycoconjugates. Human plasma *N*-glycosylation has been the subject of frequent investigation, revealing strong associations with physiological and pathological conditions. Less well-characterized is the plasma *N*-glycosylation of the mouse, the most commonly used animal model for studying human diseases, particularly with regard to differences between strains and sexes. For this reason, we used MALDI-TOF(/TOF)-MS(/MS) assisted by linkage-specific derivatization of the sialic acids to comparatively analyze the plasma *N*-glycosylation of both male and female mice originating from BALB/c, CD57BL/6, CD-1, and Swiss Webster strains. The combined use of this analytical method and the recently developed data processing software named MassyTools allowed the relative quantification of the *N*-glycan species within plasma, the distinction between α2,3- and α2,6-linked *N*-glycolylneuraminic acids (due to respective lactonization and ethyl esterification), the detection of sialic acid *O*-acetylation, as well as the characterization of branching sialylation (Neu5Gcα2,3-Hex-[Neu5Gcα2,6-]HexNAc). When analyzing the glycosylation according to mouse sex, we found that female mice present a considerably higher degree of core fucosylation (2–4-fold depending on the strain), galactosylation, α2,6-linked sialylation, and larger high-mannose type glycan species compared with their male counterparts. Male mice, on the contrary, showed on average higher α2,3-linked sialylation, branching sialylation, and putative bisection. These differences together with sialic acid acetylation proved to be strain-specific as well. Interestingly, the outbred strains CD-1 and Swiss Webster displayed considerably larger interindividual variation than inbred strains BALB/c and CD57BL/6, suggesting a strong hereditable component of the observed plasma *N*-glycome.

5.2 Introduction

Glycosylation is an important protein modification with functions ranging from protein quality control in the endoplasmic reticulum to receptor interaction and glycoprotein clearance from blood[1-3]. Because of this, the physiological and pathological conditions of an organism are highly reflected in the structure of *N*-glycans carried by individual proteins as well as within a complex biofluid such as plasma, making the study of the whole plasma *N*-glycome of interest with regard to biomarker development[4-7]. Importantly, while protein sequence and structure may be vastly conserved between species, protein glycosylation is often species-dependent and the glycan epitopes expressed in one species may be targeted by the immune system of another[8]. For instance, *N*-glycolylneuraminic acid (Neu5Gc) is the most commonly expressed sialic acid in many mammalian species, including mice, but is specifically absent and immunogenic in humans, which instead express *N*-acetylneuraminic acid (Neu5Ac)[9,10]. Further important structural features of sialic acids are their linkages (α2,3-, α2,6-, or α2,8-linked to, *e.g.*, galactose, *N*-acetylglucosamine (GlcNAc) and other sialic acids) and substituents like *O*-acetylation.

Whereas human plasma glycosylation has been well-studied, less is known about that of mouse, even though it is the most widely used animal model for in vivo experimentation. Knowledge is lacking with respect to the differences in glycosylation between commonly used mouse strains and sexes, which do display differential susceptibility to the development of disease[11]. For instance, the Swiss Webster and C57BL/6 strains are often used in cancer and diabetes research, whereas the BALB/c strain is preferentially used for investigating infectious diseases[12,13]. While glycosylation may play a role in these, a systematic study comparing the glycosylation of mouse strains and sexes has yet to be performed.

Thus, to enhance our knowledge of murine *N*-glycosylation, we analyzed by matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF)-mass spectrometry (MS) the total plasma *N*-glycomes of male and female mice belonging to the commonly used strains C57BL/6, BALB/c, CD-1, and Swiss Webster. To this end, we applied a recently developed derivatization method allowing the linkage-specific mass spectrometric detection of sialic-acid-containing *N*-glycans in reflectron positive-ion mode, with selective ethyl esterification of α2,6-linked residues and lactonization of α2,3-linked residues[14,15]. This analytical method led to the derivatization of both Neu5Ac and Neu5Gc residues, linked to either galactose or *N*-acetylglucosamine, while preserving sialic acid substituents (*i.e.*, *O*-acetylation). The obtained data were extracted and processed using MassyTools, a recently developed data processing software, prior to statistical analysis[16]. As a result, several differences in *N*-glycosylation-derived traits, especially in fucosylation, galactosylation, bisection, and sialic acid linkage, were identified when comparing male and female mice. These traits, as well as sialic acid acetylation, were also found to differ between strains.

5.3 Experimental section

Samples

Mouse plasma samples were obtained from three different suppliers. For mouse strains BALB/c, C57BL/6, CD-1, and Swiss Webster, plasma samples (disodium EDTA) from five individual males and five individual females, aged 8–12 weeks, were obtained from BioChemed Services (Winchester, VA). For the strains BALB/c, C57BL/6, CD-1, and Swiss Webster, mixed-sex (1 male and 1 female) disodium EDTA plasmas were acquired from Seralab (West-Sussex, U.K.). Mixed-sex pooled disodium EDTA plasma samples (ages 7–11 weeks) of BALB/c, C57BL/6, and CD-1 strains were furthermore obtained from Innovative Research (Novi, MI). Human pooled plasma (Visucon-F frozen normal control plasma, citrated and buffered with 0.02 M HEPES) was purchased from Affinity Biologicals (Ancaster, ON, Canada) to serve as technical control.

N-Glycan sample preparation

Throughout the study, we used ultrapure deionized water (MQ) generated from a Q-Gard 2 system (Millipore, Amsterdam, Netherlands) maintained at ≥18.2 MΩ. The analytical procedure was performed in quadruplicate for the five individual male and five individual female mice from the BALB/c, C57BL/6, CD-1, and Swiss Webster strains and in triplicate for the mixed plasmas. As controls, four human plasma samples and four blanks were included in the analysis.

N-Glycans were released from plasma samples as previously described[17]. Five µL of each plasma sample was denatured by adding 10 µL of 2% sodium dodecyl sulfate (Merck, Darmstadt, Germany) and 10 min of incubation at 60 °C. Following cool down at room temperature, the denatured samples were mixed with 10 µL of 2% Nonidet P-40 substitute (Sigma-Aldrich, Steinheim, Germany) and 0.5 mU recombinant peptide-*N*-glycosidase F (Roche Diagnostics, Mannheim, Germany) in 2.5× PBS and incubated overnight at 37 °C.

The released *N*-glycans were then derivatized by ethyl esterification and purified by hydrophilic interaction chromatography (HILIC) similar to previously described[14,18,19]. The derivatization reagent was prepared by mixing 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (Fluorochem, Hadfield, U.K.) and 1-hydroxybenzotriazole (Sigma-Aldrich) in ethanol (Merck) to a concentration of 0.25 M for both reagents. Two µL of released *N*-glycans was added to 20 µL of derivatization reagent, followed by 1 h incubation at 37 °C. The derivatized glycans were purified by HILIC using pipet tips filled with cotton as stationary phase. To this end, 20 µL of acetonitrile (ACN; Biosolve, Valkenswaard, The Netherlands) was added to the reaction mixture. The cotton tips were prewashed three times with 20 µL of MQ water and three times with 20 µL of 85% ACN, followed by pipetting 20 times up and down to load the sample. The cotton tips were

then washed by pipetting three times 20 µL of 85% ACN 1% trifluoroacetic acid (Merck) and three times 20 µL of 85% ACN, followed by elution of the glycans with 10 µL of MQ.

MALDI-TOF-MS profiling and structural elucidation

For each sample, 6 µL of HILIC eluate was mixed with 3 µL of 5 mg/mL super-DHB (Sigma-Aldrich) 1 mM NaOH (Sigma-Aldrich) in 50% ACN. Two µL of this mixture, containing the *N*-glycosylation of ~53 nL of plasma, was spotted on an MTP AnchorChip 800/384 TF MALDI target (Bruker Daltonics, Bremen, Germany) plate and left to dry by air at room temperature before proceeding to MALDI-TOF-MS measurement.

All MS analyses were performed in reflectron positive mode on an UltrafleXtreme mass spectrometer with a Smartbeam-II laser (Bruker Daltonics), controlled by flexControl 3.4 (Build 135; Bruker Daltonics). Within a window of *m/z* 1000 to 5000, spectra were recorded by combining 10 000 laser shots at a frequency of 1000 Hz, employing a full sample random walking pattern of 100 shots per raster spot. Tandem mass spectrometry (MALDI-TOF/TOF-MS/MS) was performed on the most abundant peaks of the mouse *N*-glycome via laser-induced dissociation.

To further validate the identity of highly sialylated signals, a 5 µL part of the release mixture was labeled at the reducing end with 2-aminobenzoic acid, followed by HILIC purification and nanoLC-reversed phase (RP)-electrospray (ESI)-quadrupole time-of-flight (QTOF)-MS(/MS) (Supporting Information, Experimental)[20].

Data extraction and processing

For relative quantification and quality control of the mass spectra and analytes, the raw spectra were exported from flexAnalysis 3.4 (Build 76; Bruker Daltonics) as text file (x,y) and further processed by MassyTools (version 0.1.8.0)[16], which first performed internal calibration based on a predefined list of analytes (**Table S1**). Mass spectra presenting a signal-to-noise ratio (S/N value, MinMax) of 6 or above for at least six calibration masses were included for further analysis. As a result, 14 out of 192 obtained spectra were excluded, which included all four blanks. From the remaining spectra, 120 visually established glycan compositions were extracted by integrating the areas of 80% of the theoretical isotopic envelope and subtracting from these the background detected within a window of 20 *m/z*. Additional output of MassyTools included the fraction of analyte area above S/N 9 throughout a spectrum, the highest isotope S/N, and ppm error per analyte for each spectrum as well as a value indicating the match between the observed and theoretical isotopic envelopes per glycan for each spectrum (QC). Two additional spectra were excluded from the analysis due to their “fraction of analyte area above S/N 9” below three standard deviations (SDs) of the mean.

Analytes were removed when not present in >60% of spectra of any biological group (a combination of strain and sex) with an S/N of at least 3, a ppm error below 20, and a QC value below 0.05. For the remaining 56 glycan compositions, detected with high confidence, the areas were normalized to the sum of all areas, and derived traits were calculated based on the compositional features (for the raw data see **Table S2**; for the calculations and descriptions of the derived traits, see **Table S3**).

For display purposes, mass spectra were smoothed in flexAnalysis 3.4 (SavitzkyGolay, 0.1 width, 1 cycle), baseline-subtracted (TopHat), peakpicked with the 56 glycan compositions (Snap2), and internally calibrated (cubic enhanced) using the predefined list of analytes (**Table S1**). For each strain and both sexes, the peakpicked analytes were verified with regard to area, S/N, and ppm error after compositional assignment (**Table S4**). Structural information was obtained from MALDI-TOF/TOF-MS/MS and reverse-phase LC-ESI-QTOF-MS/MS with regard to sialic acid linkage and type (by mass), antenna composition, and fucose location (**Table S5**), whereas other linkages were presumed on the basis of literature[21-24]. Figures were annotated according to CFG standards and with structures created in GlycoWorkbench (version 2.1)[25,26].

Data analysis

The variability of the measurement was established by calculating the mean, standard deviation, and coefficient of variation within each technical replicate (**Table S2**). The mean value of the technical replicates was then used for analysis of the biological variation.

Statistical analysis was performed using RStudio 0.98.1091 (RStudio, Inc.) in combination with R 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria). Unsupervised principal component analysis (PCA) was performed on either the individual glycans or the derived traits, using the mixOmics package (version 5.20) (**Table S6**). R was used to make figures of the PCA score and loading as well as for the strip plots showing the variable distribution.

Because the glycan variables were not deemed normally distributed between the sexes, strains, or strains per sex, significance testing between groups was performed by nonparametric Mann–Whitney U tests using the R function wilcox.test (**Table S7**). The derived traits of male and female mice (pooling the strains) and the different strains (pooling the sexes) were compared. Results were deemed significant if the *p*-value was below the significance threshold when Bonferroni-corrected for the total number of test performed ($\alpha = 0.05/385 = 1.3 \times 10^{-4}$). For those comparisons that proved significant, we performed additional Mann–Whitney U tests to compare the male and female individuals within each strain separately and to compare the differences per strain for each of the sexes while correcting for the number of tests within these selected traits (**Table S8**).

5.4 Results

Characterization of the mouse plasma *N*-glycosylation

The plasma *N*-glycosylation of BALB/c, C57BL/6, CD-1, and Swiss Webster mice (5 per sex, 10 per strain, 40 in total) were analyzed by reflectron positive ion mode MALDI-TOF/TOF-MS/MS after the enzymatic release of *N*-glycans and their derivatization, allowing a selective ethyl esterification of α2,6-linked residues and lactonization of α2,3-linked residues (for an example mass spectrum of a BALB/c female, see **Figure 1A** and low and high-mass sections thereof (**Figure 1B**, **C**); for representative mass spectra of male and female individuals per strain, see **Figure S1**)[14,15].

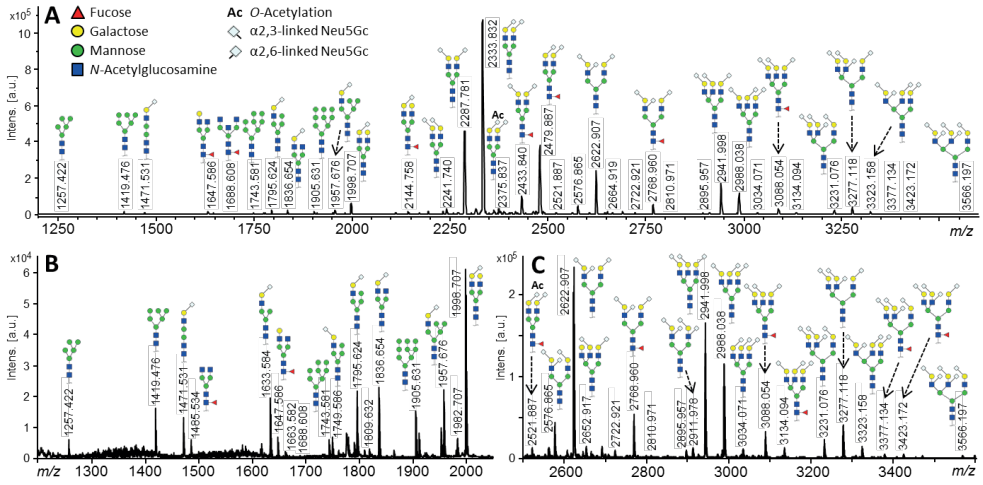


***Figure 1**. Example of a murine total plasma N-glycome profile (BALB/c female) as measured by reflectron positive mode MALDI-TOF-MS after ethyl esterification. **A)** Overall spectrum. **B)** Low mass spectrum from m/z 1200 to m/z 2050. **C)** High mass spectrum from m/z 2500 to m/z 3600. Glycan species are assigned to signals on the basis of recorded m/z values [M + Na]⁺ and MS/MS data. MS data provide N-glycolylneuraminic acid linkages, antennary compositions, and fucose locations, whereas other linkages are presumed on the basis of literature[21-24].*

The applied method led to the derivatization of both Neu5Ac and Neu5Gc residues, linked to either galactose or *N*-acetylglucosamine, while preserving sialic acid substituents (*i.e.*, *O*-acetylation). Following data processing, the obtained spectra proved to be informative for the repeatable detection of 56 glycan compositions ranging from high mannose species H5N2 (*m/z* 1257.423 [M + Na]⁺) to the highly sialylated triantennary H6N5E2L3 (*m/z* 3566.196 [M + Na]⁺) (H = hexose; N = *N*-acetylhexosamine; F = fucose; L = lactonized (α2,3-linked) Neu5Gc; E = ethyl esterified (α2,6-linked) Neu5Gc) (**Table S4**).

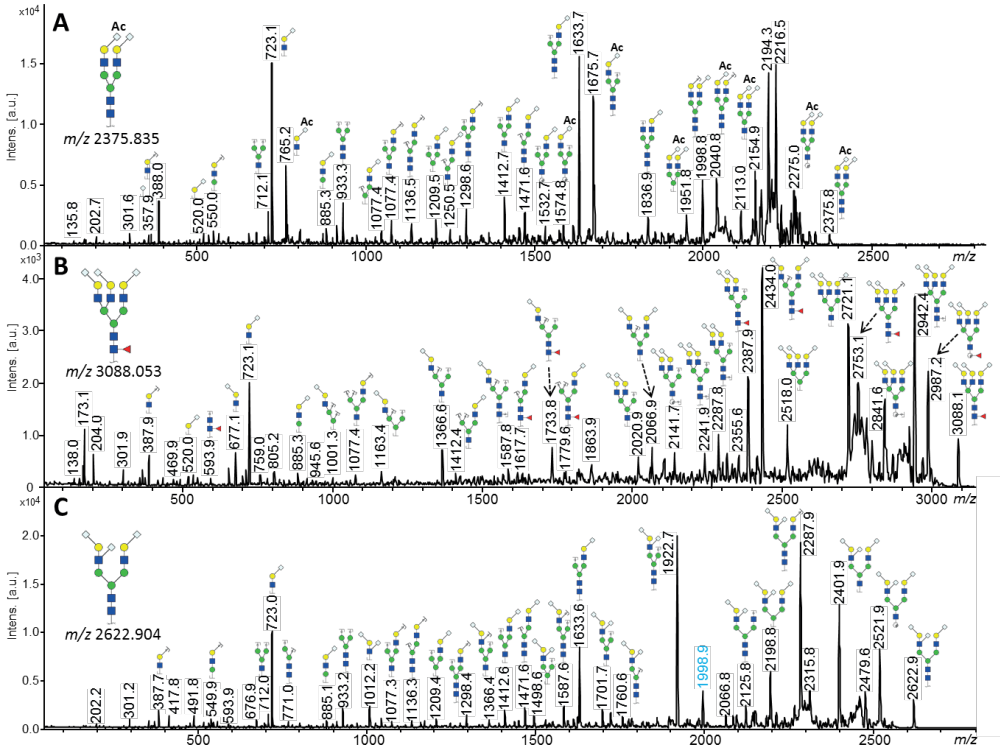


***Figure 2**. MALDI-TOF/TOF-MS/MS profiles showing the fragmentation of **A)** the signal at m/z 2375.835 (H5N4E2Ac1 [M + Na]⁺) containing N-glycolylneuraminic acid acetylation, **B)** the signal at m/z 3088.053 (H6N5F1E1L2 [M + Na]⁺) showing the core fucosylation of triantennary species, and **C)** the signal at m/z 2622.904 (H5N4E2L1 [M + Na]⁺) containing branching sialylation (Neu5Gcα2,3-Galβ1,3-[Neu5Gcα2,6-]GlcNAc). While the fragment at m/z 1998.9 (blue) would be indicative of α2,8-linked sialylation (the loss of Neu5Gcα2,8-Neu5Gcα2,6 from a Neu5Gcα2,8-Neu5Gcα2,6-Galβ1,4-GlcNAc motif), no congruent B-type fragment was observed, and the signals at m/z 1587.6 and m/z 1366.4 can only occur when the lactonized sialic is attached to a galactose (rather than another sialic acid). A mixture of the two structures was further disproved by reverse-phase LC-ESI-QTOF-MS/MS (**Figure S7**). H = hexose; N = N-acetylhexosamine; F = deoxyhexose (fucose); L = lactonized (α2,3- or α2,6-linked) N-glycolylneuraminic acid; E = ethyl-esterified (α2,6-linked) N-glycolylneuraminic acid; Ac = acetylation.*

Information on glycan structures, particularly on antenna composition and fucose location, was obtained by MALDI-TOF/TOF-MS/MS (**Figure 2**; **Table S5**). Glycan fragment ions showed that potential O-acetyl groups were associated with the sialic acids (**Figure 2A**; *e.g.*, *m/z* 765.2 indicative for acetylated Neu5Gcα2,6-Hex-HexNAc [M + Na]⁺), while fucose residues were mainly attached to the core-GlcNAc for both di- and triantennary species (**Figure 2B**; *e.g.*, *m/z* 593.9 indicative for a reducing end HexNAc-[dHex-]HexNAc [M + Na]⁺). Moreover, as previously described, MS/MS fragmentation of trisialylated diantennary *N*-glycans indicated that the disialylated antenna can be assigned to Neu5Gcα2,3-Hex-[Neu5Gcα2,6-]HexNAc structure (**Figure 2C**)[27]. Of note, we also found a specific fragment at *m/z* 1998.9

[M + Na]$^+$ (**Figure 2C**, blue; indicative of the loss of a lactonized and ethyl-esterified sialic acid), suggesting the potential presence of a disialyl motif (Neu5Gcα2,8-Neu5Gcα2,6) with a distinct fragmentation pathway originating from the same precursor. However, reverse-phase LC-ESI-QTOF-MS on 2-aminobenzoic acid-labeled *N*-glycans from the same release mixture showed a clear ion at *m/z* 511.2 (Neu5Gc-HexNAc [M + H]$^+$), whereas no disialyl oxonium ion could be observed (theoretical *m/z* 615.188 [M + H]$^+$) (**Figure S2**). As such, the presence of more sialic acids than antennae was interpreted as being due to branching sialylation.

The individual glycan signals from each spectrum were extracted by MassyTools to quantify the differences between the sexes and strains (**Table S2**)[16]. The technical variation was assessed by performing the full analytical procedure in quadruplicate, yielding an average CV of 4.7% (SD ± 2.1%) for the main peak, which was similar to previously reported for the method (**Figure S3**; **Table S2**)[14,28]. The average of the technical replicates was used for further examination (**Figure 3**), and derived traits were calculated from the individual glycans to reflect the various enzymatic steps involved in glycosylation (for calculations and descriptions see **Table S3**).

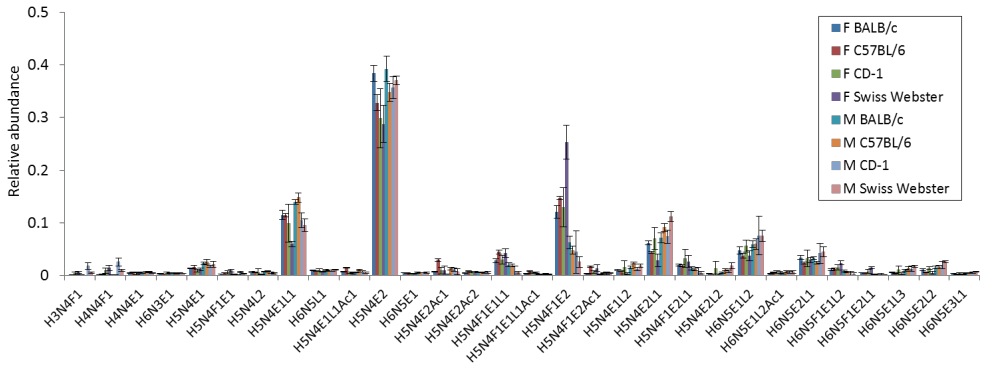


***Figure 3**. Relative abundances of the 30 most abundant N-glycans from the average mouse plasma. Shown are the means of five biological replicates, themselves the means of up to four technical replicates. Error bars show standard deviation. The variation seen in the CD-1 and Swiss Webster strains reflects the observed individual variability of the strain members.*

Sex-related differences in mouse plasma *N*-glycome

Using either the individual glycans or the derived traits, PCA was performed to assess the variation in the data (**Figures S4** and **S5**; **Table S6**). In both cases, 10 principal components were calculated, explaining, respectively, 94 and 98% of the cumulative variation. For the derived traits, in particular, the first three components could explain 43, 26, and 11% of the variance, already taking most of the variation in the data under consideration. Coloring the score plots by sex revealed a strong separation across the first principal component (**Figure S5**), notable individual variables in the male direction being α2,3-linked Neu5Gc (L)

abundance and branching sialylation (Sb), while the female direction contained α2,6-linked sialylation (E), fucosylation (F), and average size (number of mannoses) in high-mannose glycans (MM). Furthermore, considerable clustering of the biological replicates was observed, yet, while particularly true for the male and female BALB/c and C57BL/6 individuals, the CD-1 and Swiss Webster individuals (especially the CD-1 females) showed larger variation across the dimensions.

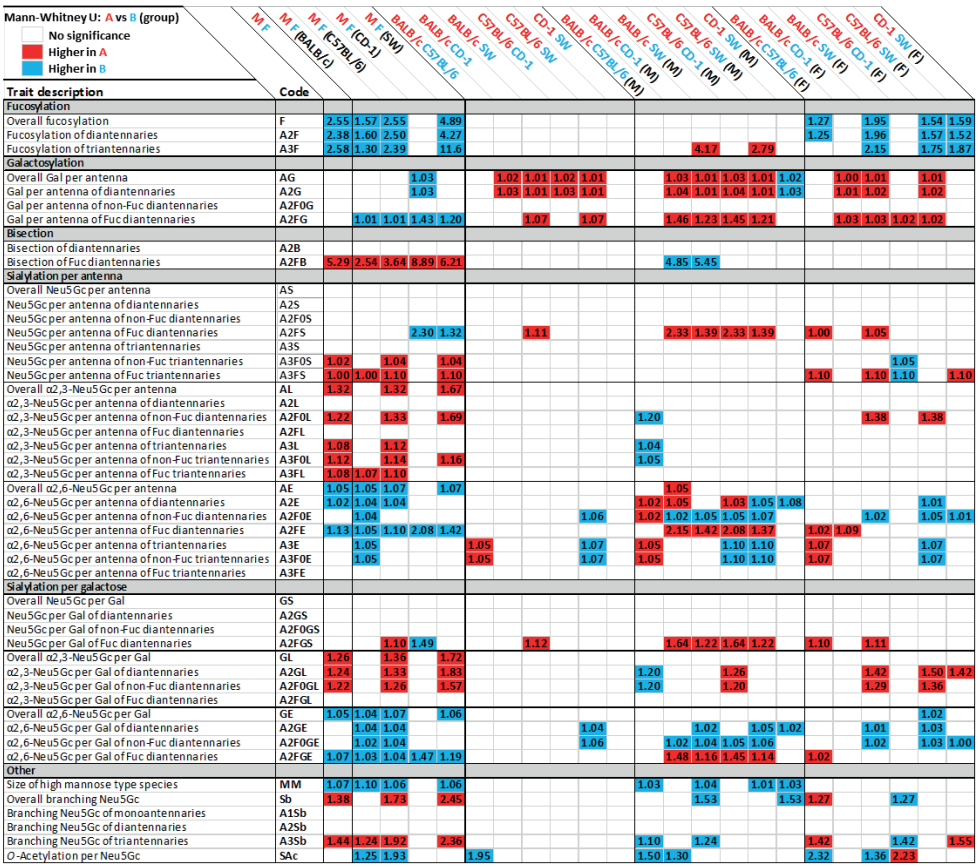


***Figure 4**. Summary of statistically significant differences (Mann–Whitney U) in derived glycosylation traits between male (M) and female (F) mice, between males and females within strains BALB/c, C57BL/6, CD-1, and Swiss Webster (SW), and between these strains (overall and separately for males and females). Compared are A (red) versus B (blue) (e.g., M vs F within BALB/c, or BALB/c vs C57BL/6 within males), with the boxes indicating which group has the higher value in corresponding color. The numerical values provide the fold difference between the group medians.*

Mann–Whitney U tests were performed to compare male and female mice (combined strains) employing a significance threshold of $\alpha = 1.3 \times 10^{-4}$ (**Table S7**), resulting in a total of 22 derived traits being significantly different between the sexes (**Figure 4**; **Table S8**). The derived traits significant in Mann–Whitney U tests were largely in agreement with the major

discriminators in PCA (**Tables S6** and **S7**). Overall, as compared with female mice, males proved to have higher percentages of bisection of diantennary fucosylated species (A2FB; a male median of 0.74% versus a female median of 0.14%; $p = 9.25 \times 10^{-7}$), lactonized sialylation of diantennary nonfucosylated structures (A2F0GL; 22% versus 18%; $p = 5.21 \times 10^{-5}$), and lactonized sialylation of triantennary structures (A3L; 57% versus 53%; $p = 1.83 \times 10^{-5}$; seen in both A3F0L (51% versus 57%; $p = 8.18 \times 10^{-6}$) and A3FL (65% versus 60%; $p = 6.91 \times 10^{-6}$)) as well as increased branching sialylation of diantennary glycans (A2Sb; 17% versus 12%; $p = 1.19 \times 10^{-4}$) and triantennary glycans (A3Sb; 23% versus 16%; $p = 1.65 \times 10^{-6}$).

Females, on the contrary, expressed larger high-mannose-type structures (MM; median average high-mannose size of 7.4 mannoses in females versus 6.9 mannoses in males; $p = 1.13 \times 10^{-6}$), an overall higher fucosylation (F; 28% versus 11%; $p = 2.02 \times 10^{-9}$; visible both in A2F (31% versus 13%; $p = 2.83 \times 10^{-9}$) and A3F (16% versus 6%; $p = 2.76 \times 10^{-10}$), as well as an increased amount of α2,6-sialylation of diantennary fucosylated structures (A2FE; 86% versus 76%; $p = 1.65 \times 10^{-6}$). Likewise, α2,6-sialylation per galactose was found to be increased for diantennary fucosylated structures (A2FGE; 88% versus 82%; $p = 5.08 \times 10^{-8}$).

Strain-related differences in mouse plasma *N*-glycome

PCA showed strains BALB/c and C57BL/6 to separate from CD-1 and Swiss Webster across the second principal component (**Figure S5**). Contributing to the BALB/c and C57BL/6 direction were galactosylation and sialylation of (fucosylated) diantennary compositions, while CD-1 and Swiss Webster were defined by bisection and α2,6-sialylation of nonfucosylated triantennary compositions.

Comparing the strains by Mann–Whitney U (combined sex; $\alpha = 1.3 \times 10^{-4}$) yielded a total of 11 statistically significant differences, again in agreement with the principal component separation (**Figure 4**; **Figure S5**; **Tables S6** and **S7**). BALB/c and C57BL/6 mice showed, as compared with CD-1 and Swiss Webster, a higher degree of galactosylation of diantennary structures (A2G; respective medians of 99, 99, 96, and 98%; $p = 1.08 \times 10^{-5}$ for BALB/c versus CD-1, BALB/c versus Swiss Webster and C57BL/6 versus CD-1 (minimum possible *p*-value for a Mann–Whitney U test for 2 groups of 10 cases) and $p = 2.17 \times 10^{-5}$ for C57BL/6 versus CD-1). Differences were even more pronounced when looking at the galactosylation of the fucosylated diantennary species (A2FG), with BALB/c and C57BL/6 (98% each) strains presenting higher levels than CD-1 (87%) and Swiss Webster (92%; $p < 7.6 \times 10^{-5}$ throughout). C57BL/6 (43%) also exhibited slightly lower degree of α2,6-linked sialylation of triantennary nonfucosylated structures (A3F0E) as compared with BALB/c (45%; $p = 4.33 \times 10^{-5}$) and Swiss Webster (46%; $p = 1.08 \times 10^{-5}$), while the comparison with CD-1 (46%; $p = 6.84 \times 10^{-3}$) revealed a nonsignificant trend after correction for multiple testing. Interestingly, we also observed a difference in the degree of sialic acid acetylation when comparing C57BL/6 and BALB/c mice (SAc; respectively, 4.3 and 2.2%; $p = 1.08 \times 10^{-5}$).

The glycosylation traits that showed significantly different between either the sexes or the strains were further examined to reveal strain-specific sex differences and sex-specific strain differences (**Figure 4**; **Tables S7** and **S8**). Notably, high mannose size (MM), fucosylation (F, A2F, A3F), and α2,6-linked sialylation of diantennary species (AE, GE, A2GE), were highest in females for every strain independently, the statistical exception being the CD-1 mice for which the interindividual variation precluded from reaching the significance threshold. Bisection (A2B, A2FB), on the contrary, showed higher in males for all strains, as well as being higher in CD-1 and Swiss Webster than in BALB/c and C57BL/6. Galactosylation of fucosylated diantennary species (A2FG) proved higher in females than in males only for CD-1 and Swiss Webster strains. With respect to acetylation of sialic acids (SAc), female individuals of the C57BL/6 strain expressed particularly high levels. The degree of branching sialylation (A2Sb, A3Sb, Sb) varied per strain but was specifically higher in males than in females for C57BL/6 and Swiss Webster (**Figure 5**; **Figure S6**).

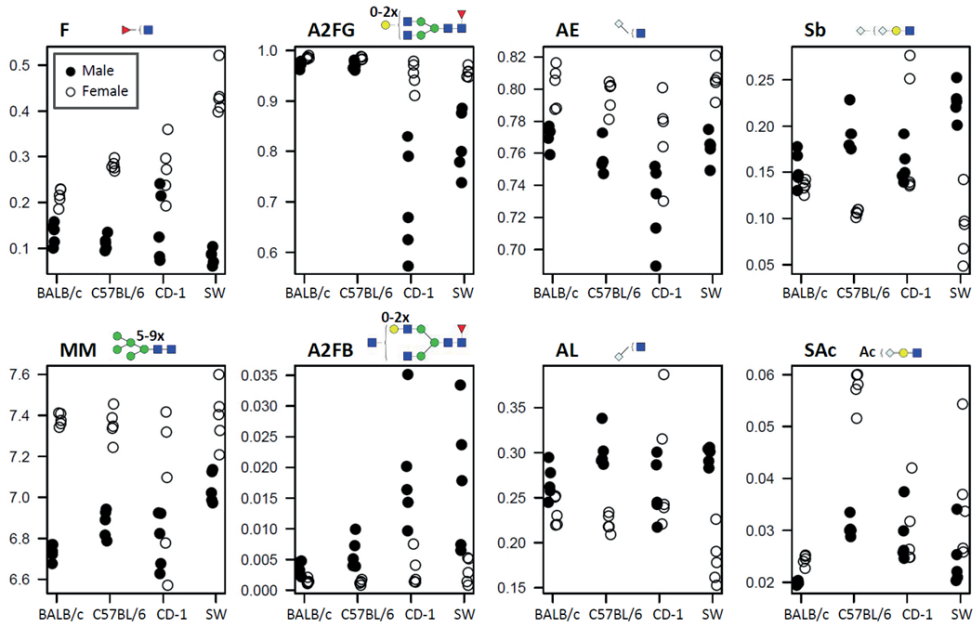


***Figure 5**. Overview of eight glycosylation traits with sex- and strain-specific differences. Each point represents the technical average of one mouse individual, with males and females indicated by respectively black and white coloring. Shown here are the overall fucosylation (F), galactosylation of diantennary fucosylated species (A2FG), overall α2,6-linked sialylation per antenna (AE), overall branching sialylation (Sb), average number of mannoses for high-mannose species (MM), bisection of diantennary fucosylated species (A2FB), overall α2,3-linked sialylation per antenna (AL), and overall acetylation per sialic acid (SAc). Swiss Webster mice are abbreviated to SW. For the derived trait figures, the minimal required composition is displayed to the right of a bracket (used as denominator for the trait calculation) and the calculated feature to the left (used as numerator). Restrictions are further indicated by numerical values. For a full list of derived trait figures, descriptions, and calculations, see **Table S3**.*

Using mixed-sex mouse plasma pools originating from a second supplier, we could confirm all differences between the inbred strains (*i.e.*, α2,6-sialylation of triantennary glycans and sialic acid acetylation between BALB/c and C57BL/6), while the triantennary α2,6-sialylation differences could be also confirmed for the other strain comparisons (**Table S2**). Diantennary galactosylation and sialylation, the main separators between the outbred strains CD-1 and Swiss Webster, proved variable between the suppliers but showed similar levels in all cases. Samples originating from a third supplier were discontinued from comparative analysis due to the abundant detection of N-acetylneuraminic acid in mouse strain C57BL/6, a sialic acid not principally endogenous to mice (**Figure S7**)[10].

5.5 Discussion

Structural characterization of mouse plasma *N*-glycoforms

In the research presented here, we used MALDI-TOF-MS to analyze the released plasma *N*-glycome profiles of individual male and female mice originating from the commonly used strains BALB/c, C57BL/6, CD-1, and Swiss Webster. To achieve this, we made use of the previously established ethyl esterification derivatization method, which exploits the propensity for α2,3- and α2,6-linked *N*-acetylneuraminic acids to, respectively, lactonize or form an ester/amide with an external nucleophile, resulting in stabilization of sialylated species in reflectron mode MALDI-TOF-MS and discrimination of the sialic acid linkage isomers by the induced mass changes[14,15,19,29,30]. Applying the method for the analysis of the mouse samples allowed the detection of Neu5Gc, the distinction between Neu5Gc and Neu5Ac, and the discrimination of their linkage isomers. Additionally, we detected masses suggesting the presence of acetyl groups that were located on Neu5Gc moieties on the basis of MS/MS fragmentation and literature[31,32]. Likewise, we demonstrated that the presence of multiple sialic acids per antenna is most likely due to the α2,6-linkage of a sialic acid to an antennary GlcNAc residue (branching sialylation) as we were unable to detect the isomeric structure with an antennary disialyl motif (Neu5Gcα2,8-Neu5Gcα2,6-Gal-GlcNAc)[27,33-36].

Polysialic acid residues are known to readily lactonize, a reaction that is further accelerated by the ethyl esterification reagents[37,38]. Because the lactonization of an α2,8-linked sialic acid yields an equal mass to that of an α2,3-linked variant, MS distinction would only be possible with subsequent MS/MS experimentation. Previously, we observed that lactonized sialic acids mainly dissociate from a precursor molecule together with their carrier monosaccharide(s) (*e.g.*, Neu5Gcα2,3-Gal or Neu5Gcα2,3-Galβ1,4-GlcNAc)[14]. We did observe the same phenomenon in the fragmentation profiles of the mouse plasma *N*-glycans and also found in several MS/MS spectra a fragment suggesting the loss of an ethyl-esterified and lactonized Neu5Gc directly from the precursor (*m/z* 624.201) indicative of such a Neu5Gcα2,8-Neu5Gcα2,6 motif. However, because we failed to detect a disialyl B-

type fragment in both MALDI-TOF/TOF-MS/MS (*m/z* 647.191 $[M + Na]^+$) and reverse-phase LC-ESI-QTOF-MS/MS, the motif was considered not present in detectable abundance.

In the case of branching sialylation, an ethyl ester was expected as reaction outcome due to the absence of a nearby intramolecular nucleophile to form a stable five- or six-membered ring structure. As predicted, we found an ethylated Neu5Gcα2,6-GlcNAc B-type fragment (*m/z* 561.190 $[M + Na]^+$) as well as the independent loss of the sialic acid directly from the precursor (*m/z* 335.122) in fragmentation spectra of most eligible precursor masses, thereby confirming the presence of Neu5Gcα2,3-Hex-[Neu5Gcα2,6-]HexNAc motif.

Profiling of mouse plasma *N*-glycosylation

Overall, the observed plasma *N*-glycomes were in good accordance with those obtained from more quantitative chromatographic methods with fluorescence detection[39]. For each strain, the spectra were dominated by the major composition H5N4E2, the differently linked variant H5N4E1L1, and these species with additional fucose (H5N4F1E2 and H5N4F1E1L1). Notably, fucosylation was predominantly found on the core *N*-acetylglucosamine for both diantennary and triantennary species. This is in contrast with human plasma glycosylation, where for tri- and tetraantennary glycans, which mostly originate from acute phase proteins, fucosylation is generally found on the antennae in the form of sialyl-Lewis X[2,3,24].

Glycosylation-derived traits were generated to look at enzymatic processes and putative receptor epitopes. As such, sialylation of diantennary structures (A2S, or A2L and A2E for, respectively, lactonized (α2,3-linked) and ethyl esterified (α2,6-linked) sialylation) informs on the level of sialylation encountered by selectins[40], while sialylation per galactose within the same antennary class (A2GS, A2GL, and A2GE) is a measure for the enzymatic attachment of sialic acids to terminal galactose residues or enzymatic desialylation in plasma[41,42]. Note that the analytical methodology used in this study did not provide information on the linkage of monosaccharides other than the sialic acids. Thus, bisection is mainly predicted from compositions having three branches (five *N*-acetylhexosamines) and incomplete galactosylation (fewer than six hexoses), losing the possibility to quantify the level of galactosylation in triantennary species and leading to possible misclassification of structures with LacdiNAc motifs. In human plasma, this provides a good measure of bisection because tri- and tetraantennary species are chiefly reported as having a near-full galactosylation of all antennae[24], but in mice this situation may be different. Similarly, we did not identify α-linked galactosylation, a murine glycan epitope with immunogenic properties in humans[8].

Mouse sex- and strain-specific plasma *N*-glycosylation differences

Interestingly, we observed large differences between male and female mice. For all strains, fucosylation was higher in females than in males, in the case of Swiss Webster even 4-fold so. Also, higher in females was the average size of the high mannose species, mainly due to

lower levels of H5N2 and H6N2 compositions. Branching sialylation was found to be higher in males than in females, which was supported by a concurrently increased degree of lactonized sialylation of di- and triantennary species (A2F0GL, A3L). The presence of these additional sialic acids has been described for many of the glycoproteins abundant in murine plasma, including fetuin, alpha-2-macroglobulin, adiponectin, vitronectin, and immunoglobulins G, M, and E[36,43]. Furthermore, when comparing the plasma *N*-glycosylation of mouse strains BALB/c, C57BL/6, CD-1, and Swiss Webster, additional differences could be found in galactosylation, sialylation, and sialic acid acetylation. Remarkably, within both BALB/c and C57BL/6, the individuals show strong clustering with regard to their glycosylation traits, whereas CD-1 and Swiss Webster glycosylation vary considerably between individuals within each group. A notable example of this is the galactosylation per antenna of diantennary fucosylated glycans (A2FG), which is ∼98% for each individual of the first two strains, but ranges from 57 to 83% in CD-1 males. One likely explanation of the differential variability is that BALB/c and C57BL/6 are inbred mice, while CD-1 and Swiss Webster are outbred mice and have a consequentially wider genetic background[44-47]. The difference in glycosylation variability between these inbred and outbred strains suggests a strong genetic influence on the expression of glycoforms or glycosylated proteins.

While sialylation and sialic acid acetylation differences between strains could be confirmed, diantennary galactosylation and sialylation proved more variable, perhaps reflecting the immune status of the mice. Interestingly, for one supplier, we could in C57BL/6 mice detect the abundant presence of *N*-acetylneuraminic acid, a sialic acid not commonly found in mice[10], whereas a second batch from the same supplier had a considerably decreased level. Because the varying level of Neu5Ac may be due to diet, sample cross-contamination, or other factors unrelated to biological variability, the samples were excluded from further comparison[48,49]. These observations, however, do demonstrate the interest in relative quantification of the sialic acid variants by the used methodology and underline the need for glycosylation analysis of biological samples prior to functional experimentation.

Overall we expect the plasma *N*-glycome differences between the sexes and strains to originate from a combination of (1) relative glycoprotein concentrations, (2) the production of selected protein glycoforms, and (3) the recycling and breakdown of selected protein glycoforms. Regarding the first option, agar electrophoresis has indicated sizable serum protein differences between mouse sexes and strains[50,51]. For instance, male individuals, when compared with females, show a 2 to 4 times higher alpha-1 globulin fraction (roughly containing the alpha-lipoproteins, alpha-1-antitrypsin, and alpha-1-acid glycoprotein) and a 1 to 2 times lower gamma-globulin fraction (roughly containing immunoglobulins G and M as well as fibrinogen), all of which contain major glycoproteins[50,52]. Similar (1.5–2-fold) globulin fraction differences were found between BALB/c, C57BL/6, and CD-1 mice,

suggesting that relative protein abundances would be a major contributor to the observed total plasma *N*-glycome profiles[51].

Second, specific regulation of individual sites of glycosylation is likely to play a large role as well. Major plasma components involved in systemic health, including immunoglobulins, acute phase proteins, and coagulation factors, are known to be functionally affected by their glycosylation and would reflect the mouse sex- and strain-specific susceptibility and response to disease[53-55]. An example of this can be the proinflammatory glycosylation phenotype (a lower degree of diantennary glycosylation and higher bisection in males) found in the CD-1 and Swiss Webster strains when compared with BALB/c and C57BL/6[7].

Third, for the observed strain and sex differences in sialylation and acetylation in particular, additional explanation may be found in the regulation of protein half-life. *N*-Glycans carried by secreted plasma glycoproteins are known to be sequentially trimmed by plasma glycosidases, the resulting glycan epitopes being recognized by endocytic lectins for recycling (*e.g.*, asialoglycoprotein, agalactoglycoprotein, and mannose receptors)[42]. Neuraminidases act at the first step of this *N*-glycan remodeling, with the shaving of sialic acids accounting for the majority of secreted protein turnover, while *O*-acetylation is reported to have an inhibitory effect on the neuraminidase action[56]. Interestingly, many diseases (*e.g.*, diabetes, cancer, cirrhosis, infection) show association with altered plasma glycosidase levels, suggesting glycan-related modulation of secreted protein aging and turnover to be an involved mechanism[42,57]. In putative association with the sex- and strain-specific susceptibility to disease, this plasma glycoprotein remodeling would also be an influential factor on the glycosylation profiles observed in this study.

