



Universiteit  
Leiden  
The Netherlands

## **Glycosidases as an analytical tool in glycomics assays**

Rebello, O.D.

### **Citation**

Rebello, O. D. (2022, October 13). *Glycosidases as an analytical tool in glycomics assays*. Retrieved from <https://hdl.handle.net/1887/3480319>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3480319>

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

# **Glycosidases as an analytical tool in glycomics assays**

ISBN: 978-90-832830-0-5

Copyright © 2022 by Osmond Rebello.

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, and at Ludger Ltd., United Kingdom.

The work was supported by the European Union Horizon 2020 Program GlySign, grant number 722095.

Cover design: Osmond Rebello

Printing: Proefschriftprinten.nl

# **Glycosidases as an analytical tool in glycomics assays**

## **Proefschrift**

ter verkrijging van  
de graad van doctor aan de Universiteit Leiden,  
op gezag van rector magnificus prof. dr. ir. H. Bijl,  
volgens besluit van het college voor promoties  
te verdedigen op donderdag 13 oktober 2022  
klokke 15:00 uur

door

**Osmond Darrel Rebello**

geboren te Dubai, U.A.E.

in 1992

**Promotor:** Prof. dr. Manfred Wuhrer

**Copromotor:** Dr. David Falck

**Leden promotiecomissie:** Prof. dr. Cornelis H. Hokke

Dr. L. Renee Ruhaak

Prof. dr. Sabine Flitsch,  
Manchester Institute of Biotechnology,  
Manchester University, Manchester,  
United Kingdom

Dr. ing. Tom Wennekes,  
Department of Chemical Biology,  
Utrecht University, Utrecht,  
The Netherlands

# Table of Contents

<b>Chapter 1 – Introduction</b>	<b>7</b>
<b>Chapter 2 – Fucosidases from the human gut symbiont <i>Ruminococcus gnavus</i></b>	<b>31</b>
<b>Chapter 3 – Crystal structure of a GH20 <i>N</i>-acetylglucosaminidase with specificity for bisecting <i>N</i>-acetylglucoamines on <i>N</i>-glycans</b>	<b>65</b>
<b>Chapter 4 – A matrix assisted laser desorption/ionization mass spectrometry assay for the relative quantitation of antennary fucosylated <i>N</i>-glycans in human plasma</b>	<b>81</b>
<b>Chapter 5 – A novel glycosidase plate-based assay for the quantification of galactosylation and sialylation on human IgG</b>	<b>109</b>
<b>Chapter 6 – Discussion and perspectives</b>	<b>133</b>
<b>Addendum</b>	<b>151</b>
Summary	<b>152</b>
Nederlandse samenvatting	<b>154</b>
Acknowledgements	<b>157</b>
Abbreviations	<b>158</b>
Curriculum vitae	<b>160</b>
List of publications	<b>161</b>



# **Chapter 1**

## **Introduction**





Analytical assay development, particularly pertaining to glycomics, is an exciting amalgam of biology, chemistry and engineering. Besides academic research in natural and medical sciences, glycomics assays have immense importance in industrial applications such as in quality control (QC) and quality assurance (QA) of glycoproteins. An up-coming industrial and clinical application is the high-throughput glycan profiling of clinical samples, such as plasma, for identifying disease associations. These glycomics assays are often based on chromatographic and mass spectrometric instrumentation. Thus, they create a requirement of instrumentation infrastructure as well as technical skills which are both not always readily available. This creates a demand in industry for the development of glycomics assays that have a low infrastructure cost as well as minimal training requirements and that are user-friendly. With these objectives in focus, this thesis develops novel exoglycosidase-based high-throughput glycomics assays for use in industrial glycan profiling. In doing so, this thesis also contributes to the development of potential products, such as glycomics kits.

## Glycosylation in human health

Protein glycosylation is a post-translational modification that can be characterised as either *N*-type or *O*-type, depending on glycan structures and their attachment to the protein backbone [1, 2]. Although *N*-glycans have diverse molecular structures, owing to variations in their composition and linkage, they have a fundamental pentasaccharide core structure composing of two *N*-acetylglucosamines (GlcNAcs) and three mannoses (Man). The core asparagine-linked GlcNAc may or may not be decorated with a fucose (Fuc) residue [3]. Variations in the outer branches, extending from the core's Man triose arms, allow for further categorising of the glycans as either, oligomannose, complex or hybrid type. Complex glycans contribute to about 97% of total plasma protein *N*-glycosylation (TPNG) [4]. These complex glycans can be di-, tri- or tetraantennary depending on the branching GlcNAc motifs which can be  $\beta(1-2/4/6)$  linked [1]. These antennary GlcNAcs can be further decorated in an orderly fashion with galactose (Gal) and sialic acid (Sia) residues, and to lesser extent Fuc residues [5]. As far as humans are concerned, these Gal residues are  $\beta(1-3/4)$  linked while the Sia residues are exclusively 5-*N*-acetylneuraminic acid (Neu5Ac) of  $\alpha(2-3/6)$  linkage [6]. Antennary fucosylation can be either  $\alpha(1-3/4)$  linked to the antennary GlcNAcs or  $\alpha(1-2)$  linked to Gal [7].

Glycosylation plays an important role in protein folding and structural maintenance [8]. Because glycans have multiple hydroxyl groups, they can influence the hydrophilicity of the protein surface which is an important physical property for protein stability [8-10]. These glycans can also form hydrogen bonds to the peptide residues which may aid in maintaining protein conformation. Besides intra-molecular functionality, glycans may also serve as ligands for inter-molecular interaction or modify protein-protein interactions and thus contribute to inter-molecular crosstalk. A previous review has described the glycosylation on a selection of 24 plasma proteins which contribute to nearly half of the plasma protein concentration [11]. Alterations in plasma protein glycosylation may influence protein function. For example, changes in IgG glycosylation during an inflammatory response are associated with the immunoregulatory functionality of the IgG molecule [12, 13].

A large repertoire of medical and clinical research has extensively demonstrated the ability to correlate plasma glycosylation to the pathophysiological state of patient populations [1]. For example, in TPNG, changes in antennary branching, sialylation and antennary fucosylation of glycans have been associated with inflammatory diseases such as diabetes [14, 15], cancers [16-18], rheumatoid arthritis [19] and inflammatory bowel disease [20]. TPNG changes have also been correlated to mental conditions such as ADHD in children [21] and Down syndrome [22]. Interestingly, TPNG changes can also be correlated to age [23, 24] and standard of living of a population [25]. These associations can be explained by external factors [25] or genetic factors [7] that cause a change to the patient's physiological state. These physiological alterations may further lead to changes in the glycosylation processes and hence a shift in the glycosylation pattern on proteins and lipids. Furthermore, since glycosylation on proteins can influence their stability or is involved in inter-molecular interactions, these glycosylation changes could sometimes further escalate the disease state [26].

In addition to the importance of glycosylation in medical and clinical sciences, it also has a significant importance in biopharmaceutical production. Glycosylation does not only influence the stability of a therapeutic protein, but can also influence the immune response of a recipient of a biopharmaceutical [27, 28]. For example, Cetuximab is a recombinant IgG1 antibody which targets the epidermal growth factor receptor and was approved for treatment of colorectal cancer. However, this antibody was associated with a high incidence of hypersensitivity among patients which was routed to the presence of a non-human Gal- $\alpha$ (1-3)-Gal motif [29]. Hence, knowing and monitoring the precise glycosylation profile of biopharmaceuticals is vital. In fact, glycosylation is a critical quality attribute (CQA) which must be monitored and engineered to ensure safety for many biopharmaceutical products [30, 31].

Glycan profiling is heavily dependent on chromatographic and spectrometric techniques. This will be elaborated in the next section. The use of exoglycosidases for glycan sequencing was and continues to be a vital approach that complements these analytical instrumentation techniques for glycan profiling [32]. Exoglycosidases are enzymes that can hydrolyse specific glycan residues from the non-reducing end of glycans, with varying degrees of monosaccharide and linkage specificity [33]. Although various workflows are published, the principle remains the same. Released glycans are treated by an assortment of exoglycosidases, followed by a liquid chromatography (LC) - fluorescence detection (FLD) analysis of the cleaved products [34, 35]. By correlating chromatographic peak shifts between exoglycosidase-treated and non-treated samples, one can sequence a glycan by consecutively removing residues. The substrate specificity of exoglycosidases, in terms of glycan residues and their linkages, can range from narrow to broad, thus creating the need to source exoglycosidase of high substrate specificity for the purpose of glycan profiling [35]. To answer this demand, one approach is bio-engineering well-characterised exoglycosidases of a certain activity to obtain an enhancement of its activity or selectivity [33, 36, 37]. Another, more widely used approach, is meta-genomic screening of complex glycan environments such as gut microbiota, for discovering exoglycosidases of desired specificities [38-40]. This approach is based on an organism's need to consume the available nutrients, in this case the need to use glycosidases to consume the glycans in their environment. This approach has proven

valuable in discovering glycosidases [40, 41] that have potential in glycan profiling [34, 35] and glycoengineering [42].

## Glycomics assays

Glycan variations are owed to compositional and linkage differences [1, 43]. Most glycomics assays are based on identifying both these compositional and linkage variations, usually in complex samples containing multiple glycan structures [44-48]. Although some glycomics assays are microarray-based binding assays that can identify glycan epitopes [49, 50], most glycomics assays rely on chromatography-, electrophoresis- and/or mass spectrometry (MS)-based techniques [51-55]. These glycomics assays have co-advanced with the development of instrumental analytical techniques. Advances in liquid chromatography (LC) column chemistry and silanol neutralisation have proven vital in chromatographic and capillary electrophoresis separations of glycan isomers [56, 57]. Additionally, advances in mass analysers and ion optics in components such as ion traps, quadrupoles, collision induced dissociation (CID) cells and ion mobility mass spectrometry (IMS) drift tubes have aided glycan structural identification by MS [58-62]. Furthermore, development of chemical derivatisation techniques such as sialic acid esterification [63], hydroxyl group methylation [64-66] and reducing-end tag conjugations [67] have complemented these instrumental techniques for glycan analysis. Finally, advances in liquid handling robotic platforms have contributed to high-throughput sample processing which is important in clinical glycomics [51, 68].

## Microarray-based binding assays

The microarray-based binding assay is a versatile tool in glycobiochemistry research, and it principally relies on glycan-ligand interactions [50]. Generally, this technique, when established in a laboratory, can be used for performing assays in a high-throughput manner with limited technical skills required. These microarray-based assays can be widely classified as either glycan arrays or lectin arrays. The former involves glycans / glycoconjugates immobilised onto a solid surface such as a microarray with the lectins or antibodies introduced in solvent phase, while the latter array is reversed in format [69]. By using known glycan structures, the assay could be used to screen for lectins or antibodies with certain specificities. In contrast, by using immobilized lectins or antibodies with known binding specificities, unknown glycan profiles in a sample can be screened. Furthermore, sandwich style lectin arrays, which use two or more lectins or antibodies of known binding specificities, have allowed for the identification of both glycan motifs as well as the associated protein or lipid on the glycoconjugates [70].

The measurements are usually performed by fluorescence spectroscopy. The solution phase components which may be either glycans or lectins are usually conjugated to a fluorophore or a molecular ligand probe such as biotin. The latter serves to bind an enzyme that could further catalyse the detectable fluorescence or chemiluminescent reaction. Label free assays have also been developed which use surface plasmon resonance (SPR) for real-time detection

of binding [71, 72]. Similar assays with MS based detection have provided a second dimension,  $m/z$  analysis, for identifying glycan structures [50].

Lectin microarrays have proven valuable in a number of research fields in glycobiology [73], but of relevance here is their use in TPNG analysis for disease association. These assays are usually applied for in-depth profiling of glycoforms for biomarker discovery [70, 74, 75]. The approach often involves identifying glycoforms on specific target proteins which require several processing steps limiting the throughput of the assay as compared to other glycomics techniques. Since the assay depends on glycan-lectin/antibody interactions, the glycan identification specificity of the assay is limited to the specificity of the detecting lectin/antibody. As mentioned above, when these assays are established in a laboratory, they may not require extensive technical skills. However, preparation and adaptation of these assays for new applications is not trivial. Adapting the assay to identify a new target glycan motif requires much effort in lectin/antibody discovery and optimisation of binding conditions.

## **Chromatography- and mass spectrometry-based assays**

The techniques of chromatography, capillary electrophoresis (CE) and MS have been around for many decades, and they have always played central roles in profiling glycan structures [60]. Table 1 compares some of the common techniques used for glycomics applications in academia and industry. Glycan profiling by LC-FLD is widely accepted in industrial laboratories for quality control (QC) and quality assurance (QA) purposes because conclusive information on glycan profiles can be obtained whilst having a relatively low infrastructural cost of instrumentation [76]. More importantly, among analytical chemists, chromatograms are easily comprehensible and can even be processed rapidly in a semi-automated manner [77]. The LC techniques commonly used are hydrophilic interaction liquid chromatography (HILIC) for glycans [56, 67] and reverse phase (RP) chromatography for glycopeptides, methylated glycans or some reducing-end conjugated glycans [78]. Additionally, analyses on some mixed-mode column-chemistries have been reported [79-81]. The molecular identification is based on retention time. In some applications this information level can be limiting. However for QC, QA and other routine applications, this may not be an issue as the profiles of the samples are usually well known. LC-FLD has also been taken up with success in some industrial service laboratories for the analysis of patient samples, for examples, TPNG [7] and IgG glycomes [82, 83]. However, for the former, the achievable information is limited by the glycome complexity of these samples. Glycan structures may be unresolved and low abundance structures could be overshadowed by co-eluting structures of higher abundance. For example, information on antennary fucosylation in TPNG [5] is usually difficult to obtain due its low abundance and overlapping peaks. However, antennary fucosylation is an important motif in association with many inflammatory diseases [7, 17, 46]. Furthermore, it is particularly challenging to identify the triantennary and tetraantennary structures of TPNG by LC-FLD owing to their low relative abundances and their structural complexity resulting in several isomers with highly similar retention times. Capillary electrophoresis (CE) - laser induced fluorescence (LIF) is an electrophoretic separation technique that has been around for several decades but is now

gaining popularity in glycomics analysis [84]. Capillary zone electrophoresis (CZE) is a CE technique which separates molecules according to their electrophoretic mobility. This mobility depends on the analyte's hydrodynamic radius and charge state in solution [85]. The separation resolution obtained with these separation methods is usually superior to that of LC methods. Capillary gel electrophoresis (CGE)-LIF is another CE-based technique that is particularly promising for industrial laboratories. The ability to multiplex sample analysis makes CGE-LIF a high-throughput electrophoretic technique unlike CZE [86, 87].

The use of MS in glycomics has been increasing over the decades, because it analytically complements LC and CE techniques for identifying and confirming glycan structures. In glycomics, since it is usually important to maintain the intact structure of the glycan during ionisation, soft ionisation techniques such as electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI), are mostly adopted [88-90]. Advances in mass analysers have led to high resolution MS which was fuelled by the need for molecular identification and has indeed also improved glycomics analysis. The mass analysers used are mostly Time of flight (TOF) [59, 91] and ion traps [92] and to a lesser extent (triple-)quadrupoles [93]. The use of some ultrahigh resolution analysers, such as Orbitraps [94] and ion cyclotron resonance (ICR) analysers [95] has further allowed for differentiating certain isobaric glycan/glycopeptide structures in the human *N*-glycome. The inclusion of scanning quadrupoles and IMS drift tubes [96, 97] in addition to fragmentation associated ion optics, such as CID cells [98] and electron capture dissociation (ECD) cells [99] have enhanced MS/MS and MS<sup>n</sup> analysis of glycans to deduce structural isomers [100]. Interestingly, in-source fragmentation, mainly on MALDI systems, has also been successful in MS<sup>n</sup> analysis [101].

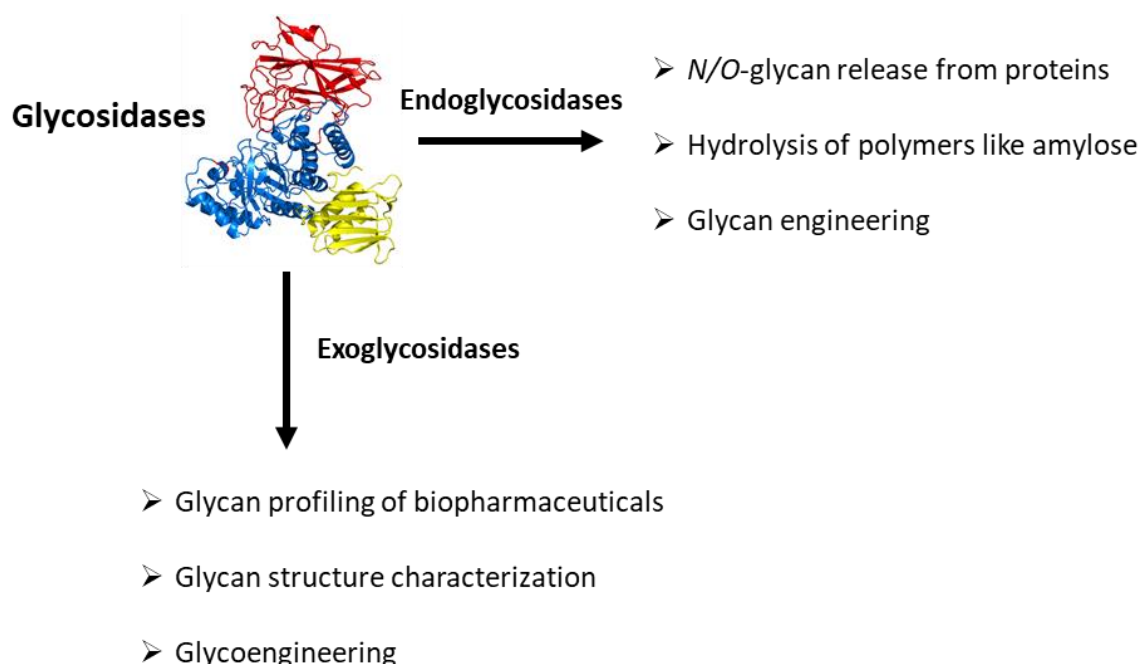
Due to the linkage variants of glycan residues, fragmentation of glycans by MS is vital in determining their isomeric structure. Method of choice is usually CID which typically breaks the glycosidic bonds of the glycan in positive ion mode. In negative ion mode, this fragmentation technique can induce cross-ring C-C bond breaks, thus providing detailed structure information [96, 102]. However, these fragmentation patterns are not always sufficiently informative or abundant to identify the glycan structure. This is especially complicated with co-eluting structures from a coupled LC/CE platform. Additionally, some motifs, such as antennary fucosylation, can undergo molecular rearrangement in the gas phase especially during CID. This can provide overlapping fragmentation patterns which complicates precise structural identification [103]. Hence, many MS-based assays rely on a two-dimensional approach of coupling separation techniques, such as LC or CE, to MS for improved structural resolution. Advantageously, these separation techniques can be easily coupled to an ESI source [83, 98], whilst additionally a few customised in-line MALDI based instruments have been reported [104]. CE-MS, in particular, is becoming popular in glycomics assays, because of its often-higher separation power compared to LC and the low amounts of sample required [84, 98]. Furthermore, the low nanoliter per minute flow rates can benefit sensitivity when using nano-ESI sources [105, 106].

**Table 1. Comparison of common chromatographic, electrophoretic and mass spectrometric techniques used for glycomics applications in for academia and industry.**

Analytical separation	Detection	Conjugation and derivatization	Application	Isomeric differentiation	Ease of use and adaptability	Throughput	Infrastructure requirements
LC	FLD	Fluorophore conjugation	Industry and academia	+	+++	++	+
	ESI-qTOF-MS/MS	Reducing end conjugation	Industry and academia	++	++	++	+++
	ESI-iontrap-MS <sup>n</sup>	Reducing end conjugation	Mostly academia	+++	++	++	+++
CE	LIF	Fluorophore conjugation	Mostly academia	++	+	+	+
	ESI-qTOF-MS/MS	Reducing end conjugation	Mostly academia	+++	+	+	+++
CGE	LIF	Fluorophore conjugation	Industry and academia	++	++	+++	++
-	MALDI-qTOF-MS	Sialic acid stabilization, Hydroxyl group methylation	Mostly academia	Possible only for sialic acid linkages	++	+++	+++

## The use of exoglycosidases in glycan profiling

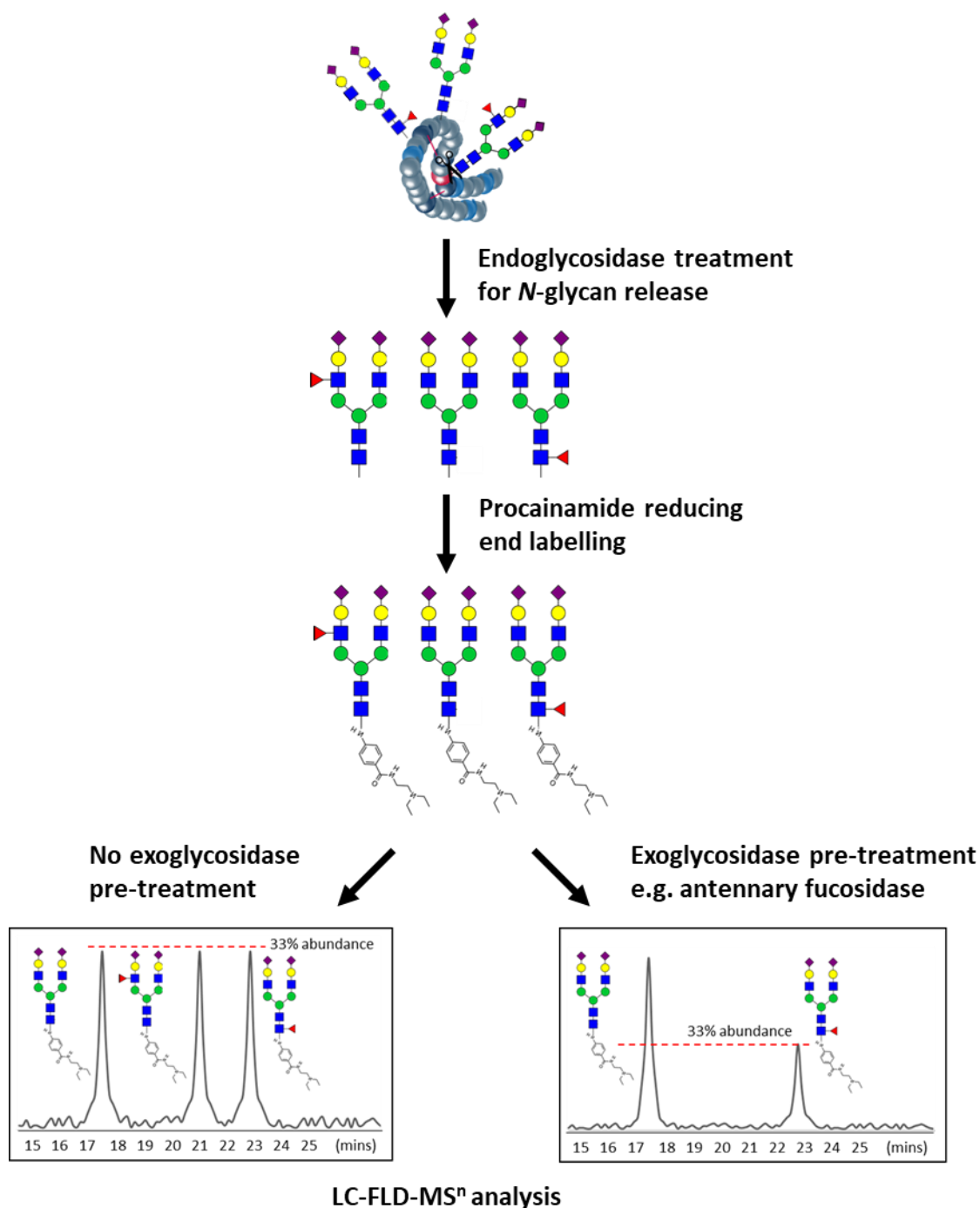
Glycosidases are enzymes that hydrolyse the glycosidic bonds within glycans and are broadly categorised as either exo- or endo glycosidases, depending on their site of hydrolysis within a glycan. Exoglycosidases can hydrolyse the glycosidic bonds at the non-reducing terminals whilst endoglycosidases hydrolyse the glycosidic bonds closer to the reducing end [33]. Owing to the substrate specificity of glycosidases, they have become a valuable tool in glycobiology research as well as for industrial applications such as glycan profiling and glycoengineering (Figure 1). The use of exoglycosidases in glycan profiling is of especial interest to this thesis.



**Figure 1. Common uses of glycosidases in industrial applications.**

Glycan profiling is heavily dependent on chromatographic and spectrometric techniques, as mentioned above. The use of exoglycosidases for glycan sequencing was and continues to be vital in complementing these analytical instrumentation techniques for glycan profiling [32]. The principle of exoglycosidase-based glycan profiling is as follows. Released glycans or glycopeptide samples are treated consecutively with an array of exoglycosidases of known specificity for identifying glycan residues as well as linkages. Such treated samples are then subjected to LC-FLD/MS analysis for the identification of the cleavage products [34, 35]. By correlating chromatographic peak shifts between exoglycosidase treated and non-treated samples, one can follow the sequential removal of residues from the non-reducing termini. This profiling approach has vital importance in biopharmaceutical characterisation and biomarker discovery for identifying glycan structures. It also serves as an alternative in confirming glycan structures that are not always conclusively identified by MS-based techniques.

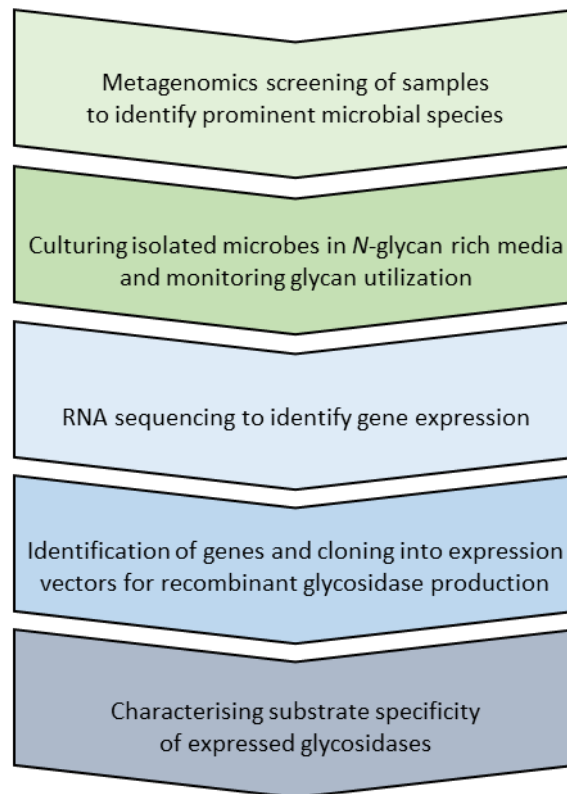




**Scheme 1. Substrate specificity characterisation of potential exoglycosidases of interest, for example an antennary fucosidase.**

The value of exoglycosidases as a profiling tool lies in the fact that they are enzymes and can have specificity for certain residues and for certain linkages. However, in practice, specificity of exoglycosidases range from narrow to broad, thus creating the need to source and characterise exoglycosidases of high substrate specificity for the purpose of glycan profiling [35]. Furthermore, owing to glycan conformation and composition complexity, the activities of some exoglycosidases can be restricted by co-occurring motifs, thus creating an additional

demand for exoglycosidases with novel activity. The latter can be exemplified by the fact that until the work of this thesis, all known antennary fucosidases had their activities restricted by steric hinderances of sialylation on the antennary fucose arm of the *N*-glycans. Scheme 1 depicts a general analytical procedure used in screening exoglycosidases for their substrate specificity.



**Scheme 2. General workflow of glycosidase discovery from metagenomics data.**

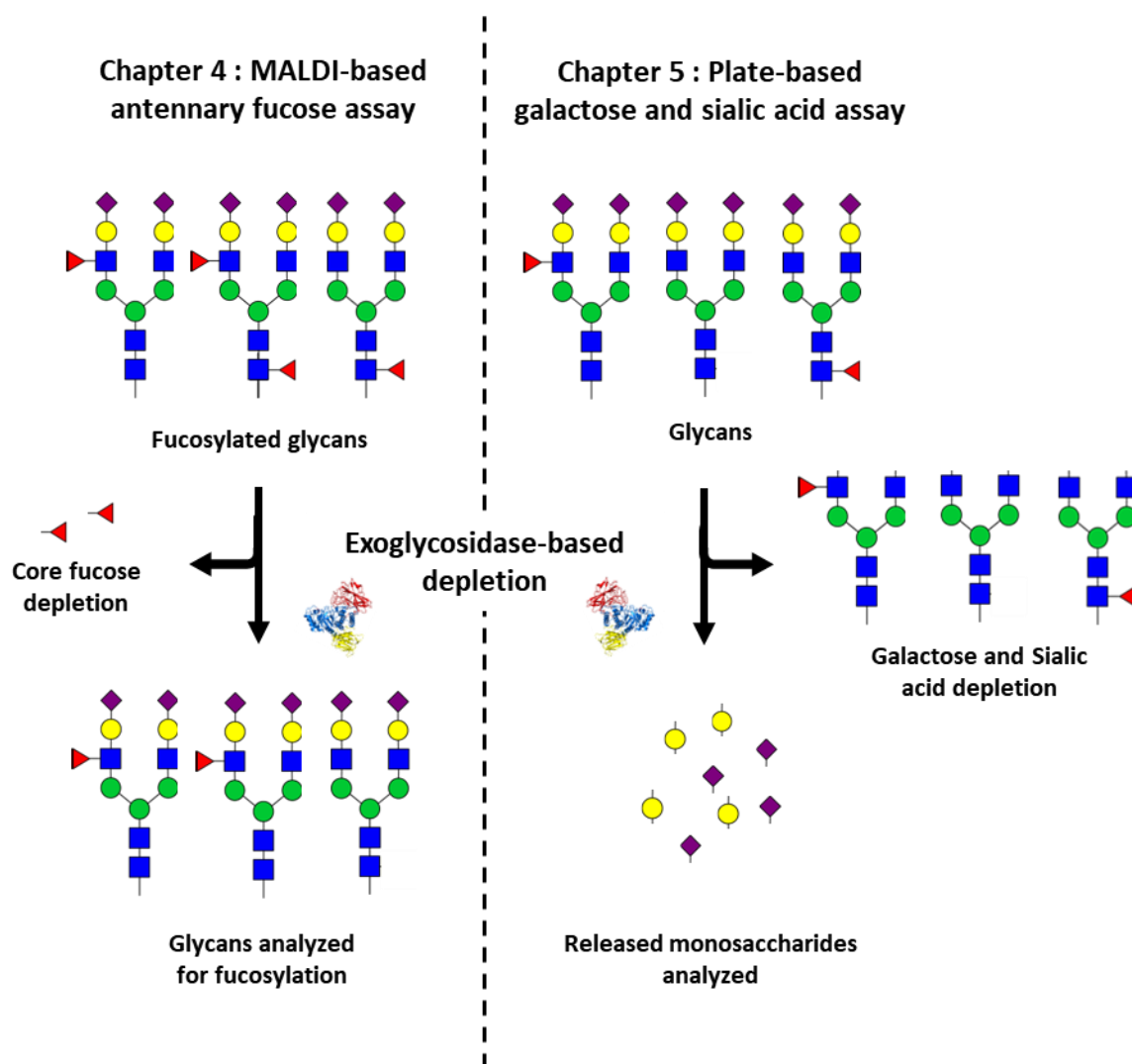
A common approach of sourcing exoglycosidases is metagenomic screening of complex glycan environments such as gut microbiota, for discovering exoglycosidases of desired specificities, as shown in Scheme 2 [38-40]. The prominent microbes in these environments are identified and their gene expression screened for potential glycosidases of interest. Basically, a successfully propagating living organism, will produce the required components to consume the available nutrients, in this case exoglycosidases to consume the glycans in their environment. Metagenomic screening for discovering novel glycosidases has proven valuable [40, 41] for identifying potential exoglycosidases for glycan profiling [34, 35] and glycoengineering [42]. A more targeted approach is bio-engineering well-characterised exoglycosidases to obtain an enhancement of activity, for example an improvement of linkage specificity [33, 36, 37]. Usually, exoglycosidases of well understood enzymatic mechanism are subjected to genetic manipulation to obtain recombinant exoglycosidases of enhanced activity and/or specificity.

## Scope

The above sections briefly describe the major analytical techniques used in glycan profiling. Many of these techniques are common to proteomics, metabolomics and to a lesser extent genomics. However, these omics fields are also complemented by biochemical assays which have a targeted approach to quantification of certain analytes. For example, enzyme linked immuno sorbent assay (ELISA) is widely used in proteomic research to identify and even quantify certain proteins of interest in a sample, while qPCR is used for quantification of certain gene sequences in a genomic sample. With the exception of lectin microarrays, glycomics assays based on the directed quantification of certain glycan epitopes have not been investigated extensively, especially in human health. Such targeted assays can meet demands in industrial applications such as QC, QA and glycome profiling where usually routine analytical queries are addressed. Additionally, focusing on quantification of certain glycan analytes of interest may simplify complicated data which could further make such assays user-friendly. The work of this thesis involves the development of novel glycomics assays that quantify glycan features of interest. A re-fashioned approach to exoglycosidase-based glycan profiling by LC-FLD/MS is taken for the development of these assays (Scheme 3). This approach is coined here as *destructive enhancement*. In short, exoglycosidases are used for cleaving certain glycan residues of certain linkages in glycome samples of TPNG or IgG. This is followed by quantifying the released desired glycan residues from the sample or profiling the remaining glycan structures that still contains the desired epitopes. Thus, by using exoglycosidases to degrade certain glycan epitopes, a destructive process, the possibility opens to the analyse of the same or different epitopes in a simple manner which would normally be tedious, leading to an enhancement of analysis (Scheme 3).

The assays developed in this thesis are heavily dependent on commercially available exoglycosidases. In addition, new exoglycosidases were characterized in this thesis to aid glycan profiling by LC-FLD-MS<sup>n</sup>. Some of the exoglycosidase activities desired for such profiling were not commercially available or mentioned in literature. The analytical chemistry work of this thesis aided metagenomic screening efforts for sourcing these desired exoglycosidases. In Chapter 2 and 3, the importance of analytical chemistry in complementing the sourcing of exoglycosidases of novel specificity is demonstrated.

Chapter 2 characterises the substrate specificity and structure of a fucosidase that is specific to antennary fucose of linkage  $\alpha(1-3/4)$ . This fucosidase, from *Ruminococcus gnavus*, was discovered in a metagenomic screening of the human gut microbiome. To our knowledge, this is the first reported antennary fucosidase which is not inhibited by sialylation such as in sialyl lewis X epitopes. This enzyme demonstrated potential in complementing the profiling of antennary fucosylation in complex glycome samples of TPNG on a LC-FLD-MS platform [80]. The analytical task of antennary fucose quantification in TPNG is especially challenging due to the low abundance of this epitope in TPNG and the phenomenon of fucose rearrangement in the gas phase. The work of this chapter has contributed to exoglycosidase tools required for enhancing this analytical task.



**Scheme 3.** The destructive enhancement approach used in assays developed in Chapters 4 and 5.

Chapter 3 is a follow-up study on a previously discovered GlcNAcase from *Bacteroides thetaiotaomicron* [40]. To our knowledge this is the first reported GlcNAcase that is specific to the bisecting  $\beta(1-4)$  linked GlcNAc on *N*-glycans. However, its activity is hindered by galactosylation and sialylation on the antennary arms. Investigating the structure of the enzyme may help us understand the basis of its substrate specificity, and thus aid in future bio-engineering efforts for enhancement of its activity. In Chapter 3, a structural characterisation of the GlcNAcase by X-ray crystallography is carried out. However, the study only provided limited insights regarding the mechanism of the enzyme due to the lack of glycan co-crystallisation. Although the catalytic residues in the active site are conserved among related GlcNAcases, the specificity to a GlcNAc linkage on a *N*-glycan is regulated by certain defining features around the entrance to the active site [40, 107]. The GlcNAcase studied here lacked such defining features and hence computer modelling of the glycan in the active site was not conclusive in determining the structural basis of its specificity.

Interestingly, a possible glycan binding activity of the C-terminal domain was proposed which has contributed to furthering the understanding of this enzyme.

The development of novel glycosidase-based assays for the quantification of certain target epitopes is carried out in Chapters 4 and 5. Chapter 4 describes the development of a novel glycosidase-based MALDI-MS assay for quantification of antennary fucosylation in TPNG. Within this chapter, quantitative superiority of this assay over a HILIC-MS<sup>n</sup> method for industrial applications is demonstrated. Quoting the *destructive enhancement* approach, a core fucosidase pre-treatment was used to destructively deplete the core fucose residues in TPNG samples before analysis on MALDI FT-ICR or MALDI-TOF platforms. This assay was able to quantify up to 19 antennary fucosylated glycans in TPNG which was a great enhancement in quantification abilities as compared to an industrial HILIC-MS<sup>n</sup> method which has long been considered the gold-standard technique for antennary fucose quantification. Furthermore, since this assay is based on MALDI-MS analysis, this makes the assay especially attractive in terms of high-throughput applications as compared to a much more lengthy HILIC-MS<sup>n</sup> analysis.

Chapter 5 introduces a novel plate-based exoglycosidase assay and demonstrates its analytical abilities for the quantification of galactosylation and sialylation in human IgG glycome. Unlike Chapter 4, Chapter 5 utilises exoglycosidases to release monosaccharide residues that are subsequently quantified rather than removing interfering linkages on a glycan. Treatments with galactosidase and galactosidase + sialidase were used in a destructive manner for releasing exposed galactose and sialic acid residues from intact IgGs purified from human plasma. The released galactose monomers were subjected to an enzyme-based redox-reaction to produce a fluorescent signal which was measured on a spectrophotometric plate reader. The signal was proportional to the galactosylation and indirectly allowed to infer sialylation of the glycans. Since the assay is also able to measure IgG amounts in the sample, absolute quantification of galactose or sialic acid residues per IgG molecule is possible. Moreover, this also helps to normalise variations in IgG amounts between patient samples. The measured readouts are available as single values with minimal data processing thus enhancing an otherwise relatively complex measurement of IgG galactosylation by an industrial HILIC-MS<sup>n</sup> method. Although, the assay provides a lower level of information as compared to LC-FLD/MS based assays, it is expected to detect associations to inflammatory diseases. In fact, the assay was successful in drawing disease association in inflammatory bowel disease (IBD) as identified previously by a LC-FLD assay [108]. Inflammatory diseases are usually associated with a change in abundance of total galactosylation and sialylation [109]. Often in LC-FLD/MS analysis of a glycome, a chromatogram is simplified to calculated derived traits which, in the case of inflammatory diseases, are total galactosylation and sialylation. The plate assay directly measures these monosaccharide units and thus serves its analytical purpose successfully. Furthermore, since this assay is based on a spectrophotometric plate reader, a relatively low infrastructure investment is required, unlike for chromatographic and/or MS based assays. This makes the assay especially attractive to laboratories that are not experienced in glycomics analysis provided they are interested in IgG galactosylation.

Chapter 6 is a detailed discussion on the current widely used techniques of chromatography, electrophoresis and mass spectrometry used in glycomics, as well as the on the analytical approaches, workflows and contributions of this thesis towards glycomics assay development. This discussion is especially pertaining to development of proof-of-concept assays that have potential of being transformed into commercial kits. The most important and critical characteristics of glycomics assays for industrial applications are discussed in great detail. Chapter 6 will also touch on outlooks for future assay development.

## References

1. Reily, C., et al., Glycosylation in health and disease. *Nature Reviews Nephrology*, 2019. 15(6): p. 346-366.
2. Rakus, J.F. and L.K. Mahal, New Technologies for Glycomic Analysis: Toward a Systematic Understanding of the Glycome. *Annual Review of Analytical Chemistry*, 2011. 4(1): p. 367-392.
3. Aebi, M., et al., N-glycan structures: recognition and processing in the ER. *Trends in Biochemical Sciences*, 2010. 35(2): p. 74-82.
4. 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. *Molecular & cellular proteomics : MCP*, 2017. 16(2): p. 228-242.
5. Rebello, O.D., et al., A Matrix-Assisted Laser Desorption/Ionization—Mass Spectrometry Assay for the Relative Quantitation of Antennary Fucosylated N-Glycans in Human Plasma. *Frontiers in Chemistry*, 2020. 8(138).
6. Royle, L., et al., HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Analytical Biochemistry*, 2008. 376(1): p. 1-12.
7. Lauc, G., et al., Genomics Meets Glycomics—The First GWAS Study of Human N-Glycome Identifies HNF1 $\alpha$  as a Master Regulator of Plasma Protein Fucosylation. *PLOS Genetics*, 2010. 6(12): p. e1001256.
8. Molinari, M., N-glycan structure dictates extension of protein folding or onset of disposal. *Nature Chemical Biology*, 2007. 3(6): p. 313-320.
9. Varki, A., Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, 1993. 3(2): p. 97-130.
10. Tams, J.W., J. Vind, and K.G. Welinder, Adapting protein solubility by glycosylation.: N-Glycosylation mutants of *Coprinus cinereus* peroxidase in salt and organic solutions. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1999. 1432(2): p. 214-221.
11. Clerc, F., et al., Human plasma protein N-glycosylation. *Glycoconjugate Journal*, 2016. 33(3): p. 309-343.
12. Tomana, M., et al., Abnormal glycosylation of serum igg from patients with chronic inflammatory diseases. *Arthritis & Rheumatism*, 1988. 31(3): p. 333-338.
13. Albrecht, S., et al., Glycosylation as a marker for inflammatory arthritis. *Cancer Biomarkers*, 2014. 14: p. 17-28.
14. Keser, T., et al., Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. *Diabetologia*, 2017. 60(12): p. 2352-2360.
15. Li, X., et al., Type 2 Diabetes Mellitus is Associated with the Immunoglobulin G N-Glycome through Putative Proinflammatory Mechanisms in an Australian Population. *OMICS: A Journal of Integrative Biology*, 2019. 23(12): p. 631-639.

16. Zhang, Z., et al., Distinguishing Benign and Malignant Thyroid Nodules and Identifying Lymph Node Metastasis in Papillary Thyroid Cancer by Plasma N-Glycomics. *Frontiers in Endocrinology*, 2021. 12(750).
17. Doherty, M., et al., Plasma N-glycans in colorectal cancer risk. *Scientific Reports*, 2018. 8(1): p. 8655.
18. de Vroome, S.W., et al., Serum N-glycome alterations in colorectal cancer associate with survival. *Oncotarget*, 2018. 9(55): p. 30610-30623.
19. Reiding, K.R., et al., Serum Protein N-Glycosylation Changes with Rheumatoid Arthritis Disease Activity during and after Pregnancy. *Frontiers in Medicine*, 2018. 4(241).
20. Clerc, F., et al., Plasma N-Glycan Signatures Are Associated With Features of Inflammatory Bowel Diseases. *Gastroenterology*, 2018. 155(3): p. 829-843.
21. Pivac, N., et al., Human Plasma Glycome in Attention-Deficit Hyperactivity Disorder and Autism Spectrum Disorders. *Molecular & Cellular Proteomics*, 2011. 10(1).
22. Borelli, V., et al., Plasma N-Glycome Signature of Down Syndrome. *Journal of Proteome Research*, 2015. 14(10): p. 4232-4245.
23. Ruhaak, L.R., et al., Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *Journal of Proteome Research*, 2011. 10(4): p. 1667-1674.
24. Pučić, M., et al., Changes in plasma and IgG N-glycome during childhood and adolescence. *Glycobiology*, 2012. 22(7): p. 975-982.
25. de Jong, S.E., et al., IgG1 Fc N-glycan galactosylation as a biomarker for immune activation. *Scientific Reports*, 2016. 6(1): p. 28207.
26. Sonneveld, M.E., et al., Antigen specificity determines anti-red blood cell IgG-Fc alloantibody glycosylation and thereby severity of haemolytic disease of the fetus and newborn. *British Journal of Haematology*, 2017. 176(4): p. 651-660.
27. Solá, R.J. and K. Griebenow, Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of Pharmaceutical Sciences*, 2009. 98(4): p. 1223-1245.
28. Kawasaki, N., et al., The Significance of Glycosylation Analysis in Development of Biopharmaceuticals. *Biological and Pharmaceutical Bulletin*, 2009. 32(5): p. 796-800.
29. Chung, C.H., et al., Cetuximab-Induced Anaphylaxis and IgE Specific for Galactose- $\alpha$ -1,3-Galactose. *New England Journal of Medicine*, 2008. 358(11): p. 1109-1117.
30. Szigeti, M., et al., Quantitative assessment of mAb Fc glycosylation of CQA importance by capillary electrophoresis. *Electrophoresis*, 2018. 39(18): p. 2340-2343.
31. Reusch, D. and M.L. Tejada, Fc glycans of therapeutic antibodies as critical quality attributes. *Glycobiology*, 2015. 25(12): p. 1325-1334.
32. Planinc, A., et al., Glycan characterization of biopharmaceuticals: Updates and perspectives. *Analytica Chimica Acta*, 2016. 921: p. 13-27.
33. Vuong, T.V. and D.B. Wilson, Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnology and Bioengineering*, 2010. 107(2): p. 195-205.
34. Shahrokh, Z., et al., Erythropoietin produced in a human cell line (Dynepo) has significant differences in glycosylation compared with erythropoietins produced in CHO cell lines. *Mol Pharm*, 2011. 8(1): p. 286-96.



35. Kobata, A., Exo- and endoglycosidases revisited. *Proceedings of the Japan Academy, Series B*, 2013. 89(3): p. 97-117.
36. Zanhporlin, L.M., et al., Structure-guided design combined with evolutionary diversity led to the discovery of the xylose-releasing exo-xylanase activity in the glycoside hydrolase family 43. *Biotechnology and Bioengineering*, 2019. 116(4): p. 734-744.
37. Kim, S.-K., et al., Engineering the N-terminal end of CelA results in improved performance and growth of *Caldicellulosiruptor bescii* on crystalline cellulose. *Biotechnology and Bioengineering*, 2017. 114(5): p. 945-950.
38. Chao, L. and S. Jongkees, High-Throughput Approaches in Carbohydrate-Active Enzymology: Glycosidase and Glycosyl Transferase Inhibitors, Evolution, and Discovery. *Angewandte Chemie International Edition*, 2019. 58(37): p. 12750-12760.
39. Rebuffet, E., et al., Discovery and structural characterization of a novel glycosidase family of marine origin. *Environmental Microbiology*, 2011. 13(5): p. 1253-1270.
40. Briliūtė, J., et al., Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology*, 2019. 4(9): p. 1571-1581.
41. Wu, H., et al., Fucosidases from the human gut symbiont *Ruminococcus gnavus*. *Cellular and Molecular Life Sciences*, 2021. 78(2): p. 675-693.
42. Benkoulouche, M., et al., Harnessing glycoenzyme engineering for synthesis of bioactive oligosaccharides. *Interface Focus*, 2019. 9(2): p. 20180069.
43. Rudd Pauline, M., et al., Glycosylation and the Immune System. *Science*, 2001. 291(5512): p. 2370-2376.
44. Rebello, O.D., et al., A Matrix-Assisted Laser Desorption/Ionization—Mass Spectrometry Assay for the Relative Quantitation of Antennary Fucosylated N-Glycans in Human Plasma. *Frontiers in Chemistry*, 2020. 8: p. 138.
45. 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. *Molecular & Cellular Proteomics*, 2017. 16(2): p. 228-242.
46. Lin, Z., et al., Mass Spectrometric Assay for Analysis of Haptoglobin Fucosylation in Pancreatic Cancer. *Journal of Proteome Research*, 2011. 10(5): p. 2602-2611.
47. de Haan, N., M. Wührer, and L.R. Ruhaak, Mass spectrometry in clinical glycomics: The path from biomarker identification to clinical implementation. *Clinical Mass Spectrometry*, 2020. 18: p. 1-12.
48. Ratner, D.M., et al., Tools for Glycomics: Mapping Interactions of Carbohydrates in Biological Systems. *ChemBioChem*, 2004. 5(10): p. 1375-1383.
49. Zhao, J., et al., Glycoprotein Microarrays with Multi-Lectin Detection: Unique Lectin Binding Patterns as a Tool for Classifying Normal, Chronic Pancreatitis and Pancreatic Cancer Sera. *Journal of Proteome Research*, 2007. 6(5): p. 1864-1874.
50. Feizi, T. and W. Chai, Oligosaccharide microarrays to decipher the glyco code. *Nature Reviews Molecular Cell Biology*, 2004. 5(7): p. 582-588.

51. Bladergroen, M.R., et al., Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *Journal of Proteome Research*, 2015. 14(9): p. 4080-4086.
52. Pučić, M., et al., High Throughput Isolation and Glycosylation Analysis of IgG Variability and Heritability of the IgG Glycome in Three Isolated Human Populations \*. *Molecular & Cellular Proteomics*, 2011. 10(10).
53. 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. *Journal of Proteome Research*, 2010. 9(12): p. 6655-6664.
54. Vanderschaeghe, D., et al., High-Throughput Profiling of the Serum N-Glycome on Capillary Electrophoresis Microfluidics Systems: Toward Clinical Implementation of GlycoHepatoTest. *Analytical Chemistry*, 2010. 82(17): p. 7408-7415.
55. Mechref, Y., et al., Quantitative Glycomics Strategies \*. *Molecular & Cellular Proteomics*, 2013. 12(4): p. 874-884.
56. Zauner, G., A.M. Deelder, and M. Wuhrer, Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *ELECTROPHORESIS*, 2011. 32(24): p. 3456-3466.
57. Stavenhagen, K., D. Kolarich, and M. Wuhrer, Clinical Glycomics Employing Graphitized Carbon Liquid Chromatography–Mass Spectrometry. *Chromatographia*, 2015. 78(5): p. 307-320.
58. Harvey, D.J., et al., Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Analytical Biochemistry*, 2008. 376(1): p. 44-60.
59. Zaia, J., Mass Spectrometry and the Emerging Field of Glycomics. *Chemistry & Biology*, 2008. 15(9): p. 881-892.
60. Zaia, J., Mass Spectrometry and Glycomics. *OMICS: A Journal of Integrative Biology*, 2010. 14(4): p. 401-418.
61. Pagel, K. and D.J. Harvey, Ion Mobility–Mass Spectrometry of Complex Carbohydrates: Collision Cross Sections of Sodiated N-linked Glycans. *Analytical Chemistry*, 2013. 85(10): p. 5138-5145.
62. Hofmann, J., et al., Identification of carbohydrate anomers using ion mobility–mass spectrometry. *Nature*, 2015. 526(7572): p. 241-244.
63. de Haan, N., et al., Glycomics studies using sialic acid derivatization and mass spectrometry. *Nature Reviews Chemistry*, 2020. 4(5): p. 229-242.
64. Zauner, G., et al., Mass spectrometric O-glycan analysis after combined O-glycan release by beta-elimination and 1-phenyl-3-methyl-5-pyrazolone labeling. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 2012. 1820(9): p. 1420-1428.
65. Schwedler, C., et al., Sialic acid methylation refines capillary electrophoresis laser-induced fluorescence analyses of immunoglobulin G N-glycans of ovarian cancer patients. *ELECTROPHORESIS*, 2014. 35(7): p. 1025-1031.

66. Parente, J.P., et al., A convenient method for methylation of glycoprotein glycans in small amounts by using lithium methylsulfinyl carbanion. *Carbohydrate Research*, 1985. 141(1): p. 41-47.
67. Keser, T., et al., Comparison of 2-Aminobenzamide, Procainamide and RapiFluor-MS as Derivatizing Agents for High-Throughput HILIC-UPLC-FLR-MS N-glycan Analysis. *Frontiers in Chemistry*, 2018. 6(324).
68. Stöckmann, H., et al., Ultrahigh Throughput, Ultrafiltration-Based N-Glycomics Platform for Ultraperformance Liquid Chromatography (ULTRA3). *Analytical Chemistry*, 2015. 87(16): p. 8316-8322.
69. Katrlík, J., et al., Glycan and lectin microarrays for glycomics and medicinal applications. *Medicinal Research Reviews*, 2010. 30(2): p. 394-418.
70. Chen, S., et al., Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nature Methods*, 2007. 4(5): p. 437-444.
71. Karamanska, R., et al., Surface plasmon resonance imaging for real-time, label-free analysis of protein interactions with carbohydrate microarrays. *Glycoconj J*, 2008. 25(1): p. 69-74.
72. Smith, E.A., et al., Surface plasmon resonance imaging studies of protein-carbohydrate interactions. *J Am Chem Soc*, 2003. 125(20): p. 6140-8.
73. Liang, P.-H., et al., Glycan arrays: biological and medical applications. *Current Opinion in Chemical Biology*, 2008. 12(1): p. 86-92.
74. Qiu, Y., et al., Plasma Glycoprotein Profiling for Colorectal Cancer Biomarker Identification by Lectin Glycoarray and Lectin Blot. *Journal of Proteome Research*, 2008. 7(4): p. 1693-1703.
75. Chen, S. and B.B. Haab, Analysis of Glycans on Serum Proteins Using Antibody Microarrays, in *Tumor Biomarker Discovery: Methods and Protocols*, M.A. Tainsky, Editor. 2009, Humana Press: Totowa, NJ. p. 39-58.
76. Xie, Y., et al., High-throughput and high-sensitivity N-Glycan profiling: A platform for biopharmaceutical development and disease biomarker discovery. *Analytical Biochemistry*, 2021. 623: p. 114205.
77. Jansen, B.C., et al., HappyTools: A software for high-throughput HPLC data processing and quantitation. *PLOS ONE*, 2018. 13(7): p. e0200280.
78. Vreeker, G.C.M. and M. Wührer, Reversed-phase separation methods for glycan analysis. *Analytical and Bioanalytical Chemistry*, 2017. 409(2): p. 359-378.
79. Mauko, L., et al., Comparison of ZIC-HILIC and graphitized carbon-based analytical approaches combined with exoglycosidase digestions for analysis of glycans from monoclonal antibodies. *Journal of Chromatography B*, 2012. 911: p. 93-104.
80. Demus, D., et al., Interlaboratory evaluation of plasma N-glycan antennary fucosylation as a clinical biomarker for HNF1A-MODY using liquid chromatography methods. *Glycoconjugate Journal*, 2021. 38(3): p. 375-386.

81. Largy, E., et al., Orthogonal LC/MS methods for the comprehensive characterization of therapeutic glycoproteins, from released glycans to intact protein level. *Journal of Chromatography A*, 2017. 1498: p. 128-146.
82. Vučković, F., et al., IgG Glycome in Colorectal Cancer. *Clinical Cancer Research*, 2016. 22(12): p. 3078.
83. Shubhakar, A., et al., High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia*, 2015. 78(5): p. 321-333.
84. Szabo, Z., et al., Improved sample preparation method for glycan analysis of glycoproteins by CE-LIF and CE-MS. *ELECTROPHORESIS*, 2010. 31(8): p. 1389-1395.
85. 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*, 2014. 13(6): p. 1598-610.
86. Hennig, R., et al., Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 2016. 1860(8): p. 1728-1738.
87. Ruhaak, L.R., et al., Total Plasma N-Glycome Changes during Pregnancy. *Journal of Proteome Research*, 2014. 13(3): p. 1657-1668.
88. Dong, X., et al., Advances in mass spectrometry-based glycomics. *ELECTROPHORESIS*, 2018. 39(24): p. 3063-3081.
89. Wuhrer, M., Glycomics using mass spectrometry. *Glycoconjugate Journal*, 2013. 30(1): p. 11-22.
90. Harvey, D.J., Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *PROTEOMICS*, 2005. 5(7): p. 1774-1786.
91. Adamczyk, B., T. Tharmalingam, and P.M. Rudd, Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 2012. 1820(9): p. 1347-1353.
92. Zhang, T., et al., Development of a 96-well plate sample preparation method for integrated N- and O-glycomics using porous graphitized carbon liquid chromatography-mass spectrometry. *Molecular Omics*, 2020. 16(4): p. 355-363.
93. White, K.Y., et al., Glycomic Characterization of Prostate-Specific Antigen and Prostatic Acid Phosphatase in Prostate Cancer and Benign Disease Seminal Plasma Fluids. *Journal of Proteome Research*, 2009. 8(2): p. 620-630.
94. Chen, B., et al., Quantitative Glycomic Analysis by Mass-Defect-Based Dimethyl Pyrimidinyl Ornithine (DiPyrO) Tags and High-Resolution Mass Spectrometry. *Analytical Chemistry*, 2018. 90(13): p. 7817-7823.
95. Vreeker, G.C.M., et al., Automated Plasma Glycomics with Linkage-Specific Sialic Acid Esterification and Ultrahigh Resolution MS. *Analytical Chemistry*, 2018. 90(20): p. 11955-11961.
96. Harvey, D.J. and W.B. Struwe, Structural Studies of Fucosylated N-Glycans by Ion Mobility Mass Spectrometry and Collision-Induced Fragmentation of Negative Ions. *Journal of The American Society for Mass Spectrometry*, 2018. 29(6): p. 1179-1193.

97. Isailovic, D., et al., Profiling of Human Serum Glycans Associated with Liver Cancer and Cirrhosis by IMS–MS. *Journal of Proteome Research*, 2008. 7(3): p. 1109-1117.
98. Lageveen-Kammeijer, G.S.M., et al., Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. *Nature Communications*, 2019. 10(1): p. 2137.
99. Zhou, W. and K. Håkansson, Electron Capture Dissociation of Divalent Metal-adducted Sulfated N-Glycans Released from Bovine Thyroid Stimulating Hormone. *Journal of the American Society for Mass Spectrometry*, 2013. 24(11): p. 1798-1806.
100. Devakumar, A., et al., Laser-induced photofragmentation of neutral and acidic glycans inside an ion-trap mass spectrometer. *Rapid Communications in Mass Spectrometry*, 2007. 21(8): p. 1452-1460.
101. Harvey, D.J., et al., Fragmentation of N-linked glycans with a matrix-assisted laser desorption/ionization ion trap time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry*, 2004. 18(24): p. 2997-3007.
102. Harvey, D.J., NEGATIVE ION MASS SPECTROMETRY FOR THE ANALYSIS OF N-LINKED GLYCANS. *Mass Spectrometry Reviews*, 2020. 39(5-6): p. 586-679.
103. Wuhrer, M., et al., Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. *Rapid Commun Mass Spectrom*, 2006. 20(11): p. 1747-54.
104. Zhan, Q., A. Gusev, and D.M. Hercules, A novel interface for on-line coupling of liquid capillary chromatography with matrix-assisted laser desorption/ionization detection. *Rapid Communications in Mass Spectrometry*, 1999. 13(22): p. 2278-2283.
105. Kammeijer, G.S.M., et al., Dopant Enriched Nitrogen Gas Combined with Sheathless Capillary Electrophoresis–Electrospray Ionization–Mass Spectrometry for Improved Sensitivity and Repeatability in Glycopeptide Analysis. *Analytical Chemistry*, 2016. 88(11): p. 5849-5856.
106. Schmidt, A., M. Karas, and T. Dülcks, Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *Journal of the American Society for Mass Spectrometry*, 2003. 14(5): p. 492-500.
107. Langley, D.B., et al., Structure of N-acetyl-beta-D-glucosaminidase (GcnA) from the endocarditis pathogen *Streptococcus gordonii* and its complex with the mechanism-based inhibitor NAG-thiazoline. *J Mol Biol*, 2008. 377(1): p. 104-16.
108. Ventham, N.T., et al., Changes to Serum Sample Tube and Processing Methodology Does Not Cause Inter-Individual Variation in Automated Whole Serum N-Glycan Profiling in Health and Disease. *PLOS ONE*, 2015. 10(4): p. e0123028.
109. Krištić, J., et al., Glycans Are a Novel Biomarker of Chronological and Biological Ages. *The Journals of Gerontology: Series A*, 2014. 69(7): p. 779-789.





# **Chapter 2**

**Fucosidases from the human gut symbiont  
*Ruminococcus gnavus***



## **Abstract**

The availability and repartition of fucosylated glycans within the gastrointestinal tract contributes to the adaptation of gut bacteria species to ecological niches. To access this source of nutrients, gut bacteria encode  $\alpha$ -L-fucosidases (fucosidases) which catalyze the hydrolysis of terminal  $\alpha$ -L-fucosidic linkages. We determined the substrate and linkage specificities of fucosidases from the human gut symbiont *Ruminococcus gnavus*. Sequence similarity network identified strain-specific fucosidases in *R. gnavus* ATCC 29149 and E1 strains that were further validated enzymatically against a range of defined oligosaccharides and glycoconjugates. Using a combination of glycan microarrays, mass spectrometry, isothermal titration calorimetry, crystallographic and saturation transfer difference NMR approaches, we identified a fucosidase with the capacity to recognize sialic acid-terminated fucosylated glycans (sialyl Lewis X/A epitopes) and hydrolyze  $\alpha(1-3/4)$  fucosyl linkages in these substrates without the need to remove sialic acid. Molecular dynamics simulation and docking showed that 3'-Sialyl Lewis X (sLeX) could be accommodated within the binding site of the enzyme. This specificity may contribute to the adaptation of *R. gnavus* strains to the infant and adult gut and has potential applications in diagnostic glycomics assays for diabetes and certain cancers.

## Introduction

The microbial community inhabiting the human gut (gut microbiota) exerts a profound effect on human health through, e.g. polysaccharide digestion, metabolite and vitamin production, maturation of the immune system and protection against pathogens [1]. The adult gut microbiota is dominated by members of Firmicutes and Bacteroidetes phyla whereas the infant gut microbiota is dominated by *Bifidobacterium* that are adapted to utilize human milk oligosaccharides (HMOs), which are one of the major glycans found in breast milk. HMOs are composed of a linear or branched backbone containing galactose (Gal), *N*-acetylglucosamine (GlcNAc) and glucose (Glc), which can be decorated with fucose (Fuc) and/or sialic acid (Sia) residues, depending on the mother's secretory status [2, 3]. In the adult colon, gut bacteria have not only access to non-digestible polysaccharides from the diet, but also to complex oligosaccharides from host mucins [4–6]. Mucins are large glycoproteins with a high carbohydrate content of up to 80%. Mucin-type *O*-glycans consists of *N*-acetylgalactosamine (GalNAc), Gal and GlcNAc, containing glycan chains modified by fucosylation, sialylation and sulfation [7–9]. The main source of glycan diversity is provided by the peripheral terminal epitopes that show considerable variation. The H1 structure,  $\alpha(1-2)$  fucose, is found in populations carrying the *secretor* gene [10], and individuals may also express the Lewis gene and the Lewis B (LeB) histoblood group antigen if they are secretors, while non-secretors express Lewis A (LeA) [11]. Another phenotype (SeW—weak secretor) is characterized by the expression of both LeA and LeB antigens [12]. The presentation of the major mucin glycan epitopes, sialic acid and fucose, varies along the gastrointestinal (GI) tract with a decreasing gradient of fucose and ABH blood group expression and an increasing gradient of sialic acid from the ileum to the colon [7]. These gradients are reversed in mice, where the small intestine is dominated by sialylated structures and the colon with those terminating in fucose [13]. These glycans provide a potential source of nutrients to members of the gut microbiota [5]. In particular,  $\alpha$ -L-fucosidases ( $\alpha$ -fucosidases) are key enzymes for the degradation and metabolism of intestinal mucin glycans or HMOs by gut microbes and therefore, contribute to shaping the composition of the gut microbiota by favoring different bacterial species and influencing health and disease. Currently,  $\alpha$ -fucosidases which catalyze the release of  $\alpha(1-2)$ ,  $\alpha(1-3)$ ,  $\alpha(1-4)$  and  $\alpha(1-6)$  linked fucose are classified into glycoside hydrolase (GH) families 29 and 95 (CAZy, [www.cazy.org](http://www.cazy.org)). All GH95 enzymes functionally characterized so far show strict substrate specificity to the terminal Fuc  $\alpha(1-2)$  Gal linkage and hydrolyze the linkage via an inverting mechanism whereas GH29 enzymes show relatively relaxed substrate specificities with hydrolysis proceeding via a retaining mechanism ([www.cazy.org](http://www.cazy.org)). It was suggested that GH29 can be divided into two subfamilies. One contains fucosidases with relaxed substrate specificities that can act on 4-nitrophenyl  $\alpha$ -L-fucopyranoside (pNP-Fuc) (referred to as GH29-A) (EC 3.2.1.51), whereas the members of the other subfamily show strict specificity for terminal  $\alpha(1-3/4)$  fucosidic linkages with little/no activity on pNP-Fuc (GH29-B) (EC 3.2.1.111) as shown for fucosidases from *Streptomyces* and *Bacteroidetes* thetaiotaomicron [14, 15]. The GH29-A subfamily includes fucosidases from *Thermotoga maritima* [16], soil metagenome [17] or bacterial pathogens [18, 19] whereas the GH29-B subfamily includes fucosidases from *Bifidobacterium bifidum* (BbAfcB) [20] *Clostridium perfringens* (CpAfc2) [21] and *Streptococcus pneumoniae* (SpGH29c) [22]. Despite the importance of fucose in regulating bacterial intestinal colonization in adults and infants, only

a limited number of fucosidases have been studied at a biochemical level from human gut symbionts.

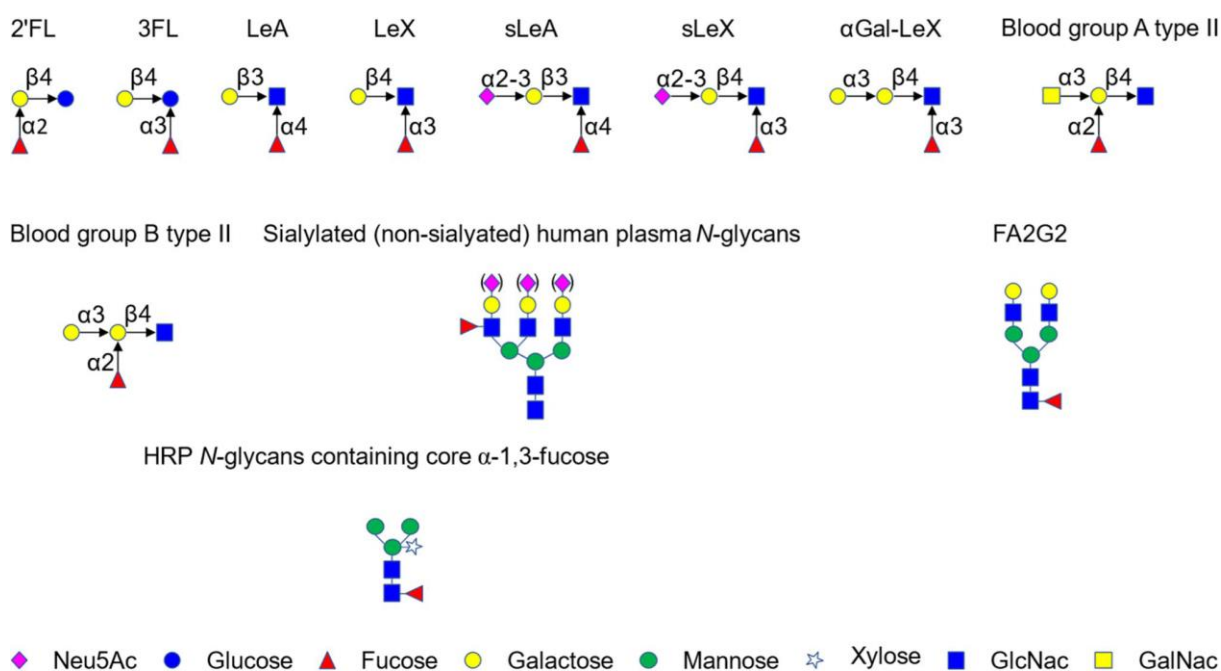
*Ruminococcus gnavus* is a prevalent member of the gut microbial community belonging to the Firmicutes division [23, 24]. *R. gnavus* is an early colonizer of the human gut [25] but persists in healthy adults where it belongs to the 57 species detected in more than 90% of human faecal samples by metagenomic sequencing [23]. In the past few years, an increasing number of studies are reporting a disproportionate representation of *R. gnavus* in diseases, such as inflammatory bowel disease [26]. In our previous work, we showed that *R. gnavus* ability to grow on HMOs or mucins was strain dependent [27, 28], underscoring the importance of analysing glycan utilization by members of the human gut microbiota at the strain level. These differences are reflected by the distribution of GH families between *R. gnavus* strains [27]. For example, *R. gnavus* E1 genome lacks a sialidase encoding gene whereas *R. gnavus* ATCC 29149 encodes a GH33 enzyme which has been functionally characterized as an intramolecular trans-sialidase [29] and is associated with a unique sialic acid metabolism pathway which forms the basis of *R. gnavus* ATCC 29149 adaptation to mucus [30]. In contrast, both *R. gnavus* E1 and *R. gnavus* ATCC 29149 genomes harbor fucosidase encoding genes belonging to GH29 or GH95 families [27], but their functional characterization has not been reported. To gain further biochemical and structural insights into *R. gnavus* strategy to utilize mucin glycans, we determined the substrate and linkage specificities of a range of fucosidases belonging to GH29 and 95 family from *R. gnavus* ATCC 29149 and E1 strains. We identified and characterized a fucosidase from *R. gnavus* E1 with the capacity to recognize fucosylated glycans capped with sialic acid and to hydrolyze  $\alpha(1-3/4)$  fucosyl linkages in these substrates without the need to remove sialic acid. This unique specificity may contribute to the adaptation of *R. gnavus* strains to distinct nutritional niches. Since changes in abundances of sialyl fucosylated epitopes on human glycans have been associated in several diseases, such as diabetes and certain cancers, these novel fucosidases may have potential in diagnostic glycomics assays.