5.6 Conclusions and perspectives

Employing a recently developed analytical method, we have analyzed by MALDI-TOF-MS the plasma *N*-glycosylation of male and female mice belonging to four different strains. In addition to the identification of Neu5Gc sialylation in various linkages, we characterized the presence of branching sialylation and sialic acid acetylation. The baseline glycosylation properties of mouse strains are an important aspect to consider when selecting an experimental animal (as are the differences with human glycosylation), and we envision this comparative research as well as the described methodology to assist in this effort. Further glycoproteomic and protein-specific studies are needed to determine to which extent the glycosylation differences observed between sexes and strains are caused by glycoprotein expression levels, or by differences in site-specific glycosylation.

Supporting information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00071.

- Supporting experimental section. (PDF)

- **Figure S1**: Assigned MALDI-TOF-MS profiles of male and female individuals from strains BALB/c, C57BL/6, CD-1, and Swiss Webster. **Figure S2**: Reverse-phase LC-ESI-QTOF-MS/MS analysis of a trisialylated diantennary *N*-glycan indicating branching sialylation. **Figure S3**: Integrated relative glycan intensities of each individual mouse quadruplicate and averages per sex and strain. **Figure S4**: PCA score and loading plots of the four principal components calculated from the individual glycans, displaying each mouse individual colored by sex, strain, or strain in a sex-specific manner. **Figure S5**: PCA score and loading plots of the four principal components calculated from the derived glycosylation traits, displaying each mouse individual colored by sex, strain, or strain in a sex-specific manner. **Figure S6**: Strip plots for all glycans and derived traits, clustered per mouse strain and sex. **Figure S7**: MALDI-TOF-MS profile showing incidental *N*-acetylneuraminic acid sialylation within C57BL/6 plasma samples. (PDF)

- **Table S1**: MS calibrants. **Table S2**: Complete data set of integrated glycosylation data. **Table S3**: Derived trait descriptions and calculations. **Table S4**: MS assignment. **Table S5**: MS/MS assignment. **Table S6**: PCA results. **Table S7**: Raw Mann–Whitney U test results. **Table S8**: Summary of Mann–Whitney U test results. (XLSX)

5.7 Acknowledgments

We acknowledge support by the European Union (Seventh Framework Programme HighGlycan project, grant number 278535). We furthermore thank Noortje de Haan for assisting with the R code for generating plots and performing Mann-Whitney U tests and David Falck for helpful chemical insights.

5.8 References

1 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130 (1993).

2 Satoh, T., Yamaguchi, T. & Kato, K. Emerging Structural Insights into Glycoprotein Quality Control Coupled with N-Glycan Processing in the Endoplasmic Reticulum. *Molecules* **20**, 2475-2491, doi:10.3390/molecules20022475 (2015).

3 Stockert, R. J., Kressner, M. S., Collins, J. C., Sternlieb, I. & Morell, A. G. Iga Interaction with the Asialoglycoprotein Receptor. *P Natl Acad Sci-Biol* **79**, 6229-6231, doi:DOI 10.1073/pnas.79.20.6229 (1982).

4 Kristic, J. *et al.* Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* **69**, 779-789, doi:10.1093/gerona/glt190 (2014).

5 Ruhaak, L. R. *et al.* Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* **10**, 1667-1674, doi:10.1021/pr1009959 (2011).

6 Lu, J. P. *et al.* Screening Novel Biomarkers for Metabolic Syndrome by Profiling Human Plasma N-Glycans in Chinese Han and Croatian Populations. *J Proteome Res* **10**, 4959-4969, doi:10.1021/pr2004067 (2011).
7 Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res Rev* **12**, 685-698, doi:10.1016/j.arr.2012.02.002 (2013).
8 Chung, C. H. *et al.* Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *New Engl J Med* **358**, 1109-1117, doi:Doi 10.1056/Nejmoa074943 (2008).
9 Varki, A. N-glycolylneuraminic acid deficiency in humans. *Biochimie* **83**, 615-622, doi:Doi 10.1016/S0300-9084(01)01309-8 (2001).
10 Raju, T. S., Briggs, J. B., Borge, S. M. & Jones, A. J. S. Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology* **10**, 477-486, doi:DOI 10.1093/glycob/10.5.477 (2000).
11 Nguyen, D. & Xu, T. The expanding role of mouse genetics for understanding human biology and disease. *Dis Model Mech* **1**, 56-66, doi:10.1242/dmm.000232 (2008).
12 Walrath, J. C., Hawes, J. J., Van Dyke, T. & Reilly, K. M. Genetically Engineered Mouse Models in Cancer Research. *Adv Cancer Res* **106**, 113-164, doi:10.1016/S0065-230x(10)06004-5 (2010).
13 Cheemarla, N. R. & Guerrero-Plata, A. Immune Response to Human Metapneumovirus Infection: What We Have Learned from the Mouse Model. *Pathogens* **4**, 682-696, doi:10.3390/pathogens4030682 (2015).
14 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).
15 de Haan, N. *et al.* Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. *Anal Chem* **87**, 8284-8291, doi:10.1021/acs.analchem.5b02426 (2015).
16 Jansen, B. C. *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* **14**, 5088-5098, doi:10.1021/acs.jproteome.5b00658 (2015).
17 Ruhaak, L. R. *et al.* Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem* **80**, 6119-6126, doi:10.1021/ac800630x (2008).
18 Selman, M. H. J., Hemayatkar, M., Deelder, A. M. & Wuhrer, M. Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal Chem* **83**, 2492-2499, doi:10.1021/ac1027116 (2011).
19 Reiding, K. R., Lonardi, E., Hipgrave Ederveen, A. L. & Wuhrer, M. Ethyl Esterification for MALDI-MS Analysis of Protein Glycosylation. *Methods Mol Biol* **1394**, 151-162, doi:10.1007/978-1-4939-3341-9_11 (2016).
20 Ruhaak, L. R., Steenvoorden, E., Koeleman, C. A. M., Deelder, A. M. & Wuhrer, M. 2-Picoline-borane: A non-toxic reducing agent for oligosaccharide labeling by reductive amination. *Proteomics* **10**, 2330-2336, doi:10.1002/pmic.200900804 (2010).
21 Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu Rev Biochem* **54**, 631-664, doi:DOI 10.1146/annurev.biochem.54.1.631 (1985).
22 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).
23 Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **13**, 2314-2327, doi:10.1021/pr401092y (2014).
24 Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-343, doi:10.1007/s10719-015-9626-2 (2016).

25 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, doi:10.1093/glycob/cwv091 (2015).
26 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).
27 Rombouts, Y. *et al.* Acute phase inflammation is characterized by rapid changes in plasma/peritoneal fluid N-glycosylation in mice. *Glycoconj J* **33**, 457-470, doi:10.1007/s10719-015-9648-9 (2016).
28 Bladergroen, M. R. *et al.* Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *J Proteome Res* **14**, 4080-4086, doi:10.1021/acs.jproteome.5b00538 (2015).
29 Wheeler, S. F., Domann, P. & Harvey, D. J. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of alpha(2 --> 3)- and alpha(2 --> 6)-isomers. *Rapid Commun Mass Spectrom* **23**, 303-312, doi:10.1002/rcm.3867 (2009).
30 Alley, W. R. & Novotny, M. V. Glycomic Analysis of Sialic Acid Linkages in Glycans Derived from Blood Serum Glycoproteins. *J Proteome Res* **9**, 3062-3072, doi:10.1021/pr901210r (2010).
31 Varki, A. & Gagneux, P. Multifarious roles of sialic acids in immunity. *Ann Ny Acad Sci* **1253**, 16-36, doi:10.1111/j.1749-6632.2012.06517.x (2012).
32 Hua, S. *et al.* Isomer-Specific LC/MS and LC/MS/MS Profiling of the Mouse Serum N-Glycome Revealing a Number of Novel Sialylated N-Glycans. *Anal Chem* **85**, 4636-4643, doi:10.1021/ac400195h (2013).
33 Lin, S. Y. *et al.* Precise mapping of increased sialylation pattern and the expression of acute phase proteins accompanying murine tumor progression in BALB/c mouse by integrated sera proteomics and glycomics. *J Proteome Res* **7**, 3293-3303, doi:10.1021/pr800093b (2008).
34 Lattova, E., Varma, S., Bezabeh, T., Petrus, L. & Perreault, H. Mass spectrometric profiling of N-linked oligosaccharides and uncommon glycoform in mouse serum with head and neck tumor. *J Am Soc Mass Spectr* **19**, 671-685, doi:10.1016/j.jasms.2008.01.016 (2008).
35 Yasukawa, Z., Sato, C. & Kitajima, K. Inflammation-dependent changes in alpha 2,3-, alpha 2,6-, and alpha 2,8-sialic acid glycotopes on serum glycoproteins in mice. *Glycobiology* **15**, 827-837, doi:10.1093/glycob/cwi068 (2005).
36 Sato, C., Yasukawa, Z., Honda, N., Matsuda, T. & Kitajima, K. Identification and adipocyte differentiation-dependent expression of the unique disialic acid residue in an adipose tissue-specific glycoprotein, adipo Q. *Journal of Biological Chemistry* **276**, 28849-28856, doi:DOI 10.1074/jbc.M104148200 (2001).
37 Lifely, M. R., Gilbert, A. S. & Moreno, C. Sialic-Acid Polysaccharide Antigens of Neisseria-Meningitidis and Escherichia-Coli - Esterification between Adjacent Residues. *Carbohyd Res* **94**, 193-203, doi:Doi 10.1016/S0008-6215(00)80717-X (1981).
38 Azurmendi, H. F. *et al.* Extracellular structure of polysialic acid explored by on cell solution NMR. *Proc Natl Acad Sci U S A* **104**, 11557-11561, doi:10.1073/pnas.0704404104 (2007).
39 Saldova, R. *et al.* Increase in Sialylation and Branching in the Mouse Serum N-glycome Correlates with Inflammation and Ovarian Tumour Progression. *PLoS One* **8**, doi:ARTN e7115910.1371/journal.pone.0071159 (2013).
40 Varki, A. Sialic acids as ligands in recognition phenomena. *Faseb J* **11**, 248-255 (1997).
41 Dall'Olio, F., Malagolini, N., Trinchera, M. & Chiricolo, M. Sialosignaling: Sialyltransferases as engines of self-fueling loops in cancer progression. *Bba-Gen Subjects* **1840**, 2752-2764, doi:10.1016/j.bbagen.2014.06.006 (2014).
42 Yang, W. H. *et al.* An intrinsic mechanism of secreted protein aging and turnover. *Proc Natl Acad Sci U S A* **112**, 13657-13662, doi:10.1073/pnas.1515464112 (2015).

43 Yasukawa, Z., Sato, C., Sano, K., Ogawa, H. & Kitajima, K. Identification of disialic acid-containing glycoproteins in mouse serum: a novel modification of immunoglobulin light chains, vitronectin, and plasminogen. *Glycobiology* **16**, 651-665, doi:10.1093/glycob/cwj112 (2006).

44 Eppig, J. T. *et al.* The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res* **43**, D726-D736, doi:10.1093/nar/gku967 (2015).

45 Aldinger, K. A., Sokoloff, G., Rosenberg, D. M., Palmer, A. A. & Millen, K. J. Genetic Variation and Population Substructure in Outbred CD-1 Mice: Implications for Genome-Wide Association Studies. *PLoS One* **4**, doi:ARTN e472910.1371/journal.pone.0004729 (2009).

46 Jacome, L. F., Burket, J. A., Herndon, A. L. & Deutsch, S. I. Genetically inbred Balb/c mice differ from outbred Swiss Webster mice on discrete measures of sociability: relevance to a genetic mouse model of autism spectrum disorders. *Autism Res* **4**, 393-400, doi:10.1002/aur.218 (2011).

47 Chia, R., Achilli, F., Festing, M. F. W. & Fisher, E. M. C. The origins and uses of mouse outbred stocks. *Nat Genet* **37**, 1181-1186, doi:10.1038/ng1665 (2005).

48 Malykh, Y. N., Schauer, R. & Shaw, L. N-glycolylneuraminic acid in human tumours. *Biochimie* **83**, 623-634, doi:Doi 10.1016/S0300-9084(01)01303-7 (2001).

49 Nguyen, D. H., Tangvoranuntakul, P. & Varki, A. Effects of natural human antibodies against a nonhuman sialic acid that metabolically incorporates into activated and malignant immune cells. *J Immunol* **175**, 228-236 (2005).

50 Espinosa, E., Canelo, E., Bravo, M. & Gonzalez, O. Sex-Associated Differences in Serum Proteins of Mice. *Science* **144**, 417-&, doi:DOI 10.1126/science.144.3617.417 (1964).

51 Zaias, J., Mineau, M., Cray, C., Yoon, D. & Altman, N. H. Reference Values for Serum Proteins of Common Laboratory Rodent Strains. *J Am Assoc Lab Anim* **48**, 387-390 (2009).

52 Jeppsson, J. O. *et al.* Agarose-Gel Electrophoresis. *Clin Chem* **25**, 629-638 (1979).

53 Zauner, G. *et al.* Glycoproteomic analysis of antibodies. *Mol Cell Proteomics* **12**, 856-865, doi:10.1074/mcp.R112.026005 (2013).

54 Hayes, J. M. *et al.* Glycosylation and Fc Receptors. *Curr Top Microbiol* **382**, 165-199, doi:10.1007/978-3-319-07911-0_8 (2014).

55 McCarthy, C. *et al.* The Role and Importance of Glycosylation of Acute Phase Proteins with Focus on Alpha-1 Antitrypsin in Acute and Chronic Inflammatory Conditions. *J Proteome Res* **13**, 3131-3143, doi:10.1021/pr500146y (2014).

56 Corfield, A. P., Sanderwewer, M., Veh, R. W., Wember, M. & Schauer, R. The Action of Sialidases on Substrates Containing O-Acetylsialic Acids. *Biol Chem H-S* **367**, 433-439, doi:DOI 10.1515/bchm3.1986.367.1.433 (1986).

57 Grewal, P. K. *et al.* Inducing host protection in pneumococcal sepsis by preactivation of the Ashwell-Morell receptor. *Proc Natl Acad Sci U S A* **110**, 20218-20223, doi:10.1073/pnas.1313905110 (2013).

Chapter 6

Human plasma *N*-glycosylation as analyzed by MALDI-FTICR-MS associates with markers of inflammation and metabolic health

Research article

Karli R. Reiding[1], L. Renee Ruhaak[2], Hae-Won Uh[3], Said el Bouhaddani[3], Erik B. van den Akker[3,4], Rosina Plomp[1], Liam A. McDonnell[1], Jeanine J. Houwing-Duistermaat[3,5], P. Eline Slagboom[6], Marian Beekman[6], Manfred Wuhrer[1]

[1]Center for Proteomics and Metabolomics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

[2]Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

[3]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

[4]Pattern Recognition & Bioinformatics, Delft University of Technology, 2600 GA Delft, The Netherlands;

[5]Department of Statistics, University of Leeds, LS2 9JT Leeds, United Kingdom;

[6]Department of Molecular Epidemiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

Reprinted and adapted with permission from Molecular & Cellular Protemics, December 8, 2016, DOI 10.1074/mcp.M116.065250

Copyright © 2017 The American Society for Biochemistry and Molecular Biology, Inc.

6.1 Abstract

Glycosylation is an abundant co- and post-translational protein modification of importance to protein processing and activity. Although not template-defined, glycosylation does reflect the biological state of an organism and is a high-potential biomarker for disease and patient stratification. However, to interpret a complex but informative sample like the total plasma *N*-glycome, it is important to establish its baseline association with plasma protein levels and systemic processes. Thus far, large-scale studies (n >200) of the total plasma *N*-glycome have been performed with methods of chromatographic and electrophoretic separation, which, although being informative, are limited in resolving the structural complexity of plasma *N*-glycans. MS has the opportunity to contribute additional information on, among others, antennarity, sialylation, and the identity of high-mannose type species.

Here, we have used matrix-assisted laser desorption/ionization (MALDI)-Fourier transform ion cyclotron resonance (FTICR)-MS to study the total plasma *N*-glycome of 2144 healthy middle-aged individuals from the Leiden Longevity Study, to allow association analysis with markers of metabolic health and inflammation. To achieve this, *N*-glycans were enzymatically released from their protein backbones, labeled at the reducing end with 2-aminobenzoic acid, and following purification analyzed by negative ion mode intermediate pressure MALDI-FTICR-MS. In doing so, we achieved the relative quantification of 61 glycan compositions, ranging from $Hex_4HexNAc_2$ to $Hex_7HexNAc_6dHex_1Neu5Ac_4$, as well as that of 39 glycosylation traits derived thereof. Next to confirming known associations of glycosylation with age and sex by MALDI-FTICR-MS, we report novel associations with C-reactive protein (CRP), interleukin 6 (IL-6), body mass index (BMI), leptin, adiponectin, HDL cholesterol, triglycerides (TG), insulin, gamma-glutamyl transferase (GGT), alanine aminotransferase (ALT), and smoking. Overall, the bisection, galactosylation, and sialylation of diantennary species, the sialylation of tetraantennary species, and the size of high-mannose species proved to be important plasma characteristics associated with inflammation and metabolic health.

6.2 Introduction

Glycosylation is a ubiquitous co- and post-translational protein modification of functional relevance to the processing and activity of the conjugate. Examples include quality control during protein folding, regulation of circulatory half-life, and modulation of receptor interactions by either providing the recognition motif or by affecting protein conformation[1-7]. Consequentially, glycosylation has been associated with a multitude of diseases and states thereof, among which the progression and metastasis of cancer and the remission of rheumatoid arthritis[8-11]. Because the process of glycosylation is not template-defined, glycosylation integrates a large series of cellular conditions such as glycosidase/glycosyltransferase abundance and activity, endoplasmic reticulum (ER)/Golgi localization and nucleotide sugar availability, and reflects the intricate biological state of an organism[1,2,12]. To establish glycosylation as biomarker for (early detection of) disease and patient stratification, analysis of an easily obtainable biofluid such as plasma is of great interest[13,14]. Observed effects in a total plasma *N*-glycome (TPNG), *i.e.*, the released *N*-glycans from all plasma proteins, are highly informative but difficult to comprehend because of the complex contributions from relative protein glycoforms and overall glycoprotein abundances[15,16].

To interpret the TPNG in a context of human health and disease it is of importance to establish the behavior of *N*-glycans, and groups of *N*-glycans, in relation to plasma protein levels and systemic processes such as inflammation and metabolism. Previous studies of suitable size (n > 200) have performed this to various degrees, finding plasma *N*-glycans to be highly associated with, *e.g.*, age, sex, inflammation, body mass index (BMI), cholesterol and lipid levels[17-24]. However, these studies have either been performed on single proteins, immunoglobulin G (IgG) being a particularly well-studied example, or predominantly by methods of liquid chromatographic and electrophoretic separation (*e.g.*, (ultra)-high-performance liquid chromatography (U)HPLC and capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF)). Although of high analytical value—the techniques can separate analytes that are the same in monosaccharide composition (*e.g.*, $Hex_4HexNAc_4dHex_1$) but differ in glycan structure (*e.g.*, α1,3-branch galactosylation versus α1,6-branch galactosylation)—the complexity of the TPNG means that observed signals generally comprise a variety of distinct compositions[25-28]. Mass spectrometry provides orthogonal information from these methodologies, as it does not distinguish isomers but instead unambiguously evaluates glycans on a compositional level[29-31]. To date, a large mass spectrometric TPNG study remains to be performed for revealing associations with markers of inflammation and metabolic health[32-34].

Here, we have used high-resolution intermediate-pressure matrix-assisted laser desorption/ionization (MALDI)-Fourier transform ion cyclotron resonance (FTICR)-mass spectrometry (MS) to profile the total plasma *N*-glycosylation of 2144 middle-aged

individuals of the Leiden Longevity Study (LLS)[35]. Although MALDI-MS is reported to lead to underestimation of sialylated glycan species because of in-source and metastable decay, a phenomenon particularly visible with reflectron-based MALDI-time-of-flight (TOF)-MS, the intermediate pressure of the here-presented method prevents the residue loss and allows for the repeatable analysis of species carrying up to four sialic acids[36-41]. The 61 plasma *N*-glycan compositions we detected by the method, as well as 39 glycosylation traits mathematically derived thereof, showed to highly associate with not only age and sex, but also with clinical markers of inflammation, liver function, cholesterol, insulin, and lipid metabolism.

6.3 Experimental procedures

Participants

The LLS, described in detail previously[35,42], is a family-based study comprising 1671 offspring of 421 nonagenarians sibling pairs of Dutch descent, and the 744 partners of these offspring. A total of 2144 individuals with clinical blood parameters available were included in the current analysis. The study protocol was approved by the Leiden University Medical Center ethical committee and an informed consent was signed by all participants prior to participation in the study.

All standard blood measurements were performed in nonfasting venous blood samples using fully automated equipment. Glucose, high-sensitivity C-reactive protein (hsCRP), triglyceride (TG), total cholesterol, and high-density lipoprotein cholesterol (HDL) levels were measured on the Hitachi Modular P800 (Roche Diagnostics, Mannheim, Germany). Free triiodothyronine (T3) levels were measured on the Modular E170 (Roche Diagnostics). Low-density lipoprotein cholesterol (LDL) levels were calculated using the Friedewald formula[43], and set to missing if plasma TG levels exceeded 4.52 mmol/L. Insulin levels were measured on the Immulite 2500 (DPC, Los Angeles, CA). Specific sandwich enzyme-linked immunosorbent assays (ELISA) were u0sed for the determination of adiponectin (R&D Systems Europe, Abingdon, UK), leptin (Diagnostics Biochem Canada, Dorchester, Canada) and interleukin 6 (IL-6) levels (Sanquin Reagents, Amsterdam, The Netherlands). Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels were measured using the NADH (with P-5'-P) methodology (Modular P800, Roche Diagnostics), and gamma-glutamyl transferase (GGT) levels using the L-gamma-glutamyl-3-carboxy-4-nitroanilide substrate methodology (Modular P800, Roche Diagnostics). Dehydroepiandrosterone sulfate (DHEA-S) levels were measured with an Architect delayed one-step immunoassay (Abbot, Wiesbaden, Germany). Hypertension was defined as having a systolic blood pressure > 140 and a diastolic blood pressure > 90. Antihypertensive medication included diuretics, beta-blockers, calcium channel blockers, and agents acting on the renin-

angiotensin system. Cytomegalovirus (CMV) serostatus was determined by ELISA using the CMV-IgG ELISA PKS assay (Medac, Wedel, Germany).

N-Glycan preparation

N-Glycans from total plasma proteins from participants of the LLS were released, labeled with 2-aminobenzoic acid (2-AA) (Sigma-Aldrich, Steinheim, Germany) to allow negative mode mass spectrometric detection, and purified using hydrophilic-interaction liquid chromatography (HILIC)-solid-phase extraction (SPE) as previously described[44]. Specifically, 20 µL of 2% sodium dodecyl sulfate (SDS) (US BioChem, Cleveland, OH) was added to 10 µL plasma, randomly distributed across twenty-seven 96-well plates, followed by protein denaturation for 10 min at 60 °C and subsequent neutralization of the SDS by 10 µL 4% Nonidet P-40 substitute (Nonidet P-40) (Sigma-Aldrich). Then, after addition of 0.5 mU peptide-*N*-glycosidase F (PNGase F; Roche Diagnostics) in 10 µL 5x phosphate buffered saline solution, the *N*-glycans were released overnight at 37 °C. Without intermediate purification, the *N*-glycans were labeled for 2 h at 65 °C with the addition of 50 µL 48 mg/mL 2-AA 63 mg/mL $NaCNBH_3$ (Merck, Darmstadt, Germany) in a 10:3 (*v/v*) mixture of dimethyl sulfoxide (DMSO; Sigma-Aldrich) and glacial acetic acid (Merck). HILIC-SPE was subsequently performed using 40 mg microcrystalline cellulose (Merck) in 96-well 0.45 µm GHP-filter plates (Pall, Ann Arbor, MI). All wells of the filter plate were washed using water and subsequently equilibrated using 80:20 (*v/v*) acetonitrile (ACN; Biosolve, Valkenswaard, The Netherlands)/water. The labeled *N*-glycan samples were then applied to the wells in 80% ACN, and the wells were washed using ACN/water (80:20 *v/v*). Purified 2-AA labeled *N*-glycans were eluted in 0.8-mL-deep well collection plates (ABgene via Westburg, Leusden, The Netherlands) using 400 µL water.

Carbon-SPE

Prior to analysis by MALDI-FTICR-MS, samples were additionally desalted using carbon SPE. To achieve this, 100 µL of graphitic porous carbon (Grace, Deerfield, IL) was applied to each well of an OF1100 96-well polypropylene filter plate with a 10 µm polyethylene frit (Orochem Technologies, Lombard, IL) using a 96-well column loader (Millipore, Billerica, MA). The stationary phase was activated and conditioned with 2 × 200 µL ACN/water (80:20 *v/v*) and 3 × 100 µL 0.1% trifluoroacetic acid (TFA; Sigma-Aldrich) in water, respectively. Of the 2-AA labeled *N*-glycans, 100 µL were loaded into the wells and washed using 3 × 100 µL 0.1% TFA in water. Slight vacuum was applied to facilitate the procedure. The 2-AA labeled *N*-glycans were eluted into a V-bottom microtiter plate (Nunc, Roskilde, Denmark) using 3 × 30 µL of freshly prepared ACN/water (80:20 *v/v*) containing 0.1% TFA by centrifugation at 500 rpm (154 mm rotational diameter).

MALDI-FTICR-MS analysis

One µL of 2-AA labeled *N*-glycans was spotted in quadruplicate on a 384-AnchorChip target plate (Bruker Daltonics, Bremen, Germany) and air-dried. Subsequently, 1 µL of 2,5-

dihydroxybenzoic acid (2,5-DHB; Bruker Daltonics) matrix (20 mg/mL in ACN/water; 50:50 (*v/v*)) was applied to the spots and left to dry. To generate microcrystals, 2 × 1 µL of ethanol was applied to the spots for recrystallization prior to mass spectrometric analysis.

The 9.4 T FTICR APEX-ultra mass spectrometer was equipped with a dual electrospray ionization (ESI)/MALDI ion source (Apollo II) incorporating a quadrupole mass filter and a smartbeam laser system. Before analysis, the instrument was externally calibrated by peptide calibration standard (Bruker Daltonics). All experiments used a laser spot size of ~150 µm, laser fluence slightly above threshold, and a laser repetition rate of 200 Hz. To allow the semiquantitative analysis for the range of expected *N*-glycans, all samples were analyzed using two methods: one optimized for lower *m/z* ions (approximately *m/z* 1000 to 2500) and another optimized for higher *m/z* ions (approximately *m/z* 2200 to 4000). The quadrupole was operated in rf-only mode with the selection masses set to *m/z* 1650 and 2500 for low-mass and high-mass measurements respectively. A customized experiment sequence (pulse program) was used, in which the multiple ICR-fill parameter was reconfigured to approximate a "random-walk" functionality[38]. Briefly, the ions produced from 50 laser shots were accumulated in a hexapole and then transferred through the rf-only quadrupole to the collision cell. The sample stage was then moved 200 µm, and fresh sample interrogated with the next 50 laser shots. This cycle was performed nine times, effectively accumulating ions from 450 laser shots in the collision cell. The accumulated ions were then transferred to the ICR cell for a mass analysis scan. Each spectrum is the sum of eight such scans. All data were acquired using ApexControl 3.0.0 software (Bruker Daltonics) in expert mode, controlled by Hystar 3.8 software (Bruker Daltonics) for automatic measurement. In total, 20,736 spectra were recorded, these being for each biological sample a quadruplicate of low- as well as high-mass measurements.

Data processing

Following acquisition, representative low- and high-mass spectra were internally calibrated in DataAnalysis 4.2 (Bruker Daltonics) using a set of expected glycan masses (**Supplemental Table S1**). Using the calibrated spectra, 37 glycan compositions were manually assigned within the low-mass measurements (H4N2 to H5N4S1), and 25 within the high-mass measurements (H5N4S1 to H7N6F1S4), using mass and parts-per-million (ppm) errors to validate the assignments (**Supplemental Table S1**) (H = hexose; N = *N*-acetylhexosamine; F = deoxyhexose (fucose); S = *N*-acetylneuraminic acid). In addition, peak widths were assessed per composition to allow precise area integration (**Supplemental Table S1**). The resulting 61 compositions (H5N4S1 being present in both the low- and high-mass spectra) were in agreement with previously reported observations, as well as knowledge of the biological synthesis of *N*-glycans[2,27,45,46].

To achieve repeated extraction of the list of glycan compositions from the 20,736 MALDI-FTICR-MS measurements (10,368 low-mass and 10,368 high-mass), the spectra were

converted to simple text based format (x,y) using msconvert from ProteoWizard 3.0.5622[47]. The raw mass spectrometric data has been made publicly available in the MassIVE repository (massive.ucsd.edu) titled "MALDI-FTICR-MS total plasma *N*-glycomics" with ID: MSV000080307. Spectrum calibration, analyte integration and spectrum curation was performed using MassyTools 0.1.5.0[48]. In short, spectra were calibrated by applying the least variance second degree polynomial fit through a set of calibration masses (**Supplemental Table S1**). During this step, spectra were excluded from further analysis when not all calibrants were detected with intensities at least 3-fold higher than the maximum signal deviation within the local noise (MinMax; roughly corresponding to a root-mean-square (RMS) signal-to-noise ratio (S/N) of 9). This led to the exclusion of 470 low-mass and 372 high-mass spectra. Glycan compositions from the analyte list were subsequently integrated by summing and grouping the areas of 95% of the theoretical isotopic envelope belonging to that composition. Prior to summation, each isotope was integrated using the peak widths previously established, and an equal width local background (within a window of 50 Thomson) was subtracted from these. To further ensure data quality, spectra were removed if more than 5% of the total analyte area was below S/N 3 (MinMax), which led to the additional exclusion of 305 low-mass and 424 high-mass spectra. After additional curation of clinical samples with no available information on age and sex, we retained a total of 16,346 spectra (8194 low-mass and 8152 high-mass), yielding glycosylation information on 2144 individuals.

To arrive at one set of glycan values per individual, the replicate spectra for the low- and high-mass spectra were averaged for that individual. In case of low-mass spectra, 1878 averages were constructed from 4 spectra, 170 from 3, 76 from 2 and 20 from 1 (**Supplemental Figure S1**). For the high-mass spectra, 1857 averages were constructed from 4 spectra, 177 from 3, 83 from 2 and 27 from 1. To reconstruct the overview of the total plasma *N*-glycome, the low- and high-mass averages were normalized on the value of their overlapping composition H5N4S1, and subsequently combined. The resulting combined pattern was normalized by dividing each glycan value by the sum of all glycan values (total area normalization). Hereof, we calculated derived glycosylation traits on basis of enzymatic steps and protein groupings (**Supplemental Table S2**). Glycan and derived trait variation within the replicate low-mass and high-mass measurements was assessed on the basis of mean, standard deviation (S.D.) and coefficient of variation (CV) (**Supplemental Figure S2**; **Supplemental Figure S3**). Mass spectrometric figures were exported from DataAnalysis 4.2 (Bruker Daltonics) and annotated with glycan depictions following the symbol nomenclature proposed by the Consortium for Functional Glycomics (CFG), created in GlycoWorkbench 2.1 build 146[49,50].

Data analysis

Throughout data analysis we employed R 3.1.2 in an environment of RStudio 0.98.1091 (RStudio Team, Boston, MA)[51]. In case of significance testing, a study-wide significance threshold was maintained of $\alpha = 1.00 \cdot 10^{-5}$, values below or equal being considered statistically significant. The value arises from being the lower bound of the order of magnitude of an $\alpha = 0.05$ significance threshold Bonferroni corrected for the total number of regression tests throughout the study (number of tests = 27 phenotypes · 100 glycan features + 26 sex comparisons + 25 age comparisons = 2751; $\alpha = 0.05/2751 = 1.82 \times 10^{-5} \approx 1.00 \times 10^{-5}$).

To limit the experimental component within the sample variability, batch correction was performed on the glycan and derived trait variables using the R package ComBat, using sample plate as batch[52]. To limit outlier influence, individual glycosylation values exceeding a 5 times S.D. value from the mean of that variable were excluded from statistical analysis. Insulin, hsCRP, IL-6, TG, adiponectin, leptin, ALT, AST, GGT, and DHEA-S levels were transformed to the natural logarithm because of nonnormal distribution of the data. In addition, to obtain interpretable estimates, the glycosylation variables were centered and scaled before analysis (*i.e.*, mean subtraction and division by S.D.).

Association of variables with age and sex

Linear and logistic regression analyses were performed to establish the association between age and sex (female = 0; male = 1) as outcome variables, and nonglycan clinical variables (**Table I**), glycans (**Table II**) and derived glycosylation traits (**Table III**) as predictor variables. As no glycosylation differences were found between LLS offspring and partners (a grouping to test predisposition for longevity), these individuals were grouped for all analyses. Furthermore, because the LLS contains multiple offspring from the same family, within-family (between-siblings) dependence was taken into account by using a sandwich estimator for the standard errors[53].

Association of glycosylation with clinical variables

To eliminate possible confounding effects, age, sex and the interaction thereof were included as covariates in further models. For these analyses, the remaining nonglycan variables were used as outcome (using linear and logistic regression for respectively continuous and dichotomous variables), whereas glycans and derived traits were used as predictor (model: nonglycan $\sim \beta_1 \cdot \text{age} + \beta_2 \cdot \text{sex} + \beta_3 \cdot \text{age*sex} + \beta_4 \cdot \text{glycan}$).

To visualize the association between nonglycans and (derived) glycan traits, the t-statistics (or Wald statistics in case of logistic regression, both $\beta_4/\text{S.E.}_4$) from the models were expressed in heatmap format. Sorting of the heatmap variables was performed using hierarchical clustering (Euclidean distance, complete linkage).

6.4 Results

To investigate the association of plasma protein *N*-glycosylation with clinical markers of metabolic health and inflammation, we analyzed the total plasma *N*-glycomes of 2144 middle-aged individuals of the LLS. *N*-Glycans were enzymatically released from their protein backbones, labeled at the reducing end with 2-AA, purified by HILIC- and carbon-SPE, and analyzed by intermediate pressure MALDI-FTICR-MS. The acidic tag 2-AA allowed the joint negative mode mass spectrometric detection and relative quantification of neutral and sialylated glycan species, whereas the intermediate pressure of the measurement limited the decay commonly observed for sialylated glycans with MALDI (**Figure 1**)[36-41].

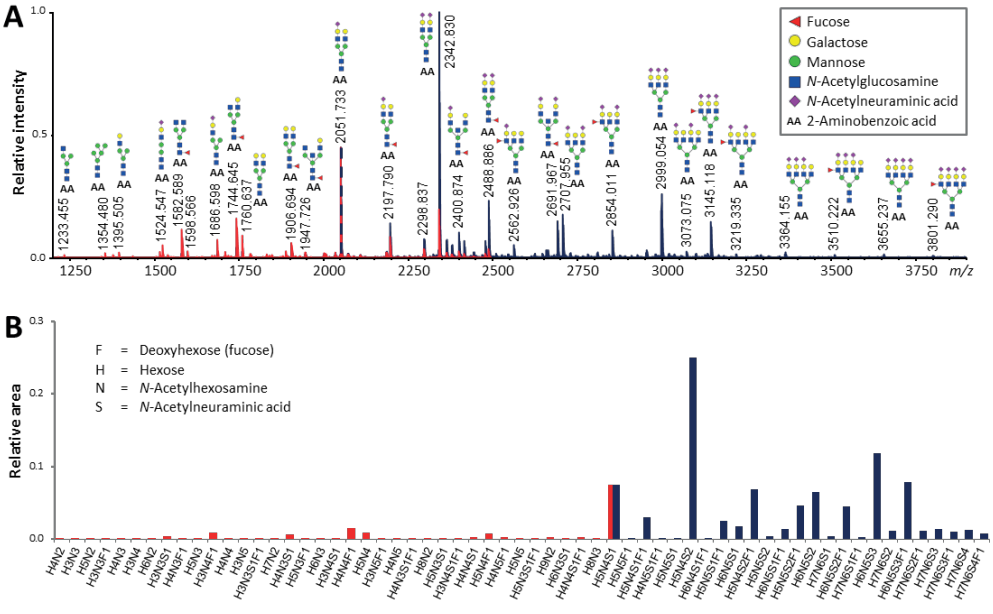


***Figure 1**. A typical total plasma N-glycome as analyzed by negative mode MALDI-FTICR-MS after enzymatic N-glycan release, 2-AA labeling, and purification. **A)** Combination of the low mass (red) and high mass (blue) mass spectra originating from a single case measurement. The relative abundances were normalized to the signal at m/z 2051.733, reflecting the N-glycan composition H5N4S1 [M-H]⁻. Whereas the glycan compositions could be established with high confidence, the displayed linkages are presumed based on literature knowledge[2,27,45,46]. **B)** Area integration of the detectable N-glycan compositions, summing the isotopes within 95% of the isotopic envelope for each species. Each composition is represented as a fraction of the total spectrum area.*

Measurement variability

Following spectrum curation, we retained a total of 16,346 mass spectra originating from low- and high-mass measurements of MALDI spotting quadruplicates for each individual. In the measurement optimized for lower masses (*m/z* 1000 to 2500) 37 *N*-glycans could be detected, ranging from H4N2 to H5N4S1, with an average absolute ppm error of 2.24 (S.D.

± 3.18). In the measurement optimized for higher masses (*m/z* 2200 to 4000) 25 additional *N*-glycans were detected, from the overlapping composition H5N4S1 to H7N6F1S4, the average absolute ppm error being 3.65 (S.D. ± 3.98) (**Supplemental Table S1**). Based on literature, the glycan compositions within the TPNG were presumed to have certain structural features[16,27,45,46]. Examples of this are the antennarity, judged as the number of *N*-acetylhexosamines minus two unless bisected, and bisection, judged to be the case if the number of *N*-acetylhexosamines equaled five and the number of hexoses five or less. Although these structural assignments are expected to represent the majority of structures contributing to an MS signal, additional structural isomers are likely to be present in the signals. For example, a composition assigned as tetraantennary may instead contain diantennary structures with two *N*-acetyllactosamine repeats, and the bisected species could be triantennary with incomplete galactosylation.

Assessing repeatability, an example quadruplicate measurement from a single individual yielded an average CV of 6.52% (S.D. ± 3.42%) for the 10 most abundant signals in the low-mass measurement (total area normalized for the mass range), and an average CV of 5.97% (S.D. ± 2.53%) for the 10 most abundant signals in the high-mass measurement (**Supplemental Figure S2A**; **Supplemental Figure S2B**). Combining the measurements by the overlapping composition H5N4S1 yielded for the 20 must abundant species an average CV of 9.29% (S.D. ± 5.88%) (**Supplemental Figure S2C**). Derived glycosylation traits, constructed to provide mathematical expressions of monosaccharide differences and groupings with structural similarity, showed a lower CV, a phenomenon previously observed for mass spectrometric plasma glycomics[54], *i.e.*, on average 1.20% (S.D. ± 0.84%) for the 20 most abundant members (**Supplemental Figure S2D**). For a listing of derived traits and their calculations see **Supplemental Table S2**.

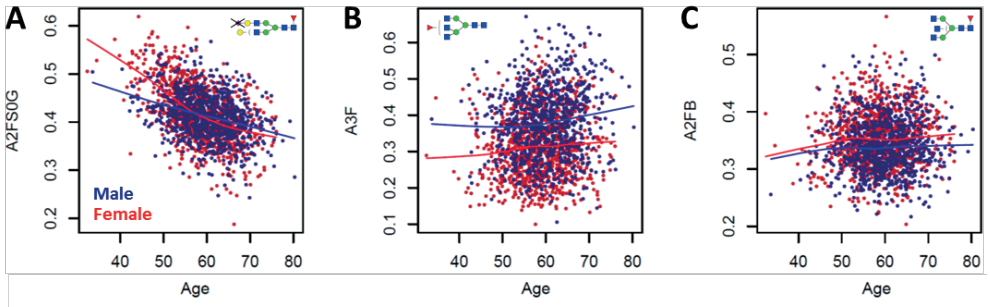


***Figure 2**. Scatterplots with local regression providing an overview of the three main derived glycosylation traits changing with age and differing by sex (males in blue, females in red). **A)** Galactosylation of nonsialylated fucosylated diantennary compositions (A2FS0G), to a large degree representative of IgG-Fc galactosylation[16]. **B)** Fucosylation of triantennary compositions (A3F). **C)** Bisection of diantennary fucosylated compositions (A2FB).*

In total, the glycosylation analysis workflow allowed for 2144 individuals the relative quantification of 61 *N*-glycan compositions and 39 derived traits. The LLS provided an additional 27 clinical variables to facilitate association analysis. Next to age and sex, measures were included on liver function (GGT, ALT, AST, AST/ALT), glucose metabolism (glucose, insulin, glucose/insulin), lipid metabolism (cholesterol, LDL-C, HDL-C, cholesterol/HDL-C, TG, lipid lowering medication, leptin, adiponectin), inflammation (hsCRP, IL-6), blood pressure (hypertension, antihypertensive medication), adrenal function (DHEA-S), thyroid function (free T3), as well as information on BMI, smoking and CMV infection, and familial propensity for longevity (**Table I**).