## Materials and methods

### Materials

All chemicals were obtained from Sigma (St Louis, MO, USA) unless otherwise stated. The structure of the oligosaccharides used in this work is shown in Figure 1. 3'-Sialyl Lewis X (sLeX) was purchased from Carbosynth Limited (Campton, UK), Lewis A (LeA),  $\alpha(1-3)$ Gal-Lewis X ( $\alpha$ Gal-LeX), Blood group A/B tetrasaccharide type II (Blood group A/B type II) were from Elicityl (Crolles, France),

Lewis X (LeX) used for activity assay was from Dextra Laboratories (Reading, UK), LeX/LeA/*N*-Acetylneuraminic acid (Neu5Ac) used for ITC were from Carbosynth Limited (Campton, UK), LeX used for STD NMR was from the Consortium for Functional Glycomics (CFG). 3'-Sialyl Lewis A (sLeA), sialylated and desialylated human plasma *N*-glycans and FA2G2 *N*-glycans were from Ludger (Oxford, UK). Horseradish peroxidase (HRP) treated with BM03341 plant specific PNGase was a kind gift from Dr Lucy Crouch (Newcastle University). *E. coli* strain (Tuner DE3 pLacI) was from Merck (Darmstadt, Germany).



**Figure 1. Fucosylated oligosaccharides used in this study.** Monosaccharide symbols follow the Symbol Nomenclature for Glycans system [98].

### Cloning, expression, mutagenesis and purification of GH29 and GH95 fucosidases

*Ruminococcus gnavus* ATCC 29149 or E1 genomic DNA (gDNA) was purified from the cell pellet of a bacterial overnight culture (1 mL) following centrifugation (5000g, 5 min) using the GeneJET Genomic DNA Purification Kit (ThermoFisher, UK), according to the manufacturer's instructions. The full-length sequence of E1\_10125 and E1\_10180, excluding the signal sequence were cloned into the pOPINF expression system [31], introducing an His-tag at the

N-terminus. The D221A mutant of E1\_10125 was produced by NZYTech (Lisbon, Portugal). The E1\_10125G260M mutant was generated using the NZY Mutagenesis kit (Lisbon, Portugal). The ATCC\_03833 and ATCC\_00842 sequences exempt of the signal sequence were cloned into pET28a with N-terminal His tag by Prozomix (Haltwhistle, UK). The E1\_10587 was synthesized by NZYTech (Lisbon, Portugal) into pHTP1 with N-terminal His tag.

The primers used are listed in Supplementary Table S1. DNA manipulation was carried out in *E. coli* DH5 $\alpha$  cells. Sequences were verified by DNA sequencing at Eurofins MWG (Ebersberg, Germany) or Earlham Institute (formerly TGAC, Norwich, UK). *E. coli* TunerDE3pLacI cells were transformed with the recombinant plasmids according to manufacturer's instructions. The expression was carried out in 1000 mL LB media growing cells at 37 °C until OD600 nm reached 0.4 to 0.6 and then induced at 16 °C for 48 h. The cells were harvested by centrifugation at 7000g for 10 min. The His-tagged proteins were purified by immobilized metal affinity chromatography (IMAC) and further purified by gel filtration (Superdex 75 and 200 columns) on an Akta system (GE Health Care Life Sciences, Little Chalfont, UK). Protein purification was assessed by standard SDS–polyacrylamide gel electrophoresis using the NuPAGE Novex 4–12% Bis–Tris (Life Technologies, Paisley, UK). Protein concentration was measured with a NanoDrop (Thermo Scientific, Wilmington, USA) and using the extinction coefficient calculated by ProtParam (ExPASy-Artimo, 2012) from the peptide sequence.

### **Glycan microarrays**

Three concentrations (5, 50 and 200  $\mu$ g/mL) of recombinant His6-tagged E1\_10125 D221A mutant were screened for binding to Core H glycan microarray glycans at the Consortium for Functional Glycomics (CFG).

### **STD NMR experiments**

An Amicon centrifuge filter unit with a 10 kDa MW cutoff was used to exchange the protein in 25 mM d19-2,2-bis(hydroxymethyl)-2,2',2''-nitrilotriethanol pH\* 7.4 (uncorrected for the deuterium isotope effect on the pH glass electrode) D<sub>2</sub>O buffer and 50 mM NaCl. All the ligands were dissolved in 25 mM d19-2,2-bis(hydroxymethyl)-2,2',2''-nitrilotriethanol pH\* 7.4, 50 mM NaCl. A concentration of 50  $\mu$ M was used for the enzyme and 2 mM for the ligands. The STD NMR spectra were performed on a Bruker Avance 800.23 MHz at 278 K. The on- and off-resonance spectra were acquired using a train of 50 ms Gaussian selective saturation pulses using a variable saturation time from 0.5 to 5 s, for binding epitope mapping determination (STD build-up curves). Residual protein resonances were filtered out using a T2 filter of 40 ms. All the spectra were performed with a spectral width of 10 KHz and 32768 data points using 256 or 512 scans. Binding epitope mappings were obtained by determining the initial slopes (STD<sub>0</sub>) calculated by performing a least-squares fitting of the following mono-exponential curve:

$$STD(t_{sat}) = STD_{max}(1 - \exp(-k_{sat}t_{sat})),$$

where  $STD(t_{sat})$  is the STD intensity for a saturation time,  $t_{sat}$ ,  $STD_{max}$  is the maximum STD intensity and  $k_{sat}$  is the rate constant for saturation transfer. In the limit,  $t_{sat} \rightarrow 0$ :

$$STD_0 = STD_{max}k_{sat}$$

Importantly,  $STD_0$  gives a value that is independent of any relaxation or rebinding effects, allowing for a more accurate binding epitope. The value of  $STD_0$  was then normalized against the proton with the largest intensity to give values in the range of 0–100%, which were then mapped onto the ligand structure to give the corresponding binding epitope mapping.

## X-ray crystallography

The E1\_10125 fucosidase His tag was removed using 3C-protease overnight at 4 °C at a mass ratio of 20:1, the His tag and 3C-protease was then removed by passing the sample over a nickel Sepharose column. The final crystallization condition was 0.2 M magnesium chloride, 25% PEG 3350, 0.1 M bis-tris pH 5.5, 10 mM 2-fucosyllactose (2'FL). Sitting drop vapour diffusion crystallization experiments of E1\_10125 or E1\_10125 D221A were set up at a concentration of 20 mg/mL and monitored using the VMXi beamline at Diamond Light Source [32]. Crystals were cryoprotected using the crystallization condition with the addition of 15% ethylene glycol. Wild-Type and D221A E1\_10125 mutant diffraction experiments were performed at Diamond Light Source on beamlines I04 (wavelength 0.9795 Å) and I03 (wavelength 0.9763 Å), respectively. The data were processed with Xia2 making use of aimless, dials and pointless [33–36]. The data were phased using PHASER using pdb 4OUE as a molecular replacement model and refined using REFMAC [37] and Coot [38] within the CCP4 software environment [39]. The PDB REDO server [40] was used to optimize the refinement parameters and the model was validated using the Molprobit server [41].

## Molecular dynamics (MD) simulation and docking of sLeX into E1\_10125 D221A

Docking calculations were run using the crystal structure of E1\_10125 D221A mutant as it showed the highest resolution. First, a MD simulation was carried out to explore the flexibility of the side chains surrounding the fucose binding subsite and the whole binding site. The input coordinates for the simulation were produced by loading the X-ray coordinates into Schrödinger Maestro [42] and processed using the protein preparation wizard [43]. Protons were added to the structure and then all buffer ions, buffer and structurally non-essential water molecules were removed. PROPKA [44] was then used to predict the ionization state of polar residues at pH 7. The OPLS force field was then used to minimize the protein structure, converging heavy atoms to a threshold of 0.3 Å.

The system was then simulated using the Amber PMEMD software [45]. The system was solvated using TIP3P water, placed within a truncated octahedron buffered to 10 Å, to a net charge of zero using Na<sup>+</sup> ions. The parameter set for the protein atoms and structural ions used was taken from the ff14SB and gaff force fields. The system was initially minimized with constraints of 20 Kcal mol<sup>-1</sup> Å<sup>-2</sup> placed on solute atoms and then minimized a second time

with no constraints placed on solute atoms. The system was then heated to a temperature of 300 K and raised to a pressure of 1 atm in two separate 500 ps steps under constraints of 20 kcal mol<sup>-1</sup> Å<sup>-2</sup> placed on solute atoms. Over the course of four 500 ps steps constraints were then released in 5 Kcal mol<sup>-1</sup> Å<sup>-2</sup> increments. The system was then simulated over the course of 500 ns with a 2 fs time step, with frame sampling every 0.5 ns. The SHAKE algorithm [46] was used to constrain bonds involving hydrogen atoms. A Berendsen barostat and a Langevin thermostat with a 5 ps<sup>-1</sup> collision frequency were used to maintain constant pressure and temperature, respectively. Non-bonding atom bond cutoff was set to 8 Å.

The trajectory file from the MD simulation was then clustered using cpptraj [47] to produce 20 average structures. The kmeans clustering algorithm within cpptraj was used with a random set of initial points. The clustering was calculated for every tenth frame and based on the distance between atoms measured using root-mean-square deviation of atomic positions (RMSD) without fitting structures to each other prior to calculating RMSD.

The 20 average structures were then imported into Schrödinger Maestro and processed using the protein preparation wizard (see above). Protons were replaced in the structure then all buffer ions and structurally non-essential water molecules introduced during MD simulation were removed. The OPLS force field was then used to minimize the protein structure, converging heavy atoms to a threshold of 0.3 Å. Protein structures showing a wide-open binding site were selected from the 20 average structures to be used for docking of the tetrasaccharide ligand. In order to perform the docking calculations, a cubic grid with a 10 Å × 10 Å × 10 Å inner box and a 30 Å × 30 Å × 30 Å outer box with the centroid placed at the middle of the binding site was generated. sLeX was built within Maestro and prepared using LigPrep [48] and low-energy conformers generated using MacroModel [49]. sLeX was then docked using Glide [50] with standard precision enhanced by 2 times without canonicalization and without sampling ring conformations.

### **Isothermal titration calorimetry (ITC)**

ITC experiments were performed using the PEAQ-ITC system (Malvern, Malvern, UK) with a cell volume of 200 µL. Prior to titration, protein samples (E1\_10125D221A) were exhaustively dialyzed into 50 mM citrate buffer pH 6. The ligand was dissolved in the dialysis buffer. The cell protein concentration was 100 µM and the syringe ligand concentration was 2 mM for all ligands tested except 20 mM for Neu5Ac. Three controls with titrant (sugar) injected into the buffer, buffer injected to protein, buffer injected into buffer, were subtracted from the data. The analysis was performed using the Malvern software, using a single-binding site model. Experiments were carried out in triplicate.

### **Activity assays and kinetics**

The optimum pH of fucosidases was determined with 0.5 mM pNP-Fuc for all fucosidase tested apart for E1\_10587 where 5 mM pNP-Fuc was used instead, in 50 mM citrate buffer (pH 3, 4, 5 and 6), 50 mM sodium phosphate buffer (pH 6, 7, 7.5 and 8) and 50 mM Tris buffer (pH 8.5 and 9.3). The final concentration of enzyme was 1 µM for ATCC\_00842, 2 µM for

E1\_10180, 20  $\mu\text{M}$  for E1\_10125, 0.015  $\mu\text{M}$  for ATCC\_03833 and 10  $\mu\text{M}$  for E1\_10587. The reaction duration was optimized to measure the reaction rates under initial conditions. After incubation at 37 °C, the reaction was stopped by adding 50  $\mu\text{L}$  of 1 M of sodium carbonate into 200  $\mu\text{L}$  of reaction (200  $\mu\text{L}$  of 1 M of sodium carbonate into 40  $\mu\text{L}$  of reaction for E1\_10587). The amount of fucose released was determined using a 96-well plate reader (BMG Labtech, Ortenberg, Germany) by measuring absorbance at 405 nm.

Kinetic studies against pNP-Fuc were performed in 50 mM citrate buffer at optimal pH (pH 6 for ATCC\_00842, E1\_10180, E1\_10125, ATCC\_03833; pH 5 for E1\_10587) with increasing amounts of pNP-Fuc and a constant enzyme concentration at 37 °C. The series of pNP-Fuc concentrations were chosen to ensure that there were at least three points below and above the  $K_m$  value. The amount of enzyme was determined to fulfil free-ligand approximation, i.e. the enzyme concentration was linear with product formation. The reaction duration was optimized to measure the reaction rates under initial conditions. A standard curve was made with a range of pNP-Fuc from 0 to 140  $\mu\text{M}$  and in the same experimental condition as the enzymatic reactions. Kinetic parameters were calculated based on the Michaelis–Menten equation using a non-linear regression analysis program (Prism 6, GraphPad, San Diego, USA). Kinetic parameters for E1\_10587 were calculated based on the Michaelis–Menten equation:

$$\frac{1}{t} \ln \frac{[S_0]}{[S_t]} = -\frac{[S_0 - S_t]}{K_m t} + \frac{V_{max}}{K_m}$$

To determine the substrate specificity of the enzymes against fucosylated oligosaccharides (2'FL, 3FL, LeA and LeX), each enzyme was incubated with the substrate (0.1 mM) at 37 °C in 50 mM citrate buffer at optimal pH. For kinetic assays with E1\_10125, 50  $\mu\text{M}$  to 350  $\mu\text{M}$  of LeX, 50  $\mu\text{M}$  to 1400  $\mu\text{M}$  of sLeX and 20  $\mu\text{M}$  to 400  $\mu\text{M}$   $\alpha\text{Gal-LeX}$  was used. Higher concentrations of LeX or  $\alpha\text{Gal-LeX}$  were not used because they contained significant amount of L-fucose, which affected the accuracy of quantification. The enzyme in the reaction was 0.01  $\mu\text{M}$  for LeX and sLeX, 1 nM for  $\alpha\text{Gal-LeX}$ . The time course of reaction used for LeX, sLeX and  $\alpha\text{Gal-LeX}$  was 9 min, 30 min and 20 min, respectively. The fucose released was quantified using the k-fucose kit (Megazyme, Wicklow, Ireland) combined with the diaphorase/resazurin assay [51]. Briefly, 40  $\mu\text{L}$  of reaction mixed with 97  $\mu\text{L}$  of mixed reagent (50  $\mu\text{L}$  of dH<sub>2</sub>O, 20  $\mu\text{L}$  of reaction buffer pH9.5, 5  $\mu\text{L}$  of NADP<sup>+</sup>, 2  $\mu\text{L}$  of L-fucose dehydrogenase suspension, 10  $\mu\text{L}$  of 1 mM resazurin solution, 10  $\mu\text{L}$  of 10U/mL solution of Diaphorase) and incubated at room temperature for 20 min before measuring the fluorescence of resorufin using a 96-wells plate reader (BMG Labtech, Ortenberg, Germany) with an excitation at 550 nm and an emission at 584 nm. One unit of activity was defined as the amount of enzyme needed to release 1  $\mu\text{mol}$  of product per min under the conditions described above. The kinetic parameters of E1\_10125 fucosidase against sLeX were calculated based on Michaelis–Menten equation using a non-linear regression analysis program (Prism 6, GraphPad, San Diego, USA). The curve of initial rate against substrate concentration for LeX and  $\alpha\text{Gal-LeX}$  was linear, which indicated that the substrate concentration was far below  $K_m$ , therefore their  $k_{cat}/K_m$  was estimated from the slope of the curve [52],  $k = (k_{cat}/K_m)[E]_0$  where  $k$  is the slope and  $[E]_0$  represents the enzyme concentration.



## LC-MS/MS analysis

10  $\mu\text{M}$  of enzyme was incubated with a range of substrates including sLeA (5  $\mu\text{M}$ ), sLeX (5  $\mu\text{M}$ ), sialylated and de-sialylated human plasma *N*-glycans released from 0.5  $\mu\text{L}$  human plasma, HRP (0.01  $\mu\text{g}/\mu\text{L}$ ), FA2G2 (5  $\text{ng}/\mu\text{L}$ ), blood group A type II (5  $\mu\text{M}$ ) or blood group B type II (5  $\mu\text{M}$ ). Reactions (20  $\mu\text{L}$ ) were performed in 50 mM citrate buffer at pH 6 and 37 °C for 24 h. Reactions were then dried down using Savant SpeedVac centrifugal evaporator (Thermo Fisher, Wilmington, USA), labelled at the reducing end with procainamide using the glycan labelling kit with sodium cyanoborohydride reductant (Ludger, Oxford, UK) and purified using S-cartridges (Ludger, Oxford, UK) to remove the excess dye. The samples were dried by speed vacuum and resuspended in 50  $\mu\text{L}$  of acetonitrile:water solvent. Then the reactions were injected onto a Waters ACQUITY UPLC Glycan BEH amide column (2.1  $\times$  150 mm, 1.7  $\mu\text{m}$  particle size, 130 Å pore size) at 40 °C on a Dionex Ultimate 3000 UHPLC instrument with a fluorescence detector ( $\lambda_{\text{ex}}$  = 310 nm,  $\lambda_{\text{em}}$  = 370 nm) coupled to a Bruker Amazon Speed ETD. A 50 mM ammonium formate solution pH 4.4 (Ludger, Oxford, UK) was used as mobile phase A and acetonitrile (Romil, UK) was used as mobile phase B. For the plasma samples, a 70 min gradient was used with mobile phase B from 76 to 51% from 0 to 53.5 min at a flow rate of 0.4 mL/min followed by mobile phase B from 51 to 0% from 53.5 min to 55.5 min at flow rate of 0.2 mL/min, and 2 min stabilization, mobile phase B from 0 to 76% from 57.5 min to 59.5 min at a flow rate 0.2 mL/min, and then last for 6 min, from 65.5 min to 66.5 min, the flow rate was changed back to 0.4 mL/min and then equilibrated for 3.5 min. HRP samples used a 75 min gradient starting from 80 to 62% mobile B. A 70–62% gradient was used for FA2G2 glycan. For the shorter fucosyl-oligosaccharides, an 85 min gradient was used from 85 to 65% mobile B. The Amazon Speed was operated in the positive ion mode using the following settings: source temperature 180 °C; gas flow 4 L/min; capillary voltage, 4500 V; ICC target, 200,000; maximum accumulation time, 50 ms; rolling average, 2; number of precursor ions selected, 3; scan mode, enhanced resolution; mass range scanned, 400 to 1700. Singly charged ions were excluded for CID except for HRP and fucosyl-oligosaccharide samples.

## Bioinformatics analyses

For sequence similarity networks (SSN) analysis, the sequences encoding GH29 and GH95 fucosidases were extracted from the Interpro database 66.0 (<https://www.ebi.ac.uk/interpro/>) after removing redundant sequences by CD-HIT Suite [53] ([https://weizhonglab.ucsd.edu/cdhit\\_suite/cgi-bin/index.cgi?cmd=cd-hit](https://weizhonglab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit)). Additional sequences included those corresponding to functionally characterized GH29 and GH95 from the CAZy database ([www.cazy.org](http://www.cazy.org)) as well as the *R. gnavus* E1 and ATCC29149 fucosidases (6 GH29 and 7 GH95). The amino acid sequences were then used to create SSN using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) [54]. The network setting for GH29 and GH95 was made to combine proteins in one node sharing over 70% and 45% identity, respectively and nodes were linked by edges when their sequences shared over 40% (the e-value threshold was 1094) and 35% identity (the e-value threshold was 10130) for GH29 and GH95, respectively. The SSN data were visualized using Cytoscape 3.6 [55]. ProtParam (ExpASy) [56] was used to determine the length, molecular weight and theoretical pI of the fucosidases under study.

LALIGN was used to do pairwise sequence alignment and obtain sequence similarities [57]. SignalP 5.0 Server was used to predict the presence and nature of signal peptides as well as cleavage sites [58]. The TMHMM Server v.2.0 was used to predict the presence of transmembrane helices [59]. CW-PRED was used for the detection of LPXTG and LPXTG-like motif and thus the prediction of cell-wall proteins in Gram-positive bacteria [60]. PSORTb v3.0.2 was used for bacterial protein subcellular localization prediction [61].

## Results

### Sequence similarity network identified strain-specific fucosidases in *R. gnavus* ATCC 29149 and E1

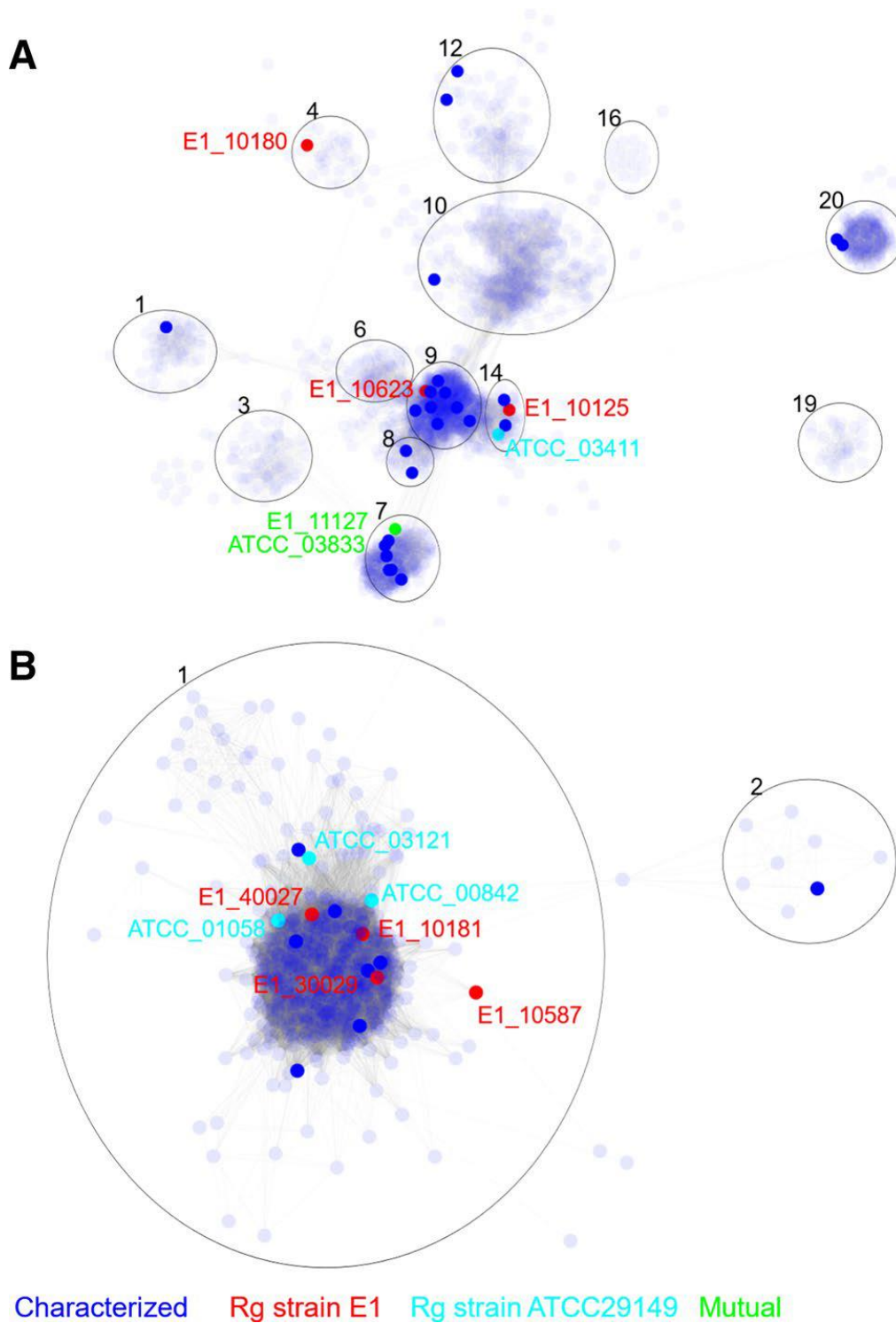
The genome of *R. gnavus* E1 encodes 4 predicted GH29 fucosidases (named E1\_10125, E1\_10180, E1\_10623 and E1\_11127) and 4 GH95 fucosidases (named E1\_10181, E1\_10587, E1\_30029, E1\_40027), whereas *R. gnavus* ATCC 29149 encodes 2 GH29 (named ATCC\_03411, ATCC\_03833) and 3 GH95 fucosidases (named ATCC\_00842, ATCC\_01058, ATCC\_03121) (for detailed information, Supplementary Table S2). A sequence similarity network (SSN) analysis was conducted to identify putative functional relationships between GH29 or GH95 fucosidases from *R. gnavus* strains and related protein sequences (Figure 2). The SSN analysis covered 6736 amino acid sequences from the GH29 family from Interpro database 66.0 and CAZy ([www.cazy.org/GH29\\_characterized.html](http://www.cazy.org/GH29_characterized.html)) and 825 GH95 sequences from Interpro IPR027414 (including 8 sequences from functionally characterized GH95 [www.cazy.org/GH95\\_characterized.html](http://www.cazy.org/GH95_characterized.html)).

For GH29, 5318 representative nodes in the SSN analysis were separated into 44 clusters according to their sequence similarity which may indicate a similar biochemical function and ligand specificity (Supplementary Figure S1A). GH29 fucosidases from *R. gnavus* E1 and ATCC 29149 strains were found in cluster 4 (E1\_10180), cluster 7 (ATCC\_03833 and E1\_11127), cluster 9 (E1\_10623) and cluster 14 (E1\_10125 and ATCC\_03411) (Figure 2A). E1\_10125 shares 94.5% sequence similarity with ATCC\_03411, whereas ATCC\_03833 is 99.8% similar to E1\_11127. The catalytic domain of E1\_10623 was 62.9% similar to Blon\_2336 from *Bifidobacterium longum* subsp. *infantis* ATCC 15697 [62]. Based on this analysis, E1\_10125, E1\_10180 and ATCC\_03833 were chosen as representatives of *R. gnavus* GH29 fucosidases for functional characterization.

The GH95 family includes fewer sequences and functionally characterized proteins ([www.cazy.org](http://www.cazy.org)). The SSN analysis of GH95 fucosidases led to the identification of 825 nodes, 627 of which were found in the cluster 1 (Figure 2B). All GH95 fucosidases from *R. gnavus* E1 and ATCC 29149 strains fall within the same cluster. Based on the pre-screening for expression (not shown), ATCC\_00842 and E1\_10587, sharing 60% sequence similarity, were selected as representative GH95 fucosidases of *R. gnavus* ATCC 29149 and E1, respectively for further characterization.

### *R. gnavus* fucosidases from GH29 and GH95 families display novel substrate specificities

The genes encoding the selected GH29 and GH95 fucosidases from *R. gnavus* ATCC 29149 and E1 strains were heterologously expressed in *E. coli* and the His6-tag recombinant proteins purified by immobilized metal ion affinity chromatography and gel filtration (see Materials and Methods section for details). *E. coli* Tuner DE3 pLacI strain was chosen as heterologous host as it does not display any endogenous  $\beta$ -galactosidase activity (due to the deletion of the LacZ gene) that may interfere with the enzymatic characterization of the recombinant enzymes. The activity of the purified enzymes was first screened against the synthetic substrate pNP-Fuc. The optimum pH of all fucosidases tested, determined using pNP-Fuc, was found to be pH 6 apart for E1\_10587, which was pH 5 (Supplementary Figure S2).



**Figure 2. The distribution of *R. gnavus* GH29 and GH95 fucosidases based on SSN analysis.** (A) Partial representation of SSN analysis of GH29 family containing fucosidases from *R. gnavus* E1 and ATCC29149 strains. (B) Representation of the SSN central cluster of GH95 family containing all GH95 from *R. gnavus* E1 and ATCC29149 strains. Blue node: sequences extracted from the CAZy database encoding functionally characterized enzymes. Red nodes sequences from *R. gnavus* E1 strain. Cyan nodes, sequences from *R. gnavus* ATCC29149 strain. Green nodes, sequences common to both *R. gnavus* E1 and ATCC29149 strains.

The kinetic parameters were determined at the optimum pH by calculating the initial rate of reaction with increasing pNP-Fuc concentrations. Fucosidase ATCC\_03833 showed the highest catalytic efficiency with a  $k_{cat}$  of  $83.6 \text{ s}^{-1}$  and a  $K_m$  of  $179.1 \text{ }\mu\text{M}$  (Table 1). These values are consistent with fucosidases belonging to GH29 subfamily A [15, 17, 63, 64]. Fucosidases ATCC\_00842 and E1\_10180 also showed activity against pNP-Fuc, but their high  $K_m$  suggest that pNP-Fuc is not a good ligand for these enzymes. Fucosidase E1\_10125 displays the lowest activity against pNP-Fuc of the characterized GH29 subfamily fucosidases as shown by its low  $k_{cat}$  of  $1.8 \cdot 10^{-3} \text{ s}^{-1}$  [15, 20–22, 65, 66]. E1\_10587 was merely active on pNP-Fuc as also reported for other GH95 fucosidases [21].

Next, the substrate specificity of the recombinant fucosidases was tested on a range of fucosylated oligosaccharides. The specific activity was determined based on fucose release against 2'FL (Fuc $\alpha$ 1,2Gal $\beta$ 1,4Glc), 3FL (Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]Glc), Lewis A (LeA, Gal $\beta$ 1-3[Fuc $\alpha$ 1-4]GlcNAc) and Lewis X (LeX, Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]GlcNAc) (Table 2). From this analysis, fucosidases E1\_10125 and E1\_10180 showed substrate specificity towards  $\alpha$ (1-3/4) fucosylated linkages while fucosidases ATCC\_00842 and ATCC\_03833 showed preference for  $\alpha$ (1-2) linkages (Table 2). E1\_10587 showed lower activity against all tested substrates, therefore, it was not possible to assess its substrate specificity.

The activity of the recombinant fucosidases was further tested on more complex oligosaccharides and glycoproteins including sialyl Lewis X (Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]GlcNAc, sLeX), sialyl Lewis A (Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3[Fuc $\alpha$ 1-4]GlcNAc, sLeA), sialylated or desialylated human plasma *N*-glycans, horseradish peroxidase *N*-glycans containing core  $\alpha$ (1-3) fucose (HRP), blood group A type II, blood group B type II and  $\alpha$ (1-6) fucosylated biantennary *N*-glycan (FA2G2) and the products of the reactions analyzed by LC–MS/MS (Supplementary Table S3). Interestingly, this screening revealed that E1\_10125 was active against  $\alpha$ (1-3) and  $\alpha$ (1-4) fucosylated substrates presenting a terminal sialic acid modification. The chromatograms clearly showed the appearance of peaks corresponding to Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3GlcNAc and Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc and the disappearance of the peaks corresponding to sLeA and sLeX (Figure 3A). The use of sialylated and de-sialylated *N*-glycans from human plasma confirmed the ability of E1\_10125 to accommodate sialyl residues in terminal location of fucosylated *N*-glycans (Figure 3B), as shown by the disappearance of the peak corresponding to fucosylated antennary *N*-glycan upon incubation with E1\_10125. In contrast, no reaction product was detected when HRP glycans (core  $\alpha$ 1,3-fucose), FA2G2 ( $\alpha$ 1,6-fucose) or blood group antigens ( $\alpha$ 1,2-fucose) were used as substrates as shown by LC–MS traces (Figures 3C–E).