***Table I**. Non-glycan descriptives for the study population and association thereof with age and sex. Displayed are mean values with standard deviation (SD) for continuous variables, and the percentage of positive cases for binary variables. To assess if the variables differ by sex (female = 0; male = 1) and with age, respective logistic and linear regression was performed. Within-family dependence was taken into account by using a sandwich estimator for the standard errors. Effect sizes for the traits are displayed as coefficient for the trait (β) with standard error (SE). Displayed in bold are the p-values considered significant at or below the study-wide significance threshold of $\alpha = 1.0 \cdot 10^{-5}$.*

Phenotype	Total mean (SD) or % n = 2144	Female mean (SD) or % n = 1170	Male mean (SD) or % n = 974	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
				β (SE)	*p*-value	β (SE)	*p*-value
Calendar age	59.2 (6.76)	58.6 (6.64)	59.8 (6.84)	0.19 (0.05)	6.1E-05	-	-
Alanine transamidase (IU/L)	24.2 (12.2)	21.8 (11.0)	27.1 (13.0)	**0.53 (0.05)**	**<2.2E-16**	-0.06 (0.14)	6.9E-01
Aspartate transamidase (IU/L)	27.0 (7.92)	25.9 (7.71)	28.3 (7.99)	**0.36 (0.05)**	**2.6E-13**	0.54 (0.14)	1.8E-04
AST ALT ratio	1.27 (0.53)	1.33 (0.52)	1.19 (0.54)	-0.33 (0.08)	5.7E-05	0.31 (0.15)	3.9E-02
Gamma-glutamyl transferase (IU/L)	31.4 (37.0)	25.7 (33.2)	38.3 (40.0)	**0.74 (0.06)**	**<2.2E-16**	0.47 (0.15)	2.2E-03
Non-fasted glucose (mmol/L)	5.87 (1.58)	5.75 (1.37)	6.01 (1.78)	0.17 (0.05)	2.4E-04	0.58 (0.15)	6.6E-05
Insulin (mU/L)	23.1 (22.0)	21.0 (17.8)	25.6 (25.9)	**0.21 (0.04)**	**1.3E-06**	0.54 (0.14)	1.8E-04
Glucose insulin ratio	0.46 (0.39)	0.49 (0.41)	0.43 (0.37)	-0.18 (0.04)	3.8E-05	-0.38 (0.14)	8.0E-03
Total cholesterol (mmol/L)	5.59 (1.17)	5.68 (1.21)	5.47 (1.12)	-0.19 (0.05)	3.4E-05	0.02 (0.16)	8.9E-01
Low-density lipoprotein cholesterol (mmol/L)	3.34 (0.96)	3.37 (0.98)	3.30 (0.94)	-0.07 (0.04)	1.1E-01	-0.10 (0.15)	5.0E-01
High-density lipoprotein cholesterol (mmol/L)	1.44 (0.46)	1.59 (0.47)	1.26 (0.36)	**-0.89 (0.06)**	**<2.2E-16**	-0.38 (0.15)	9.1E-03
Total cholesterol HDL ratio	4.20 (1.41)	3.82 (1.22)	4.65 (1.49)	**0.69 (0.05)**	**<2.2E-16**	0.40 (0.15)	7.0E-03
Triglycerides (mmol/L)	1.82 (1.16)	1.60 (0.93)	2.07 (1.34)	**0.48 (0.05)**	**<2.2E-16**	**0.74 (0.15)**	**5.3E-07**
Lipid lowering medication (%)	10.8%	8.87%	13.1%	0.14 (0.04)	1.6E-03	**1.13 (0.12)**	**<2.2E-16**
Leptin (ng/mL)	20.5 (21.6)	29.3 (24.7)	9.92 (9.83)	**-1.61 (0.09)**	**<2.2E-16**	0.17 (0.15)	2.6E-01
Adiponectin (mg/L)	6.27 (3.30)	7.49 (3.55)	4.81 (2.23)	**-1.04 (0.06)**	**<2.2E-16**	0.08 (0.15)	5.7E-01
Body mass index	25.4 (3.58)	25.1 (4.00)	25.7 (2.97)	0.18 (0.05)	7.1E-04	0.41 (0.15)	4.4E-03
Hypertension (%)	24.2%	24.3%	24.0%	-0.01 (0.05)	8.8E-01	**1.12 (0.15)**	**5.0E-14**
Antihypertensive medication (%)	20.5%	20.3%	20.7%	0.01 (0.05)	7.5E-01	**1.38 (0.15)**	**<2.2E-16**
Dehydroepiandrosterone sulfate (µmol/L)	4.45 (2.80)	3.59 (2.21)	5.49 (3.07)	**0.85 (0.06)**	**<2.2E-16**	**-1.87 (0.16)**	**<2.2E-16**
High sensitivity C-reactive protein (mg/L)	3.15 (11.0)	3.11 (10.4)	3.21 (11.7)	-0.05 (0.04)	2.9E-01	0.63 (0.15)	3.5E-05
Interleukin 6 (pg/mL)	0.68 (1.46)	0.62 (1.17)	0.75 (1.74)	0.11 (0.04)	9.8E-03	**0.77 (0.15)**	**5.6E-07**
Smoking (%)	13.6%	13.2%	14.1%	0.02 (0.05)	6.1E-01	**-0.90 (0.15)**	**2.4E-09**
Free triiodothyronine (pmol/L)	4.11 (0.72)	3.94 (0.71)	4.31 (0.67)	**0.62 (0.07)**	**<2.2E-16**	-0.55 (0.14)	9.4E-05
Cytomegalovirus infection (%)	46.9%	50.6%	42.3%	-0.17 (0.05)	5.3E-04	0.56 (0.16)	4.9E-04
Member of long-lived family (%)	69.2%	67.4%	71.4%	0.09 (0.04)	5.3E-02	0.33 (0.17)	5.0E-02

***Table II**. N-glycan descriptives and their association with age and sex. Displayed are mean values with SD. To assess if the variables differ by sex (female = 0; male = 1) and with age, respective logistic and linear regression was performed, adjusted for within-family dependence. Effect sizes for the traits are displayed as coefficient for the trait (β) with SE, all of which are representative of a 1 SD increase in the glycosylation value. Displayed in bold are the p-values considered significant at or below the study-wide significance threshold of $\alpha = 1.0 \cdot 10^{-5}$.*

Composition	*m/z* [M-H]-	Depiction	Total mean % (SD)	Female mean % (SD)	Male mean % (SD)	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
						β (SE)	*p*-value	β (SE)	*p*-value
H4N2	1192.43		0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	-0.04 (0.04)	2.9E-01	-0.19 (0.13)	1.6E-01
H3N3	1233.45		0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.09 (0.05)	4.1E-02	-0.45 (0.15)	3.5E-03
H5N2	1354.48		0.14 (0.04)	0.14 (0.04)	0.13 (0.04)	-0.10 (0.04)	2.4E-02	-0.22 (0.14)	1.2E-01
H3N3F1	1379.51		0.06 (0.02)	0.06 (0.02)	0.06 (0.02)	0.15 (0.04)	2.2E-04	0.20 (0.15)	1.8E-01
H4N3	1395.51		0.11 (0.03)	0.11 (0.03)	0.11 (0.03)	-0.02 (0.04)	5.5E-01	-0.13 (0.13)	3.1E-01
H3N4	1436.53		0.06 (0.03)	0.06 (0.03)	0.06 (0.03)	0.11 (0.05)	1.7E-02	**1.21 (0.14)**	**<2.2E-16**
H6N2	1516.53		0.20 (0.06)	0.20 (0.06)	0.20 (0.06)	-0.06 (0.04)	1.5E-01	-0.54 (0.14)	2.1E-04
H3N3S1	1524.55		0.34 (0.07)	0.34 (0.07)	0.34 (0.07)	0.03 (0.04)	4.6E-01	-0.23 (0.14)	9.0E-02
H4N3F1	1541.56		0.06 (0.02)	0.06 (0.02)	0.06 (0.02)	0.13 (0.04)	1.8E-03	**-0.97 (0.15)**	**1.1E-10**
H5N3	1557.56		0.04 (0.01)	0.04 (0.01)	0.03 (0.01)	-0.16 (0.05)	4.5E-04	-0.44 (0.14)	2.6E-03
H3N4F1	1582.59		1.18 (0.49)	1.15 (0.49)	1.23 (0.48)	0.18 (0.04)	3.1E-05	**1.50 (0.14)**	**<2.2E-16**
H4N4	1598.58		0.19 (0.06)	0.19 (0.06)	0.19 (0.06)	0.00 (0.04)	9.5E-01	-0.10 (0.15)	5.0E-01
H3N5	1639.61		0.06 (0.02)	0.06 (0.02)	0.06 (0.02)	0.10 (0.04)	2.9E-02	**1.68 (0.14)**	**<2.2E-16**
H3N3F1S1	1670.61		0.05 (0.01)	0.05 (0.01)	0.06 (0.01)	**0.21 (0.04)**	**8.3E-07**	-0.30 (0.15)	4.4E-02
H7N2	1678.58		0.08 (0.02)	0.08 (0.02)	0.07 (0.02)	-0.03 (0.04)	5.5E-01	-0.58 (0.15)	8.0E-05
H4N3S1	1686.60		0.61 (0.12)	0.60 (0.12)	0.61 (0.12)	0.08 (0.04)	4.8E-02	-0.12 (0.14)	4.0E-01
H5N3F1	1703.62		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.08 (0.04)	8.3E-02	-0.53 (0.14)	1.9E-04
H6N3	1719.61		0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.06 (0.04)	1.7E-01	**0.72 (0.15)**	**1.7E-06**
H3N4S1	1727.63		0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	-0.03 (0.04)	4.6E-01	-0.26 (0.14)	5.6E-02
H4N4F1	1744.64		1.40 (0.46)	1.36 (0.45)	1.44 (0.47)	**0.19 (0.04)**	**9.1E-06**	-0.33 (0.15)	3.1E-02
H5N4	1760.64		0.90 (0.18)	0.90 (0.18)	0.89 (0.18)	-0.06 (0.05)	1.6E-01	-0.32 (0.14)	2.4E-02
H3N5F1	1785.67		0.30 (0.11)	0.29 (0.11)	0.30 (0.11)	0.09 (0.04)	4.8E-02	**1.67 (0.14)**	**<2.2E-16**
H4N5	1801.66		0.07 (0.02)	0.06 (0.02)	0.07 (0.02)	0.05 (0.05)	2.7E-01	**0.83 (0.15)**	**1.6E-08**
H4N3F1S1	1832.66		0.10 (0.02)	0.10 (0.02)	0.10 (0.02)	0.15 (0.04)	5.0E-04	-0.46 (0.15)	2.3E-03
H8N2	1840.64		0.12 (0.03)	0.12 (0.03)	0.12 (0.03)	-0.12 (0.04)	4.2E-03	**-0.73 (0.14)**	**4.5E-07**
H5N3S1	1848.65		0.10 (0.02)	0.11 (0.02)	0.10 (0.02)	-0.20 (0.05)	1.6E-05	**0.73 (0.15)**	**1.1E-06**
H3N4F1S1	1873.69		0.03 (0.01)	0.03 (0.01)	0.03 (0.01)	-0.02 (0.04)	7.3E-01	0.13 (0.14)	3.7E-01
H4N4S1	1889.68		0.37 (0.08)	0.37 (0.09)	0.36 (0.08)	**-0.19 (0.04)**	**9.0E-06**	0.25 (0.14)	6.9E-02
H5N4F1	1906.70		0.62 (0.19)	0.61 (0.20)	0.63 (0.19)	0.14 (0.04)	1.4E-03	**-1.60 (0.15)**	**<2.2E-16**
H4N5F1	1947.72		0.29 (0.09)	0.28 (0.09)	0.29 (0.10)	0.06 (0.04)	1.6E-01	0.09 (0.15)	5.5E-01
H5N5	1963.72		0.03 (0.01)	0.03 (0.01)	0.03 (0.01)	-0.12 (0.04)	8.7E-03	-0.43 (0.15)	5.3E-03

Association of glycosylation with age and sex

Glycosylation was found to highly associate with age and sex by respectively linear and logistic regression analysis (**Figure 2**; **Table II**; **Table III**). A GEE approach was used for all statistical analyses to adjust the standard errors (S.E.) for between-sibling dependence, and a study-wide significance threshold was maintained of $\alpha = 1.0 \cdot 10^{-5}$. Changes with aging included a decrease of galactosylation of diantennary glycans, visible most specifically for the galactosylation of nonsialylated diantennaries with fucose ($\beta_{A2FS0G} = -2.82$ S.E. $\pm$ 0.13; $p_{A2FS0G} < 2.2 \times 10^{-16}$) and without fucose ($\beta_{A2F0S0G} = -1.34 \pm 0.15$; $p_{A2F0S0G} < 2.2 \times 10^{-16}$). These changes were mainly driven by the increases in glycan compositions H3N4 ($\beta_{H3N4} = 1.21 \pm 0.14$; $p_{H3N4} < 2.2 \times 10^{-16}$), H3N5 ($\beta_{H3N5} = 1.68 \pm 0.14$; $p_{H3N5} < 2.2 \times 10^{-16}$), H3N4F1 ($\beta_{H3N4F1} =$

1.50 ± 0.14; $p_{H3N4F1} < 2.2 \times 10^{-16}$), H3N5F1 ($\beta_{H3N5F1} = 1.67 \pm 0.14$; $p_{H3N5F1} < 2.2 \times 10^{-16}$) and the decreases in H5N4F1 ($\beta_{H5N4F1} = -1.60 \pm 0.15$; $p_{H5N4F1} < 2.2 \times 10^{-16}$) and H5N4F1S1 ($\beta_{H5N4F1S1} = -0.81 \pm 0.15$; $p_{H5N4F1S1} = 4.5 \times 10^{-8}$). Further increasing with age were the bisection of nonsialylated fucosylated diantennaries ($\beta_{A2FS0B} = 0.77 \pm 0.14$; $p_{A2FS0B} = 6.7 \times 10^{-8}$), sialylation per galactose of fucosylated diantennaries ($\beta_{A2FGS} = 1.14 \pm 0.15$; $p_{A2FGS} = 9.8 \times 10^{-15}$), and the fucosylation of both tri- and tetraantennary compositions ($\beta_{A3F} = 0.82 \pm 0.15$; $p_{A3F} = 7.4 \times 10^{-8}$ and $\beta_{A4F} = 0.77 \pm 0.15$; $p_{A4F} = 2.0 \times 10^{-7}$).

Table II *continued.*

Composition	*m/z* [M-H]-	Depiction	Total mean % (SD)	Female mean % (SD)	Male mean % (SD)	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
						β (SE)	*p*-value	β (SE)	*p*-value
H5N3F1S1	1994.71		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.04)	9.9E-01	-0.33 (0.15)	2.6E-02
H9N2	2002.69		0.20 (0.04)	0.20 (0.04)	0.20 (0.04)	0.12 (0.04)	6.2E-03	-0.21 (0.14)	1.4E-01
H6N3S1	2010.71		0.09 (0.02)	0.09 (0.02)	0.09 (0.02)	-0.16 (0.04)	4.9E-04	0.29 (0.14)	4.5E-02
H4N4F1S1	2035.74		0.23 (0.05)	0.23 (0.05)	0.24 (0.05)	**0.20 (0.04)**	**2.6E-06**	-0.31 (0.15)	3.6E-02
H8N3	2043.72		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.04)	6.4E-01	**-0.84 (0.14)**	**5.0E-09**
H5N4S1	2051.73		6.94 (0.88)	6.97 (0.89)	6.91 (0.87)	-0.06 (0.04)	1.5E-01	0.01 (0.14)	9.4E-01
H5N5F1	2109.77		0.15 (0.05)	0.15 (0.05)	0.15 (0.05)	-0.04 (0.04)	3.4E-01	-0.62 (0.15)	4.7E-05
H5N4F1S1	2197.79		2.14 (0.46)	2.08 (0.45)	2.20 (0.46)	**0.27 (0.05)**	**5.1E-09**	**-0.81 (0.15)**	**4.5E-08**
H4N5F1S1	2238.82		0.10 (0.04)	0.10 (0.04)	0.10 (0.04)	-0.01 (0.04)	7.2E-01	**0.95 (0.14)**	**3.4E-12**
H5N5S1	2254.81		0.16 (0.05)	0.16 (0.05)	0.16 (0.05)	-0.03 (0.04)	5.1E-01	-0.20 (0.15)	1.7E-01
H5N4S2	2342.83		24.03 (2.94)	23.9 (2.93)	24.2 (2.94)	0.09 (0.04)	2.5E-02	-0.33 (0.14)	1.7E-02
H6N4F1S1	2359.84		0.13 (0.04)	0.13 (0.04)	0.13 (0.05)	0.04 (0.04)	3.2E-01	-0.29 (0.15)	5.3E-02
H5N5F1S1	2400.87		1.74 (0.55)	1.77 (0.57)	1.72 (0.52)	-0.10 (0.04)	1.6E-02	-0.26 (0.15)	9.2E-02
H6N5S1	2416.87		1.55 (0.30)	1.64 (0.30)	1.45 (0.28)	**-0.71 (0.05)**	**<2.2E-16**	-0.51 (0.14)	3.4E-04
H5N4F1S2	2488.89		4.38 (0.92)	4.20 (0.84)	4.59 (0.95)	**0.45 (0.05)**	**<2.2E-16**	0.05 (0.14)	7.2E-01
H5N5S2	2545.91		0.26 (0.06)	0.27 (0.06)	0.25 (0.06)	-0.18 (0.04)	3.2E-05	-0.54 (0.14)	1.6E-04
H6N5F1S1	2562.92		0.93 (0.23)	0.88 (0.21)	0.99 (0.23)	**0.53 (0.05)**	**<2.2E-16**	**0.71 (0.15)**	**1.1E-06**
H5N5F1S2	2691.97		2.89 (0.87)	2.84 (0.85)	2.96 (0.88)	0.14 (0.04)	1.9E-03	0.24 (0.15)	1.1E-01
H6N5S2	2707.96		7.65 (1.54)	8.08 (1.49)	7.13 (1.43)	**-0.70 (0.05)**	**<2.2E-16**	-0.59 (0.15)	6.6E-05
H7N6S1	2782.00		0.50 (0.16)	0.54 (0.16)	0.46 (0.14)	**-0.58 (0.05)**	**<2.2E-16**	0.07 (0.13)	6.1E-01
H6N5F1S2	2854.02		4.02 (1.20)	3.73 (1.09)	4.38 (1.22)	**0.58 (0.05)**	**<2.2E-16**	**0.77 (0.15)**	**2.2E-07**
H7N6F1S1	2928.06		0.28 (0.10)	0.27 (0.10)	0.30 (0.10)	**0.38 (0.05)**	**<2.2E-16**	**0.87 (0.14)**	**1.8E-09**
H6N5S3	2999.06		14.99 (3.25)	15.8 (3.17)	14.0 (3.07)	**-0.61 (0.05)**	**<2.2E-16**	-0.67 (0.16)	1.6E-05
H7N6S2	3073.09		1.74 (0.47)	1.85 (0.48)	1.62 (0.43)	**-0.53 (0.05)**	**<2.2E-16**	0.05 (0.14)	7.2E-01
H6N5F1S3	3145.11		7.81 (2.53)	7.15 (2.31)	8.59 (2.57)	**0.61 (0.05)**	**<2.2E-16**	**0.74 (0.15)**	**8.1E-07**
H7N6F1S2	3219.15		1.08 (0.35)	1.01 (0.33)	1.17 (0.35)	**0.48 (0.05)**	**<2.2E-16**	**0.90 (0.14)**	**2.4E-10**
H7N6S3	3364.19		2.35 (0.62)	2.48 (0.62)	2.20 (0.57)	**-0.50 (0.05)**	**<2.2E-16**	-0.07 (0.15)	6.3E-01
H7N6F1S3	3510.25		1.20 (0.36)	1.12 (0.34)	1.28 (0.36)	**0.46 (0.05)**	**<2.2E-16**	**0.88 (0.15)**	**1.2E-09**
H7N6S4	3655.28		1.96 (0.47)	2.01 (0.48)	1.89 (0.46)	**-0.26 (0.05)**	**5.8E-08**	0.16 (0.15)	2.8E-01
H7N6F1S4	3801.34		1.05 (0.30)	0.99 (0.28)	1.12 (0.31)	**0.45 (0.05)**	**<2.2E-16**	**0.79 (0.14)**	**1.8E-08**

***Table III**. Derived glycosylation trait descriptives and their association with age and sex. Displayed are mean values with SD. To assess if the variables differ by sex (female = 0; male = 1) and with age, respective logistic and linear regression was performed, adjusted for within-family dependence. Effect sizes for the traits are displayed as coefficient for the trait (β) with SE, and are representative of a 1 SD increase in the glycosylation value. Displayed in bold are the p-values considered significant at or below the study-wide significance threshold of $\alpha = 1.0 \cdot 10^{-5}$.*

Derived trait	Description	Total mean % (SD)	Female mean % (SD)	Male mean % (SD)	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
					β (SE)	*p*-value	β (SE)	*p*-value
Complexity								
M	Overall high mannose type	0.78 (0.17)	0.78 (0.17)	0.78 (0.17)	-0.03 (0.04)	4.5E-01	-0.51 (0.14)	3.9E-04
MM	High mannose occupancy	690 (11.0)	689 (11.0)	691 (11.0)	0.19 (0.04)	2.1E-05	-0.12 (0.14)	4.0E-01
Hy	Overall hybrid type	0.44 (0.09)	0.45 (0.09)	0.44 (0.09)	-0.07 (0.04)	1.0E-01	0.06 (0.14)	6.6E-01
C	Overall complex type	97.3 (0.46)	97.3 (0.47)	97.3 (0.45)	-0.02 (0.04)	7.0E-01	0.36 (0.14)	1.0E-02
A1	Overall monoantennary	1.16 (0.22)	1.15 (0.23)	1.17 (0.22)	0.10 (0.04)	1.6E-02	-0.21 (0.14)	1.3E-01
A2	Overall diantennary	49.2 (5.32)	48.8 (5.30)	49.7 (5.30)	0.18 (0.04)	3.5E-05	-0.09 (0.15)	5.5E-01
A3	Overall triantennary	37.1 (4.50)	37.5 (4.46)	36.7 (4.51)	-0.18 (0.04)	2.2E-05	-0.10 (0.15)	5.0E-01
A4	Overall tetraantennary	10.2 (1.95)	10.3 (2.01)	10.1 (1.86)	-0.13 (0.04)	3.4E-03	0.52 (0.14)	2.6E-04
Fucosylation								
F	Overall	33.0 (5.58)	31.4 (5.21)	35.0 (5.37)	**0.72 (0.05)**	**<2.2E-16**	**0.78 (0.15)**	**3.9E-07**
A2F	Within A2	32.4 (4.47)	31.9 (4.48)	32.9 (4.41)	**0.22 (0.04)**	**3.3E-07**	0.19 (0.15)	1.9E-01
A3F	Within A3	34.7 (9.87)	31.7 (9.04)	38.2 (9.67)	**0.73 (0.05)**	**<2.2E-16**	**0.82 (0.15)**	**7.4E-08**
A4F	Within A4	35.6 (8.41)	33.1 (7.85)	38.7 (8.05)	**0.74 (0.05)**	**<2.2E-16**	**0.77 (0.15)**	**2.0E-07**
Bisection								
B	Overall	6.12 (1.52)	6.09 (1.54)	6.17 (1.51)	0.05 (0.04)	2.2E-01	0.18 (0.15)	2.2E-01
A2B	Within A2	12.5 (2.79)	12.5 (2.86)	12.4 (2.71)	-0.05 (0.04)	2.1E-01	0.28 (0.15)	5.8E-02
A2F0B	Within nonfucosylated A2	1.77 (0.41)	1.78 (0.41)	1.75 (0.41)	-0.09 (0.04)	3.7E-02	0.16 (0.15)	3.1E-01
A2FB	Within fucosylated A2	34.7 (4.87)	35.3 (4.84)	34.0 (4.81)	**-0.27 (0.04)**	**3.8E-10**	0.24 (0.14)	8.7E-02
A2FS0B	Within nonsialylated fucosylated A2	18.9 (2.99)	19.2 (2.94)	18.5 (3.01)	**-0.25 (0.04)**	**1.1E-08**	**0.77 (0.14)**	**6.7E-08**
A2FSB	Within sialylated fucosylated A2	40.9 (5.91)	41.6 (5.88)	40.1 (5.85)	**-0.25 (0.04)**	**4.5E-09**	0.24 (0.15)	9.4E-02
Galactosylation per antenna								
AG	Overall	96.5 (1.11)	96.6 (1.10)	96.4 (1.11)	-0.15 (0.04)	3.9E-04	**-0.94 (0.15)**	**1.3E-10**
A2G	Within A2	93.3 (1.68)	93.4 (1.71)	93.2 (1.65)	-0.10 (0.04)	1.5E-02	**-1.25 (0.14)**	**<2.2E-16**
A2F0G	Within nonfucosylated A2	98.5 (0.30)	98.5 (0.30)	98.5 (0.29)	0.05 (0.04)	2.6E-01	**-0.99 (0.14)**	**5.4E-13**
A2F0S0G	Within nonsialylated nonfucosylated A2	81.1 (3.84)	81.3 (3.94)	80.8 (3.69)	-0.14 (0.05)	2.4E-03	**-1.32 (0.15)**	**<2.2E-16**
A2FG	Within fucosylated A2	82.6 (3.97)	82.6 (4.04)	82.6 (3.89)	-0.03 (0.04)	4.8E-01	**-1.36 (0.14)**	**<2.2E-16**
A2FS0G	Within nonsialylated fucosylated A2	41.4 (5.40)	41.7 (5.80)	41.2 (4.86)	-0.10 (0.04)	2.1E-02	**-2.82 (0.13)**	**<2.2E-16**
A2FSG	Within sialylated fucosylated A2	98.5 (0.26)	98.5 (0.26)	98.5 (0.26)	0.10 (0.04)	1.7E-02	-0.50 (0.14)	3.5E-04
Sialylation per antenna								
AS	Overall	79.1 (1.71)	79.1 (1.69)	79.1 (1.72)	0.02 (0.04)	5.5E-01	-0.41 (0.15)	5.2E-03
A2S	Within A2	76.6 (2.47)	76.6 (2.45)	76.6 (2.50)	0.00 (0.04)	9.8E-01	-0.35 (0.15)	1.6E-02
A2F0S	Within nonfucosylated A2	84.5 (1.31)	84.4 (1.30)	84.6 (1.32)	0.18 (0.04)	1.0E-05	-0.56 (0.13)	3.4E-05
A2FS	Within fucosylated A2	60.1 (4.79)	60.0 (4.71)	60.3 (4.89)	0.05 (0.04)	2.3E-01	-0.20 (0.15)	1.7E-01
Sialylation per galactose								
GS	Overall	81.9 (1.18)	81.8 (1.18)	82.0 (1.17)	0.17 (0.04)	6.0E-05	0.23 (0.15)	1.2E-01
A2GS	Within A2	82.1 (1.53)	82.0 (1.53)	82.2 (1.52)	0.11 (0.04)	7.4E-03	0.63 (0.15)	2.4E-05
A2F0GS	Within nonfucosylated A2	85.8 (1.17)	85.7 (1.16)	85.8 (1.18)	0.19 (0.04)	1.2E-05	-0.34 (0.14)	1.4E-02
A2FGS	Within fucosylated A2	72.7 (3.37)	72.5 (3.38)	73.0 (3.33)	0.14 (0.04)	1.2E-03	**1.14 (0.15)**	**9.8E-15**
A3GS	Within A3	84.8 (1.01)	84.7 (0.98)	84.8 (1.05)	0.12 (0.04)	3.5E-03	-0.15 (0.14)	2.8E-01
A3F0GS	Within nonfucosylated A3	84.9 (1.09)	84.9 (1.05)	84.9 (1.14)	0.01 (0.04)	7.8E-01	-0.30 (0.14)	3.7E-02
A3FGS	Within fucosylated A3	84.2 (1.32)	84.0 (1.32)	84.5 (1.28)	**0.34 (0.04)**	**2.7E-14**	0.22 (0.14)	1.2E-01
A4GS	Within A4	71.4 (1.93)	71.3 (1.88)	71.6 (1.96)	0.16 (0.04)	9.0E-05	-0.11 (0.14)	4.3E-01
A4F0GS	Within nonfucosylated A4	71.9 (2.16)	71.6 (2.05)	72.3 (2.22)	**0.34 (0.04)**	**2.5E-14**	0.09 (0.14)	5.1E-01
A4FGS	Within fucosylated A4	70.8 (1.96)	70.8 (1.95)	70.7 (1.97)	-0.08 (0.04)	5.2E-02	-0.24 (0.14)	8.7E-02

Fucosylation of triantennary and tetraantennary structures (A3F and A4F) proved also to be a major glycosylation difference between females and males (female = 0; male = 1) (β_{A3F} =0.73 ± 0.05; $p_{A3F} < 2.2 \times 10^{-16}$ and $\beta_{A4F} = 0.74 \pm 0.05$; $p_{A4F} < 2.2 \times 10^{-16}$), driven by higher male values in all fucosylated tri- and tetraantennary compositions, such as H6N5F1S3 ($\beta_{H6N5F1S3} = 0.61 \pm 0.05$; $p_{H6N5F1S3} < 2.2 \times 10^{-16}$) and H7N6F1S4 ($\beta_{H7N6F1S4} = 0.45 \pm 0.05$; $p_{H7N6F1S4} < 2.2 \times 10^{-16}$), and significantly lower levels of all nonfucosylated tri- and tetraantennary compositions, such as H6N5S3 (β_{H6N5S3} =−0.61 ± 0.05; $p_{H6N5S3} < 2.2 \times 10^{-16}$) and H7N6S4 ($\beta_{H7N6S4}$ = −0.26 ± 0.05; $p_{H7N6S4} = 5.8 \times 10^{-8}$). Furthermore, males proved to have lower bisection of diantennary fucosylated species (β_{A2FB} = −0.27 ± 0.04; $p = 3.8 \times 10^{-10}$) when compared with females, but did have a higher sialylation per galactose of tetraantennary nonfucosylated compositions ($\beta_{A4F0GS} = 0.34 \pm 0.04$; $p = 2.5 \times 10^{-14}$).

Association glycosylation with inflammation and metabolic health

Regression analysis was used to establish the relationship between the glycosylation traits and clinical markers, adding age, sex and the interaction thereof as covariates to limit their confounding influence (**Supplemental Table S3**). The t-statistics (or Wald-statistics; both β_4/S.E.$_4$) arising from the models were expressed in clustered heatmap format (**Figure 3**; for a heatmap visualization of the results without age and sex adjustment see **Supplemental Figure S4**). Significantly associating with a selection of derived glycosylation traits were hsCRP (23 statistically significant associations out of a possible 39), GGT (14), BMI (13), leptin (13), smoking (10), TG (9), insulin (7), the ratio of total cholesterol and HDL (7), the ratio of glucose and insulin (5), adiponectin (5), HDL (4), IL-6 (3), ALT (2), the ratio of ASL and ALT (1) and lipid medication (1) (**Supplemental Table S4**; **Supplemental Table S5**). Only individual *N*-glycan associations could be proven for AST, glucose, cholesterol, LDL and free T3, whereas no associations were found for hypertension, the usage of antihypertensive medication, DHEA-S and CMV infection.

Inflammatory marker hsCRP showed the most associations with the total plasma *N*-glycome, including a positive association with tri- and tetraantennary glycans ($\beta_{A3} = 0.24 \pm 0.03$; $p_{A3} < 2.2 \times 10^{-16}$ and $\beta_{A4} = 0.20 \pm 0.03$; $p_{A4} = 2.9 \times 10^{-13}$) at the expense of high-mannose (β_M = −0.18 ± 0.02; $p_M = 1.2 \times 10^{-13}$), hybrid ($\beta_{Hy}$ = −0.24 ± 0.03; $p_{Hy} < 2.2 \times 10^{-16}$), monoantennary ($\beta_{A1}$ = −0.17 ± 0.02; $p_{A1} = 1.8 \times 10^{-12}$) and diantennary species (β_{A2} = −0.25 ± 0.03; $p_{A2} < 2.2 \times 10^{-16}$). Although fucosylation of diantennary species proved to decrease with higher hsCRP levels (β_{A2F} = −0.13 ± 0.02; $p_{A2F} = 1.5 \times 10^{-7}$), an increase was seen in the fucosylation of triantennary species (β_{A3F} = 0.13 ± 0.03; $p_{A3F} = 1.0 \times 10^{-6}$). Additional increases were found for sialylation of (most specifically) fucosylated diantennary (β_{A2FGS} = 0.25 ± 0.02; $p_{A2FGS} < 2.2 \times 10^{-16}$) and triantennary ($\beta_{A3FGS} = 0.15 \pm 0.02$; $p_{A3FGS} = 1.0 \times 10^{-9}$) species, as well as an increase in average high-mannose size ($\beta_{MM} = 0.13 \pm 0.02$; $p_{MM} = 7.9 \times 10^{-8}$) and a decrease in bisection of the nonfucosylated diantennaries in particular (β_{A2F0B} = −0.14 ± 0.02; $p_{A2F0B} = 1.4 \times 10^{-8}$). Notably, although galactosylation of nonsialylated

diantennaries without fucose increased ($\beta_{A2F0S0G}$ = 0.12 ± 0.02; $p_{A2F0S0G}$ = 2.6 × 10^{-7}), galactosylation of the same species, but with fucose, decreased instead (β_{A2FS0G} = −0.20 ± 0.03; p_{A2FS0G} = 1.7 × 10^{-12}). Interestingly, the associations observed with hsCRP could only in part be translated to the upstream cytokine IL-6[55,56], which only showed reproduction of the decreased A2F galactosylation (β_{A2FS0G} = −0.19 ± 0.03; p_{A2FS0G} = 4.6 × 10^{-12}) and increased sialylation per galactose thereof (β_{A2FGS} = 0.25 ± 0.02; p_{A2FGS} < 2.2 × 10^{-16}).

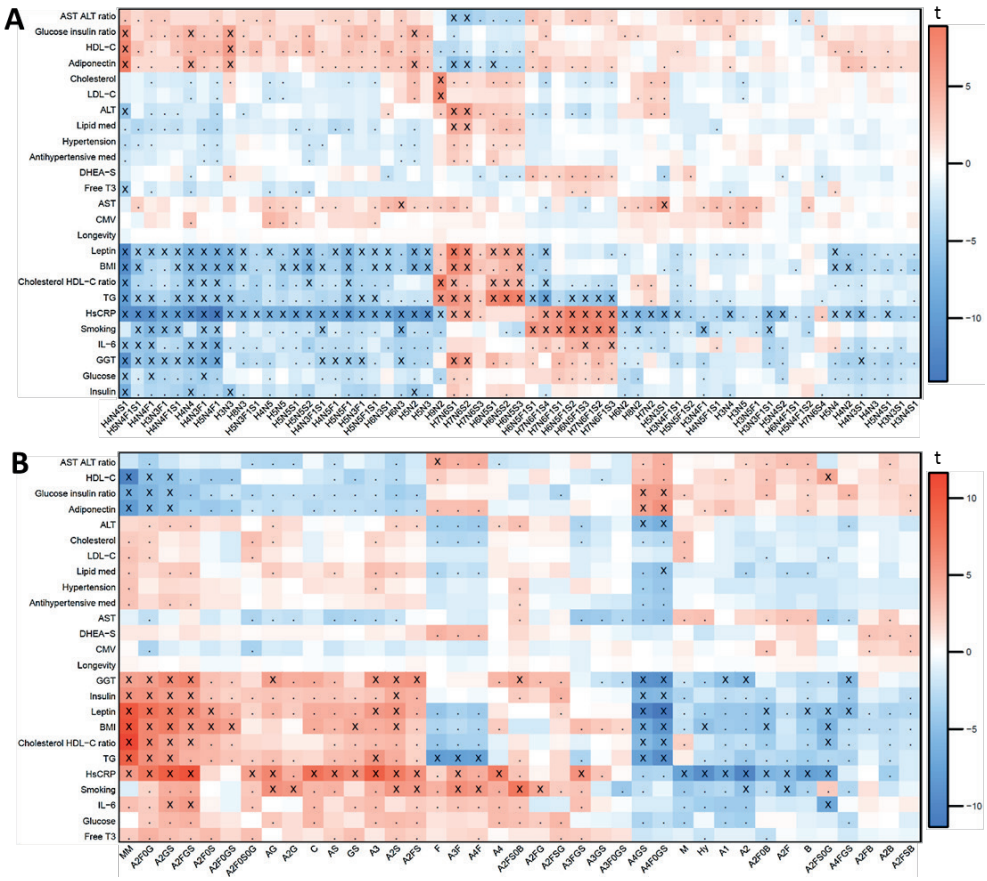


***Figure 3**. Heatmaps showing the t- (or Wald) statistics (ß/S.E.) of the associations between clinical markers of metabolic health and inflammation, and glycosylation. **A)** Single total plasma N-glycans after total area normalization. **B)** Derived glycosylation traits. All models were adjusted for age, sex, the interaction thereof, and within-family dependence. Crosses (x) indicate a statistical significance of $p <= 1.0 \times 10^{-5}$ thereby surpassing the study-wide significance threshold, whereas periods (.) indicate associations with a significance of $p <= 0.05$.*

BMI and leptin proved highly similar with respect to total plasma *N*-glycosylation associations, and showed considerable overlap with the aforementioned hsCRP as well. Taking an increase in BMI as example, galactosylation was decreased for fucosylated

diantennary species (β_{A2FS0G} = −0.55 ± 0.10; p_{A2FS0G} = 9.6 × 10^{-9}) and increased for nonfucosylated variants (β_{A2F0G} = 0.58 ± 0.09; p_{A2F0G} = 9.9 × 10^{-12}). Changes were also seen with average high-mannose size (β_{MM} = 0.89 ± 0.08; p_{MM} < 2.2 × 10^{-16}), bisection of nonfucosylated diantennary species (β_{A2F0B} = −0.39 ± 0.08; p_{A2F0B} = 2.0 × 10^{-6}), and sialylation of diantennary glycans (β_{A2GS} = 0.60 ± 0.08; p_{A2GS} = 3.1 × 10^{-13}). All of these effects were replicable within both leptin and hsCRP. However, shared with leptin but not seen with hsCRP, was the negative association of BMI with the sialylation of tetraantennary species, and specifically the nonfucosylated variants thereof (β_{A4F0GS} = −0.74 ± 0.08; p_{A4F0GS} < 2.2 × 10^{-16}). When including hsCRP and leptin as variables in the model between BMI and glycosylation, most associations between the latter two are lost with the exception of high-mannose size (β_{MM} = 0.32 ± 0.07; p_{MM} = 4.0 × 10^{-7}) and a strong remaining trend with glycan composition H4N4S1 (β_{H4N4S1} = −0.24 ± 0.07; p_{H4N4S1} = 1.8 × 10^{-4}) (**Supplemental Table S6**).

The liver marker GGT showed major associations with the total plasma *N*-glycome, whereas AST, ALT, and the ratio thereof were of only minor influence. GGT appeared similar to hsCRP in changes of increased galactosylation of nonfucosylated diantennary glycans (*e.g.*, β_{A2F0G} = 0.08 ± 0.02; p_{A2F0G} = 2.5 × 10^{-7}), sialylation of fucosylated diantennaries (*e.g.*, β_{A2FGS} = 0.09 ± 0.01; p_{A2FGS} = 9.9 × 10^{-12}), as well as an overall increased in antennarity (*e.g.*, β_{A3} = 0.07 ± 0.01; p_{A3} = 3.7 × 10^{-8}). However, the marker showed a decrease in tetraantennary sialylation similar to BMI and not hsCRP (β_{A4F0GS} = −0.11 ± 0.01; p_{A4F0GS} < 2.2 × 10^{-16}), while lacking the decreasing galactosylation of nonsialylated fucosylated diantennaries (A2FS0G) prominently seen in both BMI and hsCRP.

Clinical markers considered of beneficial metabolic nature, in our study represented by glucose/insulin, HDL-C and adiponectin[57,58], showed associations largely opposite to those established for hsCRP, BMI and GGT. Examples of this include, in case of adiponectin, the decreased galactosylation of nonfucosylated diantennaries (β_{A2F0G} = −0.06 ± 0.01; p_{A2F0G} = 3.1 × 10^{-8}), a decreased sialylation of diantennaries (β_{A2GS} = −0.05 ± 0.01; p_{A2GS} = 2.1 × 10^{-6}), a decreased average high-mannose size (β_{MM} = −0.08 ± 0.01; p_{MM} = 1.7 × 10^{-15}), and an increased sialylation of tetraantennary nonfucosylated species (β_{A4F0GS} = 0.06 ± 0.01; p_{A4F0GS} = 3.1 × 10^{-9}).

Interestingly, the only clinical markers affecting tri- and tetraantennary fucosylation (after correction for sex) proved to be smoking with a positive association (β_{A3F} = 0.50 ± 0.08; p_{A3F} = 2.6 × 10^{-10} and β_{A4F} = 0.46 ± 0.08; p_{A4F} = 2.3 × 10^{-8}), and TG levels with a negative association (β_{A3F} = −0.09 ± 0.01; p_{A3F} = 8.4 × 10^{-14} and β_{A4F} = −0.09 ± 0.01; p_{A4F} = 1.6 × 10^{-13}). In addition, smoking was also the only clinical variable to positively associate with the bisection of fucosylated nonsialylated diantennary species (β_{A2FS0B} = 0.56 ± 0.07; p_{A2FS0B} = 4.7 × 10^{-15}). Of particular interest is the high-mannose size trait (MM), which shows several of the strongest correlations, positively associating with cholesterol/HDL-C (β_{MM} = 0.08 ± 0.01; p_{MM} < 2.2 × 10^{-16}), TG levels (β_{MM} = 0.11 ± 0.01; p_{MM} < 2.2 × 10^{-16}), leptin (β_{MM} = 0.20 ± 0.02;

$p_{MM} < 2.2 \times 10^{-16}$) and BMI ($\beta_{MM} = 0.89 \pm 0.08$; $p_{MM} < 2.2 \times 10^{-16}$), whereas negatively associating with glucose/insulin ($\beta_{MM} = -0.10 \pm 0.02$; $p_{MM} = 3.0 \times 10^{-10}$), HDL ($\beta_{MM} = -0.09 \pm 0.01$; $p_{MM} < 2.2 \times 10-16$), and adiponectin ($\beta_{MM} = -0.08 \pm 0.01$; $p_{MM} = 1.7 \times 10^{-15}$). Considering the individual glycans contributing to the MM trait, the significant compositions proved to be H5N2 and H9N2.

6.5 Discussion

Here we report the MALDI-FTICR-MS analysis of the total plasma *N*-glycomes of 2144 principally healthy individuals from the LLS, and the association thereof with clinical markers of inflammation and metabolic health. In doing so, we have confirmed expectations from literature by showing a decrease of galactosylation and increase in bisection of diantennary fucosylated glycans with increasing age[59-61], and by men having a higher tri- and tetraantennary fucosylation and lower bisection than women[62,63]. Adjusting for the age and sex effects as well as the literature-reported interaction of age and sex[18,21], we proved additional associations between clinical markers for metabolic health/inflammation and plasma *N*-glycosylation characteristics including antennarity, sialylation, bisection, and galactosylation of various diantennary subgroups, and the size of high-mannose species.

Methodology

MS analysis of plasma *N*-glycosylation is not without challenges, particularly when employing MALDI ionization. A downside of this technique is the in-source and metastable loss of sialic acid residues, which is problematic given that plasma *N*-glycans are often highly sialylated (a notable exception being those from the fragment-crystallizable (Fc) region of IgG)[16,37]. Although chemical derivatization prior to MALDI-MS has been found to solve these stabilization issues (*e.g.*, by permethylation, methyl/ethyl esterification or amidation), this adds extra steps to the sample preparation workflow and often leads to byproducts[31,64-69]. Instead, here we have used MALDI-FTICR-MS equipped with an intermediate pressure source to decrease sialic acid decay[36,38]. The intermediate pressure source is known to promote ion integrity by cooling of the MALDI-generated ions[36,39-41]. In our study this has facilitated the relative quantification of *N*-glycan species containing up to four *N*-acetylneuraminic acids.

Although similarly sized investigations have analyzed plasma *N*-glycosylation by separation techniques like (U) HPLC or CGE-LIF[18-20,23,70], to our knowledge this is the largest study of its kind performed by MS[32-34]. Glycan compositions as obtained by MS provide orthogonal information to chromatographic peaks. UHPLC, for example, can separate diantennary *N*-glycan isomers with α1,3- versus α1,6-arm galactosylation and can likewise distinguish an antennary from a bisecting GlcNAc, whereas MS can provide more precise groups of di-, tri-, and tetraantennary compositions, number of sialic acids, and clear separation of high-mannose type glycans[29,71]. As such, the construction of derived glycosylation traits making

use of these features, while still biased on a compositional level, is simple to perform and provides additional insight into the complexity of glycan changes.

One example of the added information of derived traits can be found in the association of glycosylation with hsCRP. On an individual glycan level, we can only observe a relative increase in tri- and tetraantennary compositions together with the inflammation marker, and a corresponding relative decrease in all other compositions (which may be because of the total area normalization). However, when we mathematically take several individual glycans with shared biological features out of the total plasma *N*-glycome and compare them relative to each other in the form of a derived trait, we now additionally reveal, for example, a decrease in galactosylation within the subset of glycan compositions predominantly occurring on IgG-Fc (nonsialylated fucosylated diantennary species; A2FS0), a finding expected from literature[16,72,73].

Additional aspects of the analytical methodology need to be considered to allow valid interpretation of the presented findings. First, *N*-glycans are released from their protein backbones, and thus information is lost whether an observed glycan change originates from protein glycosylation or from glycoprotein abundance. Second, mass spectrometry does not distinguish isomers, meaning that a given monosaccharide mass, *e.g.*, a hexose, is assigned differently based on literature knowledge of its compositional context (*e.g.*, as mannose for H5N2 and as galactose for H4N4F1)[16,27,45,46]. Similarly, literature is the main source of information on the linkages between monosaccharides, for example presuming bisection for compositions having five *N*-acetylhexosamines but less than three galactoses (*e.g.*, H5N5F1). Although this group indeed encompasses bisection, the information within will also be confounded by triantennary structures with incomplete galactosylation, even if these are not abundant in human plasma[16,27]. Third, the methodology presented here will not provide biologically true relative ratios of glycan compositions, as the profiles will, for example, be skewed by the ionization advantage of sialylated species in negative ion mode MS. Nonetheless, the relative signal differences will still be representative for the biological directions of change, as well as providing an estimate of the magnitude.

Clinical findings

With the help of literature, we can speculate on the biological background of the observed associations. Acute inflammation, represented in our study by hsCRP and IL-6, shows to confirm previous glycomics studies with regard to the increase in antennarity, fucosylation of triantennary species, and a decrease in galactosylation of IgG glycans which is particularly well-established[9,17,21,24,59,74]. The first two may be explained by an inflammation-induced increase in plasma levels and glycosylation changes of acute phase proteins such as alpha-1-antitrypsin and alpha-1-acid glycoprotein (orosomucoid-1)[9,75]. Not only are these carriers of *N*-glycans with two or more antennae at baseline conditions, they furthermore display increased antennarity and sialyl-Lewis X upon acute inflammation[76-78]. Additional

observations include the increase of galactosylation and sialylation per galactose of diantennary fucosylated (A2F) species in general, which is notably different from the behavior of IgG. A contributor to this observation could be the level of IgM, an abundant immunoglobulin shown to increase with many autoimmune and inflammatory conditions and which carries the required highly galactosylated and sialylated A2F species[79,80].

Of particular interest is the highly significantly increased size of high-mannose glycans (MM) which is not only observed with increased hsCRP, but as well with increasing BMI, non-HDL cholesterol and TG. This glycosylation change appears to largely stem from the increase in the single glycan composition H9N2, an analyte difficult to assess by commonly used liquid chromatography with fluorescence detection[23,81]. Within the total plasma *N*-glycome, this large high-mannose glycan may predominantly originate from apolipoprotein B, the main protein constituent of most non-HDL lipoproteins (*e.g.*, VLDL, LDL)[16,82,83]. The observed correlations with H9N2 and MM can represent differing apolipoprotein B levels or glycosylation thereof, and may be indicative of an unhealthy glycosylation profile with regard to lipid transport and metabolism. The association of MM with inflammatory marker hsCRP is likely a consequence of the connection between inflammation and obesity[84], as a model corrected for BMI no longer shows significant association between high mannose size and inflammation (data not shown). The effect size being less pronounced with hsCRP than with BMI is explainable by the aforementioned increase in IgM, which contains the smaller high-mannose compositions H5N2 and H6N2 at its Asn^{279} site[80]. On the other hand, the association between MM and BMI remains true in a model adjusted for CRP and leptin, indicating the glycosylation trait may have potential to discriminate healthy from unhealthy obese.

Although largely following an inflammatory glycosylation profile, BMI, leptin, non-HDL cholesterol and TG do show a unique negative association with the sialylation of nonfucosylated tetraantennary compositions (A4F0GS), a change mainly occurring because of the relative increase of the lowly sialylated H7N6S1 and H7N6S2. Many proteins may contribute to this decrease in sialylation, a notable one being alpha-1-acid glycoprotein[85-87], which is known to bind lipophilic compounds with affinity modulated by its degree of sialylation[88-90]. This effect also applies to binding of hydrophobic drugs, possibly explaining why we similarly observed trends of decreased A4F0GS with the usage of lipid- and antihypertensive medication[89,90]. Additionally, increased levels of glycoproteins with exposed galactose residues can reflect a modulation of protein turnover, e.g. recycling by the liver-based asialoglycoprotein receptors[5,91-93]. Low sialylation would induce rapid turnover of lipid scavengers like alpha-1-acid glycoprotein, a situation which is beneficial when the blood needs to be cleared of high levels of lipophilic compounds.