**Table 1. Kinetic parameters of *R. gnavus* fucosidases towards pNP-Fuc.**

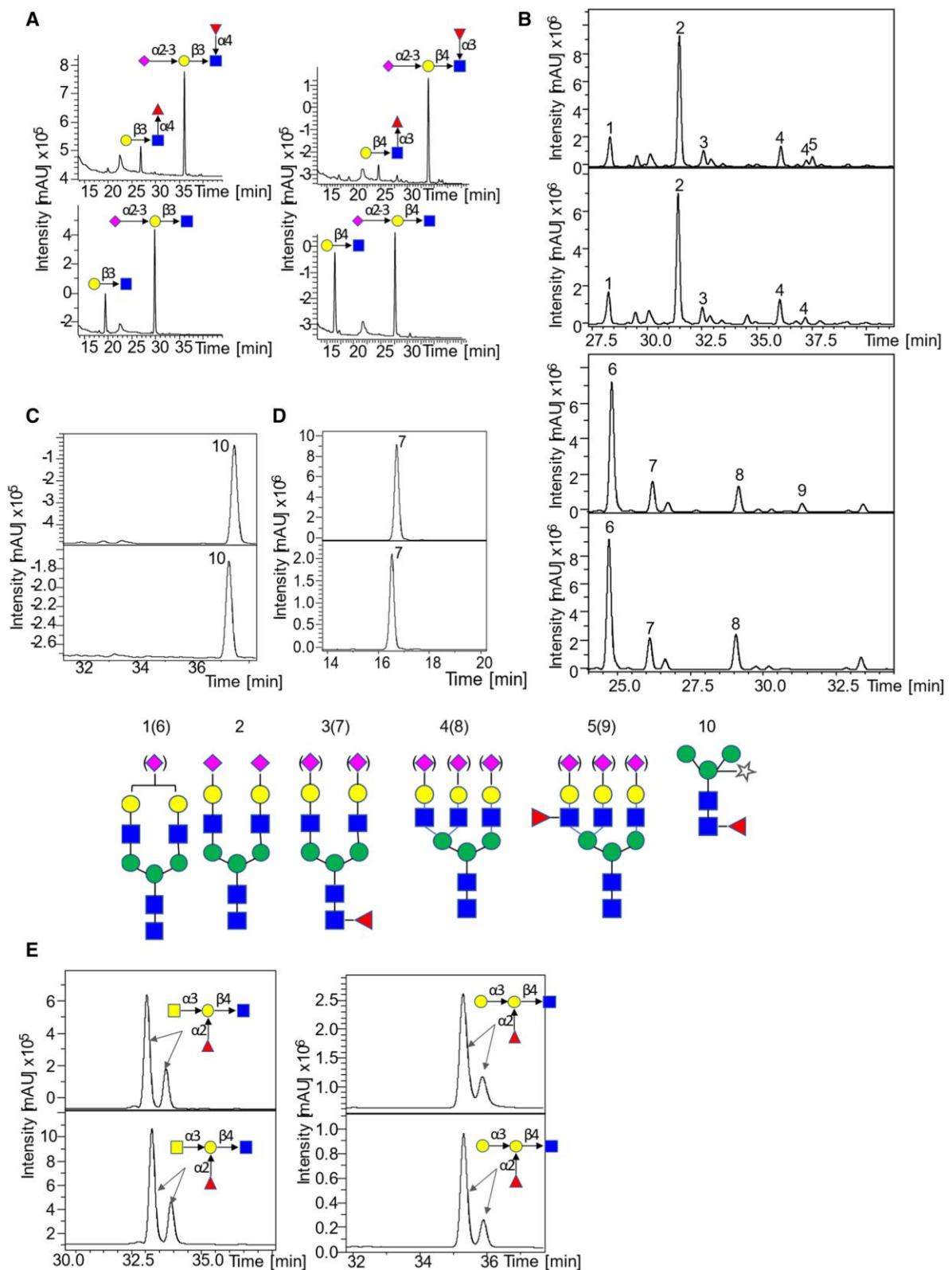
	<b>E1_10125</b>	<b>E1_10180</b>	<b>ATCC_03833</b>	<b>ATCC_00842</b>	<b>E1_10587*</b>
<b>Catalytic efficiency (<math>s^{-1} \cdot M^{-1}</math>)</b>	7.61	16.91	$4.7 \cdot 10^5$	28.14	0.72
<b><math>K_m</math> (<math>\mu M</math>)</b>	$237.9 \pm 39.69$	N.D.	$179.1 \pm 28.77$	$2958 \pm 362.8$	15010.51
<b><math>k_{cat}</math> (<math>s^{-1}</math>)</b>	$0.0018 \pm 0.000088$	N.D.	$83.6 \pm 2.97$	$0.0832 \pm 0.0032$	0.0108

*ND* could not be determined (using concentrations up to 20 mM pNP-Fuc). <sup>a</sup> Kinetic parameters were determined from the progress curve.

**Table 2. Specific activity of *R. gnavus* fucosidases towards fucosylated oligosaccharides.**

	<b>Specific activity (<math>U/\mu mol</math>)</b>			
	<b>2'FL</b>	<b>3FL</b>	<b>LeA</b>	<b>LeX</b>
<b>E1_10125</b>	$6.44E-03 \pm 7.76E-05$	$63.23 \pm 2.03$	$83.88 \pm 2.56$	$114.72 \pm 1.76$
<b>E1_10180</b>	$1.17E-03 \pm 2.24E-04$	$0.119 \pm 0.001$	$0.45 \pm 0.02$	$0.36 \pm 0.01$
<b>ATCC_03833</b>	$5.47 \pm 0.28$	$4.82E-02 \pm 7.09E-04$	$0.069 \pm 0.003$	N.D.
<b>ATCC_00842</b>	$1.13E+04 \pm 3.06E+02$	$15.96 \pm 0.67$	$1.38E-03 \pm 1.34E-04$	$1.46E-03 \pm 9.59E-05$
<b>E1_10587</b>	$5.21E-05 \pm 1.17E-05$	$1.05E-04 \pm 2.08E-05$	$1.76E-05 \pm 4.51E-06$	$2.98E-04 \pm 6.75E-06$

*ND* not detected under experimental conditions. *NS* not significant, less than  $0.01 U/\mu mol$ .



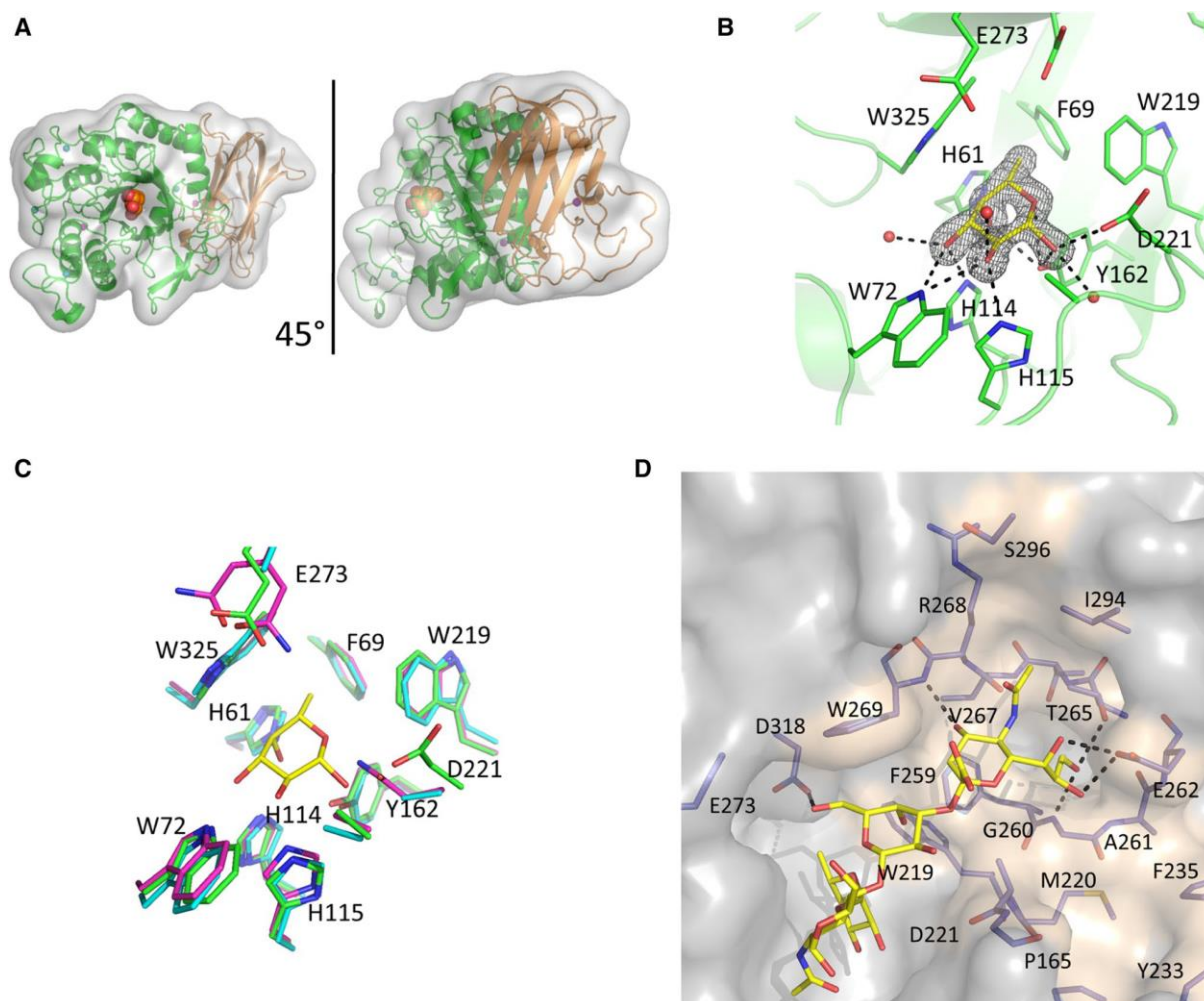
**Figure 3. LC-MS/MS analysis of *R. gnavus* GH29 fucosidase E1-10125 towards various fucosylated substrates. (A) LC-MS/MS analysis of the products released from the enzymatic reaction of E1-10125 with LeA (left) and LeX (right), the upper graph is the negative control and the lower one corresponds to the enzymatic reaction. (B) LC-MS analysis of the products**

released from the enzymatic reaction of E1-10125 with sialylated (upper) and desialylated (lower) human plasma. The negative controls are showed on top of the enzymatic reactions. **(C)** LC–MS analysis of the products released from the enzymatic reaction of E1-10125 with HRP (core  $\alpha$ 1,3-fucose) (lower). The negative control is shown in the upper panel. **(D)** LC–MS analysis of the products released from the enzymatic reaction of E1-10125 with FA2G2 ( $\alpha$ 1,6-fucose) (lower). The negative control is shown in the upper panel. **(E)** LC–MS analysis of the products released from the enzymatic reaction of E1-10125 with blood group A type II (left) and blood group B type II (right) (upper). The negative control is shown in the upper panel.

### **R. gnavus GH29 E1\_10125 fucosidase can accommodate terminal sialic acid moieties in $\alpha$ 1,3/4 antennary fucosylated substrates**

To further investigate E1\_10125 ligand specificity, the enzyme was crystallized in the presence of 2'FL, providing the crystal structure of the complex showing the  $\beta$ -fucose anomer bound in the active site (Figure 4). Data collection and refinement statistics are detailed in Table 3. Electron density maps allowed modelling of residues 23–527. The enzyme consists of two distinct domains, a catalytic domain (PF001120) domain comprising residues 46–366 in N-terminal and a F5/8 Type C domain (PF00754) covering residues 385–526 in C-terminal (Figure 4A). The catalytic domain displays a  $(\alpha/\beta)_8$  which is typical of GH29 enzymes ([www.cazy.org](http://www.cazy.org)) whereas the type C-domains shows structural homology with carbohydrate binding module (CBM) belonging to CBM32 family ([www.cazy.org](http://www.cazy.org)). The macromolecular architecture is conserved with the recently solved GH29 fucosidase enzymes from *S. pneumoniae* (RMSD: 0.969 Å) [22] and *Bifidobacterium longum* subsp. *infantis* (RMSD: 1.128 Å) [67]. Residues 23–45 wrap around the C-terminal  $\beta$ -sandwich domain. The catalytic machinery sits in a cleft in the center of the N-terminal domain (Figure 4B). By proximity to the bound fucose residue and homology to other fucosidases, Asp221 was identified as the catalytic nucleophile and Glu273 as the acid/base (amino acid numbering based on recombinant protein sequence) (Figure 4B). Trp325 creates a CH– $\pi$  interaction with the bound ligand and His61, Trp72, His114, His115 and Tyr162 provides additional hydrogen bonding interactions. Phe59 and Trp219 provide a hydrophobic pocket to accommodate the fucose C6 methyl group (Figure 4B). The fucose binding site is conserved with the *S. pneumoniae* [22] and *B. longum* GH29 fucosidases [67] and is henceforth referred to as the – 1 subsite (Figure 4C). In an attempt to obtain crystal structures in complex with the substrate, an active site mutant (E1\_10125 D221A) was generated. However, following co-crystallization experiments with 2'FL, fucose was found bound in the active site, indicating residual fucosidase activity. Clear electron density was present for both  $\alpha$ - and  $\beta$ -fucose anomers (Figure S3A and S3B). It was not possible to obtain crystals of E1\_10125 D221A in the absence of 2'FL. Attempts were made to displace the fucose molecule with other fucosylated ligands; however, these were unsuccessful, perhaps due to the substrate binding site being present at a tightly packed interface between two symmetry related protein molecules. No crystal structures could be obtained with other substrates tested including L-fucose, Neu5Ac, 3FL, LeA, LeX and sLeX.





**Figure 4. Crystal structure of *R. gnavus* GH29 fucosidase E1\_10125. (A)** Cartoon representation of E1\_10125 fucosidase, the catalytic domain is coloured green and the proposed CBM is coloured orange. A fucose residue in a sphere representation indicates the location of the active site. The views are related by a 45° rotation around the y axis. **(B)** The E1\_10125 fucose binding site. The  $\beta$ -anomer of fucose is shown in yellow with nearby active site residues shown in green. Black dashed lines indicate hydrogen-bonding interactions.  $F_o - F_c$  difference map density for the fucose residue is displayed as a black mesh, contoured at  $2\sigma$ . **(C)** The fucose binding sites of E1\_10125 (green), *S. pneumoniae* GH29 fucosidase (magenta), and *B. longum* subsp. *infantis* GH29 fucosidase (cyan) are aligned. Residue numbers refer to E1\_10125. The binding site residues are conserved across the three structures and differences present at the D221 and E273 positions are catalytic mutants. Fucose bound in the E1\_10125 is show in yellow for reference. **(D)** Model of the orientation and conformation of sLex bound to *R. gnavus* E1\_10125 proposed by MD simulations

Table 3. Data collection and refinement statistics.

Data set	WT - Fucose	D221A - Fucose
PDB identifier	6TR3	6TR4
<b>Data collection</b>		
Space group	C2	P1
a, b, c (Å)	164.2, 48.8, 132.1	49.8, 74.1, 76.5
$\alpha, \beta, \gamma$	90.0, 151.1, 90.0	82.5, 80.4, 70.4
Resolution	63.90-1.70 (1.73-1.70)	69.54-1.45 (1.47-1.45)
R <sub>merge</sub>	0.08 (0.59)	0.04 (0.11)
R <sub>meas</sub>	0.10 (0.72)	0.06 (0.16)
I/ $\sigma$	6.9 (1.1)	12.7 (3.4)
CC half	0.99 (0.58)	0.99 (0.96)
Completeness	99.1 (97.5)	92.7 (56.9)
Redundancy	3.1 (2.9)	2.0 (1.8)
<b>Refinement</b>		
Resolution	63.86-1.70	69.54-1.45
No. of reflections	55403	166349
R <sub>work</sub> /R <sub>free</sub>	0.174/0.227	0.141/0.161
Protein	4022	8134
Ligand/ion	30	48
Water	282	1507
Protein	25.22	11.5
Ligand/ion	39.0	6.7
Water	26.92	23.6
Bond lengths (Å)	0.03	0.01
Bond angle (°)	2.8	1.4

Numbers in parenthesis refer to the highest resolution shell.

Superimposition of the E1\_10125 crystal structure with that of  $\alpha$ -1,3/4-fucosidase from *B. longum* subsp. *Infantis* D172A/E217A mutant complexed with lacto-N-fucopentaose II (pdb 3UET) [67] (Supplementary Figure S4C) or with *S. pneumoniae* SpGH29CT D171N/E215Q in complex with LeA (pdb 6ORF) (Supplementary Figure S4D) or with LeX (pdb 6OR4) [62] (Supplementary Figure S3E) indicate that there are unlikely to be E1\_10125 interactions that form a distinct +1 site (GlcNAc in LeA and LeX trisaccharide antigens). E1\_10125 Trp269 is conserved with *S. pneumoniae* and *B. longum* fucosidases, maintaining the +2 site. More specifically, Trp269 is likely to form a CH- $\pi$  stacking interaction with the galactose ring, and Asp318 to form a hydrogen bond with the galactose C6 hydroxyl. In the E1\_10125 crystal structure, adjacent to the proposed +2 site is an open platform comprises primarily neutral and hydrophobic residues, which would accommodate Neu5Ac (Supplementary Figure S3F). This is in marked contrast to the *S. pneumoniae* and *B. longum* homologue structures where this region is partially occluded by incoming loops. Molecular modelling calculations were carried out to further support this hypothesis and to provide a model for the orientation of the sialic acid ring of sLeX when bound to E1\_10125 fucosidase (Figure 4D). Following 500 ns MD simulations of the E1\_10125 D221A mutant and docking of the sLeX ligand, the molecular model showed that sialic acid ring sits in the neighbouring subsite. Polar contacts are established between nearby residues, Glu262 and Trp269; in particular, and hydroxyl groups present at sialic acid C4, C8 and C9 positions. This analysis confirmed that the E1\_10125 fucosidase enzyme shows an open binding site able to accommodate the sLeX ligand.

In the absence of a complex structure of E1\_10125 with a fucosylated oligosaccharide and in order to test the hypothesis that the cavity could accommodate the sialic acid moiety, E1\_10125 R268W and E1\_10125 G260M mutants were produced in which these introduced side-chains are expected to block access to the cavity. The E1\_10125 R268W mutant showed a complete loss of activity towards all substrates tested including pNP-Fuc, 2'FL, 3FL, blood group A type II, blood group B type II, LeX and sLeX (data not shown) whereas the E1\_10125 G260M mutant showed a significant decrease in activity towards sLeX down to 28% activity while 76% activity remained towards LeX (Supplementary Table S4), suggesting that the cavity is important to accommodate terminal modifications of the fucosylated substrates.

Glycan arrays were then used to further define the ligand and linkage specificity of E1\_10125 (Supplementary Figure S5). The purified recombinant His6-tagged E1\_10125 D221A inactive mutant was screened at three protein concentrations against the Core H glycan microarray glycans at the Consortium for Functional Glycomics (CFG). Among the 585 glycans screened on the microarray, significant RFU values ( $> 300$ ) were obtained for 5 fucosylated glycans using the highest protein concentration. Glycan ID 389 with  $\alpha$ -Gal-LeA epitope displayed the highest RFU value ( $1072 \pm 47$ ) followed by two glycans, ID 249 and ID 526, containing sLeX epitopes. This recognition pattern therefore suggests that E1\_10125 could recognize fucosylated substrates with diverse terminal modifications at the reducing end.

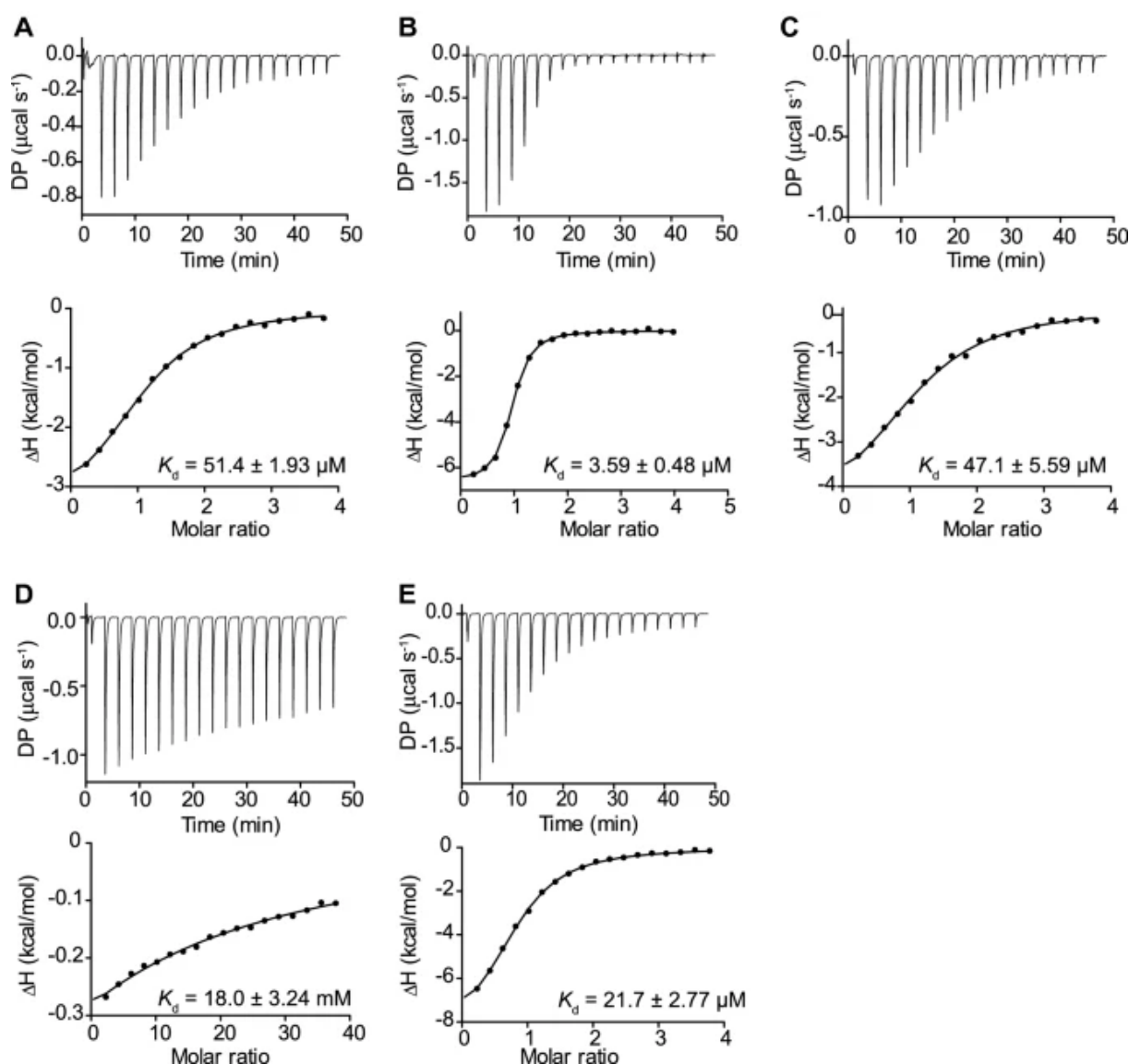
In order to further test this hypothesis, ITC was used to determine the binding parameters of E1\_10125 D221A mutant towards these ligands (Figure 5 and Supplementary Table S5). The enzyme bound to LeX with a  $K_d$  of  $51.43 \pm 1.93 \mu\text{M}$  (Figure 5A) and to sLeX with a  $K_d$  of  $3.59 \pm 0.48 \mu\text{M}$  (Figure 5B). Further, a  $K_d$  of  $47.13 \pm 5.60 \mu\text{M}$  was obtained when  $\alpha$ Gal-LeX (Figure 5c)

**Table 4. Kinetic parameters of E1\_10125 towards LeX, sLeX and  $\alpha$ Gal-LeX.**

	LeX	sLeX	$\alpha$ Gal-LeX
Catalytic efficiency ( $s^{-1}\cdot M^{-1}$ )	$1416.67 \pm 288.68$	$12874.85 \pm 1620.36$	$28888.89 \pm 2545.88$
$K_m$ ( $\mu M$ )	N.D.	$163.1 \pm 16.67$	N.D.
$K_{cat}$ ( $s^{-1}$ )	N.D.	$2.10 \pm 0.07$	N.D.

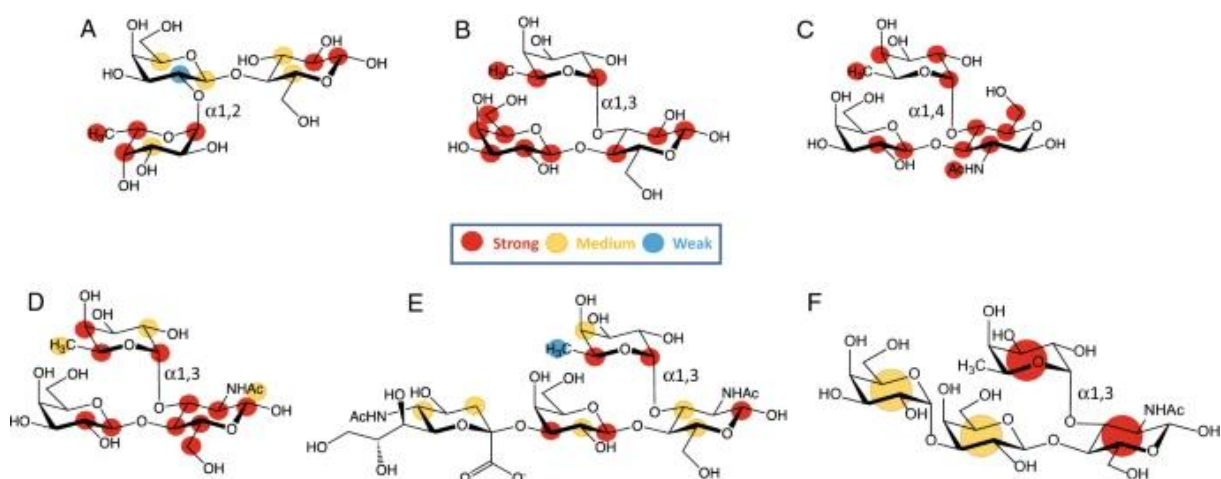
*ND could not be determined under experimental conditions.*

2



**Figure 5. ITC isotherms of *R. gnavus* GH29 fucosidase E1\_10125 binding to fucosylated ligands. (A) E1\_10125 binding to LeX. (B) E1\_10125 binding to sLeX. (C) E1\_10125 binding to  $\alpha$ Gal-LeX. (D) E1\_10125 binding to Neu5Ac. (E) E1\_10125 binding to L-fucose. DP differential power.**

was used as a ligand whereas a  $K_d$  of 17.98 mM and 21.7  $\mu$ M were obtained with the monosaccharides Neu5Ac or Fuc used as a control (Figure 5D, E). To compare the substrate specificity among LeX, sLeX and  $\alpha$ Gal-LeX, the kinetic parameters were determined against these substrates (Table 4). E1\_10125 showed strongest affinity to sLeX with a  $K_m$  of 163.1  $\mu$ M and the presence of sialic acid or galactose on the non-reducing end of LeX significantly increased the catalytic efficiency up to 20-fold, consistent with the binding parameters (Table 4).



**Figure 6. Binding epitope mapping from STD NMR spectroscopy depicting interactions of *R. gnavus* GH29 fucosidase E1-10125 with fucosylated oligosaccharides.** Normalized saturation transfer intensities (0–100%) from STD NMR experiments mapped onto the chemical structures of (A) 2'FL, (B) 3FL, (C) LeA, (D) LeX, (E) sLeX and (F)  $\alpha$ Gal-LeX. Stronger normalized STD intensities correlate with closer ligand contacts with the surface of the protein in the bound state. Legend indicates normalized STD intensities: blue, 0–24%; yellow, 25–50%, red 51–100%. The enzyme intimately recognizes 3FL and LeA, whereas looser contacts are observed for 2'FL, LeX, sLeX and  $\alpha$ Gal-LeX. For the latter, a much higher degree of proton chemical shift overlapping implied lower binding epitope resolution and a normalized STD intensity value was assigned for each ring, as an average of the STD intensities of its isolated protons.

To gain further structural insights into the unique ligand specificity of E1\_10125, STD NMR studies [68] were conducted with E1\_10125 D221A mutant in the presence of 2'FL, 3FL, LeA, LeX, sLeX and  $\alpha$ Gal-LeX (Figure 6). Transfer of magnetization as saturation from the protein to the ligand was observed for all substrates tested, in agreement with the binding of E1\_10125 to these substrates with medium-weak affinities ( $\mu$ M–mM). The enzyme intimately recognized the three sugar residues constituting 3FL and LeA (Figures 6A, B) with no significant differences in their binding epitopes. 2'FL showed binding to E1\_10125 D221A by STD NMR (Figure 6C), but the main contacts were restricted to the fucose residue, whereas loose contacts were observed with the lactose disaccharidic moiety. The binding epitope of LeX also revealed a reduction in close contacts with the enzyme, particularly at the fucose

residue. A comparison between 3FL (Figure 6A) and LeX (Figure 6D) supports an impact of the acetamido group (NHAc) at position 2 of GlcNAc on binding, which leads to changes in the contacts of the fucose ring with the protein in the bound state. This is highlighted by the reduction of the relative STD intensity of the methyl group at position 6 of this ring in the case of LeX. It is well established that the NHAc group in LeX limits the flexibility of the Fuc $\alpha$ 1–3GlcNAc linkage via steric hindrance with the adjacent fucose ring [69]. This leads to the observed changes in the contacts of the fucose ring in the bound state of LeX in comparison 3FL (Figure 6D). These structural changes together with the advantageous reduction in entropy penalty upon binding expected for LeX due to the limited inter-glycosidic flexibility, are in good agreement with the observed differences in fucosidase activities (Table 2). The STD NMR results, in alignment with the activity assays and LC–MS/MS data, confirmed that the E1\_10125 fucosidase shows a preference for  $\alpha$ (1-3/4) linkage (Figure 3, Table 2), in which the fucose is linked at the reducing glucopyranose ring of the Gal $\beta$ 1–3/4Glc(NAc) disaccharidic sequence. Interestingly, STD NMR revealed that the sialic acid moiety of sLeX makes contacts with the enzyme at C3 and C5 positions (Figure 6E), suggesting that the sialic acid moiety is in part solvent exposed, and in part surrounded by residues at the protein surface, in agreement with the crystal structure showing a cavity that could accommodate such a sialic acid residue at the non-reducing end of the ligand. The molecular model (Figure 4D) is also in excellent agreement with the experimental NMR data, as in this binding mode, protons at C3 and C5 are pointing towards the surface of the enzyme in the pocket. Likewise, in  $\alpha$ Gal–LeX, weak contacts were observed for the non-reducing  $\alpha$ Gal and  $\beta$ Gal rings in the bound state, whereas fucose was the main ligand recognition moiety, followed by GlcNAc (Figure 6F). Together, these data support the X-ray crystal structure that the binding pocket of E1\_10125 could accommodate terminal residues although with a clear preference for sialic acid.

## Discussion

Fucose decorating glycan chains in HMOs or mucins contributes to shaping the composition of the gut microbiota in adults and infants. Previous studies in mice showed that the loss of the  $\alpha$ -1,2-fucosyltransferase FUT2, and therefore fucosylated host glycans, leads to a decreased diversity and differences in intestinal microbial community [70–73], whereas an association between the composition of the intestinal microbiota and the ABO blood group or FUT2 secretor status was reported in humans [72, 74–77]. Human fetal mucins along the GI tract harbor a repertoire of *O*-glycans similar to HMOs [78, 79] which may also contribute to the differences in gut microbiota composition as compared to adults [80]. The ability to utilize fucosyllactose is a trait of early inhabitants of the human GI tract, such as *R. gnavus* [27] or various bifidobacteria species [81] as well as probiotic strains, such as *Lactobacillus casei* [82]. To access this nutrient source, gut bacteria have evolved to express a wide range of fucosidases with distinct ligand specificity, contributing to their fitness across nutritional niches [5, 6]. Furthermore, in its free form, fucose released by bacterial fucosidases may affect gut homeostasis. For example, *B. thetaiotaomicron* produces multiple fucosidases that cleave fucose from host glycans, resulting in high fucose availability in the gut lumen [83] which can then act as a signal to modulate the pathogenicity and metabolism of the pathogen enterohaemorrhagic *E. coli* (EHEC) [84].

Complexity in HMOs or mucins lies in the diversity of glycosidic bonds in these molecules, rendering a large number of potential combinations. Since fucosylation varies across and along the intestine and that fucosidase activity is dependent on the type of linkages present in the glycans or glycoconjugates, it is critical to understand the ligand specificity of the fucosidases encoded by major gut symbionts. Recently, the substrate specificities of two structure-solved GH29 fucosidases from *B. thetaiotaomicron* VPI-5482 were determined showing that the protein with locus tag BT 2970 belongs to GH29-A while BT 2192 belongs to GH29-B [15]. *R. gnavus* is a human gut symbiont of the infant and adult microbiota [23–25]. Here, we showed that *R. gnavus* strains encode a range of fucosidases belonging to GH95 and GH29-A and GH29-B families with varied specificities, highlighting the versatility of fucosylated substrates they can access, such as those present in HMOs or intestinal mucins. The enzymatic characterization of *R. gnavus* fucosidases focused on previously uncharacterized bacterial fucosidases revealed that fucosidase ATCC\_03833 belongs to GH29-A while E1\_10125 belongs to GH29-B subfamily. GH29-B E1\_10125 and E1\_10180 showed strict substrate specificity towards  $\alpha$ 1,3/4 fucosylated linkages while fucosidases GH95 ATCC\_00842 and GH29-A ATCC\_03833 showed preference for  $\alpha$ 1,2 linkages, as reported for *B. bifidum* AfcA fucosidase [85]. *O*-glycan analyses of human fetal mucins showed that fucose is present in a large variety of terminal linkages, including blood group H as well as LeA (Gal $\beta$ (1–3)[Fuc $\alpha$ 1-4]GlcNAc), LeB (Fuc $\alpha$ 1-2Gal $\beta$ 1-3[Fuc $\alpha$ 1-4]GlcNAc), LeX (Gal $\beta$ (1–4 [Fuc $\alpha$ 1-3]GlcNAc) and LeY (Fuc $\alpha$ 1-2Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]GlcNAc) determinants [79]. The diversity of fucosidases may confer *R. gnavus* strains with an advantage in colonizing the infant gut [25].