Smoking proved the only phenotype to positively associate with the fucosylation of tri- and tetraantennary species (A3F and A4F), as well as with the bisection of IgG-Fc type glycans

(A2FS0B)[20,94]. Likely these reflect the chronic response to vascular injury obtained from smoking-induced shear stress and oxidative damage[95]. Increased fucosylation of acute phase proteins, when antennary-linked in the form of sialyl-Lewis X or A, facilitates recruitment to sites of injury, *e.g.*, by interactions with selectins presented on inflammation-activated endothelial cells[16,76,96].

Although it is invalid to interpret absence of statistical significance as an absence of association, we have nevertheless failed to identify derived traits or single glycans to be predictive of DHEA-S, CMV infection or the propensity for longevity. The latter is a peculiar absence, as plasma *N*-glycosylation analysis of the same cohort by HPLC had revealed two chromatographic peaks to be predictors of the phenotype[18]. Reasons for this lack of biological reproduction may include the measurement error (HPLC tends to provide more robust measurements than mass spectrometry)[71], the chromatographic peaks representing a culmination of multiple mass spectrometric compositions that are not individually significant, the previous findings being incidental, or the inability of mass spectrometry to separate or detect the responsible analytes.

Glycosylation is of high interest for the development or improvement of biomarkers for disease and patient stratification[14,97]. In particular, the findings within this study may be of benefit to the detection and discrimination of metabolic syndrome or inflammatory disorders, and may bolster the predictability of existing biomarkers such as the Framingham Risk Score[98]. Important glycosylation phenotypes in this regard would then be the derived traits MM, A3F, A4F, A2F0G, A2F0B, A2FS0G, A2FGS, and A4F0GS, as well as the various individual glycans comprising these groups, *e.g.*, most tri- and tetraantennary compositions, high-mannose compositions H5N2, H6N2, and H9N2, and the truncated *N*-glycans suggested by compositions H3N3, H4N4, and H4N4S1. Interestingly, although this last example, H4N4S1, is a glycan composition difficult to characterize in a derived trait, it does show to be the major single glycan to positively associate with most beneficial markers of metabolic health (*e.g.*, glucose/insulin, HDL-C, adiponectin) and negatively with most detrimental ones (*e.g.*, hsCRP, BMI, GGT, hypertension, TG). The protein source of the *N*-glycan is as of yet unclear, but is of interest to study in more detail.

To summarize, we have reported a large number of associations between total plasma *N*-glycosylation as measured by MALDI-FTICR-MS and clinical markers of metabolic health and inflammation. By this, we have identified glycan compositions and derived traits indicative of overall metabolic health and inflammation, as well as finding glycosylation traits uniquely associating with single marker variables. With this knowledge, we hope to contribute to the interpretation of the plasma *N*-glycome as biomarker for health and disease, and to assist clinical translation of mass spectrometric glycosylation analysis.

6.6 Supporting information

Supplemental Figures (.pdf) - Supplemental figures.

Supplemental Tables (.xlsx) - Supplemental tables.

6.7 Acknowledgments

This work was supported by the European Union Seventh Framework Programme projects HighGlycan (278535), MIMOmics (305280), and IDEAL (259679). In addition, financial support was provided by the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology and the Netherlands Consortium for Healthy Aging (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, the Netherlands Organization for Scientific Research (NWO) and by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007).

6.8 References

1 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130 (1993).

2 Varki, A. *et al. Essentials of Glycobiology*. 2nd edn, (2009).

3 Xu, C. C. & Ng, D. T. W. Glycosylation-directed quality control of protein folding. *Nat Rev Mol Cell Bio* **16**, 742-752, doi:10.1038/nrm4073 (2015).

4 Kontermann, R. E. Strategies for extended serum half-life of protein therapeutics. *Curr Opin Biotech* **22**, 868-876, doi:10.1016/j.copbio.2011.06.012 (2011).

5 Yang, W. H. *et al.* An intrinsic mechanism of secreted protein aging and turnover. *Proc Natl Acad Sci U S A* **112**, 13657-13662, doi:10.1073/pnas.1515464112 (2015).

6 Ferrara, C. *et al.* Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fc gamma RIII and antibodies lacking core fucose. *Proc Natl Acad Sci U S A* **108**, 12669-12674, doi:10.1073/pnas.1108455108 (2011).

7 Varki, A. & Gagneux, P. Multifarious roles of sialic acids in immunity. *Ann Ny Acad Sci* **1253**, 16-36, doi:10.1111/j.1749-6632.2012.06517.x (2012).

8 Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540-555, doi:10.1038/nrc3982 (2015).

9 Arnold, J. N., Saldova, R., Hamid, U. M. A. & Rudd, P. M. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **8**, 3284-3293, doi:10.1002/pmic.200800163 (2008).

10 Axford, J. S. Glycosylation and rheumatic disease. *Bba-Mol Basis Dis* **1455**, 219-229, doi:Doi 10.1016/S0925-4439(99)00057-5 (1999).

11 Bondt, A. *et al.* Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res* **12**, 4522-4531, doi:10.1021/pr400589m (2013).

12 Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448-462, doi:10.1038/nrm3383 (2012).

13 Maverakis, E. *et al.* Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *J Autoimmun* **57**, 1-13, doi:10.1016/j.jaut.2014.12.002 (2015).

14 Ruhaak, L. R., Miyamoto, S. & Lebrilla, C. B. Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Molecular & Cellular Proteomics* **12**, 846-855, doi:10.1074/mcp.R112.026799 (2013).
15 Klein, A. Human total serum N-glycome. *Adv Clin Chem* **46**, 51-85 (2008).
16 Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-343, doi:10.1007/s10719-015-9626-2 (2016).
17 Ruhaak, L. R. *et al.* Decreased Levels of Bisecting GlcNAc Glycoforms of IgG Are Associated with Human Longevity. *PLoS One* **5**, doi:ARTN e1256610.1371/journal.pone.0012566 (2010).
18 Ruhaak, L. R. *et al.* Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* **10**, 1667-1674, doi:10.1021/pr1009959 (2011).
19 Lu, J. P. *et al.* Screening Novel Biomarkers for Metabolic Syndrome by Profiling Human Plasma N-Glycans in Chinese Han and Croatian Populations. *J Proteome Res* **10**, 4959-4969, doi:10.1021/pr2004067 (2011).
20 Knezevic, A. *et al.* Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology* **20**, 959-969, doi:10.1093/glycob/cwq051 (2010).
21 Kristic, J. *et al.* Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* **69**, 779-789, doi:10.1093/gerona/glt190 (2014).
22 Vanhooren, V. *et al.* N-glycomic changes in serum proteins during human aging. *Rejuv Res* **10**, 521-531, doi:10.1089/rej.2007.0556 (2007).
23 Igl, W. *et al.* Glycomics meets lipidomics-associations of N-glycans with classical lipids, glycerophospholipids, and sphingolipids in three European populations. *Mol Biosyst* **7**, 1852-1862, doi:10.1039/c0mb00095g (2011).
24 Knezevic, A. *et al.* Variability, Heritability and Environmental Determinants of Human Plasma N-Glycome. *J Proteome Res* **8**, 694-701, doi:10.1021/pr800737u (2009).
25 Trbojevic Akmacic, I. *et al.* Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm Bowel Dis* **21**, 1237-1247, doi:10.1097/MIB.0000000000000372 (2015).
26 Novokmet, M. *et al.* Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* **4**, 4347, doi:10.1038/srep04347 (2014).
27 Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **13**, 2314-2327, doi:10.1021/pr401092y (2014).
28 Ruhaak, L. R., Uh, H. W., Deelder, A. M., Dolhain, R. E. & Wuhrer, M. Total plasma N-glycome changes during pregnancy. *J Proteome Res* **13**, 1657-1668, doi:10.1021/pr401128j (2014).
29 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev* **18**, 349-450, doi:10.1002/(SICI)1098-2787(1999)18:6<349::AID-MAS1>3.0.CO;2-H (1999).
30 Canis, K. *et al.* Mapping the N-glycome of human von Willebrand factor. *Biochem J* **447**, 217-228, doi:10.1042/BJ20120810 (2012).
31 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).
32 Kang, P. *et al.* Glycomic alterations in the highly-abundant and lesser-abundant blood serum protein fractions for patients diagnosed with hepatocellular carcinoma. *Int J Mass Spectrom* **305**, 185-198, doi:10.1016/j.ijms.2010.11.007 (2011).
33 Borelli, V. *et al.* Plasma N-Glycome Signature of Down Syndrome. *J Proteome Res* **14**, 4232-4245, doi:10.1021/acs.jproteome.5b00356 (2015).
34 Jansen, B. C. *et al.* Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep* **6**, 23296, doi:10.1038/srep23296 (2016).

35 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* **14**, 79-84, doi:10.1038/sj.ejhg.5201508 (2006).
36 Asakawa, D., Calligaris, D., Zimmerman, T. A. & De Pauw, E. In-Source Decay during Matrix-Assisted Laser Desorption/Ionization Combined with the Collisional Process in an FTICR Mass Spectrometer. *Anal Chem* **85**, 7809-7817, doi:10.1021/ac401234q (2013).
37 Powell, A. K. & Harvey, D. J. Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption ionization mass spectrometry. *Rapid Commun Mass Sp* **10**, 1027-1032 (1996).
38 Selman, M. H. J. *et al.* Immunoglobulin G Glycopeptide Profiling by Matrix-Assisted Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* **82**, 1073-1081, doi:10.1021/ac9024413 (2010).
39 Lee, H., An, H. J., Lerno, L. A., German, J. B. & Lebrilla, C. B. Rapid profiling of bovine and human milk gangliosides by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry. *Int J Mass Spectrom* **305**, 138-150, doi:10.1016/j.ijms.2010.10.020 (2011).
40 Park, Y. M. & Lebrilla, C. B. Application of Fourier transform ion cyclotron resonance mass spectrometry to oligosaccharides. *Mass Spectrom Rev* **24**, 232-264, doi:10.1002/mas.20010 (2005).
41 O'Connor, P. B., Mirgorodskaya, E. & Costello, C. E. High pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry for minimization of ganglioside fragmentation. *J Am Soc Mass Spectr* **13**, 402-407, doi:Pii S1044-0305(02)00351-3Doi 10.1016/S1044-0305(02)00351-3 (2002).
42 Westendorp, R. G. J. *et al.* Nonagenarian Siblings and Their Offspring Display Lower Risk of Mortality and Morbidity than Sporadic Nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc* **57**, 1634-1637, doi:10.1111/j.1532-5415.2009.02381.x (2009).
43 Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge. *Clin Chem* **18**, 499 (1972).
44 Ruhaak, L. R. *et al.* Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem* **80**, 6119-6126, doi:10.1021/ac800630x (2008).
45 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).
46 Freeze, H. H. Genetic defects in the human glycome. *Nat Rev Genet* **7**, 537-551, doi:10.1038/nrg1894 (2006).
47 Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918-920, doi:10.1038/nbt.2377 (2012).
48 Jansen, B. C. *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* **14**, 5088-5098, doi:10.1021/acs.jproteome.5b00658 (2015).
49 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, doi:10.1093/glycob/cwv091 (2015).
50 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).
51 RCore Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. (2014).
52 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).

53 Liang, K. Y. & Zeger, S. L. Longitudinal Data-Analysis Using Generalized Linear-Models. *Biometrika* **73**, 13-22, doi:DOI 10.1093/biomet/73.1.13 (1986).
54 Bladergroen, M. R. *et al.* Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *J Proteome Res* **14**, 4080-4086, doi:10.1021/acs.jproteome.5b00538 (2015).
55 Heinrich, P. C., Castell, J. V. & Andus, T. Interleukin-6 and the Acute Phase Response. *Biochemical Journal* **265**, 621-636 (1990).
56 Vigushin, D. M., Pepys, M. B. & Hawkins, P. N. Metabolic and Scintigraphic Studies of Radioiodinated Human C-Reactive Protein in Health and Disease. *Journal of Clinical Investigation* **91**, 1351-1357, doi:Doi 10.1172/Jci116336 (1993).
57 O'Neill, S., Bohl, M., Gregersen, S., Hermansen, K. & O'Driscoll, L. Blood-Based Biomarkers for Metabolic Syndrome. *Trends Endocrin Met* **27**, 363-374, doi:10.1016/j.tem.2016.03.012 (2016).
58 Renaldi, O. *et al.* Hypoadiponectinemia: a risk factor for metabolic syndrome. *Acta Med Indones* **41**, 20-24 (2009).
59 Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res Rev* **12**, 685-698, doi:10.1016/j.arr.2012.02.002 (2013).
60 Shikata, K. *et al.* Structural changes in the oligosaccharide moiety of human IgG with aging. *Glycoconj J* **15**, 683-689, doi:Doi 10.1023/A:1006936431276 (1998).
61 Yamada, E., Tsukamoto, Y., Sasaki, R., Yagyu, K. & Takahashi, N. Structural changes of immunoglobulin G oligosaccharides with age in healthy human serum. *Glycoconj J* **14**, 401-405, doi:Doi 10.1023/A:1018582930906 (1997).
62 Ding, N. *et al.* Human serum N-glycan profiles are age and sex dependent. *Age Ageing* **40**, 568-575, doi:10.1093/ageing/afr084 (2011).
63 Ruhaak, L. R. *et al.* Targeted biomarker discovery by high throughput glycosylation profiling of human plasma alpha1-antitrypsin and immunoglobulin A. *PLoS One* **8**, e73082, doi:10.1371/journal.pone.0073082 (2013).
64 Johnson, S. B. & Brown, R. E. Simplified Derivatization for Determining Sphingolipid Fatty Acyl Composition by Gas-Chromatography Mass-Spectrometry. *J Chromatogr* **605**, 281-286, doi:Doi 10.1016/0021-9673(92)85248-R (1992).
65 Morelle, W. & Michalski, J. C. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* **2**, 1585-1602, doi:10.1038/nprot.2007.227 (2007).
66 Wheeler, S. F., Domann, P. & Harvey, D. J. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of alpha(2 --> 3)- and alpha(2 --> 6)-isomers. *Rapid Commun Mass Spectrom* **23**, 303-312, doi:10.1002/rcm.3867 (2009).
67 Alley, W. R. & Novotny, M. V. Glycomic Analysis of Sialic Acid Linkages in Glycans Derived from Blood Serum Glycoproteins. *J Proteome Res* **9**, 3062-3072, doi:10.1021/pr901210r (2010).
68 de Haan, N. *et al.* Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. *Anal Chem* **87**, 8284-8291, doi:10.1021/acs.analchem.5b02426 (2015).
69 Reiding, K. R., Lonardi, E., Hipgrave Ederveen, A. L. & Wuhrer, M. Ethyl Esterification for MALDI-MS Analysis of Protein Glycosylation. *Methods Mol Biol* **1394**, 151-162, doi:10.1007/978-1-4939-3341-9_11 (2016).
70 Ruhaak, L. R. *et al.* Optimized workflow for preparation of APTS-labeled N-glycans allowing high-throughput analysis of human plasma glycomes using 48-channel multiplexed CGE-LIF. *J Proteome Res* **9**, 6655-6664, doi:10.1021/pr100802f (2010).
71 Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).

72 Collins, E. S. *et al.* Glycosylation status of serum in inflammatory arthritis in response to anti-TNF treatment. *Rheumatology* **52**, 1572-1582, doi:10.1093/rheumatology/ket189 (2013).
73 Saldova, R., Wormald, M. R., Dwek, R. A. & Rudd, P. M. Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Dis Markers* **25**, 219-232 (2008).
74 Gornik, O. *et al.* Changes of serum glycans during sepsis and acute pancreatitis. *Glycobiology* **17**, 1321-1332, doi:10.1093/glycob/cwm106 (2007).
75 Peracaula, R., Sarrats, A. & Rudd, P. M. Liver proteins as sensor of human malignancies and inflammation. *Proteom Clin Appl* **4**, 426-431, doi:10.1002/prca.200900170 (2010).
76 McCarthy, C. *et al.* The Role and Importance of Glycosylation of Acute Phase Proteins with Focus on Alpha-1 Antitrypsin in Acute and Chronic Inflammatory Conditions. *J Proteome Res* **13**, 3131-3143, doi:10.1021/pr500146y (2014).
77 Higai, K., Azuma, Y., Aoki, Y. & Matsumoto, K. Altered glycosylation of alpha(1)-acid glycoprotein in patients with inflammation and diabetes mellitus. *Clinica Chimica Acta* **329**, 117-125, doi:10.1016/S0009-8981(02)00427-8 (2003).
78 Degraaf, T. W., Vanderstelt, M. E., Anbergen, M. G. & Vandijk, W. Inflammation-Induced Expression of Sialyl Lewis X-Containing Glycan Structures on Alpha-1-Acid Glycoprotein (Orosomucoid) in Human Sera. *J Exp Med* **177**, 657-666, doi:DOI 10.1084/jem.177.3.657 (1993).
79 Duarte-Rey, C., Bogdanos, D. P., Leung, P. S. C., Anaya, J. M. & Gershwin, M. E. IgM predominance in autoimmune disease: Genetics and gender. *Autoimmun Rev* **11**, A404-A412, doi:10.1016/j.autrev.2011.12.001 (2012).
80 Pabst, M. *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). *Molecular & Cellular Proteomics* **14**, 1645-1656, doi:10.1074/mcp.O114.046748 (2015).
81 Bai, L. *et al.* Plasma High-Mannose and Complex/Hybrid N-Glycans Are Associated with Hypercholesterolemia in Humans and Rabbits. *PLoS One* **11**, doi:ARTN e014698210.1371/journal.pone.0146982 (2016).
82 Garner, B. *et al.* Characterization of human apolipoprotein B100 oligosaccharides in LDL subfractions derived from normal and hyperlipidemic plasma: deficiency of alpha-N-acetylneuraminyllactosyl-ceramide in light and small dense LDL particles. *Glycobiology* **11**, 791-802 (2001).
83 Olofsson, S. O. *et al.* Apolipoprotein-B - Structure, Biosynthesis and Role in the Lipoprotein Assembly Process. *Atherosclerosis* **68**, 1-17, doi:Doi 10.1016/0021-9150(87)90088-8 (1987).
84 Ellulu, M. S., Khaza'ai, H., Rahmat, A., Patimah, I. & Abed, Y. Obesity can predict and promote systemic inflammation in healthy adults. *Int J Cardiol* **215**, 318-324, doi:10.1016/j.ijcard.2016.04.089 (2016).
85 Zhang, S., Jiang, K., Sun, C., Lu, H. J. & Liu, Y. K. Quantitative analysis of site-specific N-glycans on sera haptoglobin beta chain in liver diseases. *Acta Bioch Bioph Sin* **45**, 1021-1029, doi:10.1093/abbs/gmt110 (2013).
86 Pompach, P. *et al.* Site-specific Glycoforms of Haptoglobin in Liver Cirrhosis and Hepatocellular Carcinoma. *Molecular & Cellular Proteomics* **12**, 1281-1293, doi:10.1074/mcp.M112.023259 (2013).
87 Dage, J. L., Ackermann, B. L. & Halsall, H. B. Site localization of sialyl Lewis(x) antigen on alpha(1)-acid glycoprotein by high performance liquid chromatography electrospray mass spectrometry. *Glycobiology* **8**, 755-760, doi:DOI 10.1093/glycob/8.8.755 (1998).
88 Fournier, T., Medjoubi-N, N. & Porquet, D. Alpha-1-acid glycoprotein. *Bba-Protein Struct M* **1482**, 157-171, doi:Doi 10.1016/S0167-4838(00)00153-9 (2000).

89 Wong, A. K. L. & Hsia, J. C. Invitro Binding of Propranolol and Progesterone to Native and Desialylated Human Orosomucoid. *Canadian Journal of Biochemistry and Cell Biology* **61**, 1114-1116 (1983).

90 Ponganis, K. V. & Stanski, D. R. Factors Affecting the Measurement of Lidocaine Protein-Binding by Equilibrium Dialysis in Human-Serum. *J Pharm Sci* **74**, 57-60, doi:DOI 10.1002/jps.2600740115 (1985).

91 Morell, A. G., Gregoriadis, G., Scheinberg, I. H., Hickman, J. & Ashwell, G. The role of sialic acid in determining the survival of glycoproteins in the circulation. *J Biol Chem* **246**, 1461-1467 (1971).

92 Ashwell, G. Citation-Classic - the Role of Surface Carbohydrates in the Hepatic Recognition and Transport of Circulating Glycoproteins. *Cc/Life Sci*, 18-18 (1986).

93 Ellies, L. G. *et al.* Sialyltransferase ST3Gal-IV operates as a dominant modifier of hemostasis by concealing asialoglycoprotein receptor ligands. *Proc Natl Acad Sci U S A* **99**, 10042-10047, doi:10.1073/pnas.142005099 (2002).

94 Vasseur, J. A., Goetz, J. A., Alley, W. R. & Novotny, M. V. Smoking and lung cancer-induced changes in N-glycosylation of blood serum proteins. *Glycobiology* **22**, 1684-1708, doi:10.1093/glycob/cws108 (2012).

95 Powell, J. T. Vascular damage from smoking: disease mechanisms at the arterial wall. *Vasc Med* **3**, 21-28, doi:10.1177/1358836X9800300105 (1998).

96 Vestweber, D. & Blanks, J. E. Mechanisms that regulate the function of the selectins and their ligands. *Physiol Rev* **79**, 181-213 (1999).

97 Krishnan, S. *et al.* Combined High-Density Lipoprotein Proteomic and Glycomic Profiles in Patients at Risk for Coronary Artery Disease. *J Proteome Res* **14**, 5109-5118, doi:10.1021/acs.jproteome.5b00730 (2015).

98 Wilson, P. W. F. *et al.* Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837-1847 (1998).

Chapter 7

Serum protein *N*-glycosylation changes with rheumatoid arthritis disease activity during and after pregnancy

Research article

Karli R. Reiding[1]*, Gerda C. M. Vreeker[1], Albert Bondt[1,2], Marco R. Bladergroen[1], Johanna M. W. Hazes[3], Yuri E. M. van der Burgt[1,4], Manfred Wuhrer[1], Radboud J. E. M. Dolhain[3]

[1]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands;

[2]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands;

[3]Department of Rheumatology, Erasmus University Medical Center, Rotterdam, The Netherlands;

[4]Department of Clinical Chemistry, Leiden University Medical Center, Leiden, The Netherlands;

Reprinted and adapted under the terms of the Creative Commons Attribution License from Frontiers in Medicine, January 8, 2018, doi: 10.3389/fmed.2017.00241

Copyright © Reiding, Vreeker, Bondt, Bladergroen, Hazes, van der Burgt, Wuhrer and Dolhain

7.1 Abstract

Objective: Symptoms of rheumatoid arthritis (RA) improve during pregnancy, a phenomenon that was found to be associated with *N*-glycosylation changes of immunoglobulin G (IgG). Recent advances in high-throughput glycosylation analysis allow the assessment of the *N*-glycome of human sera as well. The aim of this study was to identify new protein *N*-glycosylation properties that associate with changes in RA disease activity during and after pregnancy.

Methods: A longitudinal cohort of serum samples was collected during 285 pregnancies (32 control individuals and 253 RA patients). Per individual one sample was collected before conception, three during pregnancy, and three after delivery. Released serum protein *N*-glycans were measured by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-TOF-MS) after employing chemical modification of the sialic acids to allow discrimination of sialic acid linkage isomers.

Results: Serum protein *N*-glycosylation was strongly modified during pregnancy, with similar changes visible in control individuals and RA pregnancies. A decrease in bisection and an increase in galactosylation in diantennary glycans were found, as well as an increase in tri- and tetraantennary species and α2,3-linked sialylation thereof. The change in RA disease activity (DAS28(3)-CRP) proved negatively associated with the galactosylation of diantennary *N*-glycans, and positively with the sialylation of triantennary fucosylated species.

Conclusion: We have identified novel changes in the sialylation of triantennary *N*-glycans (A3FGS) to be associated with RA disease activity during pregnancy. While the protein source of A3FGS is thus far unknown, future studies into the glycosylation phenotype are recommended to improve our understanding of RA disease severity.

7.2 Introduction

Rheumatoid arthritis (RA) is a prevalent autoimmune disease affecting up to 1% of the adult population in developed regions[1,2]. Its main characteristic is a symmetrical polyarthritis involving predominantly the hand- and foot joints, although every organ system may be involved. The disease is more frequently diagnosed in women than in men, and is more common with increasing age[1,3]. The potential causes of RA are diverse in nature[4,5], but a role for T-cells, antibody producing B-cells, but also monocytes/ macrophages has been suggested[1,6]. Interestingly, in many female RA patients an improvement in RA disease severity is reported during pregnancy, as well as a relapse thereof after delivery[7-10]. The reasons for these changes in disease activity are still poorly understood, but are of importance to understand the etiology of RA, and may possibly provide leads for new modes of (personalized) treatment.

Glycosylation, the process of co- and posttranslational protein modification with complex carbohydrates, plays an important role in the interaction, function, and solubility of proteins[11-13]. It is expected that more than half of all proteins is glycosylated with one or more *N*-glycans[14], and commonly observed glycoforms range from high-mannose- to complex-type with two to four antennae (branching *N*-acetylglucosamines) (**Figure 1**)[12,15]. These structures may be extended by additional monosaccharides such as a bisecting *N*-acetylglucosamine, as well as galactoses, *N*-acetylneuraminic acids (sialic acids) and fucoses in a variety of different positions and linkages. This leads to a large *N*-glycan diversity, and may also lead to the formation of specific epitopes such as sialyl-Lewis X, which can be recognized by E-selectin[12,16].

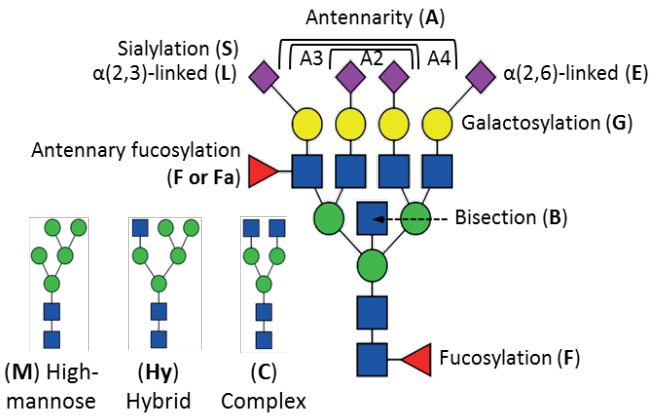


Figure 1. Schematic overview of glycosylation traits derived from human serum proteins. *N-glycan structures are generalized into high-mannose type (M), hybrid type (Hy) and complex type (C) by the number of mannoses (green circle) and antennary N-acetylglucosamines (blue square). High-mannose-type N-glycans can have up to nine mannoses, whereas each antennary N-acetylglucosamine can be terminally substituted with a galactose (yellow circle), and sialic acid (magenta diamond). Sialic acids (S) can either be α2,3-linked (L) or α2,6-linked (E). N-glycans can further be modified with a fucose (F), optionally at an antennary N-acetylglucosamine (Fa), and structures may as well be bisected (B). In case of derived traits the subject of the calculation is represented by the last letter, e.g. sialylation (S), and the group on which it is calculated by the preceding letters, e.g., triantennary fucosylated species (A3F). This, for instance, translates A3FGS into the sialylation per galactose within triantennary fucosylated species.*

Various proteins have already been found to display altered glycosylation with RA and its disease activity. For instance, the *N*-glycosylation of the fragment-crystallizable (Fc) portion of immunoglobulin G (IgG) shows to differ in galactosylation, bisection and fucosylation[17-23], and the acute-phase protein alpha-1-acid glycoprotein (also known as orosomucoid) shows differences in antennarity (*i.e.*, the ratio between di-, tri- and tetraantennary glycans) and fucosylation with RA, as well as changes throughout pregnancy[21,22]. Although such an analysis of single proteins is highly informative, additional insights may be gained by a systemic glycomics approach that covers a broad range of protein-linked glycan modifications.

The total serum *N*-glycome (TSNG) comprises the *N*-glycans from all serum proteins, which are to a large extent liver- (acute-phase proteins) and plasma cell-derived (antibodies)[15,24]. Interestingly, the TSNG has been shown, in a small sample set, to differ between healthy individuals and those with RA, and undergoes clear alteration throughout pregnancy and the following postpartum period[13,14]. However, it is hitherto unknown which TSNG characteristics are associated with the disease activity of RA, and changes thereof throughout pregnancy. Recent developments in mass spectrometry (MS)-based high-throughput glycosylation analysis have provided the opportunity to acquire information on TSNG *N*-glycan complexity, antennarity, galactosylation, fucosylation, as well as on the presence and linkage of sialic acids (α2,6- versus α2,3-linkage)[23, 24]. The latter appears to be of high immunological relevance, since the α2,3-linked sialic acids are required for sialyl-Lewis X formation implicated in the interaction with selectins[25,26].

The objective of the work presented here is to assess the changes in serum *N*-glycosylation throughout pregnancy and the postpartum period in RA patients, and to identify the glycosylation properties associated with the accompanying changes in disease activity (DAS28(3)-CRP). To this end, we studied the *N*-glycosylation of sera from 253 RA and 32 control pregnancies at seven time points before, during and after pregnancy by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS, and report the disease- and pregnancy-associated changes of 78 *N*-glycan species and 91 glycosylation traits derived thereof.

7.3 Methods

Study population and sample collection

The current research is embedded in the PARA study, a nationwide prospective cohort on pregnancy and RA[21,2721,2721, 27]. The cohort consisted of serum samples of 253 pregnancies from 219 RA patients, collected between 2002 and 2009[27]. In addition, 32 pregnancies of healthy Caucasian volunteers without adverse obstetric histories were included and followed from the 1st trimester of pregnancy. Of each patient at least one sample was obtained during pregnancy and one postpartum, with a minimum of three samples per

patient. Only completed pregnancies were included and all patients fulfilled the 1987 ACR criteria for RA. Disease activity was assessed using the disease activity score (DAS) in 28 joints, incorporating the swollen joint count, the tender joint count and the C-reactive protein level (DAS28(3)-CRP). The study was in compliance with the Helsinki Declaration and was approved by the Ethics Review Board at the Erasmus University Medical Center, Rotterdam, the Netherlands. For quality control purposes, 111 plasma standards (Visucon-F frozen normal control plasma, pooled from 20 human donors, citrated, and buffered with 0.02 M HEPES, obtained from Affinity Biologicals, Ancaster, ON) and 40 PBS blanks were distributed across the 21 96-wells sample plates.

Table 1. Study population characteristics	**Control pregnancies (n = 32)**	**Patient pregnancies (n = 253)**
Age at delivery in years, mean (SD)	32.1 (4.4)	32.8 (3.7)
Duration of pregnancy in weeks, mean (SD)	40.1 (1.4)	39.2 (1.9)
Disease duration at first visit in years, mean (SD)	-	7.7 (6.3)
ACPA positive patients, n (%)	-	153/253 (60.5)
RF positive patients, n (%)	-	161/240 (67.1)
Erosive disease, n (%)	-	149/229 (65.1)
Response during pregnancy, n (%)	-	72/137 (52.6)
Flare during postpartum period, n (%)	-	72/232 (31.0)
Disease activity score (DAS28(3)-CRP) at first trimester of pregnancy, mean (SD)	-	3.6 (1.1)

Sample preparation, measurement and analysis

For a full description of the analytical procedure, see **Supplementary Methods**. In short, *N*-glycans were enzymatically released from 6 µL of serum, derivatized by ethyl esterification, and measured by MALDI-TOF-MS using a largely automated approach[26]. Glycan compositions were assigned on the basis of literature, and data were processed and curated using MassyTools[28], resulting in a set of 78 *N*-glycans that were quantified.

Mixed linear regression was used to establish the associations between glycosylation and respectively pregnancy, RA and DAS28(3)-CRP. Linear regression was used to confirm the intra-individual DAS28-CRP associations between the time frames (ΔDAS28-CRP). For the whole study a significance threshold was maintained of $\alpha = 3.03 \cdot 10^{-2}$, allowing detection of effects under a false-discovery rate of 5%. The threshold was calculated using the Benjamini-Hochberg procedure[29].

7.4 Results

To explore the association between serum protein *N*-glycosylation and improvement of RA disease activity during pregnancy, we investigated by MALDI-TOF-MS the TSNG of 285 pregnancies embedded in the PARA study (**Table 1**)[27]. This MS methodology allowed us to obtain information on 78 glycan compositions (**Figure 2**; **Supplementary Table S1**), including discrimination between sialic acid linkage isomers, and to calculate biologically relevant ratios between subsets of glycans in the form of 91 derived traits (**Supplementary Table S2**)[25]. Data quality was confirmed by the repeated measurement of a standard sample (**Supplementary Figure S1**, **Supplementary Table S3**).

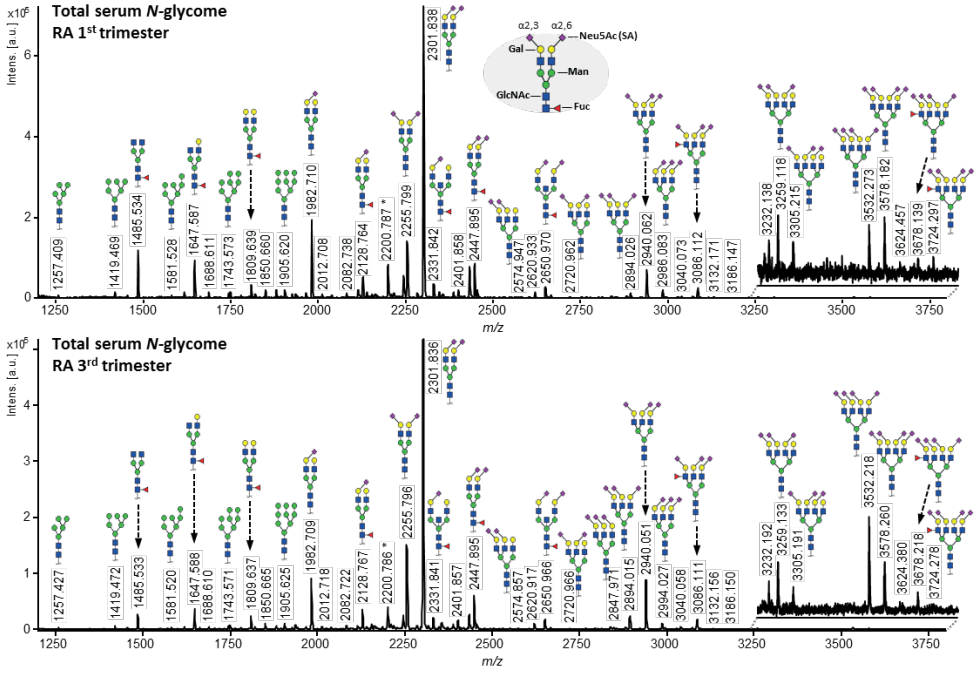


Figure 2. Comparison of MALDI-TOF-MS spectra obtained from released N-glycans from serum proteins after linkage-specific sialic acid esterification and HILIC enrichment. *The shown spectra are derived from an individual with rheumatoid arthritis who has been diagnosed with ACPA and RF, and has been classified as a responder by the EULAR response criteria. Notable differences between the 1st trimester (**top**) and 3rd trimester (**bottom**) spectra include galactosylation (e.g., between m/z 1485.5, 1647.6 and 1809.6), antennarity (e.g. m/z 2940.1 vs. 2301.8), α2,3-linked sialylation (e.g. m/z 2255.8 vs. 2301.8), and α2,6-linked sialylation (e.g. m/z 1982.7 vs. 2301.8).*

Pregnancy-associated glycosylation changes

The association of pregnancy and the postpartum period with serum protein *N*-glycosylation was established by mixed linear regression. Within individuals, comparison was made between the 1st and 3rd trimester (representative of pregnancy; respectively

coded 0 and 1; adjusted for individual), 3rd trimester and 6 weeks postpartum, and 3rd trimester and 26+ weeks postpartum (**Supplementary Table S4**).

With progression of pregnancy, a marked alteration in overall number of antennae per *N*-glycan was observed (**Figure 3**, top row). During pregnancy, within the complex-type glycans (C), the relative abundance of tri- and tetraantennary *N*-glycans (CA3, CA4) showed to increase at the expense of diantennary species (CA2) (β_{CA2} = -1.05 standard error (SE) ± 0.06; β_{CA3} = 1.08 ± 0.06; β_{CA4} = 0.70 ± 0.06), followed by a slow recovery after delivery (β_{CA2} = 0.94 ± 0.07; β_{CA3} = -0.98 ± 0.07; β_{CA4} = -0.56 ± 0.08; **Supplementary Table S4**).

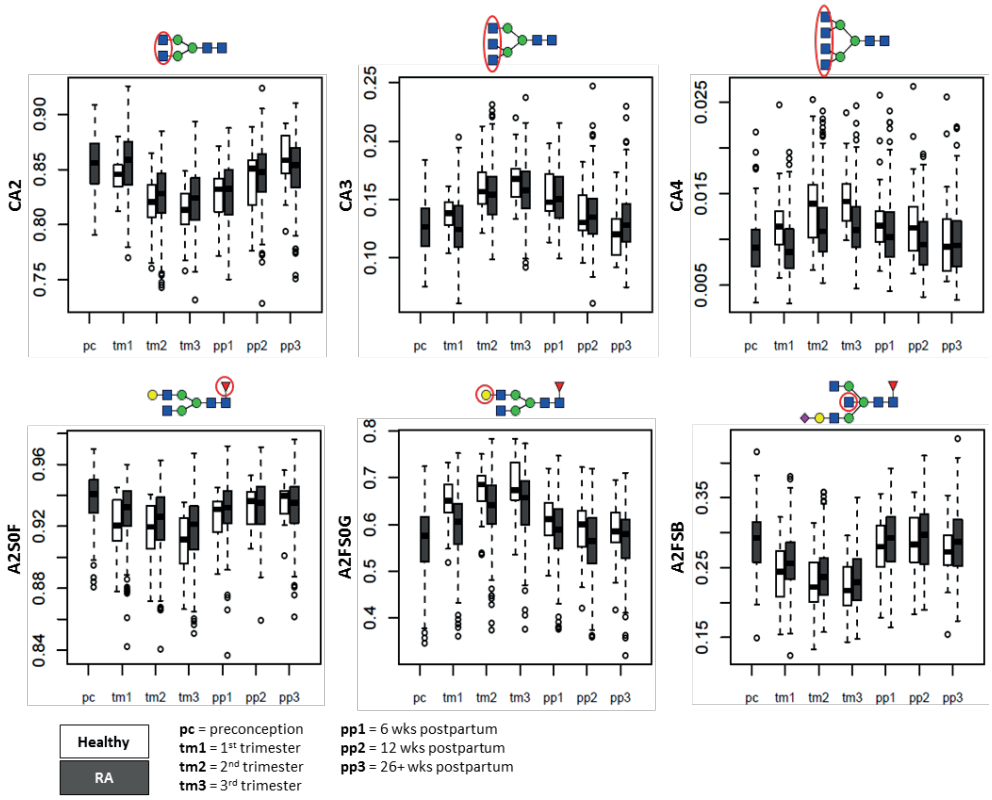


Figure 3. Overall glycosylation changes throughout pregnancy and the postpartum period shown in boxplots. *Displayed are the glycan traits related to antennarity (**top row**), and the glycosylation traits likely to be of immunoglobulin origin (**bottom row**)[15,24]. Depicted are healthy controls (in white) and patients with RA (in grey) at preconception (pc), trimesters 1 through 3 (tm1, tm2, tm3), 6 weeks postpartum (pp1), 12 weeks postpartum (pp2) and 26+ weeks postpartum (pp3). CA2 = diantennary species within complex-type; CA3 = triantennary species within complex-type; CA4 = tetraantennary species within complex-type; A2S0F = fucosylation of nonsialylated diantennary species; A2FS0G = galactosylation of nonsialylated fucosylated diantennary species; A2FSB = bisection of sialylated fucosylated diantennary species.*

For the serum *N*-glycosylation characteristics likely originating from the Fc portion of IgG (diantennary *N*-glycans without sialylation; A2S0)[15,24], we confirmed with progressing pregnancy a decrease in fucosylation (A2S0F, β_{A2S0F} = -0.59 ± 0.05) and increase in galactosylation (A2FS0G, β_{A2FS0G} = 0.59 ± 0.04), both rapidly reversing after delivery (β_{A2S0F} = 0.67 ± 0.04; β_{A2FS0G} = -0.81 ± 0.04) (**Figure 3**, **bottom row**).

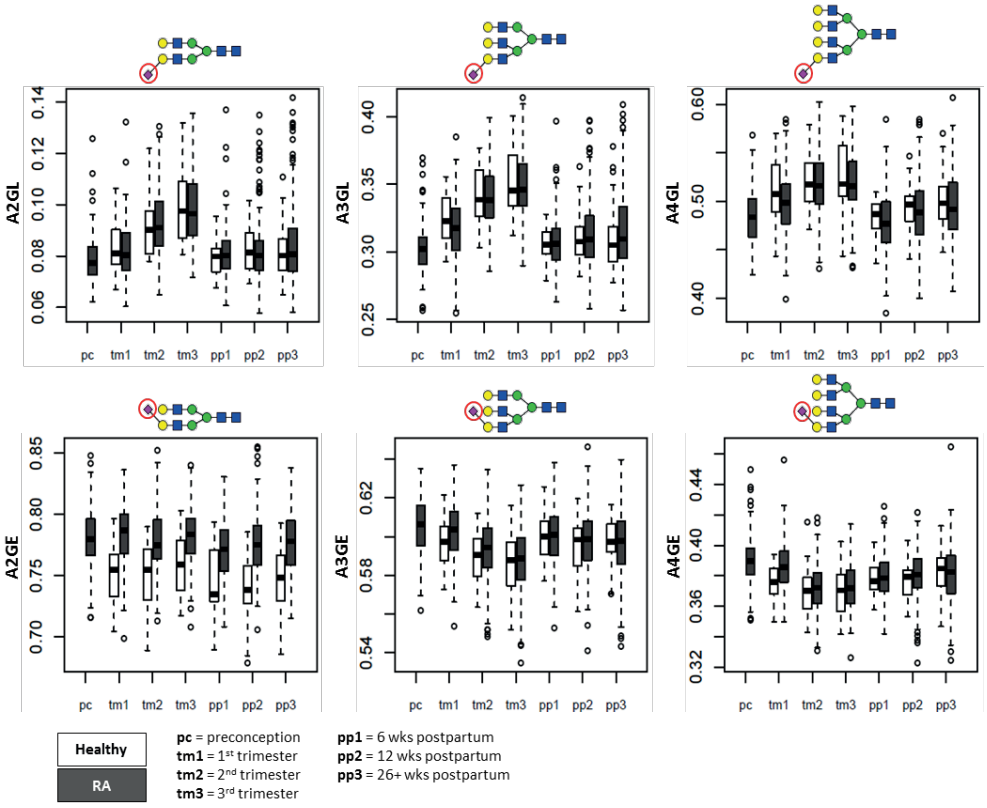


Figure 4. Sialylation changes throughout pregnancy and the postpartum period shown in boxplots. *Whereas α2,3-linked sialylation (L) displays the most prominent change throughout pregnancy (**top row**), α2,6-linked sialylation shows the most distinction between patients and controls (**bottom row**). Separation is made between pregnancies of patients with RA (in grey) and healthy controls (in white). A2GL = α2,3-linked sialylation per galactose of diantennary species; A3GL = α2,3-linked sialylation per galactose of triantennary species; A4GL = α2,3-linked sialylation per galactose of tetraantennary species; A2GE = α2,6-linked sialylation per galactose of diantennary species; A3GE = α2,6-linked sialylation per galactose of triantennary species; A4GE = α2,6-linked sialylation per galactose of tetraantennary species.*

For the serum *N*-glycosylation characteristics likely originating for a large part from IgG-Fab, IgM and IgA (fucosylated diantennary *N*-glycans with sialylation; A2FS)[15,24], we observed a decrease in bisection with trimester progression (A2FSB, β_{A2FSB} = -0.56 ± 0.04) and a rapid return to pre-pregnancy levels at 6 weeks postpartum (β_{A2FSB} = 1.11 ± 0.03).

Notably, whereas α2,6-linked sialylation remained relatively stable throughout pregnancy (**Figure 4**, **bottom row**), a drastic increase of α2,3-linked sialylation (e.g. per galactose on di- or triantennary glycans, respectively A2GL and A3GL) was observed up to the 3rd trimester (β_{A2GL} = 1.11 ± 0.05; β_{A3GL} = 1.16 ± 0.05), followed by a rapid return to baseline levels at the first time point after delivery (β_{A2GL} = -1.19 ± 0.05; β_{A3GL} = -1.42 ± 0.04) (**Figure 4**, **top row**).

Differences between healthy and RA

By mixed linear regression, glycosylation values were compared between individuals with RA and controls (healthy = 0, RA = 1; adjusted for pregnancy) (**Supplementary Table S5**). For immunoglobulin-type glycosylation we observed in RA a lower galactosylation (A2FS0G, β_{A2FS0G} = -0.44 ± 0.07), higher bisection (A2FS0B, β_{A2FS0B} = 0.55 ± 0.08), and a higher sialylation per galactose (A2FGS, β_{A2FGS} = 0.75 ± 0.08). In addition, differences were found in several non-fucosylated diantennary traits (higher A2F0G, A2F0GS and A2GE, lower A2F0B, A2F), all mainly driven by elevated levels of H5N4E2 (β_{H5N4E2} = 0.97 ± 0.07) in RA patient serum. For tri- and tetraantennary *N*-glycosylation we observe with RA a higher fucosylation (e.g. A3F, β_{A3F} = 0.34 ± 0.08; A4F, β_{A4F} = 0.56 ± 0.08) particularly within species with α2,3-linked sialylation (A3LF, β_{A3LF} = 0.35 ± 0.08; A4LF, β_{A4LF} = 0.58 ± 0.08), suggesting an increase in sialyl-Lewis X/A.

Association of glycosylation with DAS28(3)-CRP

In addition, we compared by mixed linear regression the association of glycosylation with RA disease activity, as assessed by DAS28(3)-CRP, both between and within individuals with RA (respectively adjusted for time point and individual) (**Supplementary Figure S2**, **Supplementary Table S6**). The increase in DAS28(3)-CRP associated with both the decrease in IgG-Fc-type galactosylation (*e.g.*, A2S0G, β_{A2S0G} = -0.34 ± 0.02) and the decrease in fucosylation of non-α2,3-sialylated triantennary species (A3L0F, β_{A3L0F} = -0.13 ± 0.03). Increasing with DAS28(3)-CRP were the bisection of IgG-Fc-type glycans (A2F0S0B, $\beta_{A2F0S0B}$ = 0.25 ± 0.03) as well as the sialylation and fucosylation of tri- and tetraantennary species (*e.g.*, A3FGS, β_{A3FGS} = 0.26 ± 0.02; A4F, β_{A4F} = 0.13 ± 0.03; A4GS, β_{A4GS} = 0.17 ± 0.02). Glycosylation features associating with DAS28(3)-CRP only within (and not between) individuals throughout pregnancy were the fucosylation of (α2,6-)sialylated diantennaries (A2SF, β_{A2SF} = -0.12 ± 0.02) and α2,6-sialylation of triantennary nonfucosylated species (A3F0GE, β_{A3F0GE} = 0.14 ± 0.02).

Association of glycosylation with the improvement of DAS28(3)-CRP during pregnancy

Lastly, between trimesters 1 and 3 (response timeframe) as well as between trimester 3 and 12 weeks postpartum (flare timeframe), we employed linear regression to confirm the within-individual association of DAS28(3)-CRP (ΔDAS28(3)-CRP) with glycosylation (Δglycosylation) (**Supplementary Table S7**). In both timeframes, ΔDAS28(3)-CRP showed a

negative association with the Δgalactosylation of IgG-Fc type species (ΔA2S0G, responder, $\beta_{\Delta A2S0G}$ = -0.29 ± 0.07; flare, $\beta_{\Delta A2S0G}$ = -0.35 ± 0.07; ΔA2FS0G, responder, $\beta_{\Delta A2FS0G}$ = -0.29 ± 0.07; flare, ΔA2FS0G; $\beta_{\Delta A2FS0G}$ = -0.33 ± 0.07), whereas the sialylation of triantennary fucosylated species showed in both cases a positive association (ΔA3FGS, responder, $\beta_{\Delta A3FGS}$ = 0.26 ± 0.07; flare, $\beta_{\Delta A3FGS}$ = 0.35 ± 0.07) (**Figure 5**). The flare timeframe additionally showed a positive association of ΔDAS28(3)-CRP with the (α2,6)-sialylation of (fucosylated) diantennary glycans (ΔA2GE, $\beta_{\Delta A2GE}$ = 0.27 ± 0.07; ΔA2FGS, $\beta_{\Delta A2FGS}$ = 0.24 ± 0.07).

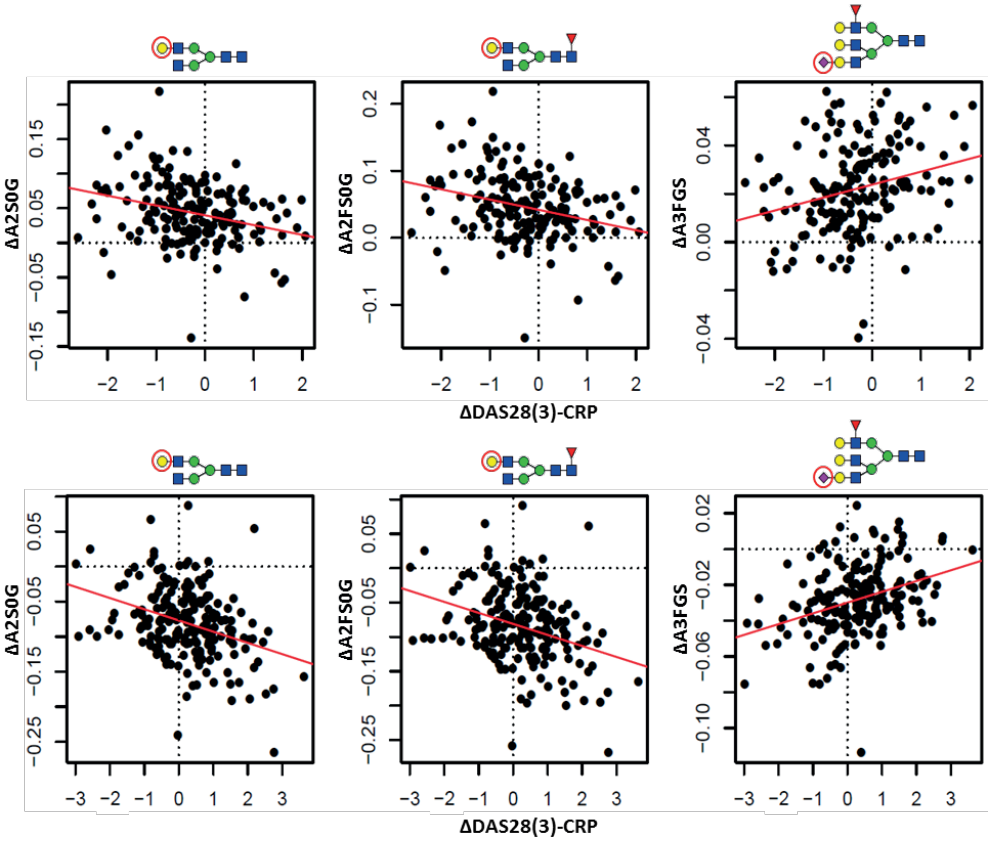


Figure 5. The relation between changing RA disease activity (ΔDAS28(3)-CRP) and changing glycosylation. *Distinction is made between the timeframes to assess responder status (the change between 1st and 3rd trimester;* ***top row****) and flare status (the change between 3rd trimester and 12 weeks postpartum;* ***bottom row****). ΔA2S0G = the change in galactosylation per antenna of diantennary nonsialylated species; ΔA2FS0G = the change in galactosylation of diantennary fucosylated nonsialylated species; ΔA3FGS = the change in sialylation per galactose of triantennary fucosylated species.*

7.5 Discussion

The decrease of RA disease activity during pregnancy and the flare following delivery are reproducible clinical observations that are mechanistically poorly understood[8-10]. To expand our understanding of protein glycosylation that associates with RA disease activity, we studied the total serum *N*-glycosylation changes occurring throughout pregnancy and the postpartum period of 253 pregnancies of patients with RA, along with 32 control pregnancies.

Interpretation of glycomics data

To interpret the information contained in this study, several aspects of our MALDI-TOF-MS analysis need to be kept in mind. First, mass spectrometry assesses glycan chemical compositions and not structures, although sialic acid-linkage information is provided by the employed derivatization technique[25]. Other structural characteristics are presumed based on a wide array of literature on biosynthetic pathways, and on experiments with enzymatic digestion, nuclear magnetic resonance spectroscopy, and MS fragmentation[12,15,30-32]. Second, the changes observed in the released *N*-glycan samples could have originated from changes in the glycosylation of proteins, but also from changes in the abundance of those glycoproteins. Derived traits have been constructed to reflect differences in biosynthesis, and our current-day understanding of the relative contribution of specific glycoproteins and tissues to the serum *N*-glycome[15,24]. Third, the mass spectrometric analysis did not provide quantitative ratios of *N*-glycosylation, but the direction and magnitude of observed changes is expected to be biologically representative, as suggested by method comparisons[33].

Glycosylation changes with pregnancy

Previously, we found glycosylation changes throughout healthy pregnancies, and have in RA patients studied the specific glycosylation of IgG-Fc throughout pregnancy[20,21,34-37]. In the current study, the first application of a similar MALDI-MS approach with automated sample preparation, we achieved the analysis of the total serum *N*-glycosylation throughout the pregnancies of both healthy controls and RA patients, in total leading to the analysis of 1770 clinical samples[26]. Since the current cohort also contains the control individuals (without RA) from prior studies, these previous results were confirmed, but we could also for the first time show that comparable TSNG changes can be observed in RA patients.

As such, we detected with the progression of pregnancy an increase of galactosylation (A2FS0G) as prior reported for the Fc part of IgG, as well as a decrease in bisection (A2FSB)[18,20,21]. Additionally, we observed with pregnancy an overall increase in glycan branching (from CA2 to CA3 and CA4) and an increase of α2,3-linked sialylation (L, in part at the expense of α2,6-sialylation, E). In all cases the postpartum period led to the return to the values before or at the beginning of pregnancy. The increased branching observed in the TSNG could have originated from the abundance and antennarity of acute-phase proteins such as alpha-1-acid glycoprotein, for which an increased serum level and *N*-glycan

branching has been reported with pregnancy[38,39]. On the other hand, we did not observe the decrease in fucosylation reported for the same protein, potentially obscured by the increased fucosylation of other proteins, and have yet to identify the source of the substantial increase of α2,3-linked sialylation up to the 3rd trimester. In literature it has been reported that IgG and alpha-1-acid glycoprotein glycosylation may be affected by estrogens[40,41]. known to change significantly throughout pregnancy[42], and it is conceivable that other proteins within the TSNG are under similar high-level control.

Glycosylation differs between RA patients and healthy controls

The total serum *N*-glycosylation changes with pregnancy showed remarkably comparable between RA patients and healthy controls, but baseline differences could be detected. For example, RA patients displayed a lower degree of IgG-Fc-type galactosylation and higher bisection when compared to controls (A2FS0G, A2FS0B). A decreased IgG-Fc galactosylation and increased bisection are well-known to associate with a variety of inflammatory conditions, including RA, inflammatory bowel disease, and aging[18,43,44], and the same glycosylation phenotypes appear detectable in our TSNG study as well[20,21]. The mechanisms by which IgG-Fc glycosylation may affect inflammatory processes remain for a large part to be elucidated, but increased galactosylation and decreased fucosylation have been implicated in increased FcγRIIa and FcγRIIIa binding and antibody-dependent cellular cytotoxicity[45-47].

Furthermore, RA patients showed in our study a higher α2,6-sialylation and lower fucosylation of diantennary glycans (A2GE, A2F) compared to healthy controls, mainly driven by the *N*-glycan composition H5N4E2, as well as a substantially higher (multi-)fucosylation of tri- and tetraantennary glycans (A3F(a), A4F(a)). The increased A3/A4 fucosylation strongly suggests the upregulation of sialyl-Lewis X, which has been reported with inflammatory arthritis for several acute-phase proteins, *e.g.*, alpha-1-acid glycoprotein, haptoglobin, alpha-1-antichymotrypsin, and transferrin[48-52]. The sialyl-Lewis X on glycans is known to bind to E-selectin, an inducible receptor expressed by endothelial cells[16]. This interaction is implicated in the homing of immune cells to a site of inflammation[53], as well as cancer metastasis[54], and may play a role in RA as well. For instance, the alpha-1-acid glycoprotein observed within RA synovial fluid is thought to be of hepatic origin[55], meaning its circulatory variant might make use of a glycan-mediated mechanism for transportation towards the inflamed synovial tissue.

Glycosylation associates with RA disease activity

Next to a negative association with the galactosylation of IgG-Fc-type *N*-glycans, we additionally report a positive association between (Δ)DAS28(3)-CRP and the sialylation of triantennary fucosylated glycans (A3FGS), notably of the α2,6-linked variety. This is in contrast to the TSNG changes throughout pregnancy, which shows marked increase of α2,3-linked sialylation but not an association with disease activity. A recent study, which

compared total plasma *N*-glycosylation with the levels of various metabolic and inflammatory markers, has indicated a link between A3FGS and C-reactive protein (CRP), suggesting that the CRP component of DAS28(3)-CRP may in part be responsible for the association found within the current study[56]. While the protein source of A3FGS remains unknown, future studies will have to reveal its biomarker potential and facilitate our understanding of RA disease severity.

7.6 Summary

To summarize, by performing MS-based total serum protein *N*-glycosylation analysis we detected 1) changes in protein glycosylation throughout pregnancy, 2) glycosylation differences between healthy individuals and RA patients, and 3) glycosylation traits coinciding with the pregnancy-associated changes in RA disease activity. While we confirmed in serum the IgG glycosylation phenotypes that were prior reported to associate with RA disease activity, our glycomics approach has additionally allowed the detection of changes that are presumably independent from IgG, namely the sialylation of fucosylated triantennary *N*-glycans (A3FGS).

7.7 Acknowledgments

This work was supported by the Dutch Arthritis Foundation (NR-10-1-411) and the European Union Seventh Framework Programme project HighGlycan (278535). We thank Florent Clerc for his invaluable assistance with conceptualising the schematic overview of glycosylation traits (Figure 1).

7.8 Additional information

Supplementary information

Supplementary information accompanies this paper.

Competing interests

The authors declare no competing financial interest.

7.9 References

1 Scott, D. L., Wolfe, F. & Huizinga, T. W. Rheumatoid arthritis. *Lancet* **376**, 1094-1108, doi:10.1016/S0140-6736(10)60826-4 (2010).

2 Cooper, G. S., Bynum, M. L. & Somers, E. C. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J Autoimmun* **33**, 197-207, doi:10.1016/j.jaut.2009.09.008 (2009).

3 Parks, C. G. *et al.* Expert panel workshop consensus statement on the role of the environment in the development of autoimmune disease. *Int J Mol Sci* **15**, 14269-14297, doi:10.3390/ijms150814269 (2014).

4 van der Helm-van Mil, A. H. & Huizinga, T. W. Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets. *Arthritis Res Ther* **10**, 205, doi:10.1186/ar2384 (2008).

5 van Oosterhout, M. *et al.* Differences in synovial tissue infiltrates between anti-cyclic citrullinated peptide-positive rheumatoid arthritis and anti-cyclic citrullinated peptide-negative rheumatoid arthritis. *Arthritis Rheum* **58**, 53-60, doi:10.1002/art.23148 (2008).
6 Choy, E. H. *et al.* Therapeutic benefit of blocking interleukin-6 activity with an anti-interleukin-6 receptor monoclonal antibody in rheumatoid arthritis: a randomized, double-blind, placebo-controlled, dose-escalation trial. *Arthritis Rheum* **46**, 3143-3150, doi:10.1002/art.10623 (2002).
7 Krause, M. L. & Makol, A. Management of rheumatoid arthritis during pregnancy: challenges and solutions. *Open Access Rheumatol* **8**, 23-36, doi:10.2147/OARRR.S85340 (2016).
8 Nelson, J. L. & Ostensen, M. Pregnancy and rheumatoid arthritis. *Rheum Dis Clin North Am* **23**, 195-212 (1997).
9 Barrett, J. H., Brennan, P., Fiddler, M. & Silman, A. J. Does rheumatoid arthritis remit during pregnancy and relapse postpartum? Results from a nationwide study in the United Kingdom performed prospectively from late pregnancy. *Arthritis Rheum* **42**, 1219-1227, doi:10.1002/1529-0131(199906)42:6<1219::AID-ANR19>3.0.CO;2-G (1999).
10 Ince-Askan, H. & Dolhain, R. J. Pregnancy and rheumatoid arthritis. *Best Pract Res Clin Rheumatol* **29**, 580-596, doi:10.1016/j.berh.2015.07.001 (2015).
11 Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3-49, doi:10.1093/glycob/cww086 (2017).
12 in *Essentials of Glycobiology* (eds A. Varki *et al.*) (2015).
13 Hart, G. W. & Copeland, R. J. Glycomics hits the big time. *Cell* **143**, 672-676, doi:10.1016/j.cell.2010.11.008 (2010).
14 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8 (1999).
15 Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-343, doi:10.1007/s10719-015-9626-2 (2016).
16 Thomas, V. H., Yang, Y. & Rice, K. G. In vivo ligand specificity of E-selectin binding to multivalent sialyl Lewisx N-linked oligosaccharides. *J Biol Chem* **274**, 19035-19040 (1999).
17 Axford, J. S. Glycosylation and rheumatic disease. *Biochim Biophys Acta* **1455**, 219-229 (1999).
18 Parekh, R. B. *et al.* Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452-457 (1985).
19 Malhotra, R. *et al.* Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med* **1**, 237-243 (1995).
20 van de Geijn, F. E. *et al.* Immunoglobulin G galactosylation and sialylation are associated with pregnancy-induced improvement of rheumatoid arthritis and the postpartum flare: results from a large prospective cohort study. *Arthritis Res Ther* **11**, R193, doi:10.1186/ar2892 (2009).
21 Bondt, A. *et al.* Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res* **12**, 4522-4531, doi:10.1021/pr400589m (2013).
22 Ercan, A. *et al.* Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis. *Arthritis Rheum* **62**, 2239-2248, doi:10.1002/art.27533 (2010).
23 Sebastian, A. *et al.* Glycan Biomarkers for Rheumatoid Arthritis and Its Remission Status in Han Chinese Patients. *Omics* **20**, 343-351, doi:10.1089/omi.2016.0050 (2016).
24 Klein, A. Human total serum N-glycome. *Adv Clin Chem* **46**, 51-85 (2008).
25 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).

26 Bladergroen, M. R. *et al.* Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *J Proteome Res* **14**, 4080-4086, doi:10.1021/acs.jproteome.5b00538 (2015).

27 de Man, Y. A., Dolhain, R. J., van de Geijn, F. E., Willemsen, S. P. & Hazes, J. M. Disease activity of rheumatoid arthritis during pregnancy: results from a nationwide prospective study. *Arthritis Rheum* **59**, 1241-1248, doi:10.1002/art.24003 (2008).

28 Jansen, B. C. *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* **14**, 5088-5098, doi:10.1021/acs.jproteome.5b00658 (2015).

29 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).

30 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).

31 Kang, P., Mechref, Y. & Novotny, M. V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun Mass Spectrom* **22**, 721-734, doi:10.1002/rcm.3395 (2008).

32 Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **13**, 2314-2327, doi:10.1021/pr401092y (2014).

33 Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).

34 Ruhaak, L. R., Uh, H. W., Deelder, A. M., Dolhain, R. E. & Wuhrer, M. Total plasma N-glycome changes during pregnancy. *J Proteome Res* **13**, 1657-1668, doi:10.1021/pr401128j (2014).

35 Jansen, B. C. *et al.* Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep* **6**, 23296, doi:10.1038/srep23296 (2016).

36 Bondt, A. *et al.* Longitudinal monitoring of immunoglobulin A glycosylation during pregnancy by simultaneous MALDI-FTICR-MS analysis of N- and O-glycopeptides. *Sci Rep* **6**, 27955, doi:10.1038/srep27955 (2016).

37 Bondt, A. *et al.* Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics* **13**, 3029-3039, doi:10.1074/mcp.M114.039537 (2014).

38 Havenaar, E. C. *et al.* Severe rheumatoid arthritis prohibits the pregnancy-induced decrease in alpha3-fucosylation of alpha1-acid glycoprotein. *Glycoconj J* **15**, 723-729 (1998).

39 van Dijk, W., Havenaar, E. C. & Brinkman-van der Linden, E. C. Alpha 1-acid glycoprotein (orosomucoid): pathophysiological changes in glycosylation in relation to its function. *Glycoconj J* **12**, 227-233 (1995).

40 Ercan, A. *et al.* Estrogens regulate glycosylation of IgG in women and men. *JCI Insight* **2**, e89703, doi:10.1172/jci.insight.89703 (2017).

41 Brinkman-Van der Linden, C. M. *et al.* Oral estrogen treatment induces a decrease in expression of sialyl Lewis x on alpha 1-acid glycoprotein in females and male-to-female transsexuals. *Glycobiology* **6**, 407-412 (1996).

42 Thomas, C. M., Corbey, R. S. & Rolland, R. Assessment of unconjugated oestradiol and progesterone serum levels throughout pregnancy in normal women and in women with hyperprolactinaemia, who conceived after bromocriptine treatment. *Acta Endocrinol (Copenh)* **86**, 405-414 (1977).

43 Trbojevic Akmacic, I. *et al.* Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm Bowel Dis* **21**, 1237-1247, doi:10.1097/MIB.0000000000000372 (2015).

44 Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res Rev* **12**, 685-698, doi:10.1016/j.arr.2012.02.002 (2013).

45 Thomann, M. *et al.* In vitro glycoengineering of IgG1 and its effect on Fc receptor binding and ADCC activity. *PLoS One* **10**, e0134949, doi:10.1371/journal.pone.0134949 (2015).

46 Thomann, M., Reckermann, K., Reusch, D., Prasser, J. & Tejada, M. L. Fc-galactosylation modulates antibody-dependent cellular cytotoxicity of therapeutic antibodies. *Mol Immunol* **73**, 69-75, doi:10.1016/j.molimm.2016.03.002 (2016).

47 Shields, R. L. *et al.* Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J Biol Chem* **277**, 26733-26740, doi:10.1074/jbc.M202069200 (2002).

48 Saroha, A., Biswas, S., Chatterjee, B. P. & Das, H. R. Altered glycosylation and expression of plasma alpha-1-acid glycoprotein and haptoglobin in rheumatoid arthritis. *J Chromatogr B Analyt Technol Biomed Life Sci* **879**, 1839-1843, doi:10.1016/j.jchromb.2011.04.024 (2011).

49 Thompson, S., Kelly, C. A., Griffiths, I. D. & Turner, G. A. Abnormally-fucosylated serum haptoglobins in patients with inflammatory joint disease. *Clin Chim Acta* **184**, 251-258 (1989).

50 Brinkman-van der Linden, E. C., de Haan, P. F., Havenaar, E. C. & van Dijk, W. Inflammation-induced expression of sialyl LewisX is not restricted to alpha1-acid glycoprotein but also occurs to a lesser extent on alpha1-antichymotrypsin and haptoglobin. *Glycoconj J* **15**, 177-182 (1998).

51 Feelders, R. A., Vreugdenhil, G., de Jong, G., Swaak, A. J. & van Eijk, H. G. Transferrin microheterogeneity in rheumatoid arthritis. Relation with disease activity and anemia of chronic disease. *Rheumatol Int* **12**, 195-199 (1992).

52 Albrecht, S., Unwin, L., Muniyappa, M. & Rudd, P. M. Glycosylation as a marker for inflammatory arthritis. *Cancer Biomark* **14**, 17-28, doi:10.3233/CBM-130373 (2014).

53 Jones, T. R., Shirasugi, N., Adams, A. B., Pearson, T. C. & Larsen, C. P. Intravital microscopy identifies selectins that regulate T cell traffic into allografts. *J Clin Invest* **112**, 1714-1723, doi:10.1172/JCI19391 (2003).

54 Hauselmann, I. & Borsig, L. Altered tumor-cell glycosylation promotes metastasis. *Front Oncol* **4**, 28, doi:10.3389/fonc.2014.00028 (2014).

55 Havenaar, E. C. *et al.* Do synovial fluid acute phase proteins from patients with rheumatoid arthritis originate from serum? *Glycoconj J* **14**, 457-465 (1997).

56 Reiding, K. R. *et al.* Human Plasma N-glycosylation as Analyzed by Matrix-Assisted Laser Desorption/Ionization-Fourier Transform Ion Cyclotron Resonance-MS Associates with Markers of Inflammation and Metabolic Health. *Mol Cell Proteomics* **16**, 228-242, doi:10.1074/mcp.M116.065250 (2017).

Chapter 8

High-throughput serum *N*-glycomics: method comparison and application to study rheumatoid arthritis and pregnancy-associated changes

Research manuscript

Karli R. Reiding[1], Albert Bondt[1,2]†, René Hennig[3,4]†, Richard A. Gardner[5]†, Roisin O'Flaherty[6]†, Irena Trbojević-Akmačić[7]†, Archana Shubhakar[5], Johanna M. W. Hazes[8], Udo Reichl[3,9], Daryl L. Fernandes[5], Maja Pučić-Baković[7], Erdmann Rapp[3,4]‡, Daniel I. R. Spencer[5]‡, Radboud J. E. M. Dolhain[3]‡, Pauline M. Rudd[6]‡, Gordan Lauc[7,10]‡, Manfred Wuhrer[1]‡

†These authors contributed equally

‡These authors contributed equally

[1]Center for Proteomics and Metabolomics, Leiden University Medical Center (LUMC), Leiden, The Netherlands;

[2]Department of Rheumatology, LUMC, Leiden, The Netherlands;

[3]Max Planck Institute (MPI) for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany;

[4]glyXera GmbH., 39120 Magdeburg, Germany;

[5]Ludger Ltd., Culham Science Centre, Abingdon, Oxfordshire, United Kingdom;

[6]GlycoScience Group, National Institute for Bioprocessing Research and Training (NIBRT), Fosters Avenue, Blackrock, Co. Dublin, Ireland;

[7]Genos Glycoscience Research Laboratory, Zagreb, Croatia;

[8]Department of Rheumatology, Erasmus University Medical Center, Rotterdam, The Netherlands;

[9]Otto von Guericke University Magdeburg, Chair of Bioprocess Engineering, 39106 Magdeburg, Germany;

[10]Faculty of Pharmacy and Biochemistry, University of Zagreb, 10000 Zagreb, Croatia;

Manuscript under review.

8.1 Abstract

N-Glycosylation is a fundamentally important protein modification with a major impact on glycoprotein characteristics such as serum half-life and receptor interaction. More than half of the proteins in human serum are glycosylated, and the relative abundances of protein glycoforms often reflect alterations in health and disease. Several analytical methods are currently capable of analyzing the total serum *N*-glycosylation in a high-throughput manner.

Here we evaluate and compare the performance of released *N*-glycome analyses by hydrophilic-interaction ultra-high-performance liquid chromatography with fluorescence detection (HILIC-UHPLC-FLD), multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (xCGE-LIF), and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) for analyzing the same panel of serum samples. The samples were obtained at multiple time points during the pregnancies and postpartum periods of healthy women and patients with rheumatoid arthritis (RA). We assessed the analytical methods based on their technical performance as well as on their ability to describe serum protein *N*-glycosylation changes throughout pregnancy, with RA, and with RA disease activity.

Overall, the methods proved to be similar in their detection and relative quantification of serum protein *N*-glycosylation. However, the non-MS methods showed superior repeatability over MALDI-TOF-MS, and allowed the best structural separation of low-complexity *N*-glycans. MALDI-TOF-MS achieved the highest throughput and provided compositional information on higher-complexity *N*-glycans. Consequentially, MALDI-TOF-MS could establish the linkage-specific sialylation differences within pregnancy and RA, whereas HILIC-UHPLC-FLD and xCGE-LIF demonstrated differences in α1,3- and α1,6-branch galactosylation. While the combination of methods proved to be the most beneficial for the analysis of total serum protein *N*-glycosylation, informed method choices can be made for the glycosylation analysis of single proteins or samples of varying complexity.

8.2 Introduction

Glycosylation is a critical and ubiquitous co- and post-translational protein modification which affects a wide variety of biological functions[1,2]. Manipulation of protein *N*-glycosylation has shown to be effective for influencing protein half-life and receptor interaction in molecules as diverse as gamma-immunoglobulins (IgG) and alpha-immunoglobulins (IgA), as well as erythropoietin[3-5]. Furthermore, glycans of pathogens and cancer cells, as well as their receptors, are promising targets for both small molecule drugs and biopharmaceuticals[6-8]. Longitudinal studies have shown a remarkable stability of the total plasma protein *N*-glycome of individuals over a several-year period[9,10], while population studies have revealed associations of protein *N*-glycosylation with aging, sex, inflammation, body-mass index, metabolism, and a variety of cancers and autoimmune disorders[11-13]. Longitudinal glycosylation profiling of single proteins and complex biofluids provides an opportunity for early detection of systemic alterations, and may serve to stratify patient populations[14-16].

The last decade saw major developments in high-throughput (HTP) analytical methodologies for achieving *N*-glycosylation analysis[17,18]. Of the methods available for glycan analysis, however, only a few have demonstrated the capability to profile the several thousands of samples making up many of the current-day clinical cohorts. The largest glycomics profiling studies thus far, all of them comprising more than 2000 cases, were performed with hydrophilic-interaction (ultra-)high-performance liquid chromatography with fluorescence detection (HILIC-(U)HPLC-FLD)[19-21], multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (xCGE-LIF)[22], and matrix-assisted laser desorption/ionization (time-of-flight) mass-spectrometry (MALDI-(TOF)-MS)[11], and have assessed the released *N*-glycans from total serum/plasma or single glycoproteins such as IgG and alpha-1-antitrypsin. However, while each of the analytical methods proved informative for the analysis of *N*-glycosylation of complex-mixtures, none of them provided full structural characterization of glycan species without follow-up experiments such as exoglycosidase digestion and/or tandem-MS[23,24].

For glycans of relatively low complexity, such as found on the fragment-crystallizable (Fc) portion of IgG, comparative analysis has revealed highly similar findings between the aforementioned analytical methods[25-27]. However, the study of IgG-Fc glycosylation does not comprise structures of higher antennarity, *i.e.* tri- and tetraantennary species, nor the high levels of terminal *N*-acetylneuraminic acids found on most serum proteins other than IgG[24,28]. In addition, many biological sources show more heterogeneous glycosylation than human IgG, and information on the comparative performance of HTP glycomics methods on such complex samples is still missing.

Here, we have studied the performance of the latest generation of HTP *N*-glycome analysis methodologies, focusing on HILIC-UHPLC-FLD, xCGE-LIF, and MALDI-TOF-MS. With this study we aimed to assess their respective suitability for total serum protein *N*-glycome (TSNG) analysis, explore the overlap and orthogonality of the information between the methods, and determine their strengths and weaknesses for revealing different types of *N*-glycan properties. To answer these questions in a clinically relevant setting, all methods were challenged with analyzing the same subset of the pregnancy-induced amelioration of rheumatoid arthritis (PARA) cohort, a longitudinal study aimed at exploration of the temporary improvement of rheumatoid arthritis (RA) severity experienced by women during pregnancy[29].

***Table 1**. Characteristics of pregnancies included in the study.*

	Pregnancies	
	Control (n = 32)	**RA (n = 36)**
Age at delivery in years, mean (SD)	32.1 (4.4)	32.5 (4.0)
Duration of pregnancy in weeks, mean (SD)	40.1 (1.4)	39.2 (1.9)
Disease duration at first visit in years, mean (SD)	-	7.3 (5.8)
ACPA positive patients, n (%)	-	22 (61%)
RF positive patients, n (%)	-	25 (69.4%)
Erosive disease, n (%)	-	14 (38.9%)
Disease activity score (DAS28(3)-CRP) at preconception, mean (SD)	-	3.8 (1.0)

***Table 2**. Participating laboratories and methodologies used in the study.*

Affiliation (participant)	Methodology	Labeling or modification	Clean-up	Automation	Annotation	References
NIBRT (1)	HILIC-UHPLC-FLD	2-AB labeling	Solid-phase hydrazide bead capture; Hypersep Diol cartridge	sample preparation	GU value with database	23,41
Genos (2)	HILIC-UHPLC-FLD	2-AB labeling	GHP filter plate	-	GU value with database	23,41
Ludger (3)	HILIC-UHPLC-FLD	2-AB labeling	LudgerClean T1 cartridge	sample preparation	GU value with database; LC-MS composition analysis	66
MPI (4)	xCGE-LIF	APTS labeling	glyXera glyXbeads	normalization; integration; database matching; quality control	Normalized migration time value with database	10,31,61
LUMC (5)	MALDI-TOF-MS	Ethyl esterification of sialic acids	GHP filter plate	sample preparation; integration; quality control	mass error; isotope distribution	35,40

8.3 Results

To qualitatively compare glycomics analytical methodologies, we measured the released TSNGs of 36 RA patient pregnancies at three time points (before conception, at the 3rd

trimester of pregnancy, and 26 weeks postpartum) and 32 pregnancies of healthy controls at two time points (at the 3rd trimester, and 26+ weeks postpartum) (**Table 1**). In addition, a repeat measurement of a standard plasma sample was included in the study to establish intra- and interplate variation. The methodologies used on the released glycan samples were as follows: HILIC-UHPLC-FLD after 2-aminobenzamide (2-AB) labeling, xCGE-LIF after 8-aminopyrene-1,3,6-trisulfonic acid (APTS) labeling, and MALDI-TOF-MS after ethyl esterification of the sialic acids (**Table 2**). The HILIC-UHPLC-FLD analysis was performed by three independent laboratories applying their standard protocols (**Supplemental Methods**).

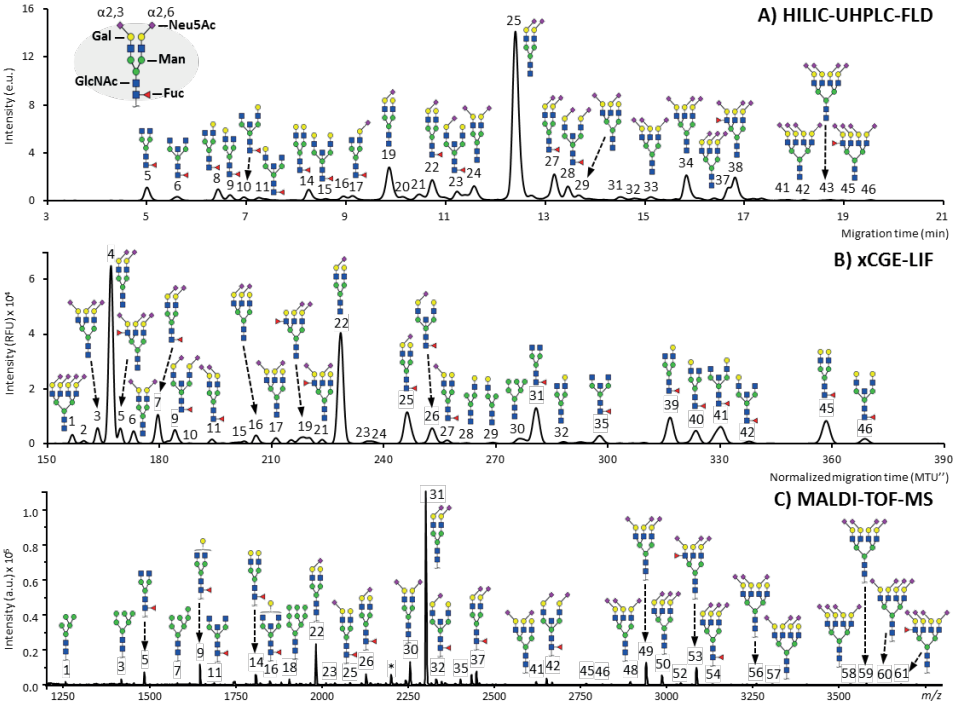


***Figure 1**. The respective profiles of the same total plasma protein N-glycan standard as recorded by HILIC-UHPLC-FLD, xCGE-LIF and MALDI-TOF-MS. **A)** Chromatogram as obtained by HILIC-UHPLC-FLD after 2-AB labeling. **B)** Electropherogram as obtained by xCGE-LIF after APTS labeling. **C)** Mass spectrum as obtained by MALDI-TOF-MS after ethyl esterification, with species assigned as $[M+Na]^+$. Signals of all recordings have been annotated to the best of knowledge, making use of the detections across the methods as well as established literature on biochemical pathways and plasma/serum N-glycosylation. The display of linkage has been restricted to the N-acetylneuraminic acids (sialic acids), which was principally acquired by MALDI-TOF-MS. Branching differences (galactose arm, bisection, fucose position) were only distinguishable by HILIC-UHPLC-FLD and xCGE-LIF. For full assignments of the signals see **Supplementary Tables S1**, **S2**, and **S3**, as well as **Supplementary Figure S1**.*

Signal detection

HILIC-UHPLC-FLD allowed the integration of 46 signals, with peak identities inferred by matching standardized retention times *via* GU values to database entries (**Supplemental Table S1**)[30]. xCGE-LIF enabled the integration of 49 signals, and peak identity was inferred by matching migration times *via* database entries (**Supplemental Table S2**)[31]. For MALDI-TOF-MS, 61 signals were detected that passed the established quality criteria (**Supplemental Figure S1**; **Supplemental Table S3**). The analyses broadly detected the same *N*-glycan species, with some variation per method in structural and compositional overlap (**Supplemental Table S4**), and the assignments were in line with previously established serum and plasma *N*-glycosylation features[32-34].

All analytical methods showed comparable results in their representation of the overall serum and plasma profile (**Figure 1**). Clearly defined for all methods were the two largest signals belonging to A2G2S[6,6]2 (H5N4E2, for HILIC-UHPLC-FLD, xCGE-LIF and MALDI-TOF-MS peaks 25, 4, and 31, respectively) and A2G2S[6]1 (H5N4E1, peaks 19, 22, and 22) (**Supplemental Table S4**). In addition, all methods demonstrated the high-definition detection of asialylated diantennary glycans, *e.g.*, FA2 (H3N4F1, peaks 5, 31, 5), and FA2G2 (H5N4F1, peaks 14, 45, 14). However, the non-MS methods allowed distinction between the galactose-linkage sites of the monogalactosylated species, *e.g.*, FA2[3]G1 (peaks 8, 39, -) and FA2[6]G1 (peaks 9, 40, -), whereas MALDI-TOF-MS could only detect the common composition belonging to these structures, namely H4N4F1 (peak 9). Similarly, while the presence of a bisecting *N*-acetylglycosamine led to unique retention/migration times, the corresponding mass spectrometric compositions also apply to triantennary structures with incomplete galactosylation. For example, the MALDI-TOF-MS composition H3N5F1 could align to FA2B (peaks 6, 35, 11) or FA3 (not detected), although these are not commonly found in high abundance in human serum[24,28]. On the other hand, mass spectrometry proved capable of individually detecting different sizes of high-mannose and hybrid-type *N*-glycans, whereas these species overlapped with diantennary glycans in the non-MS methods (**Supplemental Table S4**).

The two main species belonging to the triantennary structures were also detected by all analyses, *i.e.*, A3G3S[3,6,6]3 (H6N5L1E2, peaks 34, 3, 49) and A3F1GS[3,6,6]3 (H6N5F1L1E2, peaks 38, 5, 53). However, while the specific sialic acid linkage variants of the triantennary glycan compositions could be discriminated by MALDI-TOF-MS, *e.g.*, H6N5L3, H6N5L2E1, H6N5L1E2 and H6N5E3, the accompanying structures could not clearly be assigned to individual signals for HILIC-UHPLC-FLD and xCGE-LIF without additional steps of linkage-specific enzymatic removal of the sialic acids and sample re-measurement. This situation was similar for the tetraantennary compositions.

Based on the structural information and separation achieved by the various methods, we constructed a series of derived traits to describe single glyco-enzymatic steps and presumed protein-specific glycosylation patterns (**Supplemental Tables S5** and **S6**)[28,35].

Method repeatability

All methods showed robust detection of the plasma standard analyzed in 10 replicates, with UHPLC generally displaying the least variation and MALDI-TOF-MS the most (**Supplemental Figure S2** and **Supplemental Table S7**). Specifically, the main peak (A2G2S[6,6]2 or H5N4E2) showed for HILIC-UHPLC-FLD a coefficient of variation (CV) of 1.9% (participant 2, which displayed the lowest variation among the UHPLC methods), for xCGE-LIF of 5.1% and for MALDI-TOF-MS of 5.8%. Further examples include the CVs of FA2 (H3N4F1) respectively being 1.7%, 13.8%, and 17.7%, and of A3G3S[3,6,6]3 being 0.9%, 10.3%, and 13.2%. Overall, the average CVs of the 10 most abundant peaks for each method were 1.6%, 6.9% and 11.5%. For MALDI-TOF-MS most derived traits showed a higher repeatability than the individual peaks making up the traits, exemplified by the 50 most abundant traits having a mean CV of 2.6%, while for the non-MS methods the CVs of derived traits remained similar to their constituent peaks.

Intermethod signal correlation

Using the clinical cohort data, we could explore which signals displayed similar behavior across the orthogonal methods. To achieve this, Pearson correlation coefficients were calculated between the methods for all single signals (**Figure 2**), as well as for the derived traits (**Supplemental Figures S3, S4, S5** and **S6**). Note that the correlations obtained in this manner are the result of similarity of *N*-glycan behavior across pregnancy, RA and disease activity thereof, as well as of other biological and technical sources of variation that were unaccounted for in this study. As such, positive correlation within and between methods occurred when signals contained the same glycan structure (single isolated hit), or when different glycans underwent the same enzymatic modification (multiple hits sharing a single property such as antennary fucosylation). Negative correlation could similarly arise due the relationship between enzymatic substrates and products (*e.g.*, the process of galactosylation induces negative correlation between galactosylated and nongalactosylated species) and protein-abundance changes (*e.g.*, an increase in all diantennary species such as present on immunoglobulins will lead to a relative decrease in all tri- and tetraantennary glycans).

Between the UHPLC methods a good correlation was observed for peaks containing the same glycans, examples being peak 5 (FA2, mean r = 0.90 SD ± 0.07), peak 27 (FA2G2S[6,6]2, r = 0.93 ± 0.01) and peak 34 (A3G3S[3,6,6]3, r = 0.90 ± 0.02) (**Supplemental Figure S4A-C**). Signals with lower correlation across the methods were generally of low intensity. Strong correlation of the aforementioned signals was also visible with the corresponding xCGE-LIF annotation, *i.e.*, UHPLC peak 5 with xCGE-LIF peak 31 (FA2; r = 0.96), 27 with 7

(FA2G2S[6,6]2, r = 0.87), and 34 with 3 (A3G3S[3,6,6]3, r = 0.88) (**Supplemental Figure S4D**). Next to single signals, the derived traits showed to be highly comparable between methods as well (**Supplemental Figure S5**).

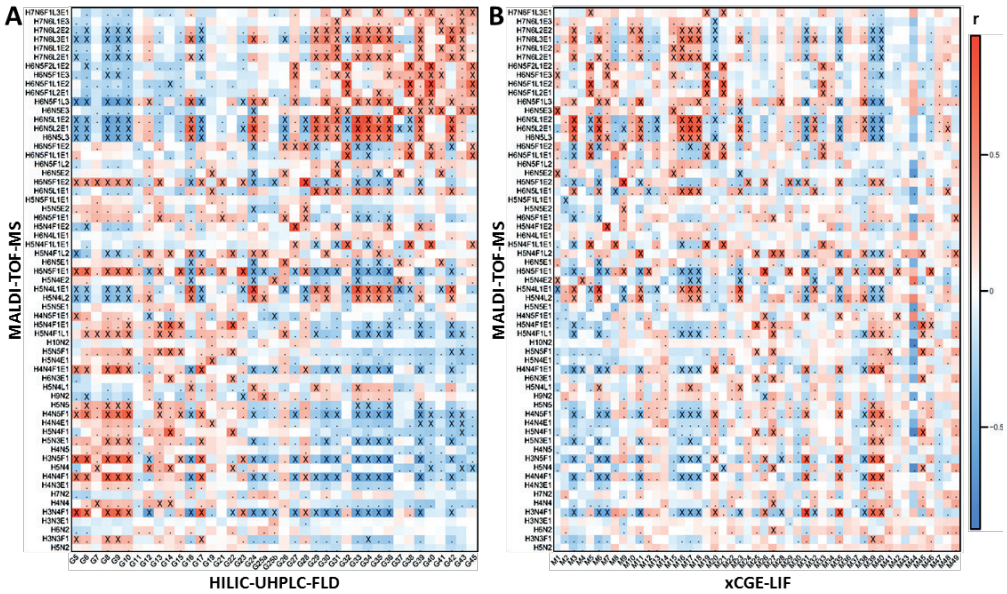


***Figure 2**. Heat map visualizing the Pearson correlation between signals from MALDI-TOF-MS, HILIC-UHPLC-FLD (participant 2) and xCGE-LIF. **A)** HILIC-UHPLC-FLD (horizontal) with MALDI-TOF-MS (vertical). **B)** xCGE-LIF (horizontal) with MALDI-TOF-MS (vertical). Signal correlations were calculated on clinical data, and represent similarity of behavior with biological phenotypes such as pregnancy and RA as well as technical variation. Crosses (X) indicate correlations significant below a p-value of $1 \cdot 10^{-5}$, whereas dots (.) indicate correlation below a p-value of 0.05. For similar heat maps between non-MS methods and derived traits see **Supplementary Figures S4**, **S5**, and **S6**. H = hexose, N = N-acetylhexosamine, F = deoxyhexose (fucose), L = (lactonized) α2,3-linked N-acetylneuraminic acid, E = (ethyl esterified) α2,6-linked N-acetylneuraminic acid.*

Complementarity of methods for signal assignment

When comparing the signals of the non-MS methods with MALDI-TOF-MS, an advantage of analyzing the sample set with two vastly orthogonal methods becomes apparent (**Figure 2**). For instance, in several cases ambiguously annotated non-MS signals could be attributed to specific glycan structures by making use of the correlation with MS. One example hereof is chromatographic peak 28, which, while theoretically encompassing FA2BG2S[6,6]2, FA2BG2S[3,6]2 and FA2BG2S[3,3]2, strongly correlated with MALDI-TOF-MS composition H5N5F1E2 (FA2BG2S[6,6]2, r = 0.92), and to a much lesser degree with H5N5F1L1E1 (FA2BG2S[3,6]2, r = 0.27) and H5N5F1L2 (FA2BG2S[3,3]2, not detected) (**Figure 2A, Supplemental Figure S3**). Similarly, by correlation xCGE-LIF peak 1 is likely to contain H6N5E3 (A3G3S[6,6,6]3, r = 0.68), even though this was not principally annotated for the

electrophoretic signal (**Figure 2B**). On the other hand, structural characteristics could be attributed to MALDI-TOF-MS compositions on basis of the assignments from the non-MS methods. For instance, the fucosylated triantennary compositions with at least one α2,3-linked sialic acid correlated strongly with antennary-fucosylated but not with core-fucosylated structures, *e.g.*, H6N5F1L1E2 with UHPLC peak 38 and xCGE-LIF peak 5 (A3F1G3S[3,6,6]3, respectively r = 0.86 and r = 0.82), and not with UHPLC peak 36 (FA3G3S[3,6,6]3, r = 0.10). While core-fucosylated structures are still likely present in the MALDI-TOF-MS composition, it does appear that the main differences observed within the cohort originated instead from antennary fucosylation.