In addition, we showed that E1\_10125 could act on LeA and LeX even when the galactose moiety was linked to a sialic acid residue or other decorations. In particular, E1\_10125 showed highest affinity towards sLeX. The ability to accommodate the sialic acid moiety, as also confirmed by STD NMR, appears to be enabled by an open region adjacent to the +2 site (Gal residue in Le antigens), which comprises residues Met220, Phe259, Gly260, Ala261 and Thr265 with additional stabilizing interactions likely to be provided by Trp269 and Glu262. This structural arrangement lacks the incoming loops present in the *S. pneumoniae* and *B. longum* GH29 enzymes [22, 67], supporting the unique specificity of *R. gnavus* E1\_10125 fucosidase. Glycan array analyses suggested that E1\_10125 could recognize fucosylated glycans with diverse terminal modifications as also supported by ITC showing binding of E1\_10125 to both sialic acid and  $\alpha$ Gal linked to Gal of LeX. Extensive differences in the glycosylation profile of mucins occur along the GI tract, characterized by the presence of decreasing gradients of fucose and ABH blood group and increasing gradients of sialic acid from ileum to rectum [7, 86]. In human colonic mucin, more than 100 complex *O*-linked oligosaccharides were identified, mostly based on the core 3 structure with sialic acid at the 6-position of the GalNAc [9]. The most abundant components were -Gal-(Fuc)GlcNAc-3(NeuAc-6)GalNAc, GalNAc-(NeuAc-)Gal-4/3GlcNAc-3(NeuAc-6)GalNAcGalNAc-3(NeuAc-6) GalNAc and GlcNAc-3(NeuAc-6)GalNAc [9]. The unusual specificity of E1\_10125 may, therefore, contribute to the fitness and spatial adaptation of *R. gnavus* strains into the adult human GI tract [23, 24].

The specificities of *R. gnavus* fucosidases could be exploited for diagnostic assays. For example, changes in the abundance of antennary fucosylation in plasma *N*-glycans have been

associated with diabetes [87, 88] and with colorectal cancer [89–91]. The quantitation of these low abundant antennary fucosylated glycans in the plasma *N*-glycome is complex due to the structural diversity of its component glycans [92, 93]. The analysis is widely performed on chromatographic platforms which generally require extensive measurement time [91–95]. However, the recent technological advances integrating the use of fucosidases or other glycosidases and analysis on a MALDI-MS platform enabled identification and quantification of glycans of specific fucose isomers [96, 97]. The antennary fucosidase specificity reported in this work could therefore be used as a discriminatory tool to identify *N*-glycan biomarkers of diseases and as a valuable tool for the purpose of glycoprofiling biopharmaceutical glycoproteins.

2

## Supplementary information

Supplementary tables and figures are available online free of charge via <https://link.springer.com/article/10.1007/s00018-020-03514-x#Sec18>

## Acknowledgments

This research was supported by the support of the Biotechnology and Biological Sciences Research Council (BBSRC); being mostly funded by the Innovate UK Biocatalyst grant Glycoenzymes for Bioindustries (BB/M029042/) with contribution from the Royal Society and the BBSRC Institute Strategic Programmes BB/J004529/1 ‘The Gut Health and Food Safety’ and BB/R012490/1 ‘Gut Microbes and Health’ and its constituent project BBS/E/F/000PR10353 (Theme 1, Determinants of microbe-host responses in the gut across life). This research was also supported by European Union (GlySign, Grant No. 722095) and the NWO (Vernieuwingsimpuls Veni Project No. 722.016.008). I would also like to thank Diamond Light Source beamlines VMXi, I03 and I04 for beamtime and assistance, as well as the crystallization facility at Harwell for access and support. Additionally, I wish to acknowledge the Consortium for Functional Glycomics grant number GM62116 and GM098791 for glycan screening.



## References

1. Thursby E, Juge N (2017) Introduction to the human gut microbiota. *Biochem J* 474(11):1823–1836.
2. Bode L (2012) Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* 22(9):1147–1162.
3. Cabrera-Rubio R, Kunz C, Rudloff S, Garcia-Mantrana I, Crehua-Gaudiza E, Martinez-Costa C, Collado MC (2019) Association of maternal secretor status and human milk oligosaccharides with milk microbiota: an observational pilot study. *J Pediatr Gastr Nutr* 68(2):256–263.
4. Koropatkin NM, Cameron EA, Martens EC (2012) How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 10(5):323–335.
5. Tailford LE, Crost EH, Kavanaugh D, Juge N (2015) Mucin glycan foraging in the human gut microbiome. *Front Genet* 6:ARTN81.
6. Ndeh D, Gilbert HJ (2018) Biochemistry of complex glycan depolymerisation by the human gut microbiota. *FEMS Microbiol Rev* 42(2):146–164.
7. Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP, Michalski JC (2003) Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. *J Biol Chem* 278(47):46337–46348.
8. Jensen PH, Kolarich D, Packer NH (2010) Mucin-type O-glycosylation—putting the pieces together. *Febs J* 277(1):81–94.
9. Larsson JM, Karlsson H, Sjoval H, Hansson GC (2009) A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* 19(7):756–766.
10. Mollicone R, Bara J, Le Pendu J, Oriol R (1985) Immunohistologic pattern of type 1 (Lea, Leb) and type 2 (X, Y, H) blood group-related antigens in the human pyloric and duodenal mucosae. *Lab Invest* 53(2):219–227
11. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB (1995) Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J Biol Chem* 270(9):4640–4649.
12. Henry SM, Jovall PA, Ghardashkhani S, Gustavsson ML, Samuelsson BE (1995) Structural and immunochemical identification of Leb glycolipids in the plasma of a group O Le(a-b-) secretor. *Glycoconj J* 12(3):309–317.
13. Larsson JMH, Thomsson KA, Rodriguez-Pineiro AM, Karlsson H, Hansson GC (2013) Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *Am J Physiol-Gastr L* 305(5):G357–G363.
14. Sano M, Hayakawa K, Kato I (1992) Purification and characterization of alpha-L-fucosidase from *Streptomyces* species. *J Biol Chem* 267(3):1522–1527
15. Sakurama H, Tsutsumi E, Ashida H, Katayama T, Yamamoto K, Kumagai H (2012) Differences in the substrate specificities and active-site structures of two alpha-L-fucosidases (glycoside hydrolase family 29) from *Bacteroides thetaiotaomicron*. *Biosci Biotechnol Biochem* 76(5):1022–1024.

16. Sulzenbacher G, Bignon C, Nishimura T, Tarling CA, Withers SG, Henrissat B, Bourne Y (2004) Crystal structure of *Thermotoga maritima* alpha-L-fucosidase. Insights into the catalytic mechanism and the molecular basis for fucosidosis. *J Biol Chem* 279(13):13119–13128.
17. Lezyk M, Jers C, Kjaerulff L, Gotfredsen CH, Mikkelsen MD, Mikkelsen JD (2016) Novel alpha-L-Fucosidases from a soil metagenome for production of fucosylated human milk oligosaccharides. *PLoS ONE* 11(1):e0147438.
18. Wongmadden ST, Landry D (1995) Purification and characterization of novel glycosidases from the bacterial genus *xanthomonas*. *Glycobiology* 5(1):19–28.
19. Megson ZA, Koerdt A, Schuster H, Ludwig R, Janesch B, Frey A, Naylor K, Wilson IB, Stafford GP, Messner P, Schaffer C (2015) Characterization of an alpha-l-fucosidase from the periodontal pathogen *Tannerella forsythia*. *Virulence* 6(3):282–292.
20. Ashida H, Miyake A, Kiyohara M, Wada J, Yoshida E, Kumagai H, Katayama T, Yamamoto K (2009) Two distinct alpha-l-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* 19(9):1010–1017.
21. Fan S, Zhang H, Chen X, Lu L, Xu L, Xiao M (2016) Cloning, characterization, and production of three alpha-L-fucosidases from *Clostridium perfringens* ATCC 13124. *J Basic Microbiol* 56(4):347–357.
22. Hobbs JK, Pluvineau B, Robb M, Smith SP, Boraston AB (2019) Two complementary alpha-fucosidases from *Streptococcus pneumoniae* promote complete degradation of host-derived carbohydrate antigens. *J Biol Chem* 294(34):12670–12682.
23. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Meta HITC, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59–65.
24. Kraal L, Abubucker S, Kota K, Fischbach MA, Mitreva M (2014) The prevalence of species and strains in the human microbiome: a resource for experimental efforts. *PLoS ONE* 9(ARTNe97279):7279.
25. Sagheddu V, Patrone V, Miragoli F, Puglisi E, Morelli L (2016) infant early gut colonization by lachnospiraceae: high frequency of *Ruminococcus gnavus*. *Front Pediatr* 4:57.
26. Hall AB, Yassour M, Sauk J, Garner A, Jiang XF, Arthur T, Lagoudas GK, Vatanen T, Fornelos N, Wilson R, Bertha M, Cohen M, Garber J, Khalili H, Gevers D, Ananthakrishnan AN, Kugathasan S, Lander ES, Blainey P, Vlamakis H, Xavier RJ, Huttenhower C (2017) A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Med* 9:ARTN103.
27. Crost EH, Tailford LE, Le Gall G, Fons M, Henrissat B, Juge N (2013) Utilisation of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain-dependent. *PLoS ONE* 8(10):ARTNe76341.

28. Crost EH, Tailford LE, Monestier M, Swarbreck D, Henrissat B, Crossman LC, Juge N (2016) The mucin-degradation strategy of *Ruminococcus gnavus*: The importance of intramolecular trans-sialidases. *Gut Microbes* 7(4):302–312.
29. Tailford LE, Owen CD, Walshaw J, Crost EH, Hardy-Goddard J, Le Gall G, de Vos WM, Taylor GL, Juge N (2015) Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat Commun* 6:7624.
30. Bell A, Brunt J, Crost E, Vaux L, Nepravishta R, Owen CD, Latousakis D, Xiao A, Li W, Chen X, Walsh MA, Claesen J, Angulo J, Thomas GH, Juge N (2019) Elucidation of a sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut. *Nat Microbiol* 4(12):2393–2404.
31. Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, Owens RJ (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* 35(6):e45
32. Sanchez-Weatherby J, Sandy J, Mikolajek H, Lobley CMC, Mazzorana M, Kelly J, Preece G, Littlewood R, Sorensen TLM (2019) VMXi: a fully automated, fully remote, high-flux in situ macromolecular crystallography beamline. *J Synchrotron Radiat* 26(Pt 1):291–301.
33. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):72–82.
34. Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr Sect D Biol Crystallogr* 69:1204–1214. <https://doi.org/10.1107/S0907444913000061>
35. Winter G, Lobley CMC, Prince SM (2013) Decision making in xia2. *Acta Crystallogr Sect D Biol Crystallogr* 69:1260–1273.
36. Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, Fuentes-Montero L, Vollmar M, Michels-Clark T, Young ID, Sauter NK, Evans G (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D* 74:85–97.
37. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Method Enzymol* 374:300–321.
38. Emsley P (2017) Tools for ligand validation in Coot. *Acta Crystallogr D* 73:203–210.
39. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* 67:235–242.
40. van Beusekom B, Joosten K, Hekkelman ML, Joosten RP, Perrakis A (2018) Homology-based loop modeling yields more complete crystallographic protein structures. *IUCrJ* 5(Pt 5):585–594.
41. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci* 27(1):293–315.
42. Schrödinger Release 2020–1: Maestro S, LLC, New York, NY, 2020.

43. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 27(3):221–234.
44. Olsson MH, Sondergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7(2):525–537.
45. Case DAIYB-S, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein DJ, Merz KM, Miao Y, Monard G, Nguyen C, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Smith J, Salomon-Ferrer R, Swails J, Walker RC, Wang J, Wei H, Wolf RM, Wu X, Xiao L, York DM, Kollman PA (2018) AMBER 2018. University of California, San Francisco
46. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23(3):327–341.
47. Roe DR, Cheatham TE 3rd (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 9(7):3084–3095.
48. Schrödinger Release 2020–1: LigPrep S, LLC, New York, NY, 2020.
49. Schrödinger Release 2020–1: MacroModel S, LLC, New York, NY, 2020.
50. Schrödinger Release 2020–1: Glide S, LLC, New York, NY, 2020.
51. Davis MI, Shen M, Simeonov A, Hall MD (2016) Diaphorase Coupling Protocols for Red-Shifting Dehydrogenase Assays. *Assay Drug Dev Techn* 14(3):207–212.
52. Matsui I, Ishikawa K, Matsui E, Miyairi S, Fukui S, Honda K (1991) Subsite structure of *Saccharomyces cerevisiae*  $\alpha$ -amylase secreted from *Saccharomyces cerevisiae*. *J Biochem* 109:566–569.
53. Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26(5):680–682.
54. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR (1854) Whalen KL (2015) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* 8:1019–1037.
55. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.
56. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RDAB (2005) Protein identification and analysis tools on the ExPASy server. In: *The proteomics protocols handbook*. Humana Press, pp 571–607.
57. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47(W1):W636–W641.
58. Armenteros JJA, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37(4):420.

59. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305(3):567–580.
60. Fimereli DK, Tsirigos KD, Litou ZI, Liakopoulos TD, Bagos SJHPG (2012) CW-PRED: A HMM-based method for the classification of cell wall-anchored proteins of Gram-positive bacteria. *Springer's Lect Notes Comput Sci* 7297(2012):285–290.
61. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26(13):1608–1615.
62. Sela DA, Garrido D, Lerno L, Wu S, Tan K, Eom HJ, Joachimiak A, Lebrilla CB, Mills DA (2012) *Bifidobacterium longum* subsp. *infantis* ATCC 15697  $\alpha$ -fucosidases are active on fucosylated human milk oligosaccharides. *Appl Environ Microbiol* 78(3):795–803.
63. Zeuner B, Muschiol J, Holck J, Lezyk M, Gedde MR, Jers C, Mikkelsen JD, Meyer AS (2018) Substrate specificity and transfucosylation activity of GH29  $\alpha$ -L-fucosidases for enzymatic production of human milk oligosaccharides. *N Biotechnol* 41:34–45.
64. Ono A, Suzuki T, Gotoh S, Kono H, Matsui M, Aoki D, Matsuda M, Kawagishi H, Ogata M (2019) Structural investigation of  $\alpha$ -L-fucosidase from the pancreas of *Patiria pectinifera*, based on molecular cloning. *Carbohydr Res* 475:27–33.
65. Shaikh FA, Lammerts van Bueren A, Davies GJ, Withers SG (2013) Identifying the catalytic acid/base in GH29  $\alpha$ -L-fucosidase subfamilies. *Biochemistry* 52(34):5857–5864.
66. Zeleny R, Leonard R, Dorfner G, Dalik T, Kolarich D, Altmann F (2006) Molecular cloning and characterization of a plant  $\alpha$ 1,3/4-fucosidase based on sequence tags from almond fucosidase I. *Phytochemistry* 67(7):641–648.
67. Sakurama H, Fushinobu S, Hidaka M, Yoshida E, Honda Y, Ashida H, Kitaoka M, Kumagai H, Yamamoto K, Katayama T (2012) 1,3–1,4- $\alpha$ -L-fucosynthase that specifically introduces Lewis a/x antigens into type-1/2 Chains. *J Biol Chem* 287(20):16709–16719.
68. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem Int* 38:1784–1788
69. Imberty A, Perez S (2000) Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem Rev* 100(12):4567–4588.
70. Meng D, Newburg DS, Young C, Baker A, Tonkonogy SL, Sartor RB, Walker WA, Nanthakumar NN (2007) Bacterial symbionts induce a FUT2-dependent fucosylated niche on colonic epithelium via ERK and JNK signaling. *Am J Physiol Gastrointest Liver Physiol* 293(4):G780–787.
71. Kashyap PC, Marcobal A, Ursell LK, Smits SA, Sonnenburg ED, Costello EK, Higginbottom SK, Domino SE, Holmes SP, Relman DA, Knight R, Gordon JI, Sonnenburg JL (2013) Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proc Natl Acad Sci USA* 110(42):17059–17064.
72. Tong M, McHardy I, Ruegger P, Goudarzi M, Kashyap PC, Haritunians T, Li X, Graeber TG, Schwager E, Huttenhower C, Fornace AJ, Sonnenburg JL, McGovern DPB, Borneman J, Braun J (2014) Reprogramming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism. *Isme J* 8(11):2193–2206.

73. Rausch P, Kunzel S, Suwandi A, Grassl GA, Rosenstiel P, Baines JF (2017) Multigenerational influences of the *fut2* gene on the dynamics of the gut microbiota in mice. *Front Microbiol* 8:ARTN991.
74. Rausch P, Rehman A, Kunzel S, Hasler R, Ott SJ, Schreiber S, Rosenstiel P, Franke A, Baines JF (2011) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and *FUT2* (Secretor) genotype. *Proc Natl Acad Sci USA* 108(47):19030–19035.
75. Wacklin P, Tuimala J, Nikkila J, Tims S, Makivuokko H, Alakulppi N, Laine P, Rajilic-Stojanovic M, Paulin L, de Vos WM, Matto J (2014) Faecal microbiota composition in adults is associated with the *FUT2* gene determining the secretor status. *PLoS ONE* 9(4):ARTNe94863.
76. Davenport ER, Goodrich JK, Bell JT, Spector TD, Ley RE, Clark AG (2016) ABO antigen and secretor statuses are not associated with gut microbiota composition in 1,500 twins. *Bmc Genom* 17:ARTN 941.
77. Turpin W, Bedrani L, Espin-Garcia O, Xu W, Silverberg MS, Smith MI, Guttman DS, Griffiths A, Moayyedi P, Panaccione R, Huynh H, Steinhart H, Aumais G, Shestopaloff K, Dieleman LA, Turner D, Paterson AD, Croitoru K (2018) *FUT2* genotype and secretory status are not associated with fecal microbial composition and inferred function in healthy subjects. *Gut Microbes* 9(4):357–368.
78. Hounsell EF (1989) Novel core, backbone and peripheral region sequences of the oligosaccharides of foetal gastrointestinal mucins present in human meconium. *Symp Soc Exp Biol* 43:149–154
79. Robbe-Masselot C, Maes E, Rousset M, Michalski JC, Capon C (2009) Glycosylation of human fetal mucins: a similar repertoire of O-glycans along the intestinal tract. *Glycoconj J* 26(4):397–413.
80. Rodriguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, Avershina E, Rudi K, Narbad A, Jenmalm MC, Marchesi JR, Collado MC (2015) The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis* 26:26050.
81. Bunesova V, Lacroix C, Schwab C (2016) Fucosyllactose and L-fucose utilization of infant *Bifidobacterium longum* and *Bifidobacterium kashiwanohense*. *Bmc Microbiol* 16:ARTN 248.
82. Rodriguez-Diaz J, Monedero V, Yebra MJ (2011) Utilization of natural fucosylated oligosaccharides by three novel alpha-L-fucosidases from a probiotic *Lactobacillus casei* strain. *Appl Environ Microbiol* 77(2):703–705.
83. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 299(5615):2074–2076.
84. Pacheco AR, Curtis MM, Ritchie JM, Munera D, Waldor MK, Moreira CG, Sperandio V (2012) Fucose sensing regulates bacterial intestinal colonization. *Nature* 492(7427):113–117.
85. Katayama T, Sakuma A, Kimura T, Makimura Y, Hiratake J, Sakata K, Yamanoi T, Kumagai H, Yamamoto K (2004) Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2-alpha-L-fucosidase (*AfcA*), a novel inverting glycosidase (glycoside hydrolase family 95). *J Bacteriol* 186(15):4885–4893.

86. Robbe C, Capon C, Coddeville B, Michalski JC (2004) Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem J* 384(Pt 2):307–316.
87. Juszcak A, Pavic T, Vuckovic F, Bennett AJ, Shah N, Pape Medvidovic E, Groves CJ, Sekerija M, Chandler K, Burrows C, Rojnic Putarek N, Vucic Lovrencic M, Cuca Knezevic J, James TJ, Gloyn AL, Lauc G, McCarthy MI, Owen KR, Gornik O (2019) Plasma fucosylated glycans and c-reactive protein as biomarkers of HNF1A-MODY in young adult-onset nonautoimmune diabetes. *Diabetes Care* 42(1):17–26.
88. Thanabalasingham G, Huffman JE, Kattla JJ, Novokmet M, Rudan I, Gloyn AL, Hayward C, Adamczyk B, Reynolds RM, Muzinic A, Hassanali N, Pucic M, Bennett AJ, Essafi A, Polasek O, Mughal SA, Redzic I, Primorac D, Zgaga L, Kokic I, Hansen T, Gasperikova D, Tjora E, Strachan MWJ, Nielsen T, Stanik J, Klimes I, Pedersen OB, Njolstad PR, Wild SH, Gyllensten U, Gornik O, Wilson JF, Hastie ND, Campbell H, McCarthy MI, Rudd PM, Owen KR, Lauc G, Wright AF (2013) Mutations in HNF1A result in marked alterations of plasma glycan profile. *Diabetes* 62(4):1329–1337.
89. Holst S, Deuss AJ, van Pelt GW, van Vliet SJ, Garcia-Vallejo JJ, Koeleman CA, Deelder AM, Mesker WE, Tollenaar RA, Rombouts Y, Wuhrer M (2016) N-glycosylation profiling of colorectal cancer cell lines reveals association of fucosylation with differentiation and caudal type homeobox 1 (CDX1)/Villin mRNA expression. *Mol Cell Proteom* 15(1):124–140.
90. de Vroome SW, Holst S, Gironde MR, van der Burgt YEM, Mesker WE, Tollenaar R, Wuhrer M (2018) Serum N-glycome alterations in colorectal cancer associate with survival. *Oncotarget* 9(55):30610–30623.
91. Doherty M, Theodoratou E, Walsh I, Adamczyk B, Stockmann H, Agakov F, Timofeeva M, Trbojevic-Akmacic I, Vuckovic F, Duffy F, McManus CA, Farrington SM, Dunlop MG, Perola M, Lauc G, Campbell H, Rudd PM (2018) Plasma N-glycans in colorectal cancer risk. *Sci Rep* 8(1):8655.
92. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, Kim YG, Henry GW, Shadick NA, Weinblatt ME, Lee DM, Rudd PM, Dwek RA (2008) HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal Biochem* 376(1):1–12.
93. Stumpo KA, Reinhold VN (2010) The N-glycome of human plasma. *J Proteome Res* 9(9):4823–4830.
94. Knezevic A, Bones J, Kracun SK, Gornik O, Rudd PM, Lauc G (2011) High throughput plasma N-glycome profiling using multiplexed labelling and UPLC with fluorescence detection. *Analyst* 136(22):4670–4673.
95. Pivac N, Knezevic A, Gornik O, Pucic M, Igl W, Peeters H, Crepel A, Steyaert J, Novokmet M, Redzic I, Nikolac M, Hercigonja VN, Curkovic KD, Curkovic M, Nedic G, Muck-Seler D, Borovecki F, Rudan I, Lauc G (2011) Human plasma glycome in attentiondeficit hyperactivity disorder and autism spectrum disorders. *Mol Cell Proteom* 10(1):M110004200.
96. Benicky J, Sanda M, Pompach P, Wu J, Goldman R (2014) Quantification of fucosylated hemopexin and complement factor H in plasma of patients with liver disease. *Anal Chem* 86(21):10716–10723.

97. Rebello OD, Nicolardi S, Lageveen-Kammeijer GSM, Nouta J, Gardner RA, Mesker WE, Tollenaar RAEM, Spencer DIR, Wuhrer M, Falck D (2020) A matrix-assisted laser desorption/ionization—mass spectrometry assay for the relative quantitation of antennary fucosylated n-glycans in human plasma. *Front Chem* 8:138.
98. Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, Prestegard JJ, Schnaar RL, Freeze HH, Marth JD, Bertozzi CR, Etzler ME, Frank M, Vliegenthart JF, Lutteke T, Perez S, Bolton E, Rudd P, Paulson J, Kanehisa M, Toukach P, Aoki-Kinoshita KF, Dell A, Narimatsu H, York W, Taniguchi N, Kornfeld S (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* 25(12):1323–1324.





# Chapter 3

**Crystal structure of a GH20  
*N*-acetylglucosaminidase with specificity  
for bisecting *N*-acetylglucosamines on *N*-glycans**

## **Abstract**

Research into how the human gut microbiota degrades different glycans and polysaccharides has led to the characterisation of numerous Carbohydrate-Active enZymes (CAZymes) of huge diversity in terms of specificity. Recently, we described the set of CAZymes employed by a prominent member of the adult human gut (*Bacteroides thetaiotaomicron*) to degrade mammalian complex *N*-glycans [1]. This report follows up on one of these enzymes, BT 0456, with a molecular structure and discusses potential applications within glycoanalytics.

## Introduction

Post-translational glycan modifications are important for a variety of functions and processes in proteins. For instance, these glycans can provide protection of the protein from degradation and can be the ligands in many cellular processes [2, 3]. Changes to the composition of particular glycan decorations are characterised in a number of disease states, which means that they can be used as a biomarker for those diseases [4, 5]. To achieve this, a variety of glycoanalytic techniques can be applied to assess glycan composition [6, 7]. One way in which these approaches are being developed involves the application of enzymes that are highly specific towards particular sugars residues and their linkages, to characterise glycans [8-11]. These enzyme tools make current glycoanalytic techniques, such as liquid chromatography (LC) for fluorescently labelled glycans, much more powerful by adding an extra level of accurate characterisation [12]. Application of these enzymes can solve or side-step any issues with poor chromatography separation or mass spectrometry (MS) fragmentation identification, which is common with highly complex or low concentration samples [11, 13]. Therefore, the characterisation and quantification of glycan biomarkers that are associated with different diseases will be greatly helped by the discovery of glycoenzymes of different specificities.

*N*-glycans are common post-translational modifications and their composition varies between type of organism, tissue, age, and disease. One sugar decoration of interest is the bisecting *N*-acetylglucosamine (GlcNAc) on *N*-glycans, which is linked to the mannose core through a  $\beta$ 1,4-linkage. Changes in the abundance of this sugar decoration is associated with several immunological diseases and cancers, for example, hepatitis B-induced liver cirrhosis [14], rheumatoid arthritis [15], type 2 diabetes [16], colorectal cancer [13], ovarian cancer [17], and prostate cancer [9]. Aside from particular diseases, the bisecting GlcNAc is also associated with an individual's metabolic age, which is indicative of overall health [19]. This sugar decoration is usually present in low abundance of approximately 6% in the human plasma *N*-glycome and currently is identified through MS/MS fragmentation. Having additional tools such as exoglycosidase that specifically cleaves these linkages will greatly improve the characterisation of this glycosylation as a potential disease biomarker [18].

Here we report the crystal structure of a glycoside hydrolase (GH) from family 20, BT 0456, originating from a human gut microbe, *Bacteroides thetaiotaomicron*. These are typically exo-acting  $\beta$ -GlcNAc'ases and BT 0456 has previously been characterised as removing the bisecting GlcNAc when all the galactose and sialic acid residues have been removed [1]. It is also capable of removing one of the antennary GlcNAc sugars from *N*-glycans, but this is not always to completion which makes this specificity inconclusive. Previous data also suggests that the  $\alpha$ 1,6-fucose commonly found on mammalian complex *N*-glycans does not prevent its activity [1].

## **Materials and Methods**

### **Cloning, expression and purification of recombinant proteins.**

DNA encoding the appropriate genes (excluding the signal sequences) was amplified from genomic DNA using appropriate primers and cloned into pET28b (Novagen) using NheI-XhoI restriction sites. These plasmids accommodated a HIS-tags which was encoded into the N-terminus of the protein. The recombinant plasmids were transformed into TUNER (Novagen) cells by heat shock treatment and then selected by culturing in 1mL of LB broth containing 10µg/mL kanamycin at 32 °C whilst shaking at 180 rpm. These cultures were then scaled up to 1 L cultures which were grown to mid-exponential phase in 2 L baffled flasks. Isopropyl B-D-thiogalactopyranoside (IPTG) was also added to these cultures at a final concentration of 0.2mM. These cells were then incubated for 16 hours at 16 °C on a rotary shaker table at 150 rpm. The cells were collected by ultracentrifugation and the cell pellets lysed by sonication on an ice bath. The recombinant His-tagged protein was purified from the cell free extracts using immobilized metal-affinity chromatography (Talon resin, Clontech) as previously described [1]. Finally, the purity and size of the proteins were analysed using SDS-PAGE and their concentration were determined by absorbance at 280nm (NanoDrop 2000c, Thermo Scientific) and their molar extinction coefficients.

### **Purification of proteins for crystallization.**

The metal-affinity purified proteins were purified by size-exclusion chromatography using a HiLoad Superdex 200 pg on an AKTA Pure FPLC system (Cytiva). The purity of the fractions was determined using SDS-PAGE and those fraction of satisfactory high purity were pooled and concentrated to 10mg/mL.

### **Crystallization.**

BT 0456 was initially screened using commercial kits (Molecular Dimensions and Hampton Research). The drops, composed of 0.1 µL or 0.2 µL of protein solution plus 0.1 µL of reservoir solution, were set up a Mosquito crystallization robot (SPT Labtech). The sitting drop method was used and the plates were incubated at 20 °C. The crystallisation condition was in 100 mM Hepes pH 7.5, 10 % PEG 8000 and 5 % ethylene glycol. Sample were cryo-protected with the addition of 20 % PEG 400 to the condition and flash-cooled in liquid nitrogen.

### **Data collection, structure solution, model building, refinement and validation.**

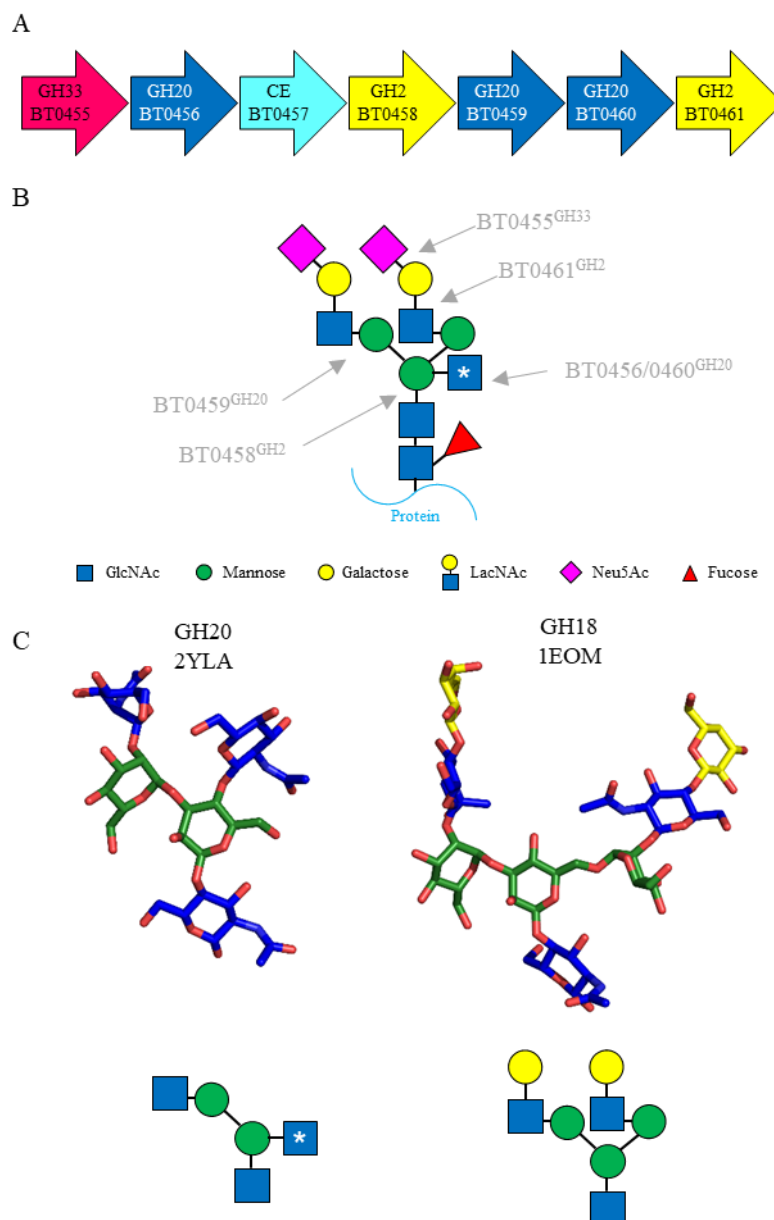
Diffraction data were collected at the synchrotron beamline I03 of Diamond light source (Didcot, UK) at a temperature of 100 K. The data set was integrated with XDS [20] and XIA2 [21]. The space group was confirmed with Pointless [22] and the data were scaled with Aimless [23]. The phase problem was solved by experimental phasing selenium SAD using ShelX [24]. The initial model was generated with the CCP4cloud [25] task CCP4build [25, 26]. The model was refined with refmac [27] and manual model building with COOT [28]. The final model was

validated with MolProbity [29] and figures were made with PyMol [30]. Validation was done with Coot and Molprobity. Other program used were from the CCP4 suite [25].

### **Bioinformatics.**

The CAZy database ([www.cazy.org](http://www.cazy.org)) was used as the main reference for carbohydrate-active enzyme activity. Dali [31] and PDBefold [32] were used to carry out structural homology searches of protein modules.