Next to single glycans, derived traits showed good overlap between MS and non-MS methods, but a larger set of these could be constructed for MS due to the unambiguous compositional assignment of signals (**Supplemental Figure S6**).

Association with pregnancy and RA

Within the PARA study, we compared for the different methods the glycosylation changes observed throughout pregnancy (assessed at preconception, the 3rd trimester of pregnancy, and 26+ weeks postpartum), between healthy controls and RA patients, as well as the association with RA disease activity as expressed by the DAS28(3)-CRP value. The comparability between methods was assessed by the effect directions of the significant findings from mixed model regression analyses, and are represented as box- and scatterplots (**Figure 3**). For all analyses, glycosylation parameters were centered to zero and scaled to represent single SD changes. Effects were deemed significant under a study-wide false-discovery rate of 5% (leading to a threshold $\alpha = 1.7 \cdot 10^{-2}$).

Pregnancy showed to have a major effect on the TSNG, and congruent effects were observed by all methods (**Supplemental Table S8**), both for the single glycans (**Figure 3A**) and derived traits (**Figure 3B**). Notable examples included the decrease of FA2 with pregnancy, as seen for HILIC-UHPLC-FLD (participant 2 throughout) peak 5 (β = -2.32 SE ± 0.38), xCGE-LIF peak 31 (β = -2.37 ± 0.39) and MALDI-TOF-MS composition H3N4F1 (β = -2.34 ± 0.43), and the consequentially increased overall galactosylation of (nonsialylated) diantennary fucosylated species representative of IgG-Fc (A2FS0G, respectively, β = 1.62 ± 0.27, β = 1.92 ± 0.31, and β = 1.67 ± 0.28)[28]. Likewise, all methods indicated a profound decrease in bisection of sialylated diantennary species, likely representative of non-IgG-Fc immunoglobulin glycosylation (A2FSB, β = -3.10 ± 0.48, β = -1.55 ± 0.26, and β = -1.81 ± 0.29)[28].

Interestingly, one of the major *N*-glycan species to change with pregnancy, namely the fully α2,3-sialylated A3F1G3S[3,3,3]3, was only uniquely separable as the MALDI-TOF-MS composition H6N5F1L3 (β = 1.61 ± 0.26), but likely also drove the changes in the ambiguously assigned UHPLC signal 33 (β = 2.92 ± 0.43) and xCGE-LIF signal 3 (β = 1.31 ±

0.24). However, due to overlap in retention and migration times, the derived trait for α2,3-linked sialylation on triantennary species (A3GL) could only be established for MALDI-TOF-MS (β = 2.57 ± 0.40) (**Figure 3C**). HILIC-UHPLC-FLD, on the other hand, uniquely showed with pregnancy a strong difference in the galactose position of monogalactosylated and bisected diantennary *N*-glycan species (*e.g.*, relative 6-arm galactosylation, FA2[r6]BG1, β = -9.41 ± 2.42), whereas xCGE-LIF could uniquely detect a change in GlcNAc position of monoantennary species (*e.g.*, relative 6-arm GlcNAc position, A1[r6]G1, β = 1.96 ± 0.31).

Differences between RA patients and healthy controls proved less pronounced than the differences during pregnancy, but were similarly detected across methods. Mixed logistic regression was used for the comparison of glycosylation parameters with RA (healthy = 0, RA = 1), correcting for pregnancy by modelling a random intercept per time point (**Supplemental Table S9**). RA patients showed to have a lower galactosylation of glycans commonly found on IgG-Fc (A2FS0G, for HILIC-UHPLC-FLD β = -1.89 ± 0.36, for xCGE-LIF β = -1.85 ± 0.36, and for MALDI-TOF-MS β = -1.78 ± 0.36), and higher bisection of sialylated fucosylated diantennary species (A2FSB, β = 1.25 ± 0.31, β = 0.53 ± 0.21, and β = 0.88 ± 0.26) (**Figure 3B**). Also detected by HILIC-UHPLC-FLD and MALDI-TOF-MS was a higher bisection with RA of afucosylated nonsialylated diantennary species (A2F0S0B, respectively β = 1.11 ± 0.28, and β = 0.69 ± 0.20). Absent from MALDI-TOF-MS but present for the non-MS methods were differences in branching for lower complexity *N*-glycans, namely for HILIC-UHPLC-FLD the relative 6-branch galactosylation of monogalactosylated and sialylated diantennary species with bisection (FA2[r6]BG1S1, β = -0.90 ± 0.30) and for xCGE-LIF the antennary-branch of monoantennary species (A1[r6]G1, β = -1.05 ± 0.28) (**Figure 3C**).

Lastly, linear regression was used to model the association of glycosylation (independent) with DAS28(3)-CRP (dependent). In addition, mixed linear regression was used with either a random intercept per individual to assess the association throughout pregnancy, or with a random intercept per time point to assess the association between individuals. These different models provided highly congruent findings (**Supplemental Table S10**). In all, for each method a strong negative association was seen between DAS28(3)-CRP and IgG-Fc-type galactosylation (A2FS0G, HILIC-UHPLC-FLD β = -0.51 ± 0.11, xCGE-LIF β = -0.51 ± 0.11, MALDI-TOF-MS β = -0.50 ± 0.11) (**Figure 3D**). Increased with disease activity proved to be the fucosylation of triantennary structures (A3F, β = 0.50 ± 0.11, β = 0.33 ± 0.12, and β = 0.38 ± 0.12), as well as the sialylation thereof (A3S, β = 0.52 ± 0.11, A3FS β = 0.23 ± 0.12 (trend), and A3FGS, β = 0.41 ± 0.12). MALDI-TOF-MS analysis specifically showed that the increases in fucosylation and sialylation were similar for the tri- and tetraantennary species (*e.g.*, A4FE, β = 0.42 ± 0.12), and revealed that there were no sialic acid linkage-biases in these increases (*e.g.*, A3EF, β = 0.39 ± 0.12 *vs.* A3LF, β = 0.40 ± 0.12). HILIC-UHPLC-FLD, on the other hand, was again able to detect significant branching differences with changing

disease activity (FA2[r6]G1S1, β = -0.34 ± 0.12), while xCGE-LIF showed higher resolution for the monoantennary species (*e.g.*, A1F, β = 0.48 ± 0.14).

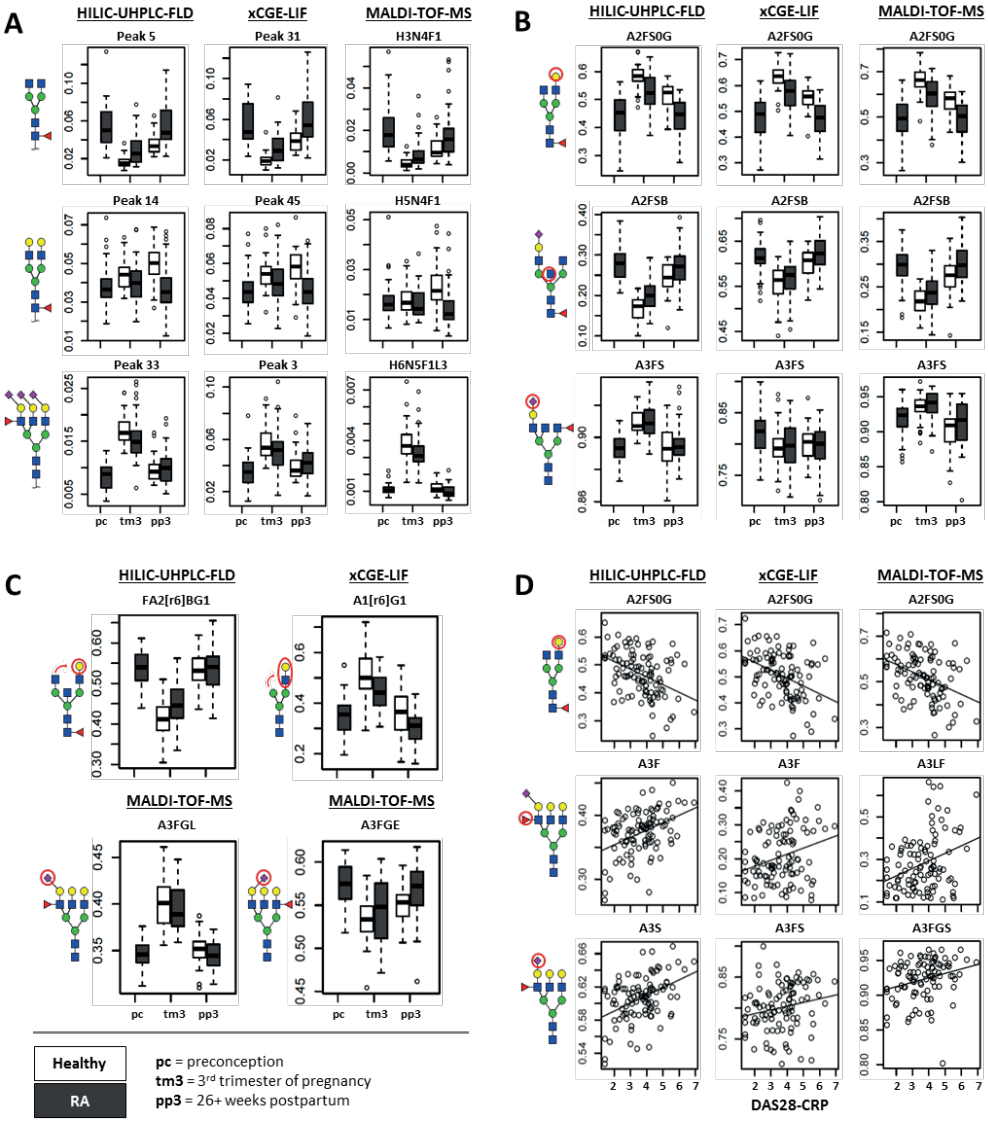


***Figure 3**. Comparability of HILIC-UHPLC-FLD (participant 2) (left), xCGE-LIF (middle) and MALDI-TOF-MS (right) for the detection of clinical characteristics of pregnancy, RA and RA disease activity. **A)** Comparability of single glycan signals (% area) with pregnancy (preconception = pc, 3rd trimester of pregnancy = tm3, 26+ weeks postpartum = pp3) and RA (healthy = white, RA = grey). **B)** Comparability of derived glycosylation traits with pregnancy and RA. **C)** Derived trait differences detected uniquely by a single method with pregnancy and RA. **D)** Association of derived glycosylation traits with RA disease activity (DAS28-CRP). For a legend of the derived traits see **Supplementary Table S6**.*

8.4 Discussion

Previous method comparisons involved the analysis of released *N*-glycans and glycopeptides of single proteins, examples being IgG and prostate-specific antigen (PSA)[25-27,36]. In our comparative study we have instead focused on the analysis of released *N*-glycans from the total pool of serum glycoproteins, which principally allows the detection of a larger variety of systemic abnormalities than would be possible from a single protein, but also displays *N*-glycosylation of considerably higher complexity[24,28,37].

We included in our comparison analytical methodologies that have displayed the throughput capacity for several thousands of serum *N*-glycome samples, as currently the case for HILIC-UHPLC-FLD, xCGE-LIF and MALDI-(TOF-)MS[11,19,20,22]. Technical considerations for these studies include analyte stability, preparation- and measurement-throughput, as well as software solutions for the integration and analysis of the consequentially large datasets, all of which have been demonstrably addressed for the indicated techniques. To appreciate the technical possibilities and constraints of the methods, their mechanisms of separation need to be considered. 1) HILIC-UHPLC-FLD separates glycan structures on basis of their hydrophilic interaction with a stationary phase, generally meaning that larger structures, as well as those with larger surface areas or charged *N*-acetylneuraminic acids, have increased retention times[38]. 2) In xCGE-LIF glycans are separated along an electric field inside a polymer-filled capillary, according to their mass/charge (*m/z*) ratio and their size/shape (hydrodynamic diameter). Analytes with lower *m/z* (higher charges and/or lower masses) migrate faster through the capillary than those with higher *m/z*[10,39]. 3) MALDI-TOF-MS separates ionized glycans (here $[M+Na]^+$) by *m/z* ratio, with the esterification procedure employed to stabilize sialic acids, to prevent unfavorable negative charges on the sialylated species and to introduce an MS-detectable mass difference between α2,3- (lactonized) and α2,6-linked (ethyl esterified) sialic acids[40].

In this comparison study, we judged the relative performance of the HTP TSNG methods on the basis of technical replicate measurements, as well as by a set of samples from a longitudinal study on the improvement of disease activity within RA patients during pregnancy[29].

Throughput and repeatability

Sample preparation throughput proved similar between the methods, each of them requiring (overnight) enzymatic *N*-glycan release, 1 – 2 h chemical derivatization at either the glycan reducing end or at the sialic acid, and HILIC solid-phase extraction (SPE) prior to analysis. Automated sample preparation was reported for both the HILIC-UHPLC-FLD and MALDI-TOF-MS workflows[35,41], but the congruencies in protocols suggest that xCGE-LIF could make use of similar strategies. Aside from sample preparation, MALDI-TOF-MS showed to have highest analytical throughput with an approximate throughput of 10 s per

sample, whereas xCGE-LIF and UHPLC runs required 40 min to an hour. A major advantage for xCGE-LIF is the multiplexing capability (with up to 96 capillaries in parallel) allowing the simultaneous analysis of up to 96 samples, which reduces the effective analysis time per sample to less than 30 seconds.

The most repeatable method proved to be HILIC-UHPLC-FLD (albeit with variation between participants), whereas MALDI-TOF-MS showed the most technical variation. Interestingly, although the lower repeatability of the MALDI-TOF-MS method did indeed lead to larger SDs on the biological effects as well as a consequential decrease in statistical significance (an increase in *p*-value) by typically one or more orders of magnitude, in practice very few findings were rejected in this study due to lack of statistical power. As such, the lower repeatability of MALDI-TOF-MS does not appear to hamper the glycomics analyses of larger cohorts, but the repeatability of the non-MS methods would definitely be of benefit for the quantification of small effect sizes.

Of interest, while MALDI-TOF-MS repeatability was higher for the derived traits, which often are groupings of chemically similar *N*-glycan species, no such improvement was observed for either HILIC-UHPLC-FLD or xCGE-LIF. This would suggest presence of an MS-specific component of measurement error, possibly relative ionization efficiency or response linearity, both of which could be controlled for with the use of internal standards[42,43].

Analyte separation and annotation

Whereas MALDI-TOF-MS does principally not separate all *N*-glycans but only those with differing chemical compositions, both HILIC-UHPLC-FLD and xCGE-LIF produce unique standardized retention times, respectively migration times, for glycans with different structures. However, while the MS analysis provided the resolution to separate the majority of possible compositions, for the non-MS methods a large portion of glycan structures were not separated from other analytes. Accordingly, only HILIC-UHPLC-FLD and xCGE-LIF were capable of revealing additional features in the low-complexity (IgG-Fc) regions of their chromatograms or electropherograms for the analyses of the TSNG, e.g. *N*-acetylglucosamine linkage (antennary or bisecting), galactose position (α1,3 or α1,6 branch), and fucose position (core or antennary). On the other hand, MALDI-TOF-MS proved the most informative for larger glycan structures (e.g. tri- and tetraantennary species), for instance on the exact number of LacNAc units (antennarity), the number of fucoses, and the number and linkage of *N*-acetylneuraminic acids. To generalize, in profiling mode HILIC-UHPLC-FLD and xCGE-LIF appeared optimal for high-density structural identification of a low-complexity sample, *e.g.*, immunoglobulin glycosylation, be it from blood or most of the commonly used recombinant production systems, whereas MALDI-TOF-MS appeared preferentially suitable for the analysis of samples with high-complexity and/or larger *N*-glycan species.

Each of the described methods may obtain additional structural information on the glycans by including more experimental dimensions. One can think of on-line hyphenation of separation techniques, examples including LC-MS(/MS), CE-MS(/MS), and ion-mobility MS[44-46], or off-line approaches such as exoglycosidase digestion[23]. While increasing the information content, these added dimensions also drastically increase analysis times and data complexity, and have thus far only been reported for relatively small sample sets. However, with the advancements in development of rapid glycan preparations protocols, laboratory automation and evolution of big data analysis methods, the idea of HTP laboratories employing these innovative approaches in glycan analysis of larger sample sets on different glycoprotein levels might not be too far way. Today, more commonly achieved for HTP applications is the thorough structural annotation of one or a few samples, with the expectation that these are representative for the set, as is for instance achieved with exoglycosidase digestion or MS/MS experiments[23,24]. Interestingly, our study suggests that parallel glycomics analysis by orthogonal HTP methods as well allows the determination of many of the sample-relevant structural characteristics, without necessarily including a step of throughput-limiting serial hyphenation. A prominent example of this is the structure FA2BG2S2 correlating with composition H5N5F1E2, for which HILIC-UHPLC-FLD and xCGE-LIF determined the core-fucosylation and bisection while MALDI-TOF-MS determined the α2,6-linkage of the *N*-acetylneuraminic acids (**Supplemental Figure S3B and C**).

Clinical observations

Associations found between *N*-glycosylation and pregnancy, RA and RA disease activity (DAS28(3)-CRP) were very much in line with previous TSNG studies[47-49](**Chapter 7**), and agreed with reports on single glycoproteins dominant in human serum[50-53]. Main findings with pregnancy included an increased antennarity, nonsialylated FA2 galactosylation, and α2,3-linked sialylation as well as a decreased bisection of sialylated FA2. Individuals with RA proved to have on average a lower FA2 galactosylation and higher bisection compared to their healthy counterparts, and disease activity of RA associated positively with A3/A4 for both fucosylation and sialylation (without particular preference in linkage), and negatively with the galactosylation of FA2.

As with all analyses of the released TSNG, the here-reported differences may originate from changes in protein glycosylation, or from differences in the relative abundances of glycoproteins in serum. Nonetheless, the changing glycosylation phenotypes seem to reflect immune modulation of either IgG-Fc (predominantly nonsialylated FA2)[28,51], IgG-Fab and other plasma-cell-derived immunoglobulins (highly sialylated FA2)[28,51], or acute phase glycoproteins such as alpha-1-antitrypsin and alpha-1-acid glycoprotein (tri- and tetraantennary species)[28,52,54,55]. In addition, the previously reported MALDI-TOF-MS association of A3FGS with DAS28(3)-CRP was reproduced (**Chapter 7**), although its protein of origin is as of yet unclear.

Interestingly, HILIC-UHPLC-FLD and xCGE-LIF allowed new findings on the glycosylation changes occurring with pregnancy and RA disease activity, in the form of branching differences on both monogalactosylated and monoantennary species. With pregnancy, for example, an increased galactosylation was detected of the α1,6-branch (as opposed to the α1,3-branch) of nonbisected monogalactosylated FA2, and a decreased α1,6-branched galactosylation of the bisected variant. The same phenotypes were inversely associated with RA disease activity, principally matching the DAS28(3)-CRP decrease observed with pregnancy[29]. These observations could be the result of altered glycosyltransferase expression or regulation, but one attractive alternative explanation might be a shift of relative IgG-subclass abundances during pregnancy[56]. As IgG2 and IgG3 display higher α1,3-branch galactosylation than α1,6-branch galactosylation, which is contrary to IgG1 and IgG4[57], the reported relative increase of IgG2 and IgG3 with pregnancy would lead to our observed linkage change in the TSNG[56,58].

Table 3. Overview of study results.

Methodology	Information	Advantages	Disadvantages	Sample suitability
HILIC-UHPLC-FLD	Structural	Best repeatability; well-established database; good separation of A2	A3/A4 signal overlap; no novel annotation[a]	Plasma cell glycoproteins or mixtures (immunoglobulins; A2 glycans)
xCGE-LIF	Structural	Throughput due to multiplexing; good separation of A2	A3/A4 signal overlap; no novel annotation[a]	Plasma cell glycoproteins or mixtures (immunoglobulins; A2 glycans)
MALDI-TOF-MS	Compositional; sialic acid linkage	Best throughput; compositional annotation; infrequent signal overlap; good separation of A3/A4; derived traits	Lowest repeatability; no structural information[a]	Hepatic cell glycoproteins (acute phase proteins; A3/A4 glycans); unknown samples

[a]Additional information may be obtained by subsequent experiments such as exoglycosidase sequencing and tandem MS.

8.5 Summary

In summary, we compared the performance of HILIC-UHPLC-FLD, xCGE-LIF and MALDI-TOF-MS for the analysis of the released serum protein *N*-glycome. Next to providing the technical and biological translation between methods, we discussed their advantages and disadvantages, including their respective throughput and repeatability (**Table 3**). In addition, we explored the differences in information content for various glycosylation types within the TSNG, and speculated upon the suitability of the methods to characterize different sample types. The combined analysis with orthogonal HTP methods proved to be highly informative for the study of the TSNG, and has led, next to confirming previous

findings, to the discovery of new glycosylation traits associated with RA and disease activity thereof.

8.6 Methods

Study population

The research presented here was performed using serum samples from the PARA study, a prospective cohort to study the interaction of pregnancy and RA[29]. In the current research, serum samples were included from 36 RA patient pregnancies, obtained prior to conception, at the 3rd trimester of pregnancy, and after 26 weeks following partum. At every time point, disease activity was assessed using the 28-joint disease activity score (DAS) with three variables based on the C-reactive protein (CRP) level (mg/L) (DAS28(3)-CRP). In addition, serum samples from 32 apparently healthy pregnancies (without adverse obstetric histories) were included to serve as controls, for these including serum collected at the 3rd trimester of pregnancy and later than 26 weeks post-partum. All pregnancies were completed, and all patients fulfilled the 1987 ACR criteria for RA. The study was in compliance with the Helsinki Declaration and was approved by the Ethics Review Board at the Erasmus University Medical Center, Rotterdam, The Netherlands.

Of each of the 178 clinical samples, 200 µL serum was distributed in randomized order across two 96-well deep-well plates (polypropylene; NUNC, Rochester, NY). Per plate, an additional 2 positions were filled with phosphate-buffered saline solution to serve as blank, 5 positions with identical plasma as technical standard (Visucon-F frozen normal control plasma; Affinity Biologicals, Ancaster, ON), and 3 positions were left to facilitate the inclusion of standards local to each laboratory. From these master plates, 40 µL sample was divided into 5 pairs of PCR plates (polypropylene; Greiner Bio-One, Frickenhausen, Germany), which were distributed among the participating laboratories.

HILIC-UHPLC-FLD analysis

Sample preparation and measurement by HILIC-UHPLC-FLD was performed by participants 1, 2 and 3, as previously reported[23,41,59,60]. The procedures are described in full detail in the supplementary information (respectively **Supplemental Method M1**, **M2**, and **M3**). To summarize, *N*-glycans were released enzymatically from their protein backbones by overnight peptide-*N*-glycosidase F (PNGase F) treatment, labeled with 2-AB by reductive amination, enriched by HILIC solid phase extraction (SPE), and subsequently analyzed by HILIC-UHPLC-FLD. For peak annotation, the sample retention times were calibrated on an external UHPLC run of dextran ladder. The hereby obtained GU values per signal were used to connect to a database of previously established assignments[30]. Signal numbers were unified between the participants on basis of the structural annotation, principally based on the nomenclature of participant 1.

xCGE-LIF analysis

xCGE-LIF sample preparation and measurement were performed as previously described (**Supplemental Method M4**)[10,31,61]. Briefly, *N*-glycans were released from the serum proteins by PNGase F, fluorescently labeled by reductive amination with APTS, enriched by HILIC-SPE, and analyzed by a multiplexed capillary gel electrophoresis system with laser-induced fluorescence detection. Each sample was internally calibrated by a co-migrating fluorescent standard, and the resulting migration times were annotated with glycan structures on basis of established database values[31]. Automated migration time normalization, peak picking, integration and database matching were performed by glyXtool[31].

MALDI-TOF-MS analysis

MALDI-TOF-MS sample preparation and measurement were performed as described previously (**Supplemental Method M5**)[35,40]. In short, after *N*-glycan release by PNGase F, an automated platform was employed to derivatize the sialic acids by ethyl esterification (of α2,6-linked sialic acids and lactonization of α2,3-linked sialic acids), to perform GHP HILIC-SPE, and to spot samples on a MALDI target[35]. MALDI-TOF-MS analysis was performed in reflectron positive mode, with an accumulation of 10,000 shots per spot in a random walking pattern. Obtained signals were annotated to be $[M+Na]^+$ glycan compositions on basis of the signal-to-noise ratio, the ppm error, and the isotopic ratio.

Data preprocessing and analysis

For each of the methods, signal areas were normalized to the total sum of area per sample (total area normalization). Derived traits were calculated on basis of known enzymatic glycosylation pathways and glycoprotein populations (for derived trait calculations see **Supplemental Table S5**, for a legend describing the derived traits see **Supplemental Table S6**)[24,28,32-34,37].

Throughout data analysis, we made use of R 3.1.2 in RStudio 0.98.1091 (RStudio Team, Boston, MA)[62]. The repeatability of each method was assessed by calculating the mean, SD and CV for each signal within the technical control samples (**Supplemental Table S7**). Correlation between glycosylation features in the clinical data was established by calculating the Pearson correlation, and the results hereof were expressed in heat map format (**Figure 2**, **Supplemental Figure S3-S5**). Boxplots and scatterplots were similarly created in R using the total-area-normalized glycosylation values.

For the association analyses with pregnancy, RA and DAS28(3)-CRP, glycosylation (and derived trait) averages were centered to 0, and scaled to represent single SD variations. Pregnancy was defined as binary variable (both preconception and 26+ weeks postpartum = 0; 3rd trimester of pregnancy = 1), as was RA (0 = control, 1 = RA). Mixed logistic regression was employed to test the association of glycosylation (independent) with pregnancy

(dependent), correcting for interindividual effects by assigning a random intercept per individual (**Supplemental Table S8**). Mixed logistic regression was additionally used to test the association between glycosylation (independent) and RA (dependent), correcting for pregnancy by modelling a random intercept per time point (**Supplemental Table S9**). Linear regression was used to test the direct association between glycosylation (independent) and the linear variable DAS28-CRP (dependent), while mixed linear regression was used to differentiate the intraindividual effects (modeling a random intercept per individual) or interindividual effects (modeling a random intercept per time point) (**Supplemental Table S10**).

The study-wide false-discovery rate was controlled to be 5% by application of the Benjamini-Hochberg procedure[63], leading to an overall significance threshold of $\alpha = 1.7 \cdot 10^{-2}$ (**Supplemental Table S11**).

Nomenclature

In text, *N*-glycan structures have followed the Oxford nomenclature[23], whereas *N*-glycan compositions have followed H = hexose, N = *N*-acetylhexosamine, F = fucose (deoxyhexose), S = unspecified *N*-acetylneuraminic acid (sialic acid), L = (lactonized) α2,3-linked *N*-acetylneuraminic acid, and E = (ethyl esterified) α2,6-linked *N*-acetylneuraminic acid. Figures have been annotated with glycan cartoons following the recommendations of the Consortium for Functional Glycomics[64] and designed using GlycoWorkbench 2.1 (build 146)[65].

Acknowledgments

This work was supported by the European Union Seventh Framework Programme project HighGlycan (278535) and the Dutch Arthritis Foundation (NR-10-1-411). Furthermore, we would like to thank Gerda Vreeker, Marco Bladergroen and Yuri van der Burgt for their assistance with measuring the MALDI-TOF-MS data.

8.7 References

1 Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3-49, doi:10.1093/glycob/cww086 (2017).

2 Hart, G. W. & Copeland, R. J. Glycomics hits the big time. *Cell* **143**, 672-676, doi:10.1016/j.cell.2010.11.008 (2010).

3 Thomann, M. *et al.* In vitro glycoengineering of IgG1 and its effect on Fc receptor binding and ADCC activity. *PLoS One* **10**, e0134949, doi:10.1371/journal.pone.0134949 (2015).

4 Rouwendal, G. J. *et al.* A comparison of anti-HER2 IgA and IgG1 in vivo efficacy is facilitated by high N-glycan sialylation of the IgA. *MAbs* **8**, 74-86, doi:10.1080/19420862.2015.1102812 (2016).

5 Oh, M. J. *et al.* Analytical platform for glycomic characterization of recombinant erythropoietin biotherapeutics and biosimilars by MS. *Bioanalysis* **5**, 545-559, doi:10.4155/bio.12.327 (2013).

6 Pancera, M. *et al.* Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. *Nat Struct Mol Biol* **20**, 804-813, doi:10.1038/nsmb.2600 (2013).
7 von Itzstein, M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418-423, doi:10.1038/363418a0 (1993).
8 Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N. & Dwek, R. A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* **343**, 1235681, doi:10.1126/science.1235681 (2014).
9 Gornik, O. *et al.* Stability of N-glycan profiles in human plasma. *Glycobiology* **19**, 1547-1553, doi:10.1093/glycob/cwp134 (2009).
10 Hennig, R. *et al.* Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. *Biochim Biophys Acta* **1860**, 1728-1738, doi:10.1016/j.bbagen.2016.03.035 (2016).
11 Reiding, K. R. *et al.* Human Plasma N-glycosylation as Analyzed by Matrix-Assisted Laser Desorption/Ionization-Fourier Transform Ion Cyclotron Resonance-MS Associates with Markers of Inflammation and Metabolic Health. *Mol Cell Proteomics* **16**, 228-242, doi:10.1074/mcp.M116.065250 (2017).
12 Albrecht, S., Unwin, L., Muniyappa, M. & Rudd, P. M. Glycosylation as a marker for inflammatory arthritis. *Cancer Biomark* **14**, 17-28, doi:10.3233/CBM-130373 (2014).
13 Vuckovic, F. *et al.* IgG Glycome in Colorectal Cancer. *Clin Cancer Res* **22**, 3078-3086, doi:10.1158/1078-0432.CCR-15-1867 (2016).
14 Kemna, M. J. *et al.* Galactosylation and Sialylation Levels of IgG Predict Relapse in Patients With PR3-ANCA Associated Vasculitis. *EBioMedicine* **17**, 108-118, doi:10.1016/j.ebiom.2017.01.033 (2017).
15 Novokmet, M. *et al.* Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* **4**, 4347, doi:10.1038/srep04347 (2014).
16 Verhelst, X. *et al.* A Glycomics-Based Test Predicts the Development of Hepatocellular Carcinoma in Cirrhosis. *Clin Cancer Res* **23**, 2750-2758, doi:10.1158/1078-0432.CCR-16-1500 (2017).
17 Shubhakar, A. *et al.* High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia* **78**, 321-333, doi:10.1007/s10337-014-2803-9 (2015).
18 Gaunitz, S., Nagy, G., Pohl, N. L. & Novotny, M. V. Recent Advances in the Analysis of Complex Glycoproteins. *Anal Chem* **89**, 389-413, doi:10.1021/acs.analchem.6b04343 (2017).
19 Lauc, G. *et al.* Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. *PLoS Genet* **6**, e1001256, doi:10.1371/journal.pgen.1001256 (2010).
20 Ruhaak, L. R. *et al.* Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* **10**, 1667-1674, doi:10.1021/pr1009959 (2011).
21 Lauc, G. *et al.* Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet* **9**, e1003225, doi:10.1371/journal.pgen.1003225 (2013).
22 Ruhaak, L. R. *et al.* Targeted biomarker discovery by high throughput glycosylation profiling of human plasma alpha1-antitrypsin and immunoglobulin A. *PLoS One* **8**, e73082, doi:10.1371/journal.pone.0073082 (2013).
23 Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **13**, 2314-2327, doi:10.1021/pr401092y (2014).
24 Stumpo, K. A. & Reinhold, V. N. The N-glycome of human plasma. *J Proteome Res* **9**, 4823-4830, doi:10.1021/pr100528k (2010).

25 Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).
26 Reusch, D. *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles--part 1: separation-based methods. *MAbs* **7**, 167-179, doi:10.4161/19420862.2014.986000 (2015).
27 Reusch, D. *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods. *MAbs* **7**, 732-742, doi:10.1080/19420862.2015.1045173 (2015).
28 Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-343, doi:10.1007/s10719-015-9626-2 (2016).
29 de Man, Y. A., Dolhain, R. J., van de Geijn, F. E., Willemsen, S. P. & Hazes, J. M. Disease activity of rheumatoid arthritis during pregnancy: results from a nationwide prospective study. *Arthritis Rheum* **59**, 1241-1248, doi:10.1002/art.24003 (2008).
30 Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A. & Rudd, P. M. GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics* **24**, 1214-1216, doi:10.1093/bioinformatics/btn090 (2008).
31 Hennig, R., Reichl, U. & Rapp, E. A software tool for automated high-throughput processing of CGE-LIF based glycoanalysis data, generated by a multiplexing capillary DNA sequencer. *Glycoconj J* **28**, 331 (2011).
32 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).
33 Freeze, H. H. Genetic defects in the human glycome. *Nat Rev Genet* **7**, 537-551, doi:10.1038/nrg1894 (2006).
34 Kang, P., Mechref, Y. & Novotny, M. V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun Mass Spectrom* **22**, 721-734, doi:10.1002/rcm.3395 (2008).
35 Bladergroen, M. R. *et al.* Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *J Proteome Res* **14**, 4080-4086, doi:10.1021/acs.jproteome.5b00538 (2015).
36 Leymarie, N. *et al.* Interlaboratory study on differential analysis of protein glycosylation by mass spectrometry: the ABRF glycoprotein research multi-institutional study 2012. *Mol Cell Proteomics* **12**, 2935-2951, doi:10.1074/mcp.M113.030643 (2013).
37 Klein, A. Human total serum N-glycome. *Adv Clin Chem* **46**, 51-85 (2008).
38 Hemstrom, P. & Irgum, K. Hydrophilic interaction chromatography. *J Sep Sci* **29**, 1784-1821 (2006).
39 Callewaert, N., Geysens, S., Molemans, F. & Contreras, R. Ultrasensitive profiling and sequencing of N-linked oligosaccharides using standard DNA-sequencing equipment. *Glycobiology* **11**, 275-281 (2001).
40 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).
41 Stockmann, H., O'Flaherty, R., Adamczyk, B., Saldova, R. & Rudd, P. M. Automated, high-throughput serum glycoprofiling platform. *Integr Biol (Camb)* **7**, 1026-1032, doi:10.1039/c5ib00130g (2015).
42 Gong, B. *et al.* N-glycosylamine-mediated isotope labeling for mass spectrometry-based quantitative analysis of N-linked glycans. *Anal Bioanal Chem* **405**, 5825-5831, doi:10.1007/s00216-013-6988-9 (2013).
43 Alvarez-Manilla, G. *et al.* Tools for glycomics: relative quantitation of glycans by isotopic permethylation using 13CH3I. *Glycobiology* **17**, 677-687, doi:10.1093/glycob/cwm033 (2007).

44 Zhou, S., Dong, X., Veillon, L., Huang, Y. & Mechref, Y. LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Anal Bioanal Chem* **409**, 453-466, doi:10.1007/s00216-016-9996-8 (2017).

45 Zhong, X. *et al.* Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry for Quantitative Analysis of Glycans Labeled with Multiplex Carbonyl-Reactive Tandem Mass Tags. *Anal Chem* **87**, 6527-6534, doi:10.1021/acs.analchem.5b01835 (2015).

46 Guttman, M. & Lee, K. K. Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry. *Anal Chem* **88**, 5212-5217, doi:10.1021/acs.analchem.6b00265 (2016).

47 Jansen, B. C. *et al.* Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep* **6**, 23296, doi:10.1038/srep23296 (2016).

48 Ruhaak, L. R., Uh, H. W., Deelder, A. M., Dolhain, R. E. & Wuhrer, M. Total plasma N-glycome changes during pregnancy. *J Proteome Res* **13**, 1657-1668, doi:10.1021/pr401128j (2014).

49 Parekh, R. B. *et al.* Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452-457 (1985).

50 Bondt, A. *et al.* Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res* **12**, 4522-4531, doi:10.1021/pr400589m (2013).

51 Bondt, A. *et al.* Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics* **13**, 3029-3039, doi:10.1074/mcp.M114.039537 (2014).

52 Brinkman-van der Linden, E. C., de Haan, P. F., Havenaar, E. C. & van Dijk, W. Inflammation-induced expression of sialyl LewisX is not restricted to alpha1-acid glycoprotein but also occurs to a lesser extent on alpha1-antichymotrypsin and haptoglobin. *Glycoconj J* **15**, 177-182 (1998).

53 Havenaar, E. C. *et al.* Severe rheumatoid arthritis prohibits the pregnancy-induced decrease in alpha3-fucosylation of alpha1-acid glycoprotein. *Glycoconj J* **15**, 723-729 (1998).

54 Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. & Altmann, F. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics* **6**, 3369-3380, doi:10.1002/pmic.200500751 (2006).

55 Lee, J. Y. *et al.* Characterization of Site-Specific N-Glycopeptide Isoforms of alpha-1-Acid Glycoprotein from an Interlaboratory Study Using LC-MS/MS. *J Proteome Res* **15**, 4146-4164, doi:10.1021/acs.jproteome.5b01159 (2016).

56 Wilson, R. *et al.* Abnormal immunoglobulin subclass patterns in women with a history of recurrent miscarriage. *Fertil Steril* **76**, 915-917 (2001).

57 Jefferis, R. *et al.* A comparative study of the N-linked oligosaccharide structures of human IgG subclass proteins. *Biochem J* **268**, 529-537 (1990).

58 Larsson, A., Palm, M., Hansson, L. O., Basu, S. & Axelsson, O. Reference values for alpha1-acid glycoprotein, alpha1-antitrypsin, albumin, haptoglobin, C-reactive protein, IgA, IgG and IgM during pregnancy. *Acta Obstet Gynecol Scand* **87**, 1084-1088, doi:10.1080/00016340802428146 (2008).

59 Akmacic, I. T. *et al.* High-throughput glycomics: optimization of sample preparation. *Biochemistry (Mosc)* **80**, 934-942, doi:10.1134/S0006297915070123 (2015).

60 Adamczyk, B., Stockmann, H., O'Flaherty, R., Karlsson, N. G. & Rudd, P. M. High-Throughput Analysis of the Plasma N-Glycome by UHPLC. *Methods Mol Biol* **1503**, 97-108, doi:10.1007/978-1-4939-6493-2_8 (2017).

61 Hennig, R. *et al.* N-Glycosylation Fingerprinting of Viral Glycoproteins by xCGE-LIF. *Methods Mol Biol* **1331**, 123-143, doi:10.1007/978-1-4939-2874-3_8 (2015).

62 RCore Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. (2014).

63 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).

64 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, doi:10.1093/glycob/cwv091 (2015).

65 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).

66 Ventham, N. T. *et al.* Changes to serum sample tube and processing methodology does not cause Intra-Individual [corrected] variation in automated whole serum N-glycan profiling in health and disease. *PLoS One* **10**, e0123028, doi:10.1371/journal.pone.0123028 (2015).

Part IV – Discussion

Chapter 9

General discussion

General discussion

While glycosylation is an important co- and post-translational modification exhibited by more than 50% of proteins[1,2], the analysis thereof is a remarkably complex affair. Underlying challenges include the non-linear arrangement of a varying number of monosaccharides, an absence of fluorescence response of native glycans, as well as heterogeneity in glycoprotein characteristics such as solubility, charge and stability[3].

This thesis mainly focuses on the method development and application within one particular branch of glycosylation analysis methodologies, namely matrix-assisted laser desorption/ionization mass spectrometry (MALDI) mass spectrometry (MS). For analysis of the ions we have made use of either time-of-flight (TOF) or Fourier-transform ion-cyclotron resonance (FTICR). The major advantage of MALDI over other ionization methods like ESI is the speed by which samples can be measured – the acquisition time for an individual sample is typically only a few seconds – making it particularly suitable for the mass spectrometric analysis of large cohorts[4-6]. What remained, and was explored in the work covered in this thesis, was to improve information content and quality, to decrease wet-lab hands-on time, to handle glycomics data integration, to streamline normalization, derived trait calculation and statistical analysis, and to apply the methodologies to challenging and informative clinical cohorts.

9.1 Developments in linkage-specific sialic acid esterification

In this thesis, two overall derivatization strategies are described for released *N*-glycans, namely the esterification of the acidic residues present on sialylated species (**Chapter 3**), and reductive amination of all *N*-glycans at their reducing ends (**Chapter 6**). The first led to the stabilization and linkage-discretization of sialic acids (both of the *N*-acetylneuraminic acid and *N*-glycolylneuraminic acid variety; **Chapter 3** and **Chapter 5**), while the second allowed high-resolution MALDI-FTICR-MS in negative mode (**Chapter 6**).

9.1.1 Application of linkage-specific sialic acid derivatization

The esterification protocol proved beneficial for the analysis of released *N*-glycans from human serum and plasma, and has contributed to a variety of case-control and longitudinal studies including on Down Syndrome[7] and pregnancy[8-10]. Furthermore, the reaction could be applied to the study of antibody glycosylation, examples including the released *N*-glycans of IgG-Fc and IgG-Fab[8], and those from engineered Her2-IgAs[11,12]. Direct comparison studies revealed that MALDI-TOF-MS with relatively simple mixtures of esterified *N*-glycans showed largely similar profiles and performances as liquid chromatography with fluorescence detection (**Chapter 2** and **Chapter 3**)[13-16]. Furthermore, our method comparison on the analysis of total serum *N*-glycosylation (TSNG) (**Chapter 8**) showed a generally higher information content (more individually detected signals) with MALDI-TOF-MS than with HILIC-UHPLC-FLD and xCGE-LIF. When applying the workflow to the analysis of murine

plasma, known to contain various glycan structures different from those found in humans[17,18], we could also establish the similarity of the reaction on α2,3- and α2,6-linked *N*-glycolylneuraminic acids, the esterification of sialic acids linked to *N*-acetylgalactosamines instead of galactoses, as well as the qualitative preservation of sialic acid *O*-acetylation due to the mild reaction conditions (**Chapter 5**)[19,20].

A downside of the ethyl esterification turned out to be the long-term stability of the esters, particularly the lactone[21], and while full reaction selectivity of the carboxylic acid modification was achieved for sialic acids, experiments on IgG glycopeptides revealed a variety of reaction products on the peptide portion[22,23]. To suit specific purposes, alternative reaction variants were designed, such as dimethylamidation / lactonization for IgG glycopeptides (showing full reaction specificity on the tryptic IgG peptide portion)[22,24], and dimethylamidation / amidation for increased long-term stability required for tissue imaging (**Table 1**)[25].