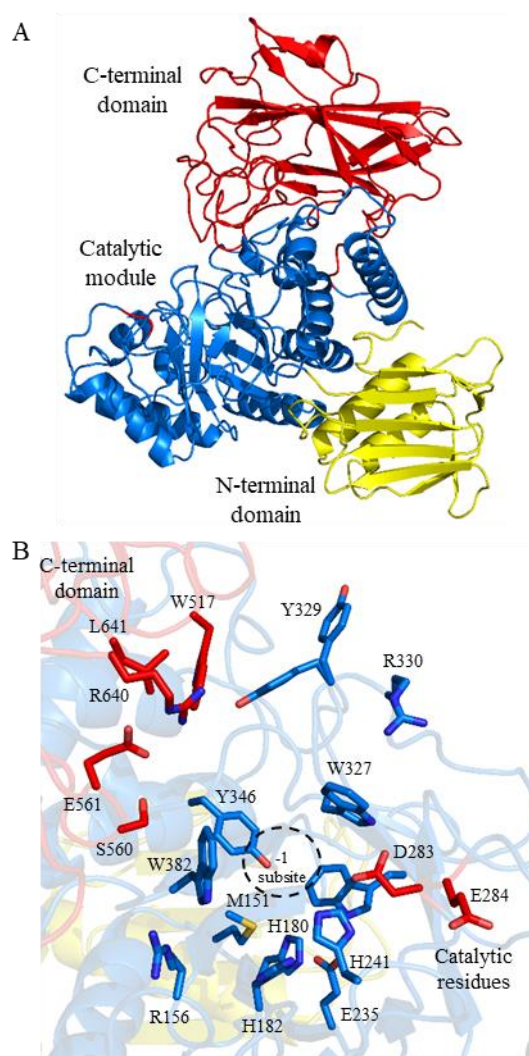
## Results and discussion



**Figure 1. Degradation of mammalian complex *N*-glycans by *Bacteroides thetaiotaomicron*.**

**(A)** A schematic of the CAZyme loci containing BT 0456. CE: carbohydrate esterase. **(B)** The core *N*-glycan pentasaccharide consists of two GlcNAc and three mannose sugars. In the case of mammalian complex *N*-glycans, this pentasaccharide is built on with *N*-acetyl-D-lactosamine (LacNAc) disaccharides, which are themselves commonly decorated with sialic acids. A biantennary complex *N*-glycan is shown here, but up to four antennae are possible. Core  $\alpha$ 1,6-fucose decoration is also common for mammalian complex *N*-glycans. The bisecting GlcNAc is always linked through a  $\beta$ 1,4-linkage (asterisk). The linkages where the CAZymes act are indicated. **(C)** The structures of two *N*-glycan fragments taken from two crystal structures to provide an indication of the steric hindrance different enzymes may face when degrading complex *N*-glycans.

$\beta$ -GlcNAc'ases are present in glycoside hydrolase (GH) families of 3, 20, 73, 84 and 85 and BT 0456 belongs to GH20 (Figure 1) [33]. To explore the specificity shown by BT 0456, its structure was solved to 2.10 Å (Figure 2). Due to the lack of adequate homologous search models (highest identify of primary sequence was 26.3%), molecular replacement strategies were not used to solve for the phase problem. Therefore, selenomethionine derived BT 4056 protein samples were used for experimental phasing with the single anomalous diffraction (SAD) method using selenium atoms. Furthermore, attempts at co-crystallising BT 0456 with substrates was unsuccessful.



**Figure 2. Crystal structure of BT0456. (A)** The overall structure of BT0456 shown as a cartoon. The N-terminal domain, catalytic module, and the C-terminal domain are yellow, blue, and red, respectively. **(B)** The active site of BT 0456. The residues forming the glycan binding site are shown as sticks, the -1 subsite is encircled, and the catalytic residues are shown in red.

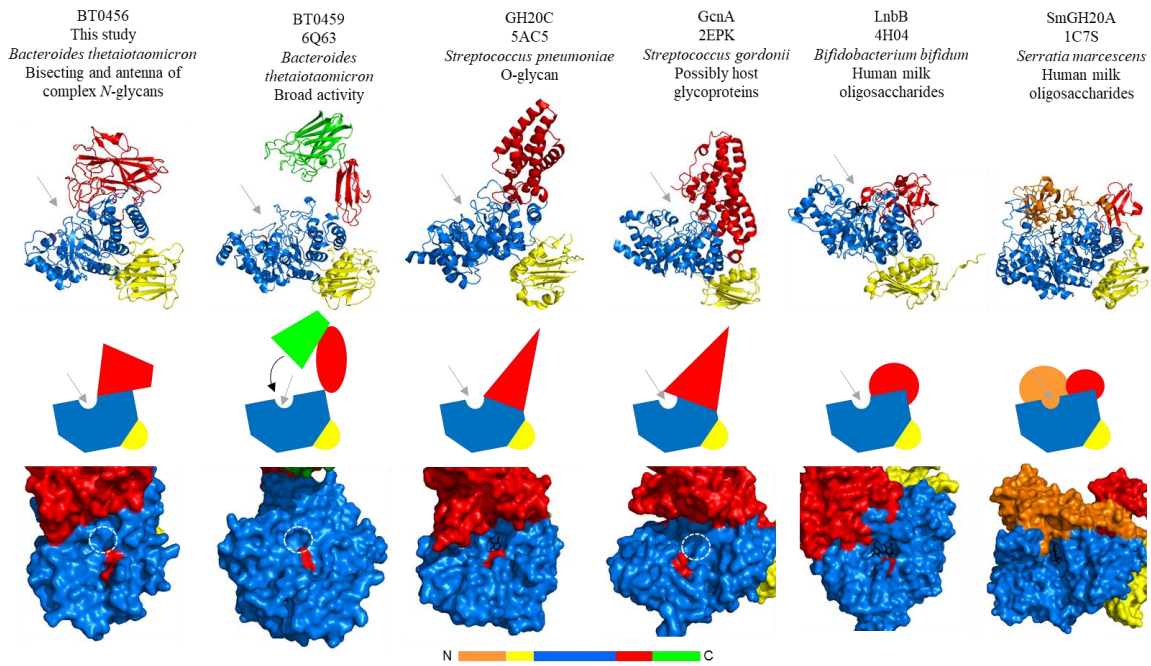


BT 0456 consist of three domains: a N-terminal domain consisting of a zincin-like fold, the catalytic domain is a ( $\beta/\alpha$ )<sub>8</sub> barrel fold typical of this family, and a C-terminal domain (Figure 2a). Crystal structures of other GH20 family members show that this N-terminal domain is a common feature, but additional accessory domains are less frequent amongst currently available crystal structures (Figure 3).

GH20 Family members typically have a HXGG(DE) catalytic motif, but in the case of BT 0456 this is HXGT(DE) [34, 35]. Interestingly, the general acid/base E284 is facing away from the active site (Figure 2B). This suggests flexibility in the active site and an element of induced fit upon substrate binding. Interestingly, the equivalent Glu from a *Streptococcus gordonii* GH20 crystal structure is also flipped away from the active site [34].

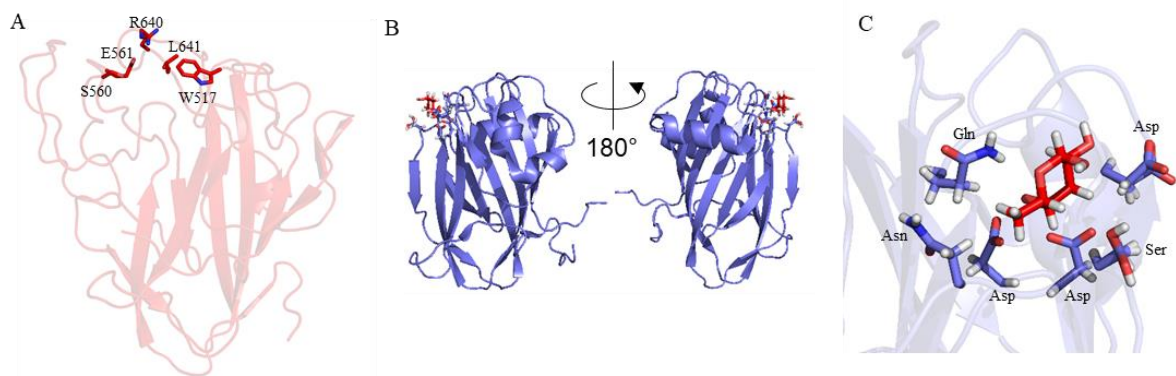
The residues surrounding the -1 subsite are largely the same when compared to other GH20 family members. The structure of BT 0459 from the same locus revealed a very open active site, which reflected its broad activity against GlcNAc substrates [1]. In comparison to this, BT 0456 has Y329 and R330 from the catalytic module contributing to the binding region in addition to residues provided by the C-terminal domain. These additional residues are likely driving the specificity of BT 0456 (Figure 2B).

There are two GH20 family members that have been previously characterised as having activity against complex *N*-glycans - GH20B [1, 34] and C [36] from *Streptococcus pneumoniae*. Both enzymes have specificity towards the antennary GlcNAcs and do not remove the bisecting GlcNAc. However GH20B can accommodate the bisecting GlcNAc to remove the antennary whereas this is not possible for GH20C. The specificity of these enzymes is conferred by residues from the catalytic module, whereas specificity drivers for BT 0456 activity likely come from both the catalytic domain and the C-terminal accessory module. The residues driving specificity in GH20B and C are not present in BT 0456. A substrate-enzyme complex would be required to outline the precise interactions used to drive specificity shown by BT 0456, but this was not successful under the conditions we explored.



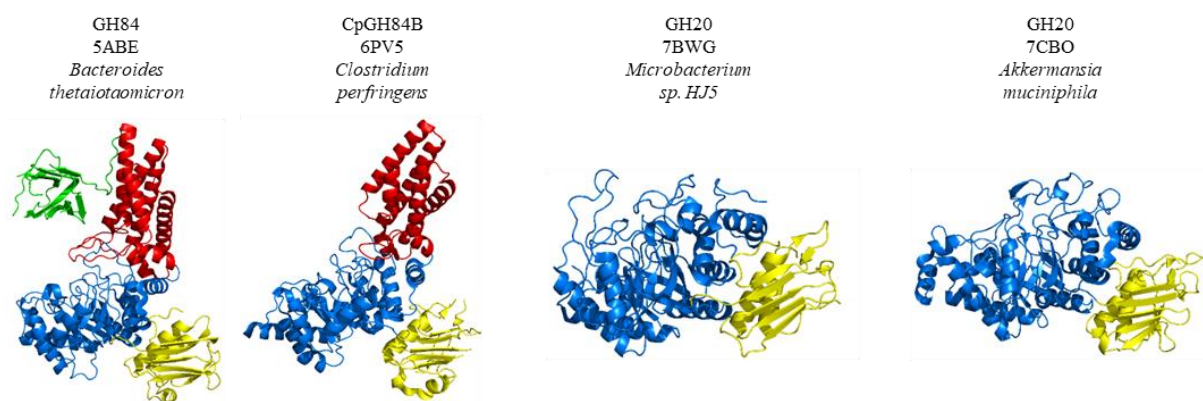
**Figure 3. Comparison of the crystal structure of BT 0456 with other GH20 family members that also have accessory modules.** The catalytic modules are in blue. The N-terminal domains are coloured yellow and a second N-terminal domain is coloured orange in the case of 1C7S. C-terminal domains are coloured red except for GH20 BT 0459, where the FN3 linker domain is coloured red and the C-terminal F5/F8 Type C domain is in green. Any bound glycans are shown as black sticks. A cartoon structure, a schematic, and a surface representation of each enzyme is provided. Grey arrows indicate where the active site is and in the case of the surface representations a white circle is the location of the -1 subsite where ligand is not present. The close association of the accessory domains with the active site make it likely that these contribute significantly to specificity. In the case of BT 0459, it is speculated that the green accessory module comes down over the active site to provide specificity or enhance binding (black arrow), which may be akin to what is seen for the SmGH20A (orange accessory domain).

One of the most striking aspects of this crystal structure is the positioning of the different domains relative to each other (Figure 2A). The C-terminal domain sits on the ( $\beta/\alpha$ )<sub>8</sub> barrel near the active site, so it is likely that the C-terminal domain contributes to the specificity of BT 0456. The residues in close proximity to this region includes S560, E561, R640, L641, and W517 (Figure 2B). The crystal structures of five other GH20 family members with accessory domains indicate that these accessory domains may also contribute to specificity based on their positioning (Figure 3).



**Figure 4. Analysis of the BT 0456 C-terminal domain.** (A) The C-terminal domain of BT 0456 shown in more detail and the residues with likely association to the active site are shown as sticks. (B) The PA14 carbohydrate-binding domain from *Marinomonas primoryensis* (MpPA14) from both sides. (C) The details of the fucose binding site of MpPA14 and the residues comprising this active site.

The sequence of the BT0456 C-terminal domain did not produce any hits in either Pfam or SMART databases. When the structure of the C-terminal domain was used to search Dali database, a PA14 carbohydrate-binding domain from *Marinomonas primoryensis* with fucose bound was highlighted (MpPA14). A comparison of these two structures revealed that the binding site of fucose in MpPA14 was positioned in the same area as the residues that are highlighted being in close proximity to the active site for the C-terminal domain of BT 0456 (Figure 4).



**Figure 5. N-terminal domain structural analysis.** The full crystal structures of the enzymes where the N-terminal domains (yellow) have close structural homology to the N-terminal domain from BT 0456. Catalytic domains are shown in blue and other accessory domains are in red and green where appropriate.

The N-terminal domain of BT0456 consist of a zincin-like fold. Structural matches of this domain were also found through Dali and include GlcNAcases from GH20 and GH84 families

(Figure 5). These N-terminal domains all show similar positioning relative to the catalytic module as that seen in BT 0456. The function of this domain is unknown but could be related to the structural stability of the overall protein.

### **Acknowledgements**

This research was supported by the European Union (GlySign, Grant No. 722095) and the NWO (Vernieuwingsimpuls Veni Project No. 722.016.008). Additionally, I also acknowledge Diamond Light Source (Oxfordshire, UK) for beamtime (proposal mx18598) and staff of beamline I03.

## References

1. Briliūtė, J., et al., Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology*, 2019. 4(9): p. 1571-1581.
2. Nakano, M., et al., Bisecting GlcNAc Is a General Suppressor of Terminal Modification of N-glycan\*, [S]. *Molecular & Cellular Proteomics*, 2019. 18(10): p. 2044-2057.
3. Reily, C., et al., Glycosylation in health and disease. *Nature Reviews Nephrology*, 2019. 15(6): p. 346-366.
4. Saldova, R., et al., Association of Medication with the Human Plasma N-Glycome. *Journal of Proteome Research*, 2012. 11(3): p. 1821-1831.
5. Everest-Dass, A.V., et al., Human disease glycomics: technology advances enabling protein glycosylation analysis – part 2. *Expert Review of Proteomics*, 2018. 15(4): p. 341-352.
6. Pučić, M., et al., High Throughput Isolation and Glycosylation Analysis of IgG Variability and Heritability of the IgG Glycome in Three Isolated Human Populations. *Molecular & Cellular Proteomics*, 2011. 10(10).
7. Vreeker, G.C.M., et al., Automated Plasma Glycomics with Linkage-Specific Sialic Acid Esterification and Ultrahigh Resolution MS. *Analytical Chemistry*, 2018. 90(20): p. 11955-11961.
8. Vanderschaeghe, D., et al., High-Throughput Profiling of the Serum N-Glycome on Capillary Electrophoresis Microfluidics Systems: Toward Clinical Implementation of GlycoHepatoTest. *Analytical Chemistry*, 2010. 82(17): p. 7408-7415.
9. Saldova, R., et al., Core fucosylation and  $\alpha$ 2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology*, 2011. 21(2): p. 195-205.
10. Benicky, J., et al., Quantification of Fucosylated Hemopexin and Complement Factor H in Plasma of Patients with Liver Disease. *Analytical Chemistry*, 2014. 86(21): p. 10716-10723.
11. Rebello, O.D., et al., A Matrix-Assisted Laser Desorption/Ionization—Mass Spectrometry Assay for the Relative Quantitation of Antennary Fucosylated N-Glycans in Human Plasma. *Frontiers in Chemistry*, 2020. 8: p. 138.
12. Royle, L., et al., HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Analytical Biochemistry*, 2008. 376(1): p. 1-12.
13. Doherty, M., et al., Plasma N-glycans in colorectal cancer risk. *Scientific Reports*, 2018. 8(1): p. 8655.
14. Liu, X.-E., et al., N-glycomic changes in hepatocellular carcinoma patients with liver cirrhosis induced by hepatitis B virus. *Hepatology*, 2007. 46(5): p. 1426-1435.
15. Magorivska, I., et al., Glycosylation of random IgG distinguishes seropositive and seronegative rheumatoid arthritis. *Autoimmunity*, 2018. 51(3): p. 111-117.
16. Wu, Z., et al., IgG Glycosylation Profile and the Glycan Score Are Associated with Type 2 Diabetes in Independent Chinese Populations: A Case-Control Study. *Journal of Diabetes Research*, 2020. 2020: p. 5041346.

17. Hecht, E.S., et al., Relative Quantification and Higher-Order Modeling of the Plasma Glycan Cancer Burden Ratio in Ovarian Cancer Case-Control Samples. *Journal of Proteome Research*, 2015. 14(10): p. 4394-4401.
18. 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. *Molecular & Cellular Proteomics*, 2017. 16(2): p. 228-242.
19. Ruhaak, L.R., et al., Decreased Levels of Bisecting GlcNAc Glycoforms of IgG Are Associated with Human Longevity. *PLOS ONE*, 2010. 5(9): p. e12566.
20. Kabsch, W., XDS. *Acta Crystallographica Section D*, 2010. 66(2): p. 125-132.
21. Winter, G., xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 2010. 43(1): p. 186-190.
22. Evans, P., Scaling and assessment of data quality. *Acta Crystallographica Section D*, 2006. 62(1): p. 72-82.
23. Evans, P.R. and G.N. Murshudov, How good are my data and what is the resolution? *Acta Crystallographica Section D*, 2013. 69(7): p. 1204-1214.
24. Sheldrick, G., A short history of SHELX. *Acta Crystallographica Section A*, 2008. 64(1): p. 112-122.
25. Kovalevskiy, O., A. Lebedev, and E. Krissinel, Helping researchers to solve their structures: automation and user guidance in CCP4 Cloud. *Acta Crystallographica Section A*, 2021. 77(a2): p. C766.
26. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, 1994. 50(Pt 5): p. 760-3.
27. Murshudov, G.N., et al., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*, 2011. 67(4): p. 355-367.
28. Emsley, P. and K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D*, 2004. 60(12 Part 1): p. 2126-2132.
29. Chen, V.B., et al., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D*, 2010. 66(1): p. 12-21.
30. Schrödinger, L., The PyMOL molecular graphics system, version 1.8. 2015.
31. Holm, L., et al., Using Dali for Structural Comparison of Proteins. *Current Protocols in Bioinformatics*, 2006. 14(1): p. 5.5.1-5.5.24.
32. Krissinel, E. and K. Henrick. Multiple Alignment of Protein Structures in Three Dimensions. in *Computational Life Sciences*. 2005. Berlin, Heidelberg: Springer Berlin Heidelberg.
33. Val-Cid, C., et al., Structural-Functional Analysis Reveals a Specific Domain Organization in Family GH20 Hexosaminidases. *PLOS ONE*, 2015. 10(5): p. e0128075.
34. Pluinage, B., et al., Inhibition of the pneumococcal virulence factor StrH and molecular insights into *N*-glycan recognition and hydrolysis. *Structure*, 2011. 19(11): p. 1603-14.
35. Prag, G., et al., Structures of chitobiase mutants complexed with the substrate Di-*N*-acetyl-*D*-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. *J Mol Biol*, 2000. 300(3): p. 611-7.

36. Robb, M., et al., A Second  $\beta$ -Hexosaminidase Encoded in the *Streptococcus pneumoniae* Genome Provides an Expanded Biochemical Ability to Degrade Host Glycans\*. *Journal of Biological Chemistry*, 2015. 290(52): p. 30888-30900.







# Chapter 4

**A matrix assisted laser desorption/ionization  
mass spectrometry assay for the relative  
quantitation of antennary fucosylated *N*-glycans  
in human plasma**

## Abstract

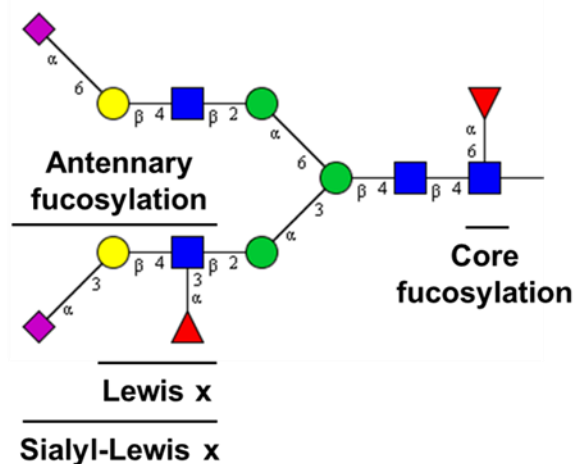
Changes in the abundance of antennary fucosylated glycans in human total plasma *N*-glycome (TPNG) have been associated with several diseases ranging from diabetes to various forms of cancer. However, it is challenging to address this important part of the human glycome. Most commonly, time-consuming chromatographic separations are performed to differentially quantify core and antenna fucosylation. Obtaining sufficient resolution for larger, more complex glycans can be challenging. We introduce a matrix-assisted laser desorption/ionization – mass spectrometry (MALDI-MS) assay for the relative quantitation of antennary fucosylation in TPNG. *N*-linked glycans are released from plasma by PNGase F and further treated with a core fucosidase before performing a linkage-informative sialic acid derivatization. The core fucosylated glycans are thus depleted while the remaining antennary fucosylated glycans are quantitated. Simultaneous quantitation of  $\alpha$ 2,3-linked sialic acids and antennary fucosylation allows an estimation of the sialyl-Lewis X motif. The approach is feasible using either ultrahigh-resolution Fourier-transform ion cyclotron resonance mass spectrometry or time-of-flight mass spectrometry. The assay was used to investigate changes of antennary fucosylation as clinically relevant marker in 14 colorectal cancer patients. In accordance with a previous report, we found elevated levels of antennary fucosylation pre-surgery which decreased after tumor resection. The assay has the potential for revealing antennary fucosylation signatures in various conditions including diabetes and different types of cancer.

## Introduction

Changes in the relative abundance of either core or antennary fucosylation have been associated with certain diseases or disease states [1, 2]. Here, we focus on antennary fucosylation in human total plasma *N*-glycome (TPNG) as a clinically relevant disease marker. For instance, the abundance of antennary fucosylation in TPNG has been correlated with 1) certain cancers such as hepatocellular carcinoma [3, 4], ovarian cancer [5] and colorectal cancer [6-8]; 2) hepatocyte nuclear factor 1 homeobox A - maturity-onset diabetes of the young (HNF1A - MODY) [9, 10]; 3) inflammatory conditions [11, 12] and 4) with attention-deficit hyperactivity disorder in children [13]. These correlations have created a need for the development of assays for quantitation of antennary fucosylation in TPNG. Quantitation of changes in core fucosylated glycans will remain relevant as well, but can be better covered by existing approaches.

The quantitation of antennary fucosylation in TPNG was also used for more specialized clinical purposes such as for prognosis or for differentiating between closely related diseases. For example, a lowered incidence of antennary fucosylation on triantennary glycans was shown to discriminate HNF1A-MODY patients from Type 1 and Type 2 diabetes [10]. Current genetic tests for diagnosing HNF1A-MODY are sometimes inconclusive, and there is a demand for additional diagnostic markers [14-16]. Plasma *N*-glycome antennary fucosylation has been reported to be regulated by HNF1A making it a proxy of HNF1A expression levels and functions [17]. Hence, quantitation of antennary fucosylation in TPNG has been found to be an HNF1A-MODY disease biomarker with potential for complementing genetic tests in disease diagnosis [10, 17]. For colorectal cancer, both better diagnosis and prediction of long-term survival are urgent clinical needs. Currently, long-term survival predictions are mostly based on tumor node metastasis classification which has low success rates [8, 18]. This negatively affects the decision making on therapy given to the patients. Antennary fucosylation was shown to be associated with the recovery of colorectal cancer patients after tumor resection [8], and thus may have the potential for long-term survival prediction.

TPNG is a rich and convenient source of valuable associations with diseases and disease states. It reflects the loss of systemic or cellular homeostasis which may affect the regulation of glycosylation pathways [1, 17, 19]. Enabling a bird's-eye view, TPNG analysis is highly complementary to the target analysis of specific proteins. Quantitation of low abundant antennary fucosylated glycans in human TPNG is complicated by the structural diversity of its component glycans. This diversity is a combination of monosaccharide composition variants and linkage isomers [20, 21]. Composed of diantennary, triantennary and tetraantennary *N*-glycans, TPNG also features bisecting *N*-acetylglucosamine and glycans with *N*-acetylglucosamine repeats, and complexity is added by different levels of galactosylation, sialylation and fucosylation [21-23]. Fucosylation can be classified as either core or antennary (Figure 1). Core fucosylation is linked by an  $\alpha(1-6)$  glycosidic bond to the reducing end *N*-acetylglucosamine. Antennary fucosylation in TPNG mainly features fucose residues that are linked by an  $\alpha(1-3)$  glycosidic bond (Lewis X epitope) to antennary *N*-acetylglucosamine residues [24].



**Figure 1. Fucose epitopes on N-glycans in human plasma.** [Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: N-acetylneuraminic acid (+45°  $\alpha$ 2,6-linked; -45°  $\alpha$ 2,3-linked). For compositions see Figure 6].

Methods for sample preparation and analysis of human TPNG have been developed on a number of analytical platforms. This includes liquid chromatography (LC) – fluorescence detection (FLD) [6, 13, 20, 25], MALDI-MS [22, 26, 27], and capillary gel electrophoresis – laser-induced fluorescence [28, 29]. Each method has its advantages and disadvantages when it comes to measuring antennary fucosylation. Capillary gel electrophoresis – laser-induced fluorescence [28] and LC-FLD [6, 13] are able to resolve some antennary fucosylated glycans in TPNG along their chromatographic dimension. However, LC-FLD often requires extensive measurement time. Nonetheless, due to its high precision and robustness, one of the most attractive techniques for routine applications is LC-FLD. MALDI-MS is gaining popularity in biomarker analysis since it features short measurement time and high molecular resolution. This, combined with linkage-informative sialic acid derivatization, is the ideal basis for a high throughput TPNG glycomics assay in a research setting [22, 26, 27]. However, due to identical mass, the differentiation between core fucosylation and antennary fucosylation in monofucosylated glycans of the same composition is not achieved by MALDI-MS. Therefore, additional experiments are needed such as tandem mass spectrometry [30-32] or exoglycosidase treatment [20]. Unfortunately, relative quantitation of fucose isomers by tandem MS is hindered by the vastly different efficiencies of the fragmentation pathways [32, 33]. Additionally, potential fucose rearrangement influences the ratios of diagnostic fragment ion(s), further interfering with the assessment of mixtures by tandem MS [30, 31, 34]. Alternatively, endoglycosidases can be used for glycomics assays [4, 35]. For example, a combination of Endo F2 and Endo F3 has been used to quantify antennary fucosylation on diantennary and triantennary glycans on hemopexin and complement factor H in patients suffering from liver diseases [4]. However, until now, this approach has not been demonstrated on TPNG. Due to the narrow specificity of these endoglycosidases in contrast to the vast structural diversity of TPNG, such an approach may not be suitable for TPNG.

We developed an assay for the relative quantitation of antennary fucosylation in human TPNG by combining MALDI-MS and exoglycosidase approaches. It can be viewed as complementing the existing approach for TPNG measurement by MALDI-MS. The assay was applied to 14 colorectal cancer patient samples. Consistent antennary fucosylation changes pre vs. post tumor resection were detected which revealed the assays potential for addressing biomedical research questions.

## Materials and Methods

### Reagents and samples

Disodium hydrogen phosphate dihydrate, potassium dihydrogen phosphate, sodium chloride, 85% phosphoric acid, 30%-33% (v/v%) ammonium hydroxide, Nonidet P-40 substitute (NP-40), 1-hydroxybenzotriazole 97% (HOBT), ammonium acetate and super-DHB (9:1 mixture of 2,5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid) were purchased from Sigma Aldrich Chemie GmbH (Steinheim, Germany). 1-Ethyl-3-(3-(dimethylamino)-propyl)carbodiimide hydrochloride (EDC) was purchased from Fluorochem (Hadfield, UK). Analytical grade ethanol, analytical grade glacial acetic acid, sodium dodecyl sulfate (SDS), trifluoroacetic acid, and potassium hydroxide were purchased from Merck KGaA (Darmstadt, Germany). HPLC-grade acetonitrile was purchased from Biosolve (Valkenswaard, The Netherlands). Girard's Reagent P (GiRP) was purchased from TCI Development Co. Ltd. (Tokyo, Japan). 5x phosphate buffered saline solution (PBS; 175 mM; pH 7.3) was prepared by dissolving 285 g of disodium hydrogen phosphate dihydrate, 23.8 g of potassium dihydrogen phosphate and 425 g of sodium chloride in 10 L deionized water. The 5x acidic PBS was prepared by adding 68  $\mu$ L of 85% phosphoric acid (14.7M) to 9.93 mL of the 5x PBS. Recombinant peptide N-glycosidase F (PNGase F) was obtained from Roche Diagnostics (Mannheim, Germany). The recombinant core fucosidase, commercially known as  $\alpha$ 1-2,4,6 Fucosidase O, was purchased from New England BioLabs (MA, USA). However, activity on the  $\alpha$ 1,4-linkage is reported as very low. Under the employed conditions, the enzyme did not noticeably act on antennary fucoses present in TPNG (see section Result and Discussion). Peptide Calibration Standard II was purchased from Bruker Daltonic (Bremen, Germany).

Human plasma standard (Visucon-F frozen normal control plasma, pooled from a minimum of 20 human donors, citrated and buffered in 0.02M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) was purchased from Affinity Biologicals (Ancaster, Ontario, Canada).

The pre-operative vs. 45 days post-operative pairs of 14 colorectal cancer patient samples were collected as part of a biobank as was previously described [8]. These serum samples were collected between October 2002 and March 2013 by the Leiden University Medical Center Surgical Oncology Biobank. This study was approved by the Medical Ethics Committee of the Leiden University Medical Center and was performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (<http://www.federa.org/>).

### **N-glycan release**

PNGase F release of human TPNG was performed similarly as previously described [22]. Briefly, 4  $\mu\text{L}$  of plasma was added to 8  $\mu\text{L}$  of 2% SDS in a polypropylene 96 well V-bottom plate (V-96 microwell, NUNC, Roskilde, Denmark). The plate was sealed (adhesive plate seals, Thermo Scientific, UK) and mildly shaken on a plate shaker for 5 minutes, before incubating at 60°C for 10 minutes. Additionally, 8  $\mu\text{L}$  of the PNGase F releasing mixture (4  $\mu\text{L}$  of 4% NP-40 solution, 4  $\mu\text{L}$  of 5x acidic PBS and 0.4  $\mu\text{L}$  of PNGase F) was then added to the plasma samples. The plate was sealed and mildly shaken on a plate shaker for 5 minutes before incubating overnight (15 – 18 hours) at 37°C.

### **Depletion of core fucosylation**

The overnight incubated plasma release mixture (5  $\mu\text{L}$ ) was diluted with 45  $\mu\text{L}$  of 1x acidic PBS solution in a 96 well V-bottom plate (V-96 microwell, Greiner Bio-one, Germany). The plate was mildly shaken on a plate shaker for 5 minutes, before transferring 1  $\mu\text{L}$  to 2  $\mu\text{L}$  of core fucosidase mixture (0.65  $\mu\text{L}$  deionized water, 0.75  $\mu\text{L}$  of 4% NP-40 solution, 0.4  $\mu\text{L}$  of 5x acidic PBS and 0.2  $\mu\text{L}$  / 0.4 Units of core fucosidase) in a 96 well V-bottom plate. After sealing of the plate, the samples were incubated overnight (15 – 18 hours) at 37°C in an enclosed, humidified chamber to prevent evaporation.

### **Linkage specific sialic acid derivatization**

Sialic acid derivatization was performed as previously described [23], but with minor alterations. With this approach, the carboxylic acid groups of  $\alpha(2,6)$ -linked sialic acids are ethyl esterified while the  $\alpha(2,3)$ -linked sialic acids are amidated. Briefly, 60  $\mu\text{L}$  of ethyl esterification reagent (solution of 0.25 M EDC and 0.25 M HOBt in ethanol) was added to the core defucosylated samples. The plate was sealed and incubated at 37°C for 30 minutes. 12  $\mu\text{L}$  of 30%-33% (v/v%)  $\text{NH}_4\text{OH}$  solution were added to the wells and the sealed plate incubated for another 30 minutes at 37°C. 72  $\mu\text{L}$  of acetonitrile was added to the wells, after which cotton hydrophilic interaction liquid chromatography (HILIC) - solid-phase extraction (SPE) microtip purification was performed immediately.

### **Cotton hydrophilic interaction liquid chromatography - solid-phase extraction microtip purification**

This purification step was performed as previously described [36] with minor modifications. The cotton HILIC- SPE microtips were prepared by inserting a cotton strand of length 3 - 4 mm into a 20  $\mu\text{L}$  capacity microtip (Mettler-Toledo, Switzerland) and pushing it into place with a stream of pressurized air/nitrogen. Conditioning was performed by pipetting 5 times 20  $\mu\text{L}$  of water, followed by an equilibration step of pipetting 3 times 20  $\mu\text{L}$  of 85% acetonitrile. The derivatized samples were loaded onto the HILIC-SPE by pipetting 20  $\mu\text{L}$  of it 20 times. Washing was performed with 3 times 20  $\mu\text{L}$  of 85% acetonitrile containing 1% trifluoroacetic acid and 3 times 20  $\mu\text{L}$  of 85% acetonitrile, consecutively. The glycans were eluted from the HILIC-SPE

by repeatedly, 10 times, pipetting 4  $\mu\text{L}$  of deionized water in a 96 well V-bottom plate (V-96 microwell, Greiner Bio-one, Germany). All steps were performed with a 12 channel multi-channel pipette.

Purified glycan samples (1 $\mu\text{L}$ ) were spotted on an MTP anchor chip 600/384 TF MALDI target plate (Bruker Daltonics, Bremen, Germany), followed by the addition of 1  $\mu\text{L}$  of the MALDI matrix solution (50% acetonitrile solution of 2.5 mg/mL super-DHB and 0.1 mM sodium hydroxide). The solutions were mixed on the plate with a pipette. The spotted samples were allowed to air dry before performing MALDI- time-of-flight -MS (MALDI-TOF-MS) or MALDI-Fourier-transform ion cyclotron resonance -MS (MALDI-FT-ICR-MS) measurements.

### **MALDI-TOF-MS analysis**

The analysis was performed on an UltrafleXtreme Mass spectrometer in reflectron positive ion mode (Bruker Daltonics, Bremen, Germany) which was operated by FlexControl version 3.4 (Build 135). A Bruker Smartbeam- II laser was used for ionization at an irradiation frequency of 1 kHz using the "small" predefined laser shot pattern. Each sample spot was irradiated by 20,000 shots with 200 shots at each laser raster. Irradiation was performed randomly over the complete sample spot. Spectra were recorded within an  $m/z$  range of 900 – 5000. Samples were measured in an automated manner using the AutoXecute function of FlexControl. Before each measurement, the instrument was calibrated with a peptide standard mix (Peptide Calibration Standard II, Bruker Daltonics).

### **MALDI-FT-ICR-MS analysis**

The analysis was performed in positive ion mode on a Bruker 15T solariX XR FT-ICR mass spectrometer equipped with a CombiSource and a ParaCell (Bruker Daltonics). The system was operated by ftmsControl version 2.2.0 (Build 150). A Bruker Smartbeam- II laser was used for ionization at an irradiation frequency of 500 Hz using the "medium" predefined laser shot pattern. Each sample spot was irradiated with a raster of 200 laser shots. 10 such scans were performed randomly over the complete sample spot. Spectra were acquired within an  $m/z$ -range 1011–5000 with 1 million data points (transient time 2.3069 s). Samples were measured in an automated manner using the AutoXecute function of ftmsControl.

### **Processing of MALDI-MS data**

The MALDI-TOF-MS spectra were internally calibrated using FlexAnalysis version 3.4 (Build 76; Bruker Daltonic). The glycan calibrants used were  $[\text{H3N4} + \text{Na}]^+$  of  $m/z$  1339.476,  $[\text{H5N4} + \text{Na}]^+$  of  $m/z$  1663.581,  $[\text{H5N4E1} + \text{Na}]^+$  of  $m/z$  1982.708,  $[\text{H5N4E2} + \text{Na}]^+$  of  $m/z$  2301.835,  $[\text{H5N5E2} + \text{Na}]^+$  of  $m/z$  2504.914,  $[\text{H6N5E2Am1} + \text{Na}]^+$  of  $m/z$  2957.078 and  $[\text{H7N6E2Am2} + \text{Na}]^+$  of  $m/z$  3612.322. The .xy files generated by Flexanalysis were further processed with MassyTools [37]. A second round of internal spectral calibration was performed with at least 4 calibrants in the low  $m/z$  range and at least 3 calibrants in the medium  $m/z$  range, having to pass signal to noise ratio (S/N) at least above 15. The  $m/z$  window for calibration, peak



detection, and spectral data integration was taken at  $\pm 0.45$  Th. The background detection windows were set to 20 Th.

The MALDI-FT-ICR-MS spectra were acquired in serial mode. Thus, a single combined file was generated. This file was split into individual compound spectra and transformed into .xy files using DataAnalysis version 5 (Bruker Daltonics). Calibration of these spectra was performed in MassyTools, for which at least 4 calibrants in the low  $m/z$  range and at least 3 calibrants in the medium  $m/z$  range, have to pass S/N at least above 50. The  $m/z$  window for calibration was taken at  $\pm 0.10$  Th. Peak detection and spectral data integration was performed using an extraction window depending on the defined glycans in the analyte list. This was done so as to exclude interferences in the spectra. The background detection windows were set at the value of 20 Th.

The calibrant list and analyte list for MALDI-TOF-MS spectra and MALDI-FT-ICR-MS spectra are shown in Supplementary Tables S1 and S2, respectively. The table of all identified and quantitated glycans by MALDI-FT-ICR-MS and MALDI-TOF-MS, respectively, are shown in Supplementary Table S3. At least 85% of the theoretical isotopic distribution of each glycan analyte was integrated. The MassyTools output was used further for analyte curation, which was based on multiple criteria. These were  $S/N \geq 9$ , isotopic pattern quality  $\leq 0.25$  and mass accuracy between  $\pm 20$  ppm (MALDI-TOF-MS) or  $\pm 10$  ppm (MALDI-FT-ICR-MS). Spectra for which the number of glycan analytes passing these criteria were less than 50% of the total number of defined glycans in the analyte list, were not considered for further processing as these spectra were deemed of low quality. The absolute area of the curated glycans was corrected to 100% of their respective isotopic distribution, before using them in total area normalization.

Derived traits, focusing on features rather than individual glycans, were calculated from the relative areas of the glycan compositions using an in-house prepared R script (Supplementary Table S4). Statistical significance of the differences between pre-operative vs. post-operative colorectal cancer patient samples was assessed by a Wilcoxon matched-pairs signed-rank test with  $\alpha=0.05$ . Multiple-testing correction was performed using a false discovery rate of 1% calculated with the Benjamini and Hochberg method. These statistical tests were performed in GraphPad Prism version 8.0.1. The box and whisker plots used for representing the features were made by an in-house prepared R script.

## **Method optimization and assay performance**

All optimization experiments of the assay were measured on the MALDI-TOF-MS (Supplementary Material `Assay optimization`). Data processing for these mentioned optimization experiments was done using the mass list of the negative control (fucosidase untreated TPNG) (Supplementary Table S2). Quantitation of residual core fucosylated glycans allowed the assessment of the completeness of the core defucosylation. Importantly, all identified antennary fucosylated glycans were also included in the mass list of the negative control (fucosidase untreated TPNG).

The intermediate precision of the antennary fucose assay for combined sample preparation and measurements on the MALDI-TOF-MS or MALDI-FT-ICR-MS were performed in three independent experiments on three different days within a period of 5 days. For each of these experiments, the assay was performed with nine replicates of glycan releases from a human plasma pool (Visucon-F, Affinity Biologicals). Data processing was performed as mentioned in the subsection 'Processing of MALDI-TOF-MS/MALDI-FT-ICR-MS data'.

### **Capillary electrophoresis - electrospray ionization mass spectrometry**

Identification of core fucosylated glycans, antennary fucosylated glycans, and composition containing both isomers was done by comparing the MALDI-FT-ICR-MS spectra from the assay (core fucosidase treated) to the negative control (fucosidase untreated). The identified antennary fucosylated glycans and mixed fucosylated isomeric glycans were structurally confirmed by targeted collision induced dissociation (CID) fragmentation on a capillary electrophoresis - electrospray ionization – tandem MS (CE-ESI-MS/MS) platform. The samples were performed in 12 replicates of which nine were used for MALDI-FT-ICR-MS measurements while the remaining three were used for CE-ESI-MS/MS measurements (n=3).

For the CE-ESI-MS/MS measurement, the replicates from the assay and the negative controls were respectively pooled together and dried down in a vacuumed centrifuge (Salm en Kipp, Breukelen, Netherlands) at 50°C. They were used for permanent cationic labeling with GiRP as previously described [23] but with certain alterations. Briefly, 10 µL of GiRP labeling solution (7.5 mg GiRP dissolved in a solution of 720 µL ethanol and 80 µL glacial acetic acid) was added to the dried samples. The plate was sealed and mildly shaken on a plate shaker for 5 minutes before incubating at 60°C for 1 hour. After incubation, the samples were dried down in a vacuumed centrifuge at 50°C and then re-suspended in 5 µL of deionized water. In total, 3.6 µL of the GiRP labeled glycans were mixed with 2.4 µL of 250 mM ammonium acetate solution as leading electrolytes (250 mM ammonium acetate solution adjusted to pH 4 with glacial acetic acid) before being transferred into a vial (nanoVial, Sciex, Framingham, USA).

All CE-ESI-MS/MS analyses were performed on a static coated neutral capillary cartridge (Neutral OptiMS cartridge, Sciex), fitted into a CESI 8000 system (Sciex). The CE system was coupled with an Impact HD UHR-QTOF-MS system (Bruker Daltonics). When the capillary was not in use, a continuous flow of water at 10 psi was applied to the separation line. Prior to usage, the separation line and reverse line were filled with the background electrolyte containing 10% acetic acid (sonicated in a water bath for 10 minutes). Prior to sample injection, the separation line was flushed with 0.1 M HCl at a pressure of 100 psi for 5 minutes. This was followed by flushing the reverse line with background electrolyte at 75 psi for 3 minutes. The separation line was filled with background electrolyte by applying a pressure of 100 psi for 10 minutes. The sample was injected into the separation line of the capillary from the nanovials via a hydrodynamic injection of 12.5 psi for 24 seconds (about 5% of the capillary volume). The tip of the separation line was washed by momentary dipping it into a vial of background electrolyte, followed by injecting a background electrolyte plug with a pressure of 2.5 psi for 15 seconds. A 20 kV voltage of normal polarity (cathode toward the

end of the capillary) was applied on the capillary for 30 minutes. During this step, a continuous flow of 2 psi was applied only on the reverse line. After 30 minutes, a flow of background electrolyte is applied to both the separation line and reverse line by applying a pressure of 2 psi on both lines for 40 minutes. Finally, the voltage on the capillary was ramped down over 5 minutes to 1 kV before termination of the run. The capillary was maintained at 30°C throughout the analysis.

For the MS analysis, a dopant enriched nitrogen gas (acetonitrile as dopant) at 0.2 bar was used for nebulization at the ESI captive sprayer. The drying gas of nitrogen at 150°C was introduced at the source at 1.2 L/minute and the internal capillary of the MS was maintained at 1200 V. A targeted CID fragmentation was performed on 20 glycan analytes of interest. This list was divided into two inclusion lists on the software otof control version 3.4 (Build 14; Bruker Daltonics) which was used for operating the MS and MS/MS analysis. Hence each sample was measured twice in-order to fragment all the glycan analytes of interest. The MS/MS fragmentation spectra were collected at a rate of 1 Hz within an  $m/z$ -range of 150 - 2000 and at an absolute intensity threshold at 2274 on the  $m/z$  values of interest. The targeted precursor ions were isolated with a width of 8 – 15 Th depending on the  $m/z$  values. The collision energies of the CID cell was set in an  $m/z$  dependent manner, ranging from 35 eV for singly charged precursor ions of  $m/z$  500 to 70 eV for singly charged precursor ions at  $m/z$  2000. Data analysis was done using DataAnalysis version 5 (Bruker Daltonics).

### **Hydrophilic interaction liquid chromatography analysis**

PNGase F released *N*-glycans from 12 replicates of human plasma samples (Visucon-F) were used for procainamide labeling and analysis on a HILIC-FLD-MS<sup>n</sup> platform as previously described [38]. Following PNGase F treatment, 8 µL of each sample was dried down and the released *N*-glycans converted to aldoses with 0.1% formic acid, filtered through a protein binding plate (LC-PBM-96, Ludger, Oxford, UK), washed twice with 100 µL of water and dried. *N*-glycans were labeled by reductive amination in 10 µL of water and 10 µL procainamide labeling solution (LT-KPROC-24 containing NaCNBH<sub>3</sub>, Ludger) and incubated at 65°C for 1 h. A HILIC-type clean-up plate (LC-PROC-96, Ludger) was used to remove unreacted procainamide dye. Procainamide labeled *N*-glycans were eluted in 300 µL of water. The samples were dried and resuspended in water (50 µL) for further analysis.

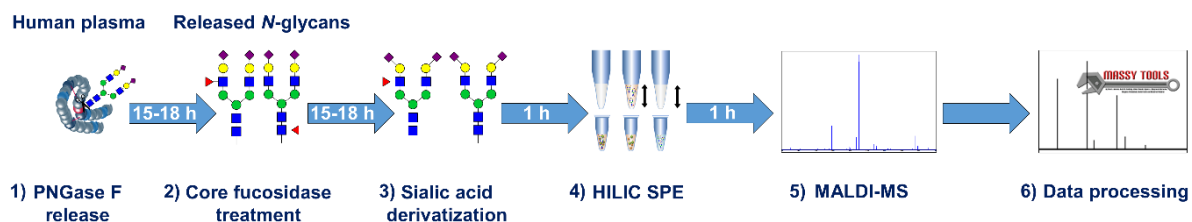
Procainamide-labeled samples were analyzed by HILIC-FLD-MS<sup>n</sup>. 12.5 µL of each sample was injected into an ACQUITY BEH Glycan column (1.7 µm, 2.1x150 mm) at 40°C on a Dionex Ultimate 3000 UHPLC instrument with a fluorescence detector (ex = 310 nm and em = 370 nm attached to an Amazon Speed ETD (Bruker Daltonics). The running conditions used were as follows: solvent A was 50 mM ammonium formate (pH 4.4) (LS-NBUFFX40, Ludger), and solvent B was acetonitrile.

Gradient conditions were as follows: 0 to 53.5 min, 76 to 51% B, 0.4 mL/min; 53.5 to 55.5 min, 51% to 0% B, 0.4 mL/min to 0.2 mL/min; 55.5 to 57.5 min, 0% B at a flow rate of 0.2 mL/min; 57.5 to 59.5 min, 0 to 76% B, 0.2 mL/min; 59.5 to 65.5 min, 76% B, 0.2 mL/min; 65.5 to 66.5 min, 76% B, 0.2 mL/min to 0.4 mL/min; 66.5 to 70.0 min, 76% B, 0.4 mL/min. The Amazon

Speed settings used were as follows: source temperature, 250°C; gas flow, 10 L/min; capillary voltage, 4500 V; ICC target, 200,000; Max. accu. time (Maximum Accumulation Time), 50.00 ms; rolling average, 2; number of precursor ions selected, 3; release after 0.2 min; positive ion mode; scan mode, enhanced resolution; mass range scanned, 400 to 1500; target mass, 900.

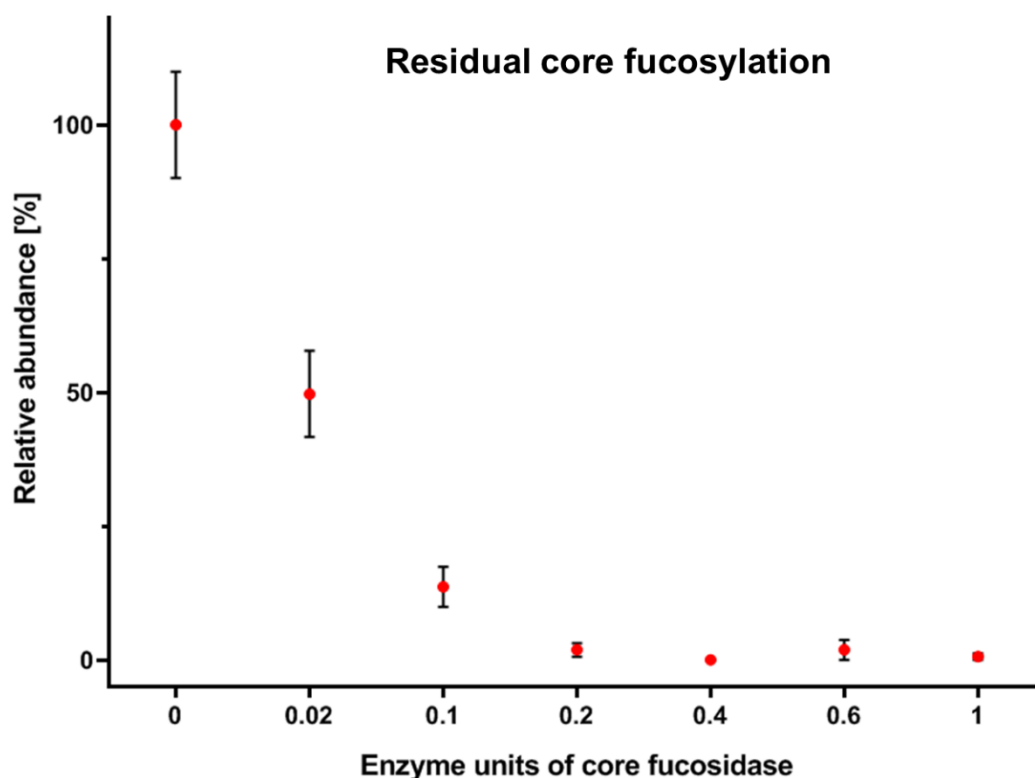
## Results and Discussions

The workflow for the antennary fucose assay is shown in Figure 2. Released *N*-glycans from human plasma were treated with an  $\alpha$ 1,6-linkage-selective fucosidase (core fucosidase) to deplete core fucosylation. Antennary fucosylation remained and was relatively quantified by MALDI-FT-ICR-MS or MALDI-TOF-MS. Information on sialic linkage is obtained in parallel, potentially improving the estimation of sialyl-Lewis X abundance. An overview of all analytical methods used during the development of the assay is shown in Supplementary Figure S1.



**Figure 2. Workflow of the antennary fucose assay for human plasma.**

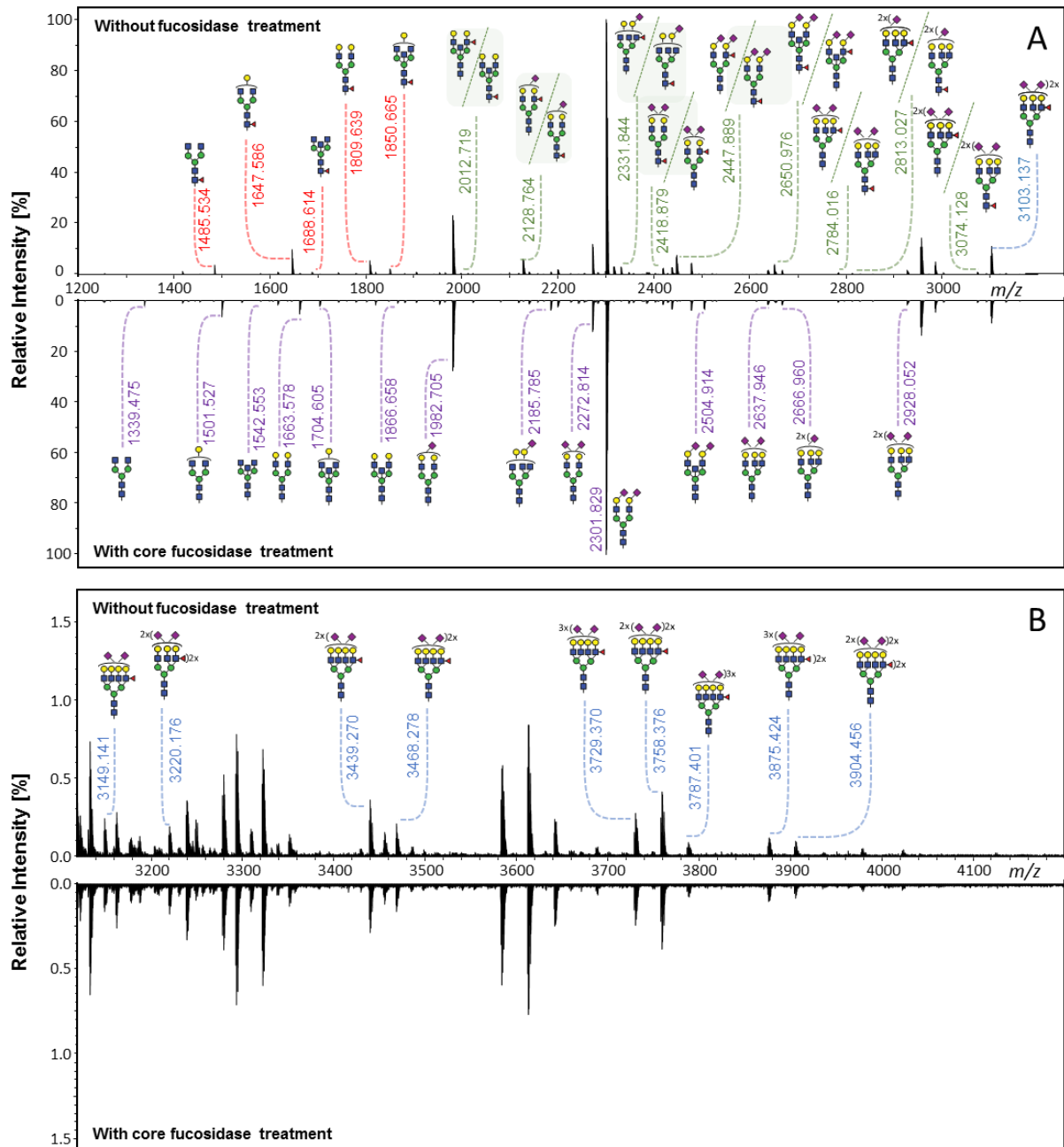
Human TPNG was  $28.3\% \pm 0.5\%$  fucosylated. Most of this is core fucosylation (Supplementary Table S5). Hence, it was important to optimize the assay with a focus on a robust and complete depletion of core fucosylation of TPNG. This aids a robust relative quantitation of the remaining antennary fucosylated glycans by MALDI-MS. The efficient use of costly reagents, specifically the core fucosidase, was also addressed. During the optimization steps, completeness of core defucosylation was judged by the inability to quantify by MALDI-MS (signal to noise ratio (S/N) < 9) some of the most abundant core fucosylated glycans namely  $[\text{H3N4F1} + \text{Na}]^+$  of  $m/z$  1485.53 and  $[\text{H4N4F1} + \text{Na}]^+$  of  $m/z$  1647.59 (Supplementary Figure S4). Details pertaining to assay optimization are described in the Supplementary Material section 'Assay optimization'. For these optimized assay conditions, a complete core defucosylation of TPNG could already be achieved with 0.2 units of core fucosidase (Figure 3). However, to facilitate robustness of the core defucosylation a 2-fold greater amount (0.4 units) was chosen for further experiments. Equally important is the preservation of antennary fucosylation. A fucosidase specific, or at least highly selective, for the  $\alpha$ 1,6-linkage over other linkages significant in TPNG (mainly  $\alpha$ 1,3-linkage) is therefore essential. The MS/MS spectra provided in the Supplementary Material demonstrate that core fucoses are removed while antennary fucoses remain (see also following sub-section) with the chosen enzyme. The integrity of the antennary fucosylation under increased enzyme concentrations (Figure 3) and incubation times (data not shown) further supports the  $\alpha$ 1,6-linkage selectivity. An absence of the oxonium ion of  $m/z$  658.26  $[\text{galactose} - \text{N-acetylglucosamine}(\text{fucose})_2 + \text{H}]^+$  excluded a significant presence of Lewis Y or Lewis B structures.



**Figure 3. Residual core fucosylation on TPNG after incubation with varying amounts of core fucosidase.** The amounts of core fucosylation on complex *N*-glycans were calculated and were normalized to their lowest and highest mean values. Error bars show the standard deviation of the mean of three replicates.

### Identification of antennary fucosylated glycans in total plasma *N*-glycome

Core fucosylated glycans and antennary fucosylated glycans in TPNG were identified by comparing the TPNG profile with and without core fucosidase treatment (Figure 4). A more detailed view can be found in Supplementary Figure S2 and Supplementary Figure S3 for MALDI-FT-ICR-MS and MALDI-TOF-MS spectra, respectively. Interestingly, some monofucosylated glycan compositions in TPNG showed a mixture of both core and antennary fucosylated isomers (Supplementary Figure S4). CID spectra, obtained on a CE-ESI-MS/MS platform, provided an orthogonal layer of evidence (Figure 5). CID spectra for all identified antennary fucosylated glycans are shown in Supplementary Figures S5 to S24. Antennary fucosylation was identified by the formation of the diagnostic B-ion of  $m/z$  512.198 assigned as [galactose-*N*-acetylglucosamine(fucose)+H]<sup>+</sup> [31]. Core fucosylation was identified by the formation of the Y-ion of  $m/z$  501.219 assigned as [*N*-acetylglucosamine(fucose)-GiRP]<sup>+</sup>. Core fucosylation also generates, to a lesser extent, the B-ion of  $m/z$  512.198 assigned as [mannose-*N*-acetylglucosamine(fucose)+H]<sup>+</sup> caused by fucose rearrangement [30, 31, 34, 39]. This complicates the assessment of mixtures, especially relative quantitation by MS/MS. Although, fucose rearrangement also limits the sensitivity of antennary fucose identification in the presence of core fucose isomers, significant contributions of antennary fucose are still readily identified (Figure 5).



**Figure 4. Identification of core fucosylated glycans and antennary fucosylated glycans by exoglycosidase and MALDI-FT-ICR-MS.** The TPNG profile without core fucosidase treatment was compared to the profile after treatment, within **(A)** the  $m/z$  range of 1200 to 3200 and **(B)** the  $m/z$  range of 3120 to 4200. Core fucosylated glycans [red  $m/z$  values] are converted to their corresponding afucosylated glycans [purple  $m/z$  values], upon core fucosidase treatment. Only the antennary fucosylated glycans [blue  $m/z$  values] and the antennary fucose isomers of the mixed fucose isomeric (core or antennary fucosylation) glycans [green  $m/z$  values] remain after core fucosidase treatment. All  $m/z$  values of annotated glycans belong to  $[M + Na]^+$  ions. The description of the glycan cartoons are shown in Figure 1.

Notably, no core fucosylation remained after core fucosidase treatment. This was assessed by the absence of the core fucosylated glycans H3N4F1 and H4N4F1 (S/N < 9) (Supplementary Figure S25). As expected, the corresponding afucosylated glycoforms increased in abundance (Figure 4). The remaining fucosylated glycans are suggested to be solely antennary fucosylated.

As similar relative abundances were maintained after core fucosidase treatment, the species, H6N5F2Am2E1, H6N5F1Am1E2, H7N6F1Am1E1, H7N6F1Am2E1, H7N6F1Am1E2, H7N6F1Am3E1, H7N6F2Am3E1, H7N6F1Am2E2, H7N6F2Am2E2, and H7N6F1Am1E3, are assigned as antennary fucosylated in human TPNG. This was confirmed by MS/MS (Supplementary Figure S14 to S23). Interestingly, we were able to prove that the difucosylated glycans, H6N5F2Am2E1, H7N6F2Am3E1 and H7N6F2Am2E2, contained only antennary fucosylation and no core fucosylation. Previously, these glycans have often been assigned as containing one core and one antennary fucose [22]. In fact, we were unable to identify any multifucosylated glycan compositions in TPNG having both core and antennary fucose residues on the same glycan. However, at low abundances (<0.5%), two such mixed, multifucosylated glycans, H6N5F2Am2E1 and H6N5F2Am1E2, have been convincingly demonstrated [23]. On haptoglobin, multifucosylated glycans having both core and antennary residues were shown to have clinical relevance in subtyping hepatocellular carcinoma patients [3]. Thus, it may be very important to differentiate between glycans with multiple antennary fucoses and glycans containing both a core and an antennary fucose. The former may be directly assessed with our assay, the latter indirectly, either via an increase of the resulting monofucosylated species or by comparison to the negative control (fucosidase untreated).

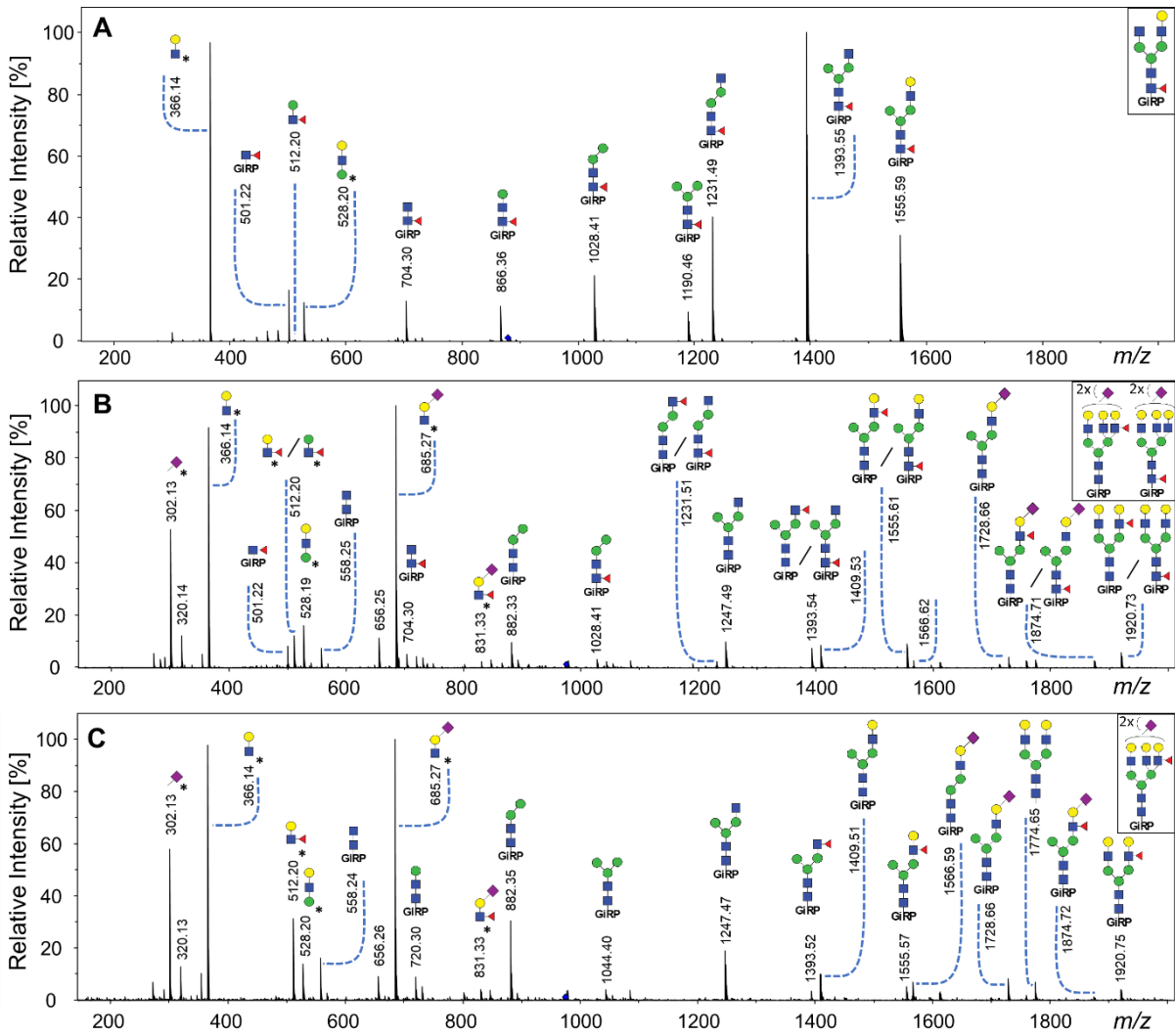
A range of monofucosylated compositions were found to be a mixture of core fucosylated and antennary fucosylated isomers, including for the following glycan compositions: H5N5F1, H5N4F1E1, H5N5F1E1, H5N4F1Am1E1, H5N4F1E2, H5N5F1E2, H6N5F1Am1E1, H6N5F1E2, and H6N5F1Am2E1. Their relative abundances were significantly lowered, but signals did not disappear after core fucosidase treatment (Supplementary Figure S4 and Table S4). These findings were further supported by their MS/MS spectra (Supplementary Figures S5 to S13). As a representative example, the MS/MS fragmentation of the monofucosylated composition H6N5F1E2 is shown without and with core defucosylation in Figures 5B and 5C, respectively. Fragmentation of this glycan from untreated human TPNG resulted in the formation of similar abundances of both a Y-ion of  $m/z$  501.219 and a B-ion of  $m/z$  512.198. This is typical of a mixture of core and antennary fucosylation (Figure 5B). Expectedly, the fragmentation of this glycan after core defucosylation, results in the formation of only the B-ion of  $m/z$  512.198 which is indicative of only antennary fucosylation (Figure 5C). Thus, H6N5F1E2 in human TPNG is comprised of a mixture of core fucose isomers and antennary fucose isomers. Such monofucosylated glycan compositions which include mixtures of core fucose and antennary fucose isomers contribute to a total abundance of 12.1%  $\pm$  1.0% in human TPNG. Our assay determined 3.1%  $\pm$  0.3% of them to be antennary fucosylated (Supplementary Table S5).

A B-ion of  $m/z$  831.325 was observed in several CID spectra which was tentatively assigned as  $[N\text{-acetylneuraminic acid(amidated)-galactose-}N\text{-acetylglucosamine(fucose)+H}]^+$ . This is an



unconventional motif because antenna fucosylation is generally observed on  $\alpha$ 2,3 sialylated antennae. The ion may be caused by fucose rearrangement between the antennae [31]. For the composition H6N5F1E2 (Figure 5C), the higher abundance of the B-ion of  $m/z$  512.198 compared to  $m/z$  831.325, indeed indicates a Lewis X structure to be more likely. Additionally, for the compositions with only  $\alpha$ 2,6 sialylated antennae, H5N4F1E2 and H5N5F1E2, we did not observe a B-ion of  $m/z$  831.325 but rather the Y-ion of  $m/z$  501.219 (Supplementary Figures S9B and S10B). Thus, these compositions are partly explained by incomplete core defucosylation. Their antennary fucosylated portion, suggested by the significant abundance of the Y-ion of  $m/z$  512.198, could be due to side-products of the esterification of  $\alpha$ (2-3) linked sialic acids [40-42].

After core fucosidase treatment, H6N5F2Am2E1 and H7N6F1Am1E2 showed a lower abundance (Supplementary Table S6). However, from the MS/MS spectra without fucosidase treatment, core fucosylation could not be confirmed on these glycans due to the lack of the Y-ion of  $m/z$  501.219 (Supplementary Figures S16A and S18A). H6N5F2Am2E1 and H7N6F2Am2E1 both contribute to < 0.25% to TPNG, and the ratio of their abundances for core fucosidase treated / untreated is 0.83 and 0.88, respectively (Supplementary Table S5). Thus, if their core fucosylated isomers are present in TPNG, they might be too low abundant ( $\leq$  0.03%) to be identified in MS/MS spectra of our CE-ESI-MS/MS platform. Furthermore, a complete depletion of core fucosylation for H5N5F1, H5N4F1E1, H5N5F1E1, H5N4F1E2 and H5N5F1E2 was not achieved as the Y-ion of  $m/z$  501.219 was still observed in their MS/MS spectra, although the B-ion of  $m/z$  512.198 is equally, if not more abundant (Supplementary Figures S5 to S7, S9 and S10). However, most of the core fucosylation was removed (Supplementary Figure S4). Due to the relatively low abundance of the affected glycans, this small overestimation is not likely to have a significant impact on the measurements. A pessimistic estimate is an 0.1% bias in relative quantitation of total antennary fucosylation (<3% for affected glycans, such as H5N4F1E1).