Table 1. *Sialic acid masses before and after linkage-specific derivatization.*

Modification	Nucleophile	Solvent	Δmass	Neu5Ac Δmass	Neu5Gc Δmass	Refs
α2,3-linkage-specific						
Lactonization	-	Methanol; Ethanol; DMSO	-18.0106	273.0849	289.0798	16,26
Amidation	Ammonium hydroxide	DMSO	-0.9840	290.1114	306.1063	25
Methylamidation	Methylamine	DMSO	13.0316	304.1271	320.1220	27,28
α2,6-linkage-specific						
Methyl esterification	Methanol	Methanol	14.0157	305.1111	321.1060	16,26
Ethyl esterification	Ethanol	Ethanol	28.0313	319.1267	335.1216	16
Dimethylamidation	Dimethylamine	DMSO	27.0473	318.1427	334.1376	22,25
Non-derivatized						
Native	-	-	-	291.0954	307.0903	-
Sodiated [M-H+Na]	-	-	21.9819	313.0774	329.0723	-
Potassiated [M-H+K]	-	-	37.9559	329.0513	345.0462	-

The ethyl esterification protocol, encompassing the chemical derivatization of glycans, HILIC enrichment and MALDI-TOF-MS measurement, proved automatable, but required modifications in the reaction temperature, the HILIC stationary phase, and the matrix solvents (**Chapter 4**)[29]. The resulting shift of the analytical bottleneck to data analysis was addressed by the development of software packages for integration and quality control,

such as the recently-published MassyTools[30]. Notably, the data analysis tools could make use of the enhanced resolution of the now-attained reflectron-mode MALDI-TOF-MS measurements. This includes quality control like high-precision ppm error determination on the now well-isolated monoisotopic signals, and comparison of theoretical with observed isotopic envelopes. The resulting high-throughput *N*-glycomics workflows were applied to several of the larger mass spectrometric total serum *N*-glycomics studies to date (**Chapter 6** and **Chapter 7**)[31].

9.2 Persistent challenges in mass spectrometric *N*-glycomics

While this thesis describes advances in MALDI-MS *N*-glycomics throughput and structural information (sialic acid linkage), the methodology is at current not capable of unravelling the full complexity of glycosylation in biological systems. Taking human glycosylation as an example, several layered challenges still need to be addressed. These include a larger control of technical variability, an increase in structural information, a traceability of glycans to specific glycosylation sites, as well as a coupling of glycomics with the analysis of other post-translational modifications (ideally in a protein-specific manner).

9.2.1 Controlling technical variability

So far, identification of glycomics biomarkers has been made difficult by relatively low correlation coefficients, indicating a lot of uncontrolled variability in the data. While part of this may arise from biological sources (which need to be understood), a large contribution may come from measurement errors as well. Most chapters in this thesis describe coefficients of variation based on positive controls in the experimental run (**Chapter 3-8**)[16,19,29,31], typical values ranging from 1% to 6% for the highest signal but increasing with decreasing signal intensity. While quite decent from an analytical perspective, this nonetheless means that additional effort may be useful to reduce this component of the cumulative error.

One complicating factor in *N*-glycan analysis is the variable abundance of the reducing-end glycosylamine which is present for glycans upon PNGase F release and converts over time to a regular aldose in aqueous conditions[32]. Performing the ethyl esterification protocol in the presence of these glycosylamines can lead to the unwanted reducing-end labeling with EDC or a fragment thereof, which in turn diminishes the signal intensity of the desired product. Fortunately, variability in PNGase F digestion and glycosylamine decay may be reduced, for example by increasing protocol incubation times, as reactions will principally be most robust when performed to completion. However, elongating incubation steps also leads to a conflict with method throughput. Recent reports describe attractive reducing-end labeling methods that make use of rapid PNGase F digestion and subsequent glycosylamine labeling, with complete deglycosylation of samples (IgG) reported within 10 minutes[32,33]. However, uniform glycosylamine production will presumably be hard to

achieve for all glycan species in a complex mixture – the most stable end-product being the aldose – making the strategy particularly risky when following up with ethyl esterification.

Another source of analytical variability lies in the sample purification step. Sample purity is often a requirement for mass spectrometric glycomics, as glycans (and glycopeptides) show a large ionization disadvantage over non-glycosylated peptides, and may be ion-suppressed in case of mixtures[34]. In addition, many release and derivatization protocols introduce high concentrations of reagents and detergents, preventing MS analysis of the crude reaction mixture[16,35,36]. One way to ensure the necessary sample purity is by performing a step of solid-phase extraction (SPE). In glycan analysis this typically involves making use of stationary phases (*e.g.*, C18-functionalized beads, PGC, a sugar polymer), and a mobile phase comprising water, an organic solvent, or a mixture thereof[31,37,38]. However, using SPE comes with several challenges as well. It is very difficult to keep the mobile phase or elution buffer concentrations constant during a multiple-sample SPE-session, as selective evaporation of volatile organics will inevitably occur. Unfortunately, even small changes in the organic solvent concentration of a mobile phase or elution buffer will influence glycan recovery in a species-dependent manner. Also challenging for this reason is elution, as residual washing buffer present on a stationary phase will be variable and cause differences in analyte recovery from one sample to the next.

With manual preparation it is possible to make small adjustments to a protocol to compensate, *e.g.*, keeping solvents at cold temperature and in closed bottles to limit selective evaporation or applying a slightly longer vacuum step when residual solvent is visibly present, but for an automated protocol (**Chapter 4**) every step needs to be predefined (and is immutable). As such, to allow sufficient repeatability for an automated variant of the ethyl esterification protocol (**Chapter 4**), the established workflow (**Chapter 3**) had to be adjusted in several ways. Evaporation was limited by covering the fluid containers with plastic foil (to be pierced at the first pipetting step) and by placing two covers on each reaction plate when incubating. Vacuum washes at concentration-sensitive steps were elongated to limit solvent residuals, and elution after SPE was performed with an automated centrifuge. One particular head-ache of protocol automation, namely the formation of bubbles in mixing steps, was prevented by increasing organic solvent concentrations, most notably of the MALDI matrix solution.

A profoundly different method of clean-up may lie in the covalent immobilization of glycans or glycoproteins to a stationary phase, of which applications have already been presented[39-41]. An advantage of this is the less strict requirement of washing solution organics concentrations, and the protocols can be directly combined with reducing-end labeling of the glycans[41]. Having said this, covalent methods of purification have not yet been reported to have lower variability than SPE methods of purification, indicating that further optimization might still be necessary[39-41].

An interesting observation throughout our MALDI-MS studies is that derived glycosylation traits, biology-inspired mathematical expressions encompassing chemically similar glycan species, show a lower technical variation than the single glycan traits (**Chapter 4** and **Chapter 8**)[29]. Next to grouping by chemical similarity, derived traits cluster *N*-glycan species that are found in a similar *m/z* range, *e.g.*, triantennary species (± *m/z* 2700 – 3200 when esterified $[M + Na]^+$) and tetraantennary species (± *m/z* 3200 – 3800). In addition, the derived traits cluster species with similar intensity, for example the di-, tri, and tetraantennary species which differ in intensity by approximately an order of magnitude [**Chapters 3**, **4**, **6**, and **8**][16,29,31]. Part of the uncontrolled variation may thus be caused by sample preparation aspects that differentially affect the chemically divert (and massed) glycan structures, or by the limited dynamic range of the MS instrumentation (*i.e.*, the range in which analyte concentrations show a linear relationship with signal intensity)[6]. For the derived traits, compared to the native glycans, the influence of both sources of variation would then be limited.

One interesting solution to limit variation on the experimental side of MALDI-MS could be the inclusion of internal standards to a sample, for instance by heavy-isotope reducing-end labeling[42-44], permethylation[15,45], or sialic acid derivatization[46]. In addition to decreasing the dependence on mathematical adjustments such as the commonly employed total area normalization and derived trait calculation (**Chapter 5**), this internal standardization would provide a (semi-)absolute measure of analyte quantity.

9.2.2 Stereoisomerism

As exemplified by **Chapter 8**, MALDI-MS principally does not discriminate between glycan compositional and constitutional isomers. The differences in monosaccharide enantiomers (D and L), diastereomers (mannose and galactose), anomers (α and β), and linkages (usually 1,3, 1,4 or 1,6, with 2,8 and 2,9 possible for sialic acids) therefore stay invisible. For screening experiments, MALDI-MS is limited to providing the monosaccharide composition in a number of hexoses, *N*-acetylhexosamines, deoxyhexoses, and sialic acids (*e.g.*, H5N4F1S2). Nonetheless, several chapters have made assumptions on *N*-glycan structural features (**Chapters 5**, **6**, and **7**), primarily based on literature knowledge gathered with, *e.g.*, nuclear magnetic resonance spectroscopy, lectin interaction assays, and a variety of glycoanalytical methods combined with exoglycosidase digestion[47-52]. For example, throughout this thesis the composition H5N4 has been interpreted as a structure with all HexNAcs being GlcNAcs and the five hexoses comprising three mannoses and two galactoses, *i.e.*, a diantennary glycan with two galactoses. However, for this composition a biologically valid alternative is a hybrid structure without galactoses and two antennary GlcNAcs. In addition, one of these HexNAcs might be a bisected GlcNAc, or instead be a GalNAc in a LacdiNAc motif (GalNAcβ1,4-GlcNAc). Because of this, a derived trait describing the “galactosylation of diantennary species” may be confounded by bisection, hybrid-type

glycosylation, and any other type of structure with the same monosaccharide composition. Part of the scientific process is to build on previously established knowledge, and structural presumptions are necessary to efficiently produce the link between profiling experiments and functional studies. However, for any given experiment, particularly when performed on samples for which limited prior knowledge is available, these presumptions may turn out to be false.

Mass spectrometry does provide several opportunities to characterize glycans on a structural level. Reports have been made on diagnostic ions found in negative mode MS/MS (*e.g.*, the D-ion)[53], MS/MS of sodium adducts of permethylated glycans[54,55], and differentiation between GlcNAc and GalNAc in high-energy fragmentation[56]. In addition, MS can be hyphenated to various forms of chromatography, CE, and ion mobility for an extra experimental dimension that can benefit structural interpretation[57-59]. However, while hyphenation and orthogonal analysis can usually pinpoint the major structures for a given signal, in complex mixtures the ratios may be hard to relatively quantify, and for large sample sets the workload may be too much to handle. Typically, thorough analysis is performed on a limited set of samples from a cohort (or a pool thereof), which then provides a possibility-space for annotation.

9.2.3 Site-specific glycosylation analysis

The associations between total plasma (or serum) *N*-glycosylation and other clinical variables reported throughout this thesis (**Chapters 6** and **7**) can often be explained by a change in plasma glycoprotein levels rather than glycosylation changes within a specific glycoprotein, but very little objective discrimination can be made between the two[31]. The analysis of glycopeptides rather than released glycans would provide protein- and site-specificity, but while mass spectrometry is a technique that is principally capable of doing this, there is still a lot of ground to cover. Current glycopeptidomics approaches can be grouped into two overall strategies, namely a glycomics-based variant that attempts optimal recognition of distinct glycoforms for a small set of peptides, or a proteomics-based variant that attempts optimal identification of the peptides but usually provides low information on the glycan portion.

For the first approach, glycomics-based, a protein of interest is typically isolated from a complex mixture, digested with trypsin, LysC or another protease to yield distinct peptides, some of which are glycopeptides[60-63]. These glycopeptides can then be analyzed by MALDI-MS, LC-MS or other methodologies. This approach has for example been extensively used for the analysis of tryptic IgG-Fc glycopeptides, making distinction between different subclasses of the antibody (IgG1-4)[60,64]. Tandem MS has shown effective in obtaining additional structural information in these situations. Ion trap instruments may, for example, make use of collision-induced dissociation (CID) and electron-capture dissociation (ECD) to fragment the peptide portion of glycopeptides, or of electron-transfer dissociation (ETD) to

preferentially fragment the glycan portion[65,66]. Comprehensive glycopeptide information may be obtained, among other methods, by a combination of the above[67], by the application of multiple collisional energy levels[63], by the higher-energy collisional dissociation (HCD) methodology that is available for C-trap-equipped Orbitrap instruments[68], or by negative mode MSn approaches[69]. Next to this, a variety of derivatization strategies have been proposed to enhance separation of the glycopeptides, for example by derivatization to allow linkage-specific sialic acid analysis, to enhance ionization efficiency, or to allow isobaric labeling[22,70,71].

The second overall strategy, proteomics-based, attempts to broaden the scope of peptides that can be detected, but for the glycan part usually ends at binary recognition of site-occupancy[72-74]. Reported approaches include proteolytic digestion of the complex mixtures of proteins (*e.g.*, blood plasma, spinal fluid) and subsequent endoglycosidase treatment (PNGase F, PNGase A, Endo H)[72-74]. Readout of the location of glycosylation is provided by the asparagine deamidation as produced by PNGase F or PNGase A after glycolysis, or the detectable presence of a single (fucosylated) *N*-acetylglucosamine as provided by endoglycosidases H or F[72-74]. As is clear, this strategy may lead to the identification of glycosylation sites, and in some instances may be indicative of the core-fucosylation of these sites, but no further information on microheterogeneity is provided[74].

9.3 Developments human plasma *N*-glycosylation analysis

However limited our understanding of human *N*-glycosylation may seem to be, we have also made several advances in the processing and interpretation of complex mixtures like the *N*-glycosylation of blood plasma and serum.

9.3.1 Mathematical developments

For the serum/plasma *N*-glycosylation we have analyzed throughout the thesis (**Chapters 3-8**), we always represent each signal as a fraction of the total analyte area (total area normalization). This form of normalization, also used for UHPLC-FLD and CGE-LIF glycomics, is highly beneficial for the repeatability and data quality of a run[75,76]. However, the downside of such an approach is the introduction of potentially spurious correlation, as the relative increase in one peak will mathematically lead to a decrease in all the others. This conversion works remarkably well for samples with a defined number of glycosylation sites, *e.g.*, IgG-Fc with one *N*-glycan site per C_H2 domain at Asn297[77,78], as the negative correlation between glycoforms will approach reality; for a single glycosylation site the higher abundance of one species necessitates the lower abundance of the others. However, in case of multiple proteins, such as found in plasma, the number of glycosylation sites will differ by the abundance of those glycoproteins. This means that for total plasma *N*-glycome (TPNG) analysis changes in glycosylation values are representative of both changes in protein glycosylation and glycoprotein abundance.

With a reasonable understanding of glycan processing[3,79,80], it is possible to create mathematical constructs to express this in TPNG data, as explored in **Chapter 4** and used within each subsequent chapter. For example, one can relate a nongalactosylated glycan species to a galactosylated species, and thereby obtain a ratio of galactosylation for that species. However, the situation is complicated by the fact that glycans are not independent, but related to many other glycans. For instance, while a nongalactosylated glycan may be processed into a monogalactosylated variant, the latter might be further processed into a digalactosylated galactose thereby hiding the change in the monogalactosylated one[3,74,75]. This means that single enzymatic events may not always be reflected in the direct ratios of product and educt, and the need arises to see glycan processing in a wider framework. An example of a derived trait taking this into account is the calculation of the "degree of galactosylation within all diantennary glycans", yielding a higher number for each additional galactose.

Furthermore, each monosaccharide addition (or removal, but let us regard addition) is dependent on the presence of an appropriate substrate, *e.g.*, galactosylation on the availability of an appropriately linked *N*-acetylglucosamine, and sialylation on the presence of a galactose[3,79,80]. This dependency may also influence biological insight into disease mechanisms. For example, sialylation associates with RA disease activity while sialylation-per-galactose does not, implying that galactosylation and not sialylation might be the important factor to consider in terms of dysregulation[81]. For the derived traits it then is logical to consider these dependencies as well.

In addition, with an estimate of the contribution of plasma proteins to the observed total plasma *N*-glycome[47,48], theoretical protein groupings of glycosylation can be designed. The most prominent example of this is the glycosylation of the Fc portion of IgG, which, to our knowledge, is by far the main contributor to the nonsialylated complex glycans[48]. Looking within plasma at a derived trait describing the galactosylation of nonsialylated glycans (supposed to be mainly IgG-Fc), we see a remarkable congruency with the effects described for IgG itself(**Chapter 6** and **7**)[81,82].

The concept of derived traits has been prior explored in literature, mostly for the HPLC analysis of IgG[83,84]. This protein contains an envelope of glycosylation consisting of glycan species that are largely well-separated by chromatography, which allows for the construction of reasonably well-defined traits[83,84]. However, as demonstrated in **Chapter 8** of this thesis, the chromatographic separation of the *N*-glycosylation from plasma is more structurally ambiguous, particularly in the high antennarity regions of a profile, and poses some challenges for the construction of derived traits. MALDI-MS, while lacking in structural information (*e.g.*, core- *vs.* antennary-fucosylation, α1,3- *vs.* α1,6-branch galactosylation), does show for plasma *N*-glycans an excellent separation on a compositional level and can therefore benefit from derived traits based on these compositions. **Chapters 5** through **8**

describe a large number of compositionally derived traits for serum/plasma *N*-glycosylation analysis, in some cases outnumbering the single observed species. Favorably, these derived traits performed very well for the analysis of plasma *N*-glycosylation, displaying lower methodological variability, yielding stronger statistical associations, and overall providing information that is better clinically interpretable than single *N*-glycan compositions.

9.3.2 Serum and plasma protein grouping

As it is, several broad protein categories were suggested throughout this thesis to influence the serum/plasma *N*-glycome (**Chapter 6** and **7**)[31]. Most clearly we can discriminate a set of glycans also observed on the Fc portion of IgG, secondly we can detect the glycans commonly found on other plasma-cell-produced immunoglobulins together with IgG-Fab, thirdly we can identify the glycosylation also detected on the plasma-abundant acute-phase proteins, and finally we seem to isolate the glycosylation of lipoproteins reflecting lipid transport[31,47,48].

When compared to the overall *N*-glycan structures found in human plasma, the Fc-portion of IgG shows complex-type species with a rather low state of post-translational processing. Whereas most complex *N*-glycans have a near-complete (sub-)terminal galactosylation and sialylation[31,47,48], the *N*-glycans on IgG-Fc typically shows incomplete galactosylation and around 20% of structures carrying sialic acids[8,77]. Plasma cells, the cells that produce a major part of the IgG found in plasma[77,85], are fully capable of high degrees of galactosylation and sialylation. This is apparent from the larger diantennary *N*-glycans observable on IgM, IgA and other plasma-cell derived immunoglobulins[47,48,86,87], from the glycans expressed on the Fab portion of IgG[8], and also from reports on *in vitro* stimulation of B cells to achieve altered Fc-glycosylation[88]. In all, the evidence indicates a protein- and site-specific regulation of the glycosylation of IgG-Fc, which is of biological relevance due to the large influence of the carbohydrate on Fc-gamma receptor (FcγR) binding[77,89]. Molecular modelling of IgG reveals that the Fc-glycans of IgG are inwards oriented, meaning that the low complexity may arise from steric hindrance, *i.e.*, a decreased accessibility of the glycan for glycosyltransferases[90,91]. Another interesting hypothesis is that the glycan may interact with the peptide backbone at specific stages of processing[92]. For example, as soon as the α1,6-branch *N*-acetylglucosamine is galactosylated this residue may form hydrogen bonds with a nearby lysine (Lys246) and threonine (Thr260), thereby limiting further sialylation[92-94]. By whichever means the glycosylation of IgG-Fc is generated, it is sufficiently different from the glycosylation of other abundant plasma proteins that its influence can be seen when looking at glycosylation subgroups in the total plasma/serum *N*-glycome[47,48]. For example, the decrease of IgG-Fc galactosylation with age is a repeated observation with a variety of analytical platforms[83,95,96]. Using the IgG-Fc-specific derived trait in our healthy human plasma study (**Chapter 6**), we could detect the strong association with age as well.

The other main contributors to the plasma/serum *N*-glycome are the acute-phase proteins, coagulation factors (in which lies the difference between plasma and serum), and lipid transport proteins[47,48,97]. Fibrinogen, transferrin, haptoglobin, alpha-1-antitrypsin and alpha-1-acid glycoprotein, all involved in coagulation and the acute-phase response, display typically di- to tetraantennary glycan structures with near-complete galactosylation and sialylation (both α2,3- and α2,6-linked), and show a preference of antennary- over core-fucosylation (in case of α2,3-linked sialylation, thereby forming the glycan epitopes sialyl-Lewis X or A)[16,48,49,98-101]. A common observation in our plasma/serum *N*-glycome studies (**Chapters 6**, **7** and **8**) is that injury and inflammation lead to an altogether upregulation of tri- and tetraantennary glycan species, a phenomenon that would be most easily explained by an upregulation of the abundance of the aforementioned acute-phase proteins, but regulation of glycan microheterogeneity will likely also play a role.

The last plasma protein grouping observed in our studies involved lipid transport (**Chapter 6**)[47,48]. Of interest is that the larger high-mannose glycan structures (*e.g.*, Man8 and Man9) are not commonly found on glycoproteins that contribute to the total plasma *N*-glycome[47,48]. In fact, the only protein we have identified to have sufficient influence on the Man9 content in the TPNG is apolipoprotein B-100, a structural protein in the formation of (V)LDL particles[102-104]. In a fascinating parallel, our study regarding metabolic markers in healthy individuals (**Chapter 6**) shows one of the strongest associations between glycosylation and lipid metabolism to be with the size of high-mannose glycans, notably by the upregulation of Man8 and Man9. The simplest biological explanation of this association would be the increased abundance of LDL particles and accompanying high-mannose-glycosylated apolipoprotein B-100, but also here specific regulation may play a role.

In all, it has to be noted that many of the prior statements have been logically induced from joining the known protein concentrations in human fluids with the overall understanding of the glycosylation thereof[47,48], but this may of course be untrue in any particular case. An important step for future research would be the actual quantification of the proteins from plasma, combined with the glycosylation thereof. This may also provide information on whether our repeated observations come from the change in glycoprotein quantity or from an altered expression of protein glycoforms. Furthermore, not all high-abundant plasma protein could be included in the theoretical calculation. For instance, there are several acute phase proteins that are, by the discrepancy between their conceptual protein mass and actual observed mass, likely carriers of high degrees of tri- and tetraantennary types of glycosylation. Examples of this are histidine-rich glycoprotein and kininogen-1[105,106], but so far these have not been characterized with respect to glycan macro- and microheterogeneity[48].

9.3.3 Cellular origin of plasma *N*-glycosylation

Altogether, it appears highly likely that the total plasma *N*-glycome represents the glycosylation characteristics of a limited set of cell populations, the main contributors being plasma cells and hepatocytes. These populations appear to have discrete glycosylation-type outputs on their secreted glycoproteins[48]. For instance, plasma cells are rather unique in their abundance of core-fucosylated *N*-glycans, their processing of glycans to at best diantennary complex-type, the near-exclusivity of α2,6-linked sialylation, and the occurrence of bisection[48]. Hepatocytes, on the other hand, readily show the capability to generate tri- and tetraantennary *N*-glycans, LacNAc repeats, antennary fucosylation, both α2,6- and α2,3-linked sialylation, and so far no bisected structures have been identified. For cell-line engineering it was previously shown that the cell type has a large influence on the glycosylation type that can be expected. Examples of this are CHO cells and HEK cells, each providing their unique glycosylation characteristics[107,108]. Protein-specific attributes such as steric hindrance and glycan-protein interaction can then fine-tune the glycosylation per site, but the overall regime appears defined by the cell type.

Nonetheless, plasma will also contain glycoproteins originating from other cell populations, such as kidney cells and thyroid cells[109,110]. Of most clinical interest would be the plasma glycoproteins that originate from aberrant cell populations, such as tumor tissue. With sufficient analytical sensitivity, it may be possible to see the minor amounts of uncommon glycan structures that originate from these pathological tissues, potentially providing the simple biomarker with high sensitivity and specificity that is coveted by clinical practice. The most promising biomarkers are commonly of low abundance[111], signifying the need for analytical tools of high sensitivity.

9.3.4 Clinical associations within the plasma/serum *N*-glycome

Chapters 6 and **7** have explored the relationship between plasma as well as serum *N*-glycosylation and physiological and clinical parameters like age, sex, lipid levels, markers of inflammation, and a variety of others. These studies are not the first ones to associate glycosylation with clinical variables, but are exceptional by the large-scale use of mass spectrometry, which provides information orthogonal to that of other glycoanalytical methods (**Chapter 8**).

Several total plasma *N*-glycosylation traits show to differ between the sexes. One prominent example of this is tri- and tetraantennary fucosylation (presumable antennary-bound in the form of sialyl-Lewis X), which show on average to be higher in males than in females[31,112,113]. While it is difficult to attribute these glycosylation differences to one cause or another, one attractive direction of future research may be hormonal regulation, *e.g.*, by estrogen as indicated in the literature[114,115]. Association with hormone levels may also be the driving force behind several of the pregnancy-associated glycosylation changes we observed in the total serum *N*-glycome throughout pregnancy (**Chapter 7**).

Other clear associations seen within the total plasma *N*-glycome appear to largely stem from various aspects of inflammation. **Chapter 6** confirmed with mass spectrometry the strong negative association between calendar age and galactosylation of IgG-Fc-type *N*-glycans[31,83], a widely-reported phenomenon thought to be related to the build-up of chronic inflammation with age[95]. Furthermore, as prior mentioned, a large part of the tri- and tetraantennary structures within plasma stem from acute-phase proteins that differ in level upon inflammation[116], and may alter their glycosylation to affect their function and recognition[116,117]. However, when assessing the biomarker potential of glycosylation for the diagnosis of specific diseases, it has to be kept in mind that these markers of general inflammation will likely be picked up as well and will have limited specificity in clinical screening.

Furthermore, lipid metabolism, marked in its simplest form by BMI, but also by measures of, *e.g.*, cholesterol, leptin, adiponectin, and triglycerides, show in part the same glycosylation associations as inflammation, but show in addition also several unique effects (**Chapter 6**). One particular glycosylation characteristic, highlighted in the metabolic study of **Chapter 6**, is the glycan composition H4N4S1 which provides discrimination between healthy and unhealthy metabolic profiles, and is an interesting target for future research.

9.4 Future prospects of plasma *N*-glycosylation analysis by MALDI-MS

9.4.1 Study design for plasma/serum *N*-glycome analysis

This thesis reports on the glycomics measurement of several large cohorts (**Chapters 6**, **7**, and **8**), which, together with the literature, provides the opportunity to make some overarching statements. As it seems, findings that are presented in glycomics studies are typically small in effect size, and associate, next to the outcome of interest, as well with a broad range of other phenomena. For instance, a multitude of (inflammatory) disorders is marked by the galactosylation behavior of IgG-Fc[95,118-120], and many of the acute phase protein glycosylation behaviors are expected to be representative of overall mechanisms of acute inflammation rather than something that is disease-specific[116,121,122].

To find disease-specific biomarkers (and markers for stratification of patients) within the total plasma *N*-glycome, it makes sense to always include in the study the clinical information on age, sex, BMI and a marker for inflammation such as the level of C-reactive protein (CRP; or interleukin-6, but in our study CRP showed to be the better predictor) (**Chapter 6**). Without these measures it will be very hard to provide sensible interpretation of new clinical associations, and the specificity of new biomarkers would be limited as a consequence. Interestingly, while we have performed the largest mass spectrometric inquiry between clinical health markers and plasma *N*-glycosylation to date (**Chapter 6**), the majority of the glycomics variance is still not explained by clinical factors. This suggests

plenty of room still for future research into the cause and effect of glycan profiles, for example into the direction of genetic and epigenetic regulation[123-125]. One potential aspect of (epi)genetic regulation is the repeated observation that glycosylation is more variable between individuals than within individuals (in homeostasis) (**Chapter 7** and **8**)[51,126]. For the clinical exploitation of glycomics this means that a disease biomarker could preferentially lie in the repeat measurement of the glycans of an individual throughout their lifetime. A deviation from a standard profile of a person will then provide the warning call to proceed to more specific modes of diagnosis. In this possible future, MALDI-MS will have the benefit of throughput (**Chapter 8**)[5] – a useful feature for the analytical workload of population screening – but it remains to be seen if the relevant glycosylation characteristics can be detected by MS.

One example of the benefit of longitudinal analysis can already be found in this thesis, where we demonstrate the rather drastic upregulation of α2,3-linked sialylation with pregnancy on an individual level (**Chapter 7**). However, in general still the most basic of longitudinal analyses need to be performed with regard to glycomics, encompassing intraday variation (*e.g.*, dietary intake, tissue damage, acute inflammation, physical or mental activity), short-term interday variation (*e.g.*, menstrual cycle, virus infection), and the longer time-frames which have the most clinical relevance.

9.4.2 Functional studies

The analysis of plasma or serum *N*-glycosylation (**Chapters 5**, **6**, **7** and **8**) is a fundamentally phenotypical study. Whereas genetics can be presumed to be upstream to many diseases (a clear genetic defect leaves little speculation on causality), glycans observed in plasma or even single proteins are the end product of a large series of biological processes, making it challenging to determine cause and consequence. Factors influencing glycosylation are incredibly encompassing, from the abundance of glycosyltransferases and nucleotide sugars, *via* the time spent in the ER and Golgi, to the activity of intra- and extracellular recycling mechanisms[3,127,128].

Much of the functional research in glycomics, the targeted manipulation of glycosylation properties and observing their effects, has been performed on IgG[129-131]. It is an attractive object of study, because the protein is involved in several immunological processes (adaptive immune response and complement activation), and has altered function depending on its glycosylation[77,129-131]. In addition, pharmaceutical companies are interested in the modulation of IgG glycosylation to, *e.g.*, decrease toxicity, increase half-life, and increase ADCC[89,129,132]. However, functional studies outside of IgG have been relatively scarce.

One factor that is complicating functional studies is the translation from cell lines and animal models to the human situation, as glycosylation appears to be species-specific and differs

from one cell line to the next[3,18,107,108,133]. Mouse *N*-glycosylation, for which we have studied total plasma and IgG-Fc in healthy individuals (**Chapter 5**)[19], differs in sever ways from that of humans. For one, the mouse expresses predominantly *N*-glycolylneuraminic acid (as opposed to *N*-acetylneuraminic acid in humans)[3,18], and its *N*-glycans may carry α-linked galactoses (which are normally absent in humans)[20,134]. Further differences include a partial acetylation of sialic acids, the presence of GlcNAc-linked sialic acids for the triantennary species (possibly at the expense of the antennary fucosylation commonly found in humans) , and a relatively low level of IgG-Fc type glycans in plasma (**Chapter 5**)[19,20]. Next to this, mouse strains show to differ in glycosylation properties, which is most prominently visible by the degree of fucosylation in female individuals (ranging from 10% to 50% depending on the strain) (**Chapter 5**)[19].

For cell line work there are similar concerns, as cell lines such as CHO, HEK, and others differ in their respective glycosylation[107,108,133,135]. Fortunately, several strategies have been published that can provide control over protein glycoforms by *in vitro* or *in vivo* manipulation[131,136]. Isolated proteins can be modified by glycosidases and glycosyltransferases in presence of the appropriate nucleotide sugar[131], and cell systems have been described that make use of genetic knock-outs and knock-ins, and transient expression of the appropriate genes[131,135-137].

9.5 Concluding remarks

To summarize, this thesis describes several contributions to the high-throughput analysis of complex glycan samples by MALDI-MS, covering strategies of chemical derivatization (**Chapter 3**), automation (**Chapter 4**), and computational solutions for data preprocessing and statistical analysis (**Chapter 4** and **6**). The functionality of the developed platforms was demonstrated by several applications, including the characterization of mouse-strain plasma *N*-glycosylation (**Chapter 5**), the profiling of principally healthy human donors (**Chapter 6**), and an assessment of the intraindividual changes in RA patients throughout pregnancy (**Chapter 7** and **8**). Lastly, the workflow developed in this thesis for MALDI-TOF-MS with derivatized sialic acid residues was compared with other contemporary methods for high-throughput glycan analysis (**Chapter 8**).

A next large step for MALDI-MS method development would lie in the inclusion of internal standardization methods, both to decrease analytical variability and to limit the mathematical co-dependence of glycan signals. Analytical method hyphenation, parallel running of orthogonal methods for the same samples, and/or further advances in glycan characterization by MS(/MS) are still required to obtain the remaining relevant structural information of glycans in high-throughput fashion. Clinical glycomics would benefit from the translation of plasma-wide observations to particular proteins or groups of proteins, which can then subsequently be isolated for in-depth analysis. Lastly, the advent of longitudinal

glycomics studies is expected to be of considerable benefit to the translation of analytical methodology to clinical diagnostics and to facilitate patient stratification.

9.6 References

1 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8 (1999).

2 Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* **1**, doi:10.1038/srep00090 (2011).

3 Varki, A. *et al. Essentials of Glycobiology*. 2nd edn, (2009).

4 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev* **18**, 349-450, doi:10.1002/(SICI)1098-2787(1999)18:6<349::AID-MAS1>3.0.CO;2-H (1999).

5 Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).

6 de Hoffmann, E. & Stroobant, V. *Mass spectrometry: principles and applications*. 3rd edn, (2007).

7 Borelli, V. *et al.* Plasma N-Glycome Signature of Down Syndrome. *J Proteome Res* **14**, 4232-4245, doi:10.1021/acs.jproteome.5b00356 (2015).

8 Bondt, A. *et al.* Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics* **13**, 3029-3039, doi:10.1074/mcp.M114.039537 (2014).

9 Jansen, B. C. *et al.* Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep* **6**, 23296, doi:10.1038/srep23296 (2016).

10 Jansen, B. C. *et al.* MALDI-TOF-MS reveals differential N-linked plasma- and IgG-glycosylation profiles between mothers and their newborns. *Sci Rep* **6**, 34001, doi:10.1038/srep34001 (2016).

11 Rouwendal, G. J. *et al.* A comparison of anti-HER2 IgA and IgG1 in vivo efficacy is facilitated by high N-glycan sialylation of the IgA. *MAbs* **8**, 74-86, doi:10.1080/19420862.2015.1102812 (2016).

12 Meyer, S. *et al.* Improved in vivo anti-tumor effects of IgA-Her2 antibodies through half-life extension and serum exposure enhancement by FcRn targeting. *MAbs* **8**, 87-98, doi:10.1080/19420862.2015.1106658 (2016).

13 Reusch, D. *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles--part 1: separation-based methods. *MAbs* **7**, 167-179, doi:10.4161/19420862.2014.986000 (2015).

14 Reusch, D. *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods. *MAbs* **7**, 732-742, doi:10.1080/19420862.2015.1045173 (2015).

15 Shubhakar, A. *et al.* High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia* **78**, 321-333, doi:10.1007/s10337-014-2803-9 (2015).

16 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).

17 Galili, U., Shohet, S. B., Kobrin, E., Stults, C. L. & Macher, B. A. Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. *J Biol Chem* **263**, 17755-17762 (1988).

18 Varki, A. Diversity in the sialic acids. *Glycobiology* **2**, 25-40 (1992).

19 Reiding, K. R., Hipgrave Ederveen, A. L., Rombouts, Y. & Wuhrer, M. Murine Plasma N-Glycosylation Traits Associated with Sex and Strain. *J Proteome Res* **15**, 3489-3499, doi:10.1021/acs.jproteome.6b00071 (2016).
20 de Haan, N. *et al.* The N-Glycosylation of Mouse Immunoglobulin G (IgG)-Fragment Crystallizable Differs Between IgG Subclasses and Strains. *Front Immunol* **8**, 608, doi:10.3389/fimmu.2017.00608 (2017).
21 Reiding, K. R., Lonardi, E., Hipgrave Ederveen, A. L. & Wuhrer, M. Ethyl Esterification for MALDI-MS Analysis of Protein Glycosylation. *Methods Mol Biol* **1394**, 151-162, doi:10.1007/978-1-4939-3341-9_11 (2016).
22 de Haan, N. *et al.* Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. *Anal Chem* **87**, 8284-8291, doi:10.1021/acs.analchem.5b02426 (2015).
23 Gomes de Oliveira, A. G. *et al.* A systematic study of glycopeptide esterification for the semi-quantitative determination of sialylation in antibodies. *Rapid Commun Mass Spectrom* **29**, 1817-1826, doi:10.1002/rcm.7287 (2015).
24 de Haan, N., Reiding, K. R. & Wuhrer, M. Sialic Acid Derivatization for the Rapid Subclass- and Sialic Acid Linkage-Specific MALDI-TOF-MS Analysis of IgG Fc-Glycopeptides. *Methods Mol Biol* **1503**, 49-62, doi:10.1007/978-1-4939-6493-2_5 (2017).
25 Holst, S. *et al.* Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues. *Anal Chem* **88**, 5904-5913, doi:10.1021/acs.analchem.6b00819 (2016).
26 Wheeler, S. F., Domann, P. & Harvey, D. J. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of alpha(2 --> 3)- and alpha(2 --> 6)-isomers. *Rapid Commun Mass Spectrom* **23**, 303-312, doi:10.1002/rcm.3867 (2009).
27 Li, H., Gao, W., Feng, X., Liu, B. F. & Liu, X. MALDI-MS analysis of sialylated N-glycan linkage isomers using solid-phase two step derivatization method. *Anal Chim Acta* **924**, 77-85, doi:10.1016/j.aca.2016.04.023 (2016).
28 Nishikaze, T. *et al.* Differentiation of Sialyl Linkage Isomers by One-Pot Sialic Acid Derivatization for Mass Spectrometry-Based Glycan Profiling. *Anal Chem* **89**, 2353-2360, doi:10.1021/acs.analchem.6b04150 (2017).
29 Bladergroen, M. R. *et al.* Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *J Proteome Res* **14**, 4080-4086, doi:10.1021/acs.jproteome.5b00538 (2015).
30 Jansen, B. C. *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* **14**, 5088-5098, doi:10.1021/acs.jproteome.5b00658 (2015).
31 Reiding, K. R. *et al.* Human Plasma N-glycosylation as Analyzed by Matrix-Assisted Laser Desorption/Ionization-Fourier Transform Ion Cyclotron Resonance-MS Associates with Markers of Inflammation and Metabolic Health. *Mol Cell Proteomics* **16**, 228-242, doi:10.1074/mcp.M116.065250 (2017).
32 Gong, B. *et al.* N-glycosylamine-mediated isotope labeling for mass spectrometry-based quantitative analysis of N-linked glycans. *Anal Bioanal Chem* **405**, 5825-5831, doi:10.1007/s00216-013-6988-9 (2013).
33 Lauber, M. A. *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* **87**, 5401-5409, doi:10.1021/acs.analchem.5b00758 (2015).
34 Zaia, J. Mass Spectrometry and Glycomics. *OMICS : a Journal of Integrative Biology* **14**, 401-418, doi:10.1089/omi.2009.0146 (2010).
35 Morelle, W. & Michalski, J. C. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* **2**, 1585-1602, doi:10.1038/nprot.2007.227 (2007).

36 Ruhaak, L. R., Steenvoorden, E., Koeleman, C. A. M., Deelder, A. M. & Wuhrer, M. 2-Picoline-borane: A non-toxic reducing agent for oligosaccharide labeling by reductive amination. *Proteomics* **10**, 2330-2336, doi:10.1002/pmic.200900804 (2010).
37 Ruhaak, L. R. *et al.* Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem* **80**, 6119-6126, doi:10.1021/ac800630x (2008).
38 Selman, M. H. J., Hemayatkar, M., Deelder, A. M. & Wuhrer, M. Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal Chem* **83**, 2492-2499, doi:10.1021/ac1027116 (2011).
39 Yang, S., Hu, Y., Sokoll, L. & Zhang, H. Simultaneous quantification of N- and O-glycans using a solid-phase method. *Nat Protoc* **12**, 1229-1244, doi:10.1038/nprot.2017.034 (2017).
40 Zhou, H. *et al.* Universal Solid-Phase Reversible Sample-Prep for Concurrent Proteome and N-Glycome Characterization. *J Proteome Res* **15**, 891-899, doi:10.1021/acs.jproteome.5b00865 (2016).
41 Miura, Y. *et al.* BlotGlycoABCTM, an integrated glycoblotting technique for rapid and large scale clinical glycomics. *Mol Cell Proteomics* **7**, 370-377, doi:10.1074/mcp.M700377-MCP200 (2008).
42 Prien, J. M., Prater, B. D., Qin, Q. & Cockrill, S. L. Mass Spectrometric-Based Stable Isotopic 2-Aminobenzoic Acid Glycan Mapping for Rapid Glycan Screening of Biotherapeutics. *Anal Chem* **82**, 1498-1508, doi:10.1021/ac902617t (2010).
43 Walker, S. H., Budhathoki-Uprety, J., Novak, B. M. & Muddiman, D. C. Stable-Isotope Labeled Hydrophobic Hydrazide Reagents for the Relative Quantification of N-Linked Glycans by Electrospray Ionization Mass Spectrometry. *Anal Chem* **83**, 6738-6745, doi:10.1021/ac201376q (2011).
44 Zhang, P. *et al.* Relative quantitation of glycans using stable isotopic labels 1-(d0/d5) phenyl-3-methyl-5-pyrazolone by mass spectrometry. *Anal Biochem* **418**, 1-9, doi:10.1016/j.ab.2011.07.006 (2011).
45 Atwood, J. A., 3rd *et al.* Quantitation by isobaric labeling: applications to glycomics. *J Proteome Res* **7**, 367-374, doi:10.1021/pr070476i (2008).
46 Shah, P. *et al.* Mass Spectrometric Analysis of Sialylated Glycans Using Solid Phase Labeling of Sialc Acids. *Anal Chem* **85**, 3606-3613, doi:10.1021/ac3033867 (2013).
47 Klein, A. Human total serum N-glycome. *Adv Clin Chem* **46**, 51-85 (2008).
48 Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-343, doi:10.1007/s10719-015-9626-2 (2016).
49 Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. & Altmann, F. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics* **6**, 3369-3380, doi:10.1002/pmic.200500751 (2006).
50 Zhang, L., Luo, S. & Zhang, B. The use of lectin microarray for assessing glycosylation of therapeutic proteins. *MAbs* **8**, 524-535, doi:10.1080/19420862.2016.1149662 (2016).
51 Hennig, R. *et al.* Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. *Biochim Biophys Acta* **1860**, 1728-1738, doi:10.1016/j.bbagen.2016.03.035 (2016).
52 Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **13**, 2314-2327, doi:10.1021/pr401092y (2014).
53 Harvey, D. J. 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 Spectr* **16**, 622-630, doi:https://doi.org/10.1016/j.jasms.2005.01.004 (2005).

54 Mechref, Y., Kang, P. & Novotny, M. V. Differentiating structural isomers of sialylated glycans by matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry. *Rapid Commun Mass Spectrom* **20**, 1381-1389, doi:10.1002/rcm.2445 (2006).
55 Stumpo, K. A. & Reinhold, V. N. The N-glycome of human plasma. *J Proteome Res* **9**, 4823-4830, doi:10.1021/pr100528k (2010).
56 Halim, A. *et al.* Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J Proteome Res* **13**, 6024-6032, doi:10.1021/pr500898r (2014).
57 Zhou, S., Dong, X., Veillon, L., Huang, Y. & Mechref, Y. LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Anal Bioanal Chem* **409**, 453-466, doi:10.1007/s00216-016-9996-8 (2017).
58 Zhong, X. *et al.* Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry for Quantitative Analysis of Glycans Labeled with Multiplex Carbonyl-Reactive Tandem Mass Tags. *Anal Chem* **87**, 6527-6534, doi:10.1021/acs.analchem.5b01835 (2015).
59 Guttman, M. & Lee, K. K. Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry. *Anal Chem* **88**, 5212-5217, doi:10.1021/acs.analchem.6b00265 (2016).
60 Selman, M. H. J. *et al.* Immunoglobulin G Glycopeptide Profiling by Matrix-Assisted Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* **82**, 1073-1081, doi:10.1021/ac9024413 (2010).
61 Bondt, A. *et al.* Longitudinal monitoring of immunoglobulin A glycosylation during pregnancy by simultaneous MALDI-FTICR-MS analysis of N- and O-glycopeptides. *Sci Rep* **6**, 27955, doi:10.1038/srep27955 (2016).
62 Moh, E. S., Lin, C. H., Thaysen-Andersen, M. & Packer, N. H. Site-Specific N-Glycosylation of Recombinant Pentameric and Hexameric Human IgM. *J Am Soc Mass Spectr* **27**, 1143-1155, doi:10.1007/s13361-016-1378-0 (2016).
63 Hinneburg, H. *et al.* The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *J Am Soc Mass Spectr* **27**, 507-519, doi:10.1007/s13361-015-1308-6 (2016).
64 Selman, M. H. J. *et al.* Fc specific IgG glycosylation profiling by robust nano-reverse phase HPLC-MS using a sheath-flow ESI sprayer interface. *J Proteomics* **75**, 1318-1329, doi:10.1016/j.jprot.2011.11.003 (2012).
65 Wuhrer, M., Catalina, M. I., Deelder, A. M. & Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **849**, 115-128, doi:10.1016/j.jchromb.2006.09.041 (2007).
66 Dodds, E. D. Gas-phase dissociation of glycosylated peptide ions. *Mass Spectrom Rev* **31**, 666-682, doi:10.1002/mas.21344 (2012).
67 Alley, W. R., Jr., Mechref, Y. & Novotny, M. V. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* **23**, 161-170, doi:10.1002/rcm.3850 (2009).
68 Nilsson, J. Liquid chromatography-tandem mass spectrometry-based fragmentation analysis of glycopeptides. *Glycoconj J* **33**, 261-272, doi:10.1007/s10719-016-9649-3 (2016).
69 Nishikaze, T., Kawabata, S. & Tanaka, K. Fragmentation characteristics of deprotonated N-linked glycopeptides: influences of amino acid composition and sequence. *J Am Soc Mass Spectr* **25**, 988-998, doi:10.1007/s13361-014-0854-7 (2014).
70 Nishikaze, T., Nakamura, T., Jinmei, H. & Amano, J. Negative-ion MALDI-MS2 for discrimination of α2,3- and α2,6-sialylation on glycopeptides labeled with a pyrene derivative. *Journal of Chromatography B* **879**, 1419-1428, doi:http://dx.doi.org/10.1016/j.jchromb.2010.10.032 (2011).
71 Nishikaze, T., Kawabata, S. & Tanaka, K. In-Depth Structural Characterization of N-Linked Glycopeptides Using Complete Derivatization for Carboxyl Groups Followed by Positive- and

Negative-Ion Tandem Mass Spectrometry. *Anal Chem* **86**, 5360-5369, doi:10.1021/ac500340t (2014).