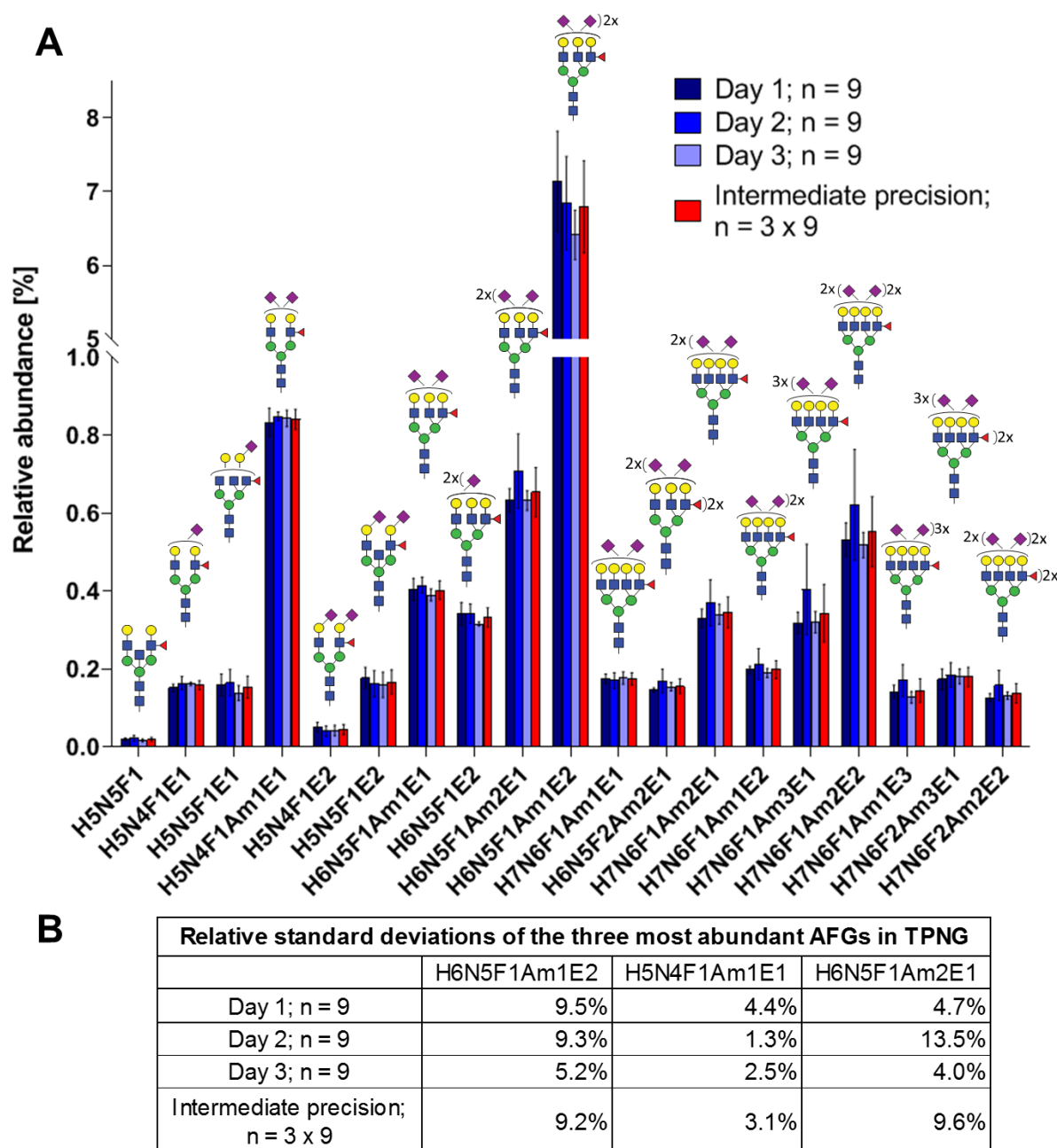
**Figure 5.** CE-ESI-MS/MS spectra of sialic acid derivatized and GiRP labeled TPNG for the confirmation of antennary, mixed isomeric (core or antennary fucosylation) and core fucosylated glycans. **(A)** The  $[M+H]^{2+}$  ion of H4N4F1 ( $m/z$  879.838) is fragmented for fucosidase untreated TPNG. Its fragment ions are indicative of core fucosylation. **(B)** The  $[M+2H]^{3+}$  ion of H6N5F1E2 ( $m/z$  975.709) [second isotope] is fragmented for fucosidase untreated TPNG. Its fragment ions are indicative of both core fucosylated and antennary fucosylated isomers. **(C)** The ion  $[M+2H]^{3+}$  of H6N5F1E2 ( $m/z$  975.709) [second isotope] is again fragmented, but for core fucosidase treated TPNG. Its fragmentation pattern is indicative of only antennary fucosylation. The descriptions of the glycan cartoons are shown in Figure 1. GiRP represents Girard's reagent P label; \* represents an oxonium ion.

## Assay performance

Intermediate precision was assessed by three independent experiments on different days each with nine replicates of glycan releases from a human plasma pool measured by MALDI-FT-ICR-MS. 70 glycans were quantified (Supplementary Table S7, Supplementary Figure S26), which included 19 antennary fucosylated glycans (Figure 6). These antennary fucosylated glycans make up  $11.8\% \pm 0.9\%$  of total abundance in human TPNG, with the three most abundant antennary fucosylated glycans, H6N5F1Am1E2, H5N4F1Am1E1 and H6N5F1Am2E1, contributing  $6.80\% \pm 0.63\%$ ,  $0.84\% \pm 0.03\%$  and  $0.66\% \pm 0.06\%$ , respectively (Figure 6B). The abundance of H6N5F1Am1E2 is consistent with previous quantitation using MALDI-FT-ICR-MS [22]. The median of intermediate precisions for the 19 quantitated antennary fucosylated glycans is 12.4% (9.1% to 18.5% interquartile range; Supplementary Figure S27), which is in-line with the ca. 10% previously described for MALDI-FT-ICR-MS analysis of all TPNG glycans [22]. Notably, neither the focus on low abundant antennary fucosylated glycans (< 1%, except H6N5F1Am1E2) nor the additional processing steps resulted in a marked loss of precision.

Previous research quantified 21 antennary fucosylated glycans in human TPNG using a MALDI-FT-ICR-MS platform [22]. These antennary fucosylated glycans are consistent with our findings. However, we were also able to identify a mixture of both core fucose isomers and antennary fucose isomers for nine of these monofucosylated glycan compositions (Supplementary Figure S4). For example, we have identified H6N5F1E2 and H5N4F1Am1E1 as being a mixture of core fucose isomers and antennary fucose isomers. Previously, these glycans were assumed to be antennary fucosylated (H6N5F1E2) and core fucosylated (H5N4F1Am1E1), respectively. The specific measurement of antennary fucosylated glycans using our assay may increase the accuracy of the relative quantitation of antennary fucose.

To demonstrate the accuracy of the quantitation of antennary fucosylated glycans, procainamide labeled human TPNG was analyzed on a HILIC-FLD-MS<sup>n</sup> platform [38]. This analytical platform was chosen for its accuracy and precision of measurements. Antennary fucosylated glycan peaks were identified from their CID spectra while relative quantitation was performed from the FLD chromatograms. By using the FLD chromatogram instead of the MS spectra for quantification, we overcome the ionization bias resulting from charge differences conferred by underivatized sialic acids (which applies to most antennary fucosylated glycans), as compared to the neutral glycans [43]. As the HILIC-FLD-MS<sup>n</sup> platform did not identify antennary fucosylated tetraantennary glycans, we based the comparison to our MALDI-MS method on antennary fucosylation of triantennary glycans only. This includes nearly 70% of the total abundance of antennary fucosylated glycans in TPNG (Supplementary Table S7). With  $32.4 \pm 1.1\%$  result from our MALDI-MS method were highly comparable results to the  $34.7 \pm 1.3\%$  quantified with the HILIC-FLD-MS<sup>n</sup> reference method. Thus, our assay is capable of accurately quantifying antennary fucosylation in TPNG.



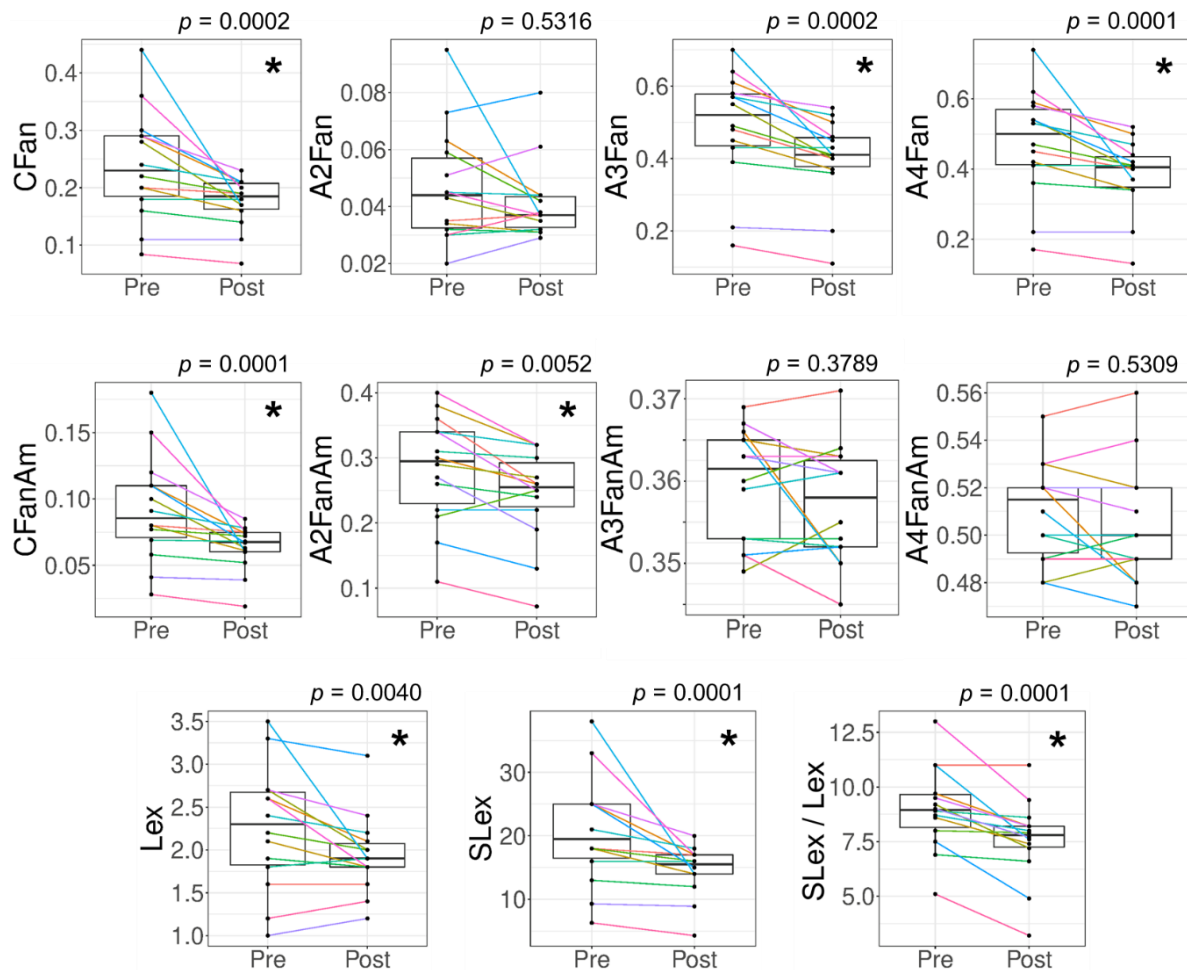
**Figure 6. Intermediate precision of sample preparation for assay and MALDI-FT-ICR-MS measurements of antennary fucosylated glycans in human TPNG. (A)** The mean relative abundances of the 19 quantified antennary fucosylated glycans are shown with the error bars representing standard deviation ( $n = 9$ ). **(B)** The relative standard deviations of the three most abundant antennary fucosylated glycans are shown. The description of the glycan cartoons are as described in Figure 1. [H = hexose, N = *N*-acetylhexosamine, F = deoxyhexose (fucose), Am = amidated *N*-acetylneuraminic acid ( $\alpha 2,3$ -linked), E = ethyl esterified *N*-acetylneuraminic acid ( $\alpha 2,6$ -linked)].

MALDI-FT-ICR-MS is not widely available. Therefore, we also demonstrated that the assay can be measured with a somewhat more widespread MALDI-TOF-MS instrument (Supplementary

Figure S28). In total, 58 glycans could be relatively quantitated of which 15 were antennary fucosylated glycans with a total abundance of  $7.3\% \pm 0.7\%$  using MALDI-TOF-MS (Supplementary Table S8). In contrast, 19 antennary fucosylation glycans out of 70 glycans in total were quantified with a total abundance of  $11.8\% \pm 0.9\%$  using MALDI-FT-ICR-MS (Supplementary Table S7). Thus, fewer antennary fucosylated glycans were quantified with MALDI-TOF-MS due to the expectedly lower sensitivity of the instrument. Furthermore, the differences in quantified antennary fucosylation between the instruments can be accounted for by the different efficiencies of ionization and ion transfer to the detector over the  $m/z$  range; the MALDI-FT-ICR-MS having been more efficiently tuned for good sensitivity in the high mass range. Consistent with the overall trend, the three most abundant antennary fucosylated glycans also show lower (or equal) values in the MALDI-TOF-MS, H6N5F1Am1E2, H5N4F1Am1E1, and H6N5F1Am2E1, contributing to  $3.9\% \pm 0.4\%$ ,  $0.81\% \pm 0.07\%$  and  $0.41\% \pm 0.05\%$  of the total abundance, respectively (Supplementary Figure S28B). The abundance of H6N5F1Am1E2 is consistent with previous MALDI-TOF-MS analysis of TPNG [26, 27]. The median of intermediate precisions for the 15 quantitated antennary fucosylated glycans is 12.5% (10.6% to 13.6% interquartile range; Supplementary Figure S29), which is virtually identical to our MALDI-FT-ICR-MS measurements of the 19 antennary fucosylated glycans. The intermediate precision of all *N*-glycans quantitated by MALDI-TOF-MS is shown in Supplementary Figure S30. Thus, the assay measured on either instrument can be used for detailed quantitation of antennary fucosylation in human TPNG.

### **Quantitation of antennary fucosylation in colorectal cancer patient samples**

To demonstrate the applicability of the developed antennary fucose assay and especially its ability to reveal clinically relevant markers of antennary fucosylation, total serum *N*-glycome (TSNG) samples were analyzed from colorectal cancer patients pre and post tumor resection. This is of specific interest, as colorectal cancer has been associated with an increase in antennary fucosylation and a decrease in core fucosylation, next to an increase in *N*-glycan antennarity and sialylation [6, 8]. Previously, TSNG has been analyzed by MALDI-TOF-MS on sample pairs (pre-operative vs. post-operative) of 61 colorectal cancer patients from the same cohort [8]. The derived traits that were a proxy for antennary fucosylation on *N*-glycans are especially relevant to our study. Multifucosylation on triantennary glycans and  $\alpha$ 2,3-sialylation per antenna in fucosylated triantennary glycans were significantly lowered in the post-operative patient samples as compared to pre-operative samples [8]. The latter derived trait was used as a proxy for sialyl-Lewis X epitopes in TSNG. These changes were thought to be associated with the recovery of the patients since values were closer to healthy controls in the post-operative than in the pre-operative samples.



**Figure 7. Alteration in features related to antennary fucosylated glycans for colorectal cancer patient samples.** Derived traits for glycans were calculated to evaluate antennary fucosylation changes between the 14 pairs of pre-operative (Pre) vs. post-operative (Post) colorectal cancer patient samples. The patient samples were measured with the antennary fucose assay (MALDI-FT-ICR-MS readout). Significant changes were observed for antennary fucosylation in complex *N*-glycans (CFan), antennary fucosylation in triantennary glycans (A3Fan), antennary fucosylation in tetraantennary glycans (A4Fan),  $\alpha$ 2,3-sialylation per antenna of total antennary fucosylated glycans (CFanAm),  $\alpha$ 2,3-sialylation per antenna of diantennary antennary fucosylated glycans (A2FanAm), sialyl-Lewis X abundance (relative area%) in TPNG (SLex), Lewis X abundance in TPNG (Lex) and the ratio of sialyl-Lewis X to Lewis X abundances in TPNG (SLex / Lex). No significant changes were observed for antennary fucosylation in diantennary glycans (A2Fan),  $\alpha$ 2,3-sialylation per antenna of triantennary antennary fucosylated glycans (A3FanAm) and  $\alpha$ 2,3-sialylation per antenna of tetraantennary antennary fucosylated glycans (A4FanAm). The *p*-values shown are from a Wilcoxon matched-pairs signed-rank test with confidence level taken as 95%. Multiple-testing correction was performed using a false discovery rate of 1% calculated with the Benjamini and Hochberg method. The *p*-values < 0.0073 are considered significant and are represented with an asterisk (\*).

The feasibility to analyze clinically relevant markers of antennary fucosylation with our assay was demonstrated on sample pairs (pre-operative vs. post-operative) of 14 colorectal cancer patients. Glycosylation changes in colorectal cancer patient samples are shown in Figure 7 and Supplementary Figure S31. Details on features and their calculations are shown in Supplementary Table S4. In line with previous findings, we were able to show that post-operative patient samples had a significantly lower total antennary fucosylation on total complex *N*-glycans (CFan) as compared to pre-operative samples (Figure 7). Unlike previously, no significant differences were observed in  $\alpha$ 2,3-sialylation per antenna of antennary fucosylated triantennary glycans (A3FanAm), but rather a significantly lowered  $\alpha$ 2,3-sialylation per antenna of antennary fucosylation diantennary glycans (A2FanAm) in post-operative patient samples as compared to pre-operative samples (Figure 7). Since we used only a quarter of the samples, compared to the previous study, missing statistical power provides a simple explanation for missing the sialylation effect on fucosylated triantennary glycans. However, when we focused on the quantitation of sialyl-Lewis X epitopes, it was possible to reproduce the finding. We are able to calculate highly specific derived traits based on antennary fucosylated glycans structures, allowing us to study more specific glycan features. For example, assuming  $\alpha$ 2,6-sialylated antennae are not fucosylated, there is no preference for fucosylation of  $\alpha$ 2,3-sialylated or asialylated antennae and multiple fucoses are on different arms, the relative abundances of sialyl-Lewis X can be calculated. This is indeed lowered in the post-operative samples (Figure 7). The discovery of a novel association with A2FanAm is also easily explained by the increased specificity of our assay. The trait is largely composed of compositions representing a mixture of core and antennary fucose isomers before core defucosylation. Hence, in a regular TPNG / TSNG MALDI-MS analysis unrelated variations in core fucosylation would interfere with the detection of changes in antennary fucosylation of diantennary glycans.

The changes in CFan are mainly contributed by the triantennary glycans (A3Fan) and tetraantennary glycans (A4Fan) rather than the diantennary glycans (A2Fan) and bisecting diantennary glycans (A2BFan) (A2BFan shown in Supplementary Figure S31). We did not observe a change in A2Fan and A2BFan. Without core defucosylation, these traits would largely measure core fucosylation (Supplementary Figure S4). Furthermore, we also observed a significantly lowered multiantennary fucosylation (CFm\_an) in post-operative patient samples as compared to pre-operative samples (Supplementary Figure S31). Finally, we also approximated the abundance of glycans having Lewis X or sialyl-Lewis X epitopes in TPNG. sialyl-Lewis X / Lewis X ratio was significantly lowered in post-operative patient samples as compared to pre-operative samples. This change was mainly contributed by a decreased abundance of sialyl-Lewis X in post-operative samples, although the Lewis X abundance was also lowered. This may be associated with a decreased inflammatory state of the recovering patient [11, 12].

All studied features, except for Lewis X ( $p = 0.0513$ ), were replicated for the MALDI-TOF-MS measurements of the samples (Supplementary Figure 32). Thus, MALDI-TOF-MS measurements are sufficient to detect many of the clinical changes.

## Conclusion

We developed an assay for the relative quantitation of antennary fucosylation and approximation of Lewis X and sialyl-Lewis X abundances in TPNG, based on high-throughput MALDI-MS analysis. This assay is compatible with high sensitivity and ultrahigh-resolution MALDI-FT-ICR-MS or with MALDI-TOF-MS. In total, 19 antennary fucosylated glycans were relatively quantified with precision and accuracy expectable of a MALDI-MS approach. Furthermore, the assay was applied to measuring biomedically relevant changes in antennary fucosylation in colorectal cancer patients pre vs. post tumor resection. Next to previous findings that could be repeated, despite the reduced sample size, the increased specificity enabled the discovery of novel associations. Additionally, we were able to investigate more specialized features based on antennary fucosylation which would not be possible with regular TPNG analysis. The next steps would include further automatization of the assay and perform a high throughput analysis on a large set of patient samples.

## Supplementary information

Supplementary tables and figures are available online free of charge via

<https://www.frontiersin.org/articles/10.3389/fchem.2020.00138/full#h11>

4

## Acknowledgements

This research was supported by the European Union (GlySign, Grant No. 722095) and the NWO (Vernieuwingsimpuls Veni Project No. 722.016.008).



## References

1. Blomme, B., et al., Alteration of protein glycosylation in liver diseases. *J Hepatol*, 2009. 50(3): p. 592-603.
2. Testa, R., et al., N-glycomic changes in serum proteins in type 2 diabetes mellitus correlate with complications and with metabolic syndrome parameters. *PLoS One*, 2015. 10(3): p. e0119983.
3. Zhu, J., et al., Analysis of serum haptoglobin fucosylation in hepatocellular carcinoma and liver cirrhosis of different etiologies. *J Proteome Res*, 2014. 13(6): p. 2986-97.
4. Benicky, J., et al., Quantification of fucosylated hemopexin and complement factor H in plasma of patients with liver disease. *Anal Chem*, 2014. 86(21): p. 10716-23.
5. Saldova, R., et al., Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology*, 2007. 17(12): p. 1344-56.
6. Doherty, M., et al., Plasma N-glycans in colorectal cancer risk. *Sci Rep*, 2018. 8(1): p. 8655.
7. 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*, 2016. 15(1): p. 124-40.
8. de Vroome, S.W., et al., Serum N-glycome alterations in colorectal cancer associate with survival. *Oncotarget*, 2018. 9(55): p. 30610-30623.
9. Juszczak, A., et al., Plasma Fucosylated Glycans and C-Reactive Protein as Biomarkers of HNF1A-MODY in Young Adult-Onset Nonautoimmune Diabetes. *Diabetes Care*, 2019. 42(1): p. 17-26.
10. Thanabalasingham, G., et al., Mutations in HNF1A result in marked alterations of plasma glycan profile. *Diabetes*, 2013. 62(4): p. 1329-37.
11. Higai, K., et al., Glycosylation of site-specific glycans of alpha1-acid glycoprotein and alterations in acute and chronic inflammation. *Biochim Biophys Acta*, 2005. 1725(1): p. 128-35.
12. Brinkman-van der Linden, E.C., et al., 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*, 1998. 15(2): p. 177-82.
13. Pivac, N., et al., Human plasma glycome in attention-deficit hyperactivity disorder and autism spectrum disorders. *Mol Cell Proteomics*, 2011. 10(1): p. M110.004200.
14. Thanabalasingham, G. and K.R. Owen, Diagnosis and management of maturity onset diabetes of the young (MODY). *Bmj*, 2011. 343: p. d6044.
15. Schober, E., et al., Phenotypical aspects of maturity-onset diabetes of the young (MODY diabetes) in comparison with Type 2 diabetes mellitus (T2DM) in children and adolescents: experience from a large multicentre database. *Diabet Med*, 2009. 26(5): p. 466-73.
16. Hattersley, A.T. and K.A. Patel, Precision diabetes: learning from monogenic diabetes. *Diabetologia*, 2017. 60(5): p. 769-777.
17. 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*, 2010. 6(12): p. e1001256.

18. Cserni, G., Nodal staging of colorectal carcinomas and sentinel nodes. *J Clin Pathol*, 2003. 56(5): p. 327-35.
19. Nairn, A.V., et al., Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem*, 2008. 283(25): p. 17298-313.
20. Royle, L., et al., HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal Biochem*, 2008. 376(1): p. 1-12.
21. Stumpo, K.A. and V.N. Reinhold, The N-glycome of human plasma. *J Proteome Res*, 2010. 9(9): p. 4823-30.
22. Vreeker, G.C.M., et al., Automated Plasma Glycomics with Linkage-Specific Sialic Acid Esterification and Ultrahigh Resolution MS. *Anal Chem*, 2018. 90(20): p. 11955-11961.
23. Lageveen-Kammeijer, G.S.M., et al., Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. *Nat Commun*, 2019. 10(1): p. 2137.
24. Staudacher, E., et al., Fucose in N-glycans: from plant to man. *Biochim Biophys Acta*, 1999. 1473(1): p. 216-36.
25. Knezevic, A., et al., High throughput plasma N-glycome profiling using multiplexed labelling and UPLC with fluorescence detection. *Analyst*, 2011. 136(22): p. 4670-3.
26. Reiding, K.R., et al., High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem*, 2014. 86(12): p. 5784-93.
27. 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*, 2015. 14(9): p. 4080-6.
28. Vanderschaeghe, D., et al., High-throughput profiling of the serum N-glycome on capillary electrophoresis microfluidics systems: toward clinical implementation of GlycoHepatoTest. *Anal Chem*, 2010. 82(17): p. 7408-15.
29. 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*, 2010. 9(12): p. 6655-64.
30. Chen, X. and G.C. Flynn, Analysis of N-glycans from recombinant immunoglobulin G by on-line reversed-phase high-performance liquid chromatography/mass spectrometry. *Anal Biochem*, 2007. 370(2): p. 147-61.
31. Wuhrer, M., A.M. Deelder, and Y.E. van der Burgt, Mass spectrometric glycan rearrangements. *Mass Spectrom Rev*, 2011. 30(4): p. 664-80.
32. Lattova, E., J. Skrickova, and Z. Zdrahal, Applicability of Phenylhydrazine Labeling for Structural Studies of Fucosylated N-Glycans. *Anal Chem*, 2019. 91(13): p. 7985-7990.
33. Banazadeh, A., et al., Characterization of glycan isomers using magnetic carbon nanoparticles as a MALDI co-matrix. *RSC Adv*, 2019. 9(35): p. 20137-20148.
34. Harvey, D.J., et al., "Internal residue loss": rearrangements occurring during the fragmentation of carbohydrates derivatized at the reducing terminus. *Anal Chem*, 2002. 74(4): p. 734-40.

35. Vanderschaeghe, D., et al., Endoglycosidase S Enables a Highly Simplified Clinical Chemistry Procedure for Direct Assessment of Serum IgG Undergalactosylation in Chronic Inflammatory Disease. *Mol Cell Proteomics*, 2018. 17(12): p. 2508-2517.
36. Selman, M.H., et al., Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal Chem*, 2011. 83(7): p. 2492-9.
37. 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*, 2015. 14(12): p. 5088-98.
38. Kozak, R.P., et al., Comparison of procainamide and 2-aminobenzamide labeling for profiling and identification of glycans by liquid chromatography with fluorescence detection coupled to electrospray ionization-mass spectrometry. *Anal Biochem*, 2015. 486: p. 38-40.
39. Lettow, M., et al., The role of the mobile proton in fucose migration. *Anal Bioanal Chem*, 2019. 411(19): p. 4637-4645.
40. Suzuki, N., T. Abe, and S. Natsuka, Quantitative LC-MS and MS/MS analysis of sialylated glycans modified by linkage-specific alkylamidation. *Anal Biochem*, 2019. 567: p. 117-127.
41. Toyoda, M., et al., Quantitative derivatization of sialic acids for the detection of sialoglycans by MALDI MS. *Anal Chem*, 2008. 80(13): p. 5211-8.
42. Pongracz, T., M. Wuhrer, and N. de Haan, Expanding the Reaction Space of Linkage-Specific Sialic Acid Derivatization. *Molecules*, 2019. 24(19): p. 3617.
43. Wheeler, S.F., P. Domann, and D.J. Harvey, 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*, 2009. 23(2): p. 303-12.





# **Chapter 5**

**A novel glycosidase plate-based assay for the  
quantification of galactosylation and sialylation  
on human IgG**

## **Abstract**

Changes in human IgG galactosylation and sialylation have been associated with several inflammatory diseases which are a major burden on the health care system. A large body of work on well-established glycomics and glycopeptidomics assays has repeatedly demonstrated inflammation-induced changes in IgG glycosylation. However, these assays are usually based on specialized analytical instrumentation which could be considered a technical barrier for uptake by some laboratories. Hence there is a growing demand for simple biochemical assays for analyzing these glycosylation changes. We have addressed this need by introducing a novel glycosidase plate-based assay for the absolute quantification of galactosylation and sialylation on IgG. IgG glycoproteins are treated with specific exoglycosidases to release the galactose and/or sialic acid residues. The released galactose monosaccharides are subsequently used in an enzymatic redox reaction that produces a fluorescence signal that is quantitative for the amount of galactosylation and, in-turn, sialylation on IgG. The glycosidase plate-based assay has the potential to be a simple, initial screening assay or an alternative assay to the usage of high-end analytical platforms such as HILIC-FLD-MS<sup>n</sup> when considering the analysis of galactosylation and sialylation on IgG. We have demonstrated this by comparing our assay to an industrial established HILIC-FLD-MS<sup>n</sup> glycomics analysis of 15 patient samples and obtained a Pearson's *r* correlation coefficient of 0.8208 between the two methods.

## Introduction

Changes in IgG glycosylation have been associated with certain inflammatory diseases such as rheumatoid arthritis [1, 2], systemic lupus erythematosus [3], type 2 diabetes [4] and inflammatory bowel disease (IBD) [5]. During an inflammatory state, a loss of systemic and/or cellular homeostasis may trigger a change in the glycosylation of IgG molecules [6, 7]. These changes have been shown to alter the IgG glycoprotein interactions to Fc $\gamma$ R receptors among other proteins, and thus tuning the immunoregulatory responses [8-10]. For example, the afucosylation of Fc glycans induces a pro-inflammatory response through interaction with Fc $\gamma$ RIIIA receptor [11], whilst their hyper-galactosylation further strengthens this interaction [12, 13]. Masking exposed galactose epitopes on Fc glycans by increased sialylation could induce an anti-inflammatory response through interactions with the immune suppressing DC-SIGN receptor [14, 15]. However, increased agalactosylation of these glycans counteracts this immune suppressive ability by directly reducing the abundance of sialylated epitopes. Additionally, these agalactosylated glycans further induce a pro-inflammatory response via activation of the complement system through interaction with the mannose binding proteins [16]. Although it is clear that traits of fucosylation, galactosylation and sialylation have their own important function in an immunoregulatory response, we are especially interested in galactosylation and sialylation since they are widely known for having mutually opposing immunoregulatory effects, whilst being dependent on each other for their exposure at the non-reducing termini of the glycans. Hence, we focus here on the quantification of galactosylation and sialylation on IgG as they are indicative of the pro-inflammatory and anti-inflammatory states of the patient, respectively.

Glycomics assays are valuable analytical tools to draw associations between an *N*-glycome and the disease state of patients for diagnostic and stratification purposes. These assays are often complicated by the composition and linkage diversity of the glycans and hence they rely heavily on high-end analytical instrumentation [17-19]. The IgG glycome especially, has been widely studied using assays based on the modern iterations of liquid chromatography (LC) with fluorescence detection (FLD) [1, 20, 21], capillary gel electrophoresis (CGE) with laser-induced fluorescence (LIF) detection [22, 23] and matrix assisted laser desorption mass spectrometry (MALDI-MS) [24, 25]. These analytical platforms have demonstrated a multitude of times the ability to perform detailed and high-throughput analyses of IgG glycosylation, and have deeply progressed our understanding of the IgG glycome. It is due to this work that the IgG glycome is now widely regarded as a simple glycome consisting of only 37 *N*-glycans over the 4 subclasses of IgG [24]. These glycans are predominantly biantennary complex type glycans which are mostly core fucosylated [26] while a smaller proportion have bisecting GlcNAc motifs (~13% abundance) [24]. Furthermore, the sialylation and galactosylation are exclusively  $\alpha$ (2-6) and  $\beta$ (1-4) linked, respectively. Thus rather than analyzing all glycan variants by using any of the aforementioned analytical platforms, the sole quantification of the inflammatory relevant galactose and sialic acid epitopes could be potentially valuable in drawing associations between a patient's inflammatory state and the immune modulation by IgG glycosylation. Our research focused on developing such a method as a simple microtitre plate-based biochemical assay that could bypass the use of the high-end analytical platforms. This is an advantage as there is often a frequent high barrier to entry



into the use of such analytical platforms as they usually require a substantial amount of time, skill and investment to establish in a laboratory. Furthermore, the commonly used industry platform, LC-FLD, often requires extensive measurement time and complex data processing.

Several sandwich immuno/lectin-based biochemical assays have been previously investigated in the quantification of simple glycan epitopes such as sialylation, galactosylation and fucosylation, to more complex epitopes such as the Lewis epitopes [27-31]. These assays often work on the principle of immuno capturing specific glycoproteins from a sample such as serum or plasma onto a microtiter plate, microarray slide or microfluidic cell, followed by detecting their glycosylation using probing lectins or antibodies that bind to specific glycan epitopes