72 Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V. & Wisniewski, J. R. Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **4**, 454-465, doi:10.1002/pmic.200300556 (2004).

73 Liu, T. *et al.* Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res* **4**, 2070-2080, doi:10.1021/pr0502065 (2005).

74 Jia, W. *et al.* A strategy for precise and large scale identification of core fucosylated glycoproteins. *Mol Cell Proteomics* **8**, 913-923, doi:10.1074/mcp.M800504-MCP200 (2009).

75 Warrack, B. M. *et al.* Normalization strategies for metabonomic analysis of urine samples. *J Chromatogr B Analyt Technol Biomed Life Sci* **877**, 547-552, doi:10.1016/j.jchromb.2009.01.007 (2009).

76 Wu, Y. & Li, L. Sample normalization methods in quantitative metabolomics. *Journal of chromatography. A* **1430**, 80-95, doi:10.1016/j.chroma.2015.12.007 (2016).

77 Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol* **25**, 21-50, doi:10.1146/annurev.immunol.25.022106.141702 (2007).

78 Zauner, G. *et al.* Glycoproteomic analysis of antibodies. *Mol Cell Proteomics* **12**, 856-865, doi:10.1074/mcp.R112.026005 (2013).

79 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).

80 Freeze, H. H. Genetic defects in the human glycome. *Nat Rev Genet* **7**, 537-551, doi:10.1038/nrg1894 (2006).

81 Bondt, A. *et al.* Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res* **12**, 4522-4531, doi:10.1021/pr400589m (2013).

82 Plomp, R. *et al.* Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health. *Sci Rep* **7**, 12325, doi:10.1038/s41598-017-12495-0 (2017).

83 Kristic, J. *et al.* Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* **69**, 779-789, doi:10.1093/gerona/glt190 (2014).

84 Trbojevic Akmacic, I. *et al.* Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm Bowel Dis* **21**, 1237-1247, doi:10.1097/MIB.0000000000000372 (2015).

85 Borghesi, L. & Milcarek, C. From B cell to plasma cell: regulation of V(D)J recombination and antibody secretion. *Immunol Res* **36**, 27-32, doi:10.1385/IR:36:1:27 (2006).

86 Arnold, J. N. *et al.* Human serum IgM glycosylation - Identification of glycoforms that can bind to mannan-binding lectin. *Journal of Biological Chemistry* **280**, 29080-29087, doi:10.1074/jbc.M504528200 (2005).

87 Mattu, T. S. *et al.* The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fcalpha receptor interactions. *J Biol Chem* **273**, 2260-2272 (1998).

88 Wang, J. *et al.* Fc-glycosylation of IgG1 is modulated by B-cell stimuli. *Mol Cell Proteomics* **10**, M110 004655, doi:10.1074/mcp.M110.004655 (2011).

89 Reusch, D. & Tejada, M. L. Fc glycans of therapeutic antibodies as critical quality attributes. *Glycobiology* **25**, 1325-1334, doi:10.1093/glycob/cwv065 (2015).

90 Wormald, M. R. *et al.* Variations in oligosaccharide-protein interactions in immunoglobulin G determine the site-specific glycosylation profiles and modulate the dynamic motion of the Fc oligosaccharides. *Biochemistry-Us* **36**, 1370-1380, doi:10.1021/bi9621472 (1997).

91 Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. *Nature reviews. Drug discovery* **8**, 226-234, doi:10.1038/nrd2804 (2009).
92 Raymond, C. l. *et al.* in *Glycosylation* (ed Stefana Petrescu) Ch. 17 (InTech, 2012).
93 Wormald, M. R. *et al.* Variations in oligosaccharide-protein interactions in immunoglobulin G determine the site-specific glycosylation profiles and modulate the dynamic motion of the Fc oligosaccharides. *Biochemistry-Us* **36**, 1370-1380, doi:Doi 10.1021/Bi9621472 (1997).
94 Yamaguchi, Y. *et al.* Dynamics of the carbohydrate chains attached to the Fc portion of immunoglobulin G as studied by NMR spectroscopy assisted by selective C-13 labeling of the glycans. *J Biomol NMR* **12**, 385-394, doi:Doi 10.1023/A:1008392229694 (1998).
95 Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res Rev* **12**, 685-698, doi:10.1016/j.arr.2012.02.002 (2013).
96 Catera, M. *et al.* Identification of novel plasma glycosylation-associated markers of aging. *Oncotarget* **7**, 7455-7468, doi:10.18632/oncotarget.7059 (2016).
97 Van Slyke, D. D. *et al.* THE ESTIMATION OF PLASMA PROTEIN CONCENTRATION FROM PLASMA SPECIFIC GRAVITY. *Journal of Biological Chemistry* **183**, 331-347 (1950).
98 Adamczyk, B., Struwe, W. B., Ercan, A., Nigrovic, P. A. & Rudd, P. M. Characterization of Fibrinogen Glycosylation and Its Importance for Serum/Plasma N-Glycome Analysis. *J Proteome Res* **12**, 444-454, doi:10.1021/pr300813h (2013).
99 Fu, D. & van Halbeek, H. N-glycosylation site mapping of human serotransferrin by serial lectin affinity chromatography, fast atom bombardment-mass spectrometry, and 1H nuclear magnetic resonance spectroscopy. *Anal Biochem* **206**, 53-63 (1992).
100 Pompach, P. *et al.* Site-specific glycoforms of haptoglobin in liver cirrhosis and hepatocellular carcinoma. *Mol Cell Proteomics* **12**, 1281-1293, doi:10.1074/mcp.M112.023259 (2013).
101 Imre, T. *et al.* Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals. *J Proteomics* **71**, 186-197, doi:10.1016/j.jprot.2008.04.005 (2008).
102 Olofsson, S. O. *et al.* Apolipoprotein B: structure, biosynthesis and role in the lipoprotein assembly process. *Atherosclerosis* **68**, 1-17 (1987).
103 Olofsson, S. O. & Boren, J. Apolipoprotein B secretory regulation by degradation. *Arterioscler Thromb Vasc Biol* **32**, 1334-1338, doi:10.1161/ATVBAHA.112.251116 (2012).
104 Harazono, A., Kawasaki, N., Kawanishi, T. & Hayakawa, T. Site-specific glycosylation analysis of human apolipoprotein B100 using LC/ESI MS/MS. *Glycobiology* **15**, 447-462, doi:10.1093/glycob/cwi033 (2005).
105 Lottspeich, F., Kellermann, J., Henschen, A., Foertsch, B. & Muller-Esterl, W. The amino acid sequence of the light chain of human high-molecular-mass kininogen. *Eur J Biochem* **152**, 307-314 (1985).
106 Poon, I. K., Patel, K. K., Davis, D. S., Parish, C. R. & Hulett, M. D. Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. *Blood* **117**, 2093-2101, doi:10.1182/blood-2010-09-303842 (2011).
107 Croset, A. *et al.* Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J Biotechnol* **161**, 336-348, doi:10.1016/j.jbiotec.2012.06.038 (2012).
108 Bohm, E. *et al.* Expression of recombinant human coagulation factors VII (rFVII) and IX (rFIX) in various cell types, glycosylation analysis, and pharmacokinetic comparison. *BMC Proc* **5 Suppl 8**, P23, doi:10.1186/1753-6561-5-S8-P23 (2011).
109 West, M. B. *et al.* Analysis of Site-specific Glycosylation of Renal and Hepatic γ-Glutamyl Transpeptidase from Normal Human Tissue. *Journal of Biological Chemistry* **285**, 29511-29524 (2010).
110 Ikegami, K. *et al.* Tissue-specific posttranslational modification allows functional targeting of thyrotropin. *Cell Rep* **9**, 801-810, doi:10.1016/j.celrep.2014.10.006 (2014).

111 Gramolini, A., Lau, E. & Liu, P. P. Identifying Low-Abundance Biomarkers. *Circulation* **134**, 286 (2016).
112 Ruhaak, L. R. *et al.* Targeted biomarker discovery by high throughput glycosylation profiling of human plasma alpha1-antitrypsin and immunoglobulin A. *PLoS One* **8**, e73082, doi:10.1371/journal.pone.0073082 (2013).
113 Ruhaak, L. R. *et al.* Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* **10**, 1667-1674, doi:10.1021/pr1009959 (2011).
114 Ercan, A. *et al.* Estrogens regulate glycosylation of IgG in women and men. *JCI Insight* **2**, e89703, doi:10.1172/jci.insight.89703 (2017).
115 Brinkman-Van der Linden, C. M. *et al.* Oral estrogen treatment induces a decrease in expression of sialyl Lewis x on alpha 1-acid glycoprotein in females and male-to-female transsexuals. *Glycobiology* **6**, 407-412 (1996).
116 Albrecht, S., Unwin, L., Muniyappa, M. & Rudd, P. M. Glycosylation as a marker for inflammatory arthritis. *Cancer Biomark* **14**, 17-28, doi:10.3233/CBM-130373 (2014).
117 Ceciliani, F. & Pocacqua, V. The acute phase protein alpha1-acid glycoprotein: a model for altered glycosylation during diseases. *Curr Protein Pept Sci* **8**, 91-108 (2007).
118 Theodoratou, E. *et al.* Glycosylation of plasma IgG in colorectal cancer prognosis. *Sci Rep* **6**, 28098, doi:10.1038/srep28098 (2016).
119 Miyoshi, E. *et al.* Role of aberrant IgG glycosylation in the pathogenesis of inflammatory bowel disease. *Proteomics Clin Appl* **10**, 384-390, doi:10.1002/prca.201500089 (2016).
120 Lemmers, R. F. H. *et al.* IgG glycan patterns are associated with type 2 diabetes in independent European populations. *Biochim Biophys Acta* **1861**, 2240-2249, doi:10.1016/j.bbagen.2017.06.020 (2017).
121 McCarthy, C. *et al.* The role and importance of glycosylation of acute phase proteins with focus on alpha-1 antitrypsin in acute and chronic inflammatory conditions. *J Proteome Res* **13**, 3131-3143, doi:10.1021/pr500146y (2014).
122 Peracaula, R., Sarrats, A. & Rudd, P. M. Liver proteins as sensor of human malignancies and inflammation. *Proteomics Clin Appl* **4**, 426-431, doi:10.1002/prca.200900170 (2010).
123 Huffman, J. E. *et al.* Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum Mol Genet* **20**, 5000-5011, doi:10.1093/hmg/ddr414 (2011).
124 Zoldos, V. *et al.* Epigenetic silencing of HNF1A associates with changes in the composition of the human plasma N-glycome. *Epigenetics-Us* **7**, 164-172, doi:10.4161/epi.7.2.18918 (2012).
125 Menni, C. *et al.* Glycosylation of immunoglobulin g: role of genetic and epigenetic influences. *PLoS One* **8**, e82558, doi:10.1371/journal.pone.0082558 (2013).
126 Gornik, O. *et al.* Stability of N-glycan profiles in human plasma. *Glycobiology* **19**, 1547-1553, doi:10.1093/glycob/cwp134 (2009).
127 Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448-462, doi:10.1038/nrm3383 (2012).
128 Yang, W. H. *et al.* An intrinsic mechanism of secreted protein aging and turnover. *Proc Natl Acad Sci U S A* **112**, 13657-13662, doi:10.1073/pnas.1515464112 (2015).
129 Thomann, M. *et al.* In vitro glycoengineering of IgG1 and its effect on Fc receptor binding and ADCC activity. *PLoS One* **10**, e0134949, doi:10.1371/journal.pone.0134949 (2015).
130 Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N. & Dwek, R. A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* **343**, 1235681, doi:10.1126/science.1235681 (2014).
131 Dekkers, G. *et al.* Decoding the Human Immunoglobulin G-Glycan Repertoire Reveals a Spectrum of Fc-Receptor- and Complement-Mediated-Effector Activities. *Front Immunol* **8**, 877, doi:10.3389/fimmu.2017.00877 (2017).

132 Thomann, M., Reckermann, K., Reusch, D., Prasser, J. & Tejada, M. L. Fc-galactosylation modulates antibody-dependent cellular cytotoxicity of therapeutic antibodies. *Mol Immunol* **73**, 69-75, doi:10.1016/j.molimm.2016.03.002 (2016).

133 Holst, S. *et al.* N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homebox 1 (CDX1)/Villin mRNA Expression. *Mol Cell Proteomics* **15**, 124-140, doi:10.1074/mcp.M115.051235 (2016).

134 Chung, C. H. *et al.* Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *New Engl J Med* **358**, 1109-1117, doi:Doi 10.1056/Nejmoa074943 (2008).

135 Yang, Q. *et al.* Glycan Remodeling of Human Erythropoietin (EPO) Through Combined Mammalian Cell Engineering and Chemoenzymatic Transglycosylation. *ACS Chem Biol* **12**, 1665-1673, doi:10.1021/acschembio.7b00282 (2017).

136 Yang, Z. *et al.* Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat Biotechnol* **33**, 842-844, doi:10.1038/nbt.3280 (2015).

137 Hamilton, S. R. *et al.* Production of Complex Human Glycoproteins in Yeast. *Science* **301**, 1244 (2003).

Addendum

English summary

Protein glycosylation, the enzymatic attachment of a complex carbohydrate structure to a peptide backbone, is a wide-spread co- and posttranslational modification to be found within nearly all forms of life. Human *N*-linked and *O*-linked protein glycosylation can modulate many aspects of the glycoconjugate, including, among other things, receptor interaction, solubility, half-life, susceptibility to proteolytic digestion, and intra- and interspecies recognition – aspects that apply to human glycosylation as well as to biopharmaceutical glycoproteins. Furthermore, human *N*-glycosylation is known to be readily altered in response to disease (and in some cases causal to it), and aberrant cells such as found in tumor tissue may produce glycan species that differ from those of healthy cells. In all, to study the effect of *N*-glycosylation on human proteins, to ensure uniform and beneficial glycosylation of biopharmaceuticals, and to screen the *N*-glycosylation of accessible biofluids to establish biomarkers for early disease detection and patient stratification, it is of importance to have suitable analysis methodology.

One particularly high-throughput technique to study *N*-glycosylation is matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS). This technique can assess monosaccharide compositions in rapid succession from one sample to the next, making the technique ideal for population studies. However, analysis by MALDI-MS cannot distinguish structural aspects of *N*-glycans without subsequent investigation by tandem MS or by orthogonal analytical methodologies. Furthermore, the study of sialylated *N*-glycan species is complicated by the carboxylic acid function of the sialic acid residues, which facilitates in-source and metastable decay, alters the ionization propensity of the molecule in relation to the nonsialylated *N*-glycan species, and may cause the splitting of MS signals into a variety of different salts variants. Chemical derivatization of *N*-glycans has been employed to overcome these analytical challenges, commonly-used workflows incorporating permethylation or targeted modification of the sialic acid carboxyl groups. **Chapter 2** provides an overview of these existing methodologies for high-throughput glycosylation analysis.

This thesis mainly explores the use of one branch of *N*-glycan derivatization for MALDI-MS analysis, namely the stereospecific esterification of sialic acid residues with either an external nucleophile (in case of α2,6-linked sialylation), or by lactonization (in case of α2,3-linked sialylation). **Chapter 3** details on the development of this protocol, and the exploration of a variety of carboxylic acid activators, catalysts and nucleophiles to arrive at a combination of 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC) and 1-hydroxybenzotriazole (HOBt) in an environment of ethanol. Next to the neutralization of the carboxyl group, the employed chemistry led to the introduction of a mass difference between the sialic acid linkage variants, enabling their mass-based discrimination by MS. The validity of the protocol was confirmed by derivatizing and analyzing sialyllactose

standards with known sialic acid linkages, and by comparing the *N*-glycosylation profiles of fibrinogen between MALDI-TOF-MS and liquid chromatography with fluorescence detection. Finally, we applied the method to the analysis of the challenging human plasma *N*-glycome.

The linkage-specific esterification proved automatable with a few modifications in the protocol, a process that is described in **Chapter 4**. Starting from released *N*-glycan samples, the automated process proved capable of performing the esterification of released complex *N*-glycan samples, of post-derivatization clean-up by hydrophilic-interaction liquid chromatography solid-phase extraction, and of co-crystallizing the samples with matrix on a MALDI target. In all, the process took approximately 2.5 hours for one 96-well plate, while each additional sample plate added an extra hour to the protocol. Measuring nine 96-well plates of a single mixture of released plasma *N*-glycome (three per day across three days) revealed a high repeatability of the protocol, and even more so for derived glycosylation traits, *e.g.*, fucosylation, galactosylation, or the α2,3-sialylation of triantennary species.

A commonly-used animal model to study human disease is the mouse, but not often are the differences between human and murine glycosylation considered in these studies. Furthermore, very little is known about the glycosylation differences between male and female mice, and the variation between mouse strains. **Chapter 5** attempts to ameliorate this situation by comparing by MALDI-TOF-MS the plasma *N*-glycosylation profiles of male and female individuals of the commonly-used mouse strains BALB/c, CD57BL/6, CD-1 and Swiss Webster. In the process, the esterification protocol could be established for sialylation features not commonly found in humans, *i.e.*, *N*-glycolylneuraminic acid (whereas human glycans mainly incorporate *N*-acetylneuraminic acid), sialic acid *O*-acetylation, and *N*-acetylglucosamine-linked sialylation. Male and female mice showed to differ in glycosylation features like fucosylation, galactosylation, and sialylation, and these differences proved to be strain-dependent as well.

While chemical derivatization of sialic acid residues showed beneficial for limiting in-source and metastable decay upon MALDI-MS analysis, increased stability can also be achieved by alternative means. **Chapter 6** describes the application of intermediate-pressure MALDI Fourier-transform ion cyclotron resonance (FTICR) MS to perform the largest mass spectrometric profiling of plasma *N*-glycosylation to date. Although the lack of sialic acid derivatization meant no information on the sialic acid linkage could be obtained, the collisional cooling induced by the intermediate-pressure source did mean that highly sialylated *N*-glycan species could be detected as well. Indexed were the plasma *N*-glycosylation profiles of more than 2000 principally healthy individuals from the Leiden Longevity Study, and the association thereof established with age, sex, and blood markers of inflammation and metabolic health.

The first large-scale application of the automated ethyl esterification protocol with subsequent MALDI-TOF-MS analysis can be found in **Chapter 7**, with the study of the pregnancy-induced amelioration of rheumatoid arthritis (RA). RA is a prevalent autoimmune disease of which the etiology is not yet fully understood, but one remarkable observation is that the severity of the disease lessens during pregnancy and resurges after delivery. As immunoglobulin G (IgG) glycosylation is known to associate with this temporary alleviation of disease activity, this provides reason for a broader study into glycosylation. We profiled the serum *N*-glycosylation of more than 200 women at multiple time points before, during and after pregnancy, and could establish the *N*-glycan features that associated with pregnancy, that differed between healthy individuals and those with RA, and that associated with the temporary amelioration of disease activity during pregnancy.

While MALDI-TOF-MS of esterified glycans provides structural information on sialylation isomers the method is otherwise limited in the discrimination of *N*-glycan structural aspects. For a given glycan sample it is therefore warranted to make use of orthogonal analytical methodologies. **Chapter 8** compares MALDI-TOF-MS (with ethyl esterification) with two other contemporary methods for high-throughput serum *N*-glycosylation analysis, namely hydrophilic-interaction ultra-high-performance liquid chromatography with fluorescence detection (HILIC-UHPLC-FLD; of 2-aminobenzamide-labeled glycans), and multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (xCGE-LIF; of 8-aminopyrene-1,3,6-trisulfonic acid-labeled glycans). The study was performed in an international collaboration ensuring that each method was applied by those laboratories with proven expertise. Furthermore, each partner analyzed the same subset of the prior-mentioned rheumatoid arthritis cohort, meaning that results between laboratories could be directly compared. In all, HILIC-UHPLC-FLD and xCGE-LIF proved more informative for the analysis of diantennary *N*-glycosylation as found on IgG and other immunoglobulins, elucidating structural features such as bisection and 3- *vs.* 6-branch galactosylation, while MALDI-TOF-MS proved more informative for the discrimination of tri- and tetraantennary species and their sialylation properties.

The thesis is concluded by **Chapter 9**, the general discussion. This chapter places the performed research in context to the literature, elaborates on pros and cons of the applied analysis methodologies, and discusses the current state and potential future of the field of glycomics.

Nederlandse samenvatting

Eiwitglycosylering, de enzymatische koppeling van complexe suikerstructuren aan eiwitten, is een wijdverspreide co- en posttranslationele modificatie die nagenoeg binnen elke vorm van leven te vinden is. Zowel *N*- als *O*-glycosylering kunnen vele aspecten van een glycoconjugaat beïnvloeden, zoals de interactie met receptoren, de oplosbaarheid, de halfwaardetijd in bloed, de vatbaarheid voor proteolytische digestie, en de herkenning binnen en tussen diersoorten – allemaal aspecten die zowel gelden voor humane glycosylering als voor farmaceutische eiwitten. Hiernaast is het bekend dat menselijke *N*-glycosylering kan veranderen onder invloed van ziekten (en in sommige gevallen de veroorzaker hiervan kan zijn), en ongezonde cellen, zoals bijvoorbeeld te vinden in tumoren, kunnen een eigen type glycosylering produceren dat anders is dan die van gezonde cellen. Alles bij elkaar genomen, om het effect van *N*-glycosylering op menselijke eiwitten te bestuderen, om zeker te zijn van uniforme en gunstige glycanen op farmaceutische eiwitten, en om biomarkers te vinden voor vroege detectie en onderverdeling van ziekten, is het van belang om de juiste onderzoeksmethodologie te gebruiken.

Een analysemethode met een buitengewoon hoge doorvoersnelheid voor het analyseren van *N*-glycosylering is MALDI-TOF-MS, wat een afkorting is voor *matrix-assisted laser desorption/ionization time-of-flight* massaspectrometrie. Deze techniek kan in zeer korte tijd monosaccharidecomposities bepalen in opeenvolgende monsters, wat ideaal is voor populatiestudies. MALDI-TOF-MS kan daarentegen geen structurele eigenschappen van *N*-glycanen bepalen zonder gebruik te maken van fragmentatieanalyse of andersoortige analytische chemie. Daarnaast is het bestuderen van gesialyleerde *N*-glycanen een lastige bezigheid wegens de carbonzuurgroepen van de siaalzuren, welke de oorzaak kunnen zijn van metastabiel verval, kunnen leiden tot een andere ionisatie dan die van neutrale glycanen, en welke ertoe kunnen leiden dat een enkel molecuul opsplitst in verschillende zoutvarianten. Chemische modificatie van *N*-glycanen, waarvan permethylering of de specifieke modificatie van de carbonzuurgroepen veelgebruikte varianten zijn, biedt de mogelijkheid om deze analytische problemen weg te nemen. **Hoofdstuk 2** schetst een overzicht van de hedendaagse methoden van glycosyleringsanalyse met hoge doorvoersnelheid.

Dit proefschrift houdt zich voornamelijk bezig met een specifieke tak van *N*-glycaan derivatisering voor MALDI-MS, namelijk de stereospecifieke verestering van siaalzuren met ofwel een externe nucleofiel (in het geval van α2,6-gebonden siaalzuren), ofwel door lactonisering (in het geval van α2,3-gebonden siaalzuren). In **Hoofdstuk 3** zijn de details beschreven van het ontwikkelen van dit protocol, en het uittesten van carbonzuuractivators, katalysatoren en nucleofielen om uiteindelijk uit te komen op een combinatie van de chemicaliën 1-ethyl-3-(3-(dimethylamino)propyl)carbodi-imide (EDC) en 1-hydroxybenzotriazool (HOBt) in ethanol. Naast het neutraliseren van de carbonzuren

leidde de toegepaste chemie tot een massaverschil tussen de verschillende siaalzuurisomeren, hetgeen met massaspectrometrie waar te nemen is. De validiteit van het protocol werd bevestigd door sialyllactosemonsters met bekende siaalzuurisomeren te derivatiseren en analyseren, en door profielen van fibrinogeenglycosylering van MALDI-TOF-MS en vloeistofchromatografie met fluorescentiedetectie te vergelijken.

Het protocol voor stereospecifieke verestering kon met behulp van enkele aanpassingen geautomatiseerd worden, een proces dat beschreven is in **Hoofdstuk 4**. Het automatische proces stelde ons in staat om de verestering toe te passen op losse *N*-glycanen, om de glycanen op te zuiveren met hydrofiele interactie-vloeistofchromatografie vastefase-extractie, en om de monsters met matrix te kristalliseren op een MALDI plaat. Alles bij elkaar duurde het proces 2,5 uur voor een enkele 96-putjesplaat, en een uur extra voor iedere additionele plaat. De goede herhaalbaarheid van het protocol werd bevestigd door negen van dit soort platen te meten, ieder gevuld met hetzelfde plasmamonster, en afgeleide glycosyleringseigenschappen (zoals fucosylering, galactosylering, en α2,3-gebonden sialylering van structuren met drie antennes) bleken een nog betere herhaalbaarheid te hebben.

De muis is een veelgebruikt proefdiermodel voor het bestuderen van menselijke ziekten, maar binnen dit soort studies wordt niet vaak rekening gehouden met de glycosyleringsverschillen tussen mensen en muizen. Hiernaast is er ook nog vrij weinig bekend over hoe glycosylering zich verhoudt tussen de geslachten, en of er verschil is tussen muizenstammen. **Hoofdstuk 5** tracht bij te dragen aan deze kennis door met MALDI-TOF-MS de plasmaglycosyleringsprofielen te bepalen voor zowel mannetjes- als vrouwtjesmuizen van de veelgebruikte stammen BALB/c, CD57BL/6, CD-1 en Swiss Webster. De studie gaf ook de mogelijkheid om de effecten te bestuderen van het veresteringsprotocol op niet-menselijke glycosylering, *i.e.*, *N*-glycolylneuraminezuren (mensen hebben *N*-acetylneuraminezuren), *O*-acetylering, en siaalzuren verbonden met *N*-acetylglucosamines. Mannetjes- en vrouwtjesmuizen bleken onder andere te verschillen in fucosylering, galactosylering en sialylering, en de verschillen bleken afhankelijk te zijn van de muizenstam.

Alhoewel chemische aanpassing van siaalzuren de stabiliteit van glycanen tijdens MALDI-MS ten goede komt, is dit niet de enige manier waarop stabiliteit gewaarborgd kan worden. **Hoofdstuk 6** beschrijft het gebruik van MALDI *Fourier-transform* ion cyclotron resonantie MS voor het uitvoeren van de – tot nog toe grootste – massaspectrometrische studie naar plasma *N*-glycosylering. Het gebrek aan siaalzuurmodificatie betekende ook een gebrek aan verbindingsinformatie, maar het gebruik van een ionenbron met hogere druk (en de resulterende koeling die ionen daarin ondervinden) leidden ertoe dat hooggesialyleerde moleculen nog steeds gedetecteerd konden worden. Binnen de studie werd de plasma *N*-glycosylering van meer dan 2000 mensen bekeken, en de resultaten konden worden

vergeleken met leeftijd, geslacht, en bloedwaarden voor ontsteking en metabolische gezondheid.

De eerste grootschalige toepassing van het geautomatiseerde veresteringsprotocol (gevolgd door MALDI-TOF-MS) is te vinden in **Hoofdstuk 7**, waar een studie wordt beschreven naar de spontane verbetering van reumatoïde artritis (RA) tijdens zwangerschap. RA is een veelvoorkomende auto-immuunziekte waarvan de oorzaak nog niet geheel duidelijk is, maar waarvan de ernst van de ziekte opmerkelijk genoeg afneemt tijdens de zwangerschap en weer toeneemt na de bevalling. Het is bekend dat de glycosylering van immunoglobuline G (IgG) associeert met deze tijdelijke verlichting van ziekteactiviteit, wat de aanleiding was om een bredere studie betreffende glycosylering uit te voeren. Daarom hebben wij de serum *N*-glycosylering bestudeerd van meer dan 200 vrouwen op verschillende tijdpunten voor, tijdens en na zwangerschap, en konden we de *N*-glycaanpatronen vaststellen die veranderden met zwangerschap, die verschilden tussen gezonde mensen en mensen met RA, en die associeerden met de tijdelijke afname van ziekteactiviteit gedurende zwangerschap.

Alhoewel de MALDI-TOF-MS analyse van veresterde glycanen structurele informatie oplevert omtrent de siaalzuurisomeren, is de methode beperkt informatief voor andere structurele eigenschappen van glycosylering. Het kan daarom wenselijk zijn om gebruik te maken van alternatieve analysemethodologie. **Hoofdstuk 8** vergelijkt MALDI-TOF-MS (van veresterde glycanen) met twee andere hedendaagse methoden voor *N*-glycaananalyse met hoge doorvoersnelheid, namelijk hydrofiele-interactie ultrahogedruk vloeistofchromatografie met fluorescentiedetectie (HILIC-UHPLC-FLD; van glycanen gelabeld met 2-aminobenzamide), en multiplex capillaire gelelektroforese met laser-geïnduceerde fluorescentiedetectie (xCGE-LIF; van glycanen gelabeld met 8-aminopyreen-1,3,6-trisulfonzuur). De studie werd uitgevoerd in het kader van een internationale samenwerking, wat inhield dat iedere methode kon worden uitgevoerd door een laboratorium met aantoonbare expertise op het gebied van de gebruikte methode. En, omdat iedere samenwerkingspartner dezelfde monsters analyseerde, was het mogelijk om de resultaten van de verschillende methoden direct met elkaar te vergelijken. Alles bij elkaar genomen bleken HILIC-UHPLC-FLD en xCGE-LIF het meest informatief voor de analyse van glycanen met twee antennes of minder, zoals te vinden op IgG en andere immunoglobulinen (door bijvoorbeeld 3- en 6-tak galactosylering van elkaar te onderscheiden, evenals tweesplijtende *N*-acetylglucosamines), terwijl MALDI-TOF-MS beter in staat was om glycanen met drie en vier antennes en de sialylering daarvan te onderscheiden.

Het proefschrift wordt afgesloten door **Hoofdstuk 9**, de algemene discussie. Dit hoofdstuk plaatst het beschreven onderzoek in het kader van de literatuur, beschrijft de voor- en

nadelen van de gebruikte analysemethoden, en bediscussieert de huidige staat en mogelijke toekomst van massaspectrometrisch glycosyleringsonderzoek.

Acknowledgements

First of all, I would like to thank Manfred, my promotor and guide. I could not have wished for better supervision. You have allowed me to find my place in the group and in research, and for that, I am very grateful.

I also want to thank André, my second promotor, who has shown faith in me from the first day. Although retirement has limited our eventual interaction, I very much appreciate your time and dedication to bring me to the finish line.

Of course I would like to thank the fantastic group in Leiden (although whose constituents have changed and grown to such a degree that individual naming is becoming quite the affair). I can hardly believe my luck to have stumbled upon you. I regard many of you as friend, and because of you, I rarely consider the things I do to be work.

Thank you, very much, those of you that wanted me at their side during their own defenses, Niels, Rosina, Albert, and Bas. Paying it forward, I am grateful to my own paranymfs Noortje and David, whose scientific and personal lifes kept crossing with mine.

Furthermore, I would like to thank my collaborators, and the many people I was allowed to work with abroad, which took me in and showed me the larger world of science. Rene, Erdmann and the others in Magdeburg, Maja, Genadij, Gordan and the others in Zagreb, and Mili and Daryl and the others in Culham. It is great to be a part of the international community you have created.

Op persoonlijk vlak wil ik mijn reisgenoten bedanken: Niels (dezelfde), Remco, Mark, Timo, Thijs en Julius. Onze expedities door de wildernis hebben een belangrijke indruk op mij achtergelaten, inclusief faliekante misrekeningen in de voedselvoorraad, geothermisch smeltende schoenen, en de viering van het verdaguurtje. Alleen het vangen van die orka is er nog niet van gekomen.

Vanzelfsprekend kan dit dankwoord niet compleet zijn zonder mijn ouders te noemen. Jullie hebben hard gewerkt om mij, onzelfzuchtig, alle kansen in het leven te geven, en daar ben ik jullie zeer dankbaar voor. Het boekje dat voor jullie ligt is minstens zoveel jullie verdienste als het mijne.

Bovenal wil ik Wietske bedanken, zonder wiens aanwezigheid dit hele wetenschappelijke avontuur waarschijnlijk niet had plaatsgevonden. Wat ik doe, dat doe ik voor jou. Jij bent de drijvende kracht in mijn leven.

Curriculum Vitae

Karli Robert Reiding was born October 7th 1985 in Rotterdam, The Netherlands. He quickly developed a broad interest in understanding the world, and enjoyed his explorations into the natural sciences, history, and philosophy.

After finishing his primary and secondary education he studied Biomedical Sciences at Leiden University, obtaining his Bachelor's degree in 2008 and his Master's degree in 2011. The education was broadly focused on different aspects of biomedical research, including biological knowledge and technical skills, as well as study design, statistical analysis, and the art of writing and presenting. During his Master's, Karli specialized into the research track. As optional courses, he found an interest in bioinformatics and biostatistics, giving him experience with several of the programming languages commonly found in science, including R and Python. In addition, he followed a minor in science-based business, which detailed on project management, strategic marketing, financial management, and other aspects of business administration.

During his internships, Karli had the opportunity to gain hands-on experience with a wide variety of subjects. His graduation internship focused on the identification of new genes involved in mismatch repair at the Leiden University Medical Center, Department of Toxicology, within the group of Prof.dr. Marcel Tijsterman. Two other internships covered the (epi)genetic mechanisms behind protein S deficiency (Thrombosis and Hemostasis, Prof.dr. Pieter H. Reitsma), and HLA-G expression in throphoblast cells (Immunohematology and Blood Transfusion, Prof.dr. Peter J. van den Elsen).

Karli proceeded to do his PhD education at the Leiden University Medical Center Biomolecular Mass Spectrometry Unit (later the Center for Proteomics and Metabolomics) under the supervision of Prof.dr. André M. Deelder and Prof.dr. Manfred Wuhrer. His research focused on the development and application of high-throughput methodology for mass spectrometric glycosylation analysis, which has led to more than two dozen publications in peer-reviewed journals, as well as a patent application on linkage-specific sialic acid derivatization for MALDI-MS analysis.

Currently, Karli is employed as a post-doctoral researcher at Biomolecular Mass Spectrometry and Proteomics, Utrecht University, within the group of Prof.dr. Albert Heck, where he is deepening his understanding of proteomics and mass spectrometric native (glyco)protein analysis.

Publications and submitted manuscripts

1. **Reiding, K. R.**, Blank, D., Kuijper, D. M., Deelder, A. M., and Wuhrer, M. (2014) High-throughput profiling of protein *N*-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* 86, 5784-5793.

2. Bondt, A., Rombouts, Y., Selman, M. H., Hensbergen, P. J., **Reiding, K. R.**, Hazes, J. M., Dolhain, R. J., and Wuhrer, M. (2014) Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics* 13, 3029-3039.

3. **Reiding, K. R.**, Bladergroen, M. R., Hipgrave Ederveen, A. L., Vreeker, G. C., Clerc, F., Holst, S., Bondt, A., Wuhrer, M., and van der Burgt, Y. E. (2015) Automation of high-throughput mass spectrometry-based plasma *N*-glycome analysis with linkage-specific sialic acid esterification. *J Proteome Res* 14, 4080-4086.

4. Borelli, V., Vanhooren, V., Lonardi, E., **Reiding, K. R.**, Capri, M., Libert, C., Garagnani, P., Salvioli, S., Franceschi, C., and Wuhrer, M. (2015) Plasma *N*-glycome signature of Down Syndrome. *J Proteome Res* 14, 4232-4245.

5. de Haan, N., **Reiding, K. R.**, Haberger, M., Reusch, D., Falck, D., and Wuhrer, M. (2015) Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. *Anal Chem* 87, 8284-8291.

6. Jansen, B. C., **Reiding, K. R.**, Bondt, A., Hipgrave Ederveen, A. L., Palmblad, M., Falck, D., and Wuhrer, M. (2015) MassyTools: a high-throughput targeted data processing tool for relative quantitation and quality control developed for glycomic and glycoproteomic MALDI-MS. *J Proteome Res* 14, 5088-5098.

7. Shubhakar, A., **Reiding, K. R.**, Gardner, R. A., Spencer, D. I., Fernandes, D. L., and Wuhrer, M. (2015) High-throughput analysis and automation for glycomics studies. *Chromatographia* 78, 321-333.

8. Breen, L. D., Pucic-Bakovic, M., Vuckovic, F., **Reiding, K.R.**, Trbojevic-Akmacic, I., Srajer Gajdosik, M., Cook, M. I., Lopez, M. J., Wuhrer, M., Camara, L. M., Andjelkovic, U., Dupuy, D. E., and Josic, D. (2016) IgG and IgM glycosylation patterns in patients undergoing image-guided tumor ablation. *Biochim Biophys Acta* 1860, 1786-1794.

9. **Reiding, K. R.**, Lonardi, E., Hipgrave Ederveen, A. L., and Wuhrer, M. (2016) Ethyl esterification for MALDI-MS analysis of protein glycosylation. *Methods Mol Biol* 1394, 151-162.

10. Clerc, F., **Reiding, K. R.**, Jansen, B. C., Kammeijer, G. S., Bondt, A., and Wuhrer, M. (2016) Human plasma protein *N*-glycosylation. *Glycoconj J* 33, 309-343.

11. de Haan, N., **Reiding, K. R.**, Driessen, G., van der Burg, M., and Wuhrer, M. (2016) Changes in healthy human IgG Fc-glycosylation after birth and during early childhood. *J Proteome Res* 15, 1853-1861.

12. Jansen, B. C., Bondt, A., **Reiding, K. R.**, Lonardi, E., de Jong, C. J., Falck, D., Kammeijer, G. S., Dolhain, R. J., Rombouts, Y., and Wuhrer, M. (2016) Pregnancy-associated serum *N*-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep* 6, 23296.

13. Jansen, B. C., Bondt, A., **Reiding, K. R.**, Scherjon, S. A., Vidarsson, G., and Wuhrer, M. (2016) MALDI-TOF-MS reveals differential *N*-linked plasma- and IgG-glycosylation profiles between mothers and their newborns. *Sci Rep* 6, 34001.

14. Klasic, M., Kristic, J., Korac, P., Horvat, T., Markulin, D., Vojta, A., **Reiding, K. R.**, Wuhrer, M., Lauc, G., and Zoldos, V. (2016) DNA hypomethylation upregulates expression of the MGAT3 gene in HepG2 cells and leads to changes in *N*-glycosylation of secreted glycoproteins. *Sci Rep* 6, 24363.

15. Kozak, R. P., Urbanowicz, P. A., Punyadeera, C., **Reiding, K. R.**, Jansen, B. C., Royle, L., Spencer, D. I., Fernandes, D. L., and Wuhrer, M. (2016) Variation of human salivary O-glycome. *PLoS One* 11, e0162824.

16. Meyer, S., Nederend, M., Jansen, J. H., **Reiding, K. R.**, Jacobino, S. R., Meeldijk, J., Bovenschen, N., Wuhrer, M., Valerius, T., Ubink, R., Boross, P., Rouwendal, G., and Leusen, J. H. (2016) Improved in vivo anti-tumor effects of IgA-Her2 antibodies through half-life extension and serum exposure enhancement by FcRn targeting. *MAbs* 8, 87-98.

17. **Reiding, K. R.**, Hipgrave Ederveen, A. L., Rombouts, Y., and Wuhrer, M. (2016) Murine plasma *N*-glycosylation traits associated with sex and strain. *J Proteome Res* 15, 3489-3499.

18. Shubhakar, A., Kozak, R. P., **Reiding, K. R.**, Royle, L., Spencer, D. I., Fernandes, D. L., and Wuhrer, M. (2016) Automated high-throughput permethylation for glycosylation analysis of biologics using MALDI-TOF-MS. *Anal Chem* 88, 8562-8569.

19. Rombouts, Y., Jonasdottir, H. S., Hipgrave Ederveen, A. L., **Reiding, K. R.**, Jansen, B. C., Freysdottir, J., Hardardottir, I., Ioan-Facsinay, A., Giera, M., and Wuhrer, M. (2016) Acute phase inflammation is characterized by rapid changes in plasma/peritoneal fluid *N*-glycosylation in mice. *Glycoconj J* 33, 457-470.

20. Rouwendal, G. J., van der Lee, M. M., Meyer, S., **Reiding, K. R.**, Schouten, J., de Roo, G., Egging, D. F., Leusen, J. H., Boross, P., Wuhrer, M., Verheijden, G. F., Dokter, W. H., Timmers, M., and Ubink, R. (2016) A comparison of anti-HER2 IgA and IgG1 in vivo efficacy is facilitated by high *N*-glycan sialylation of the IgA. *MAbs* 8, 74-86.

21. de Haan, N., **Reiding, K. R.**, and Wuhrer, M. (2017) Sialic Acid Derivatization for the rapid subclass- and sialic acid linkage-specific MALDI-TOF-MS analysis of IgG Fc-glycopeptides. *Methods Mol Biol* 1503, 49-62.

22. **Reiding, K. R.**, Ruhaak, L. R., Uh, H. W., El Bouhaddani, S., van den Akker, E. B., Plomp, R., McDonnell, L. A., Houwing-Duistermaat, J. J., Slagboom, P. E., Beekman, M., and Wuhrer, M. (2017) Human plasma *N*-glycosylation as analyzed

by MALDI-FTICR-MS associates with markers of inflammation and metabolic health. *Mol Cell Proteomics* 16, 228-242.

23. de Haan, N., **Reiding, K. R.**, Kristic, J., Hipgrave Ederveen, A. L., Lauc, G., Wuhrer, M. (2017) The *N*-glycosylation of mouse IgG-Fc differs between IgG subclasses and strains. *Front Immunol* 8, 608.

24. Plomp, R., Ruhaak, L. R., Uh, H. W., **Reiding, K. R.**, Selman, M. H., Houwing-Duistermaat, J. J., Slagboom, P. E., Beekman, M., Wuhrer, M. (2017) Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health. *Sci Rep* 7, 12325.

25. Ferrantelli, E., Farhat, K., Hipgrave Ederveen, A. L., **Reiding, K. R.**, Beelen, R. H. J., van Ittersum, F. J., Wuhrer, M., Dotz, V. (2018) Effluent and serum protein *N*-glycosylation is associated with inflammation and transport characteristics in peritoneal dialysis patients. *Sci Rep* 8, 979.

26. **Reiding, K. R.**, Vreeker, G. C. M., Bondt, A., Bladergroen, M. R., Hazes, J. M. W., van der Burgt, Y. E. M., Wuhrer, M., Dolhain, R. J. E. M. (2018) Serum protein *N*-glycosylation changes with rheumatoid arthritis disease activity during and after pregnancy. *Front Med* 4, 241.

27. Jacobino, S. R., Nederend, M., Reijneveld, J. F., Augustijn, D., Jansen, J. H. M., Meeldijk, J., **Reiding, K. R.**, Wuhrer, M., Coenjaerts, F. E. J., Hack, C. E., Bont, L. J., Leusen, J. H. W. (2018) Reformatting palivizumab and motavizumab from IgG to human IgA impairs their efficacy against RSV infection *in vitro* and *in vivo*. *MAbs*. *Manuscript accepted*.

28. **Reiding, K. R.**, Bondt, A., Hennig, R., Gardner, R. A., O'Flaherty, R., Trbojević-Akmačić, I., Shubhakar, A., Hazes, J. M. W., Reichl, U., Fernandes, D. L., Pučić-Baković, M., Rapp, E., Spencer, D. I. R., Dolhain, R. J. E. M., Rudd, P. M., Lauc, G., Wuhrer, M. High-throughput serum *N*-glycomics: method comparison and application to study rheumatoid arthritis and pregnancy-associated changes. *Manuscript under review*.

29. Clerc, F., Novokmet, M., Dotz, V., **Reiding, K. R.**, de Haan, N., Kammeijer, G. S. M., Dalebout, H., Bladergroen, M. R., Vukovic, F., Rapp, E., IBD-BIOM consortium, Targan, S. R., Barron, G., Manetti, N., Latiano, A., McGovern, D. P. B., Annese, V., Lauc, G., Wuhrer, M. Plasma *N*-glycan signatures of inflammatory bowel diseases. *Manuscript submitted*.

30. de Haan, N., Boeddha, N. P., Ekinci, E., **Reiding, K. R.**, Emonts, M., Hazelzet, J. A., Wuhrer, M., Driessen, G. J. Differences in IgG Fc glycosylation are associated with outcome of pediatric meningococcal sepsis. *Manuscript submitted*.

31. Dotz, V., Lemmers, R. F. H., **Reiding, K. R.**, Hipgrave Ederveen, A. L., Lieverse, A. G., Mulder, M. T., Sijbrands, E. J. G., Wuhrer, M. van Hoek, M. Plasma protein *N*-glycan signatures of type 2 diabetes. *Manuscript submitted*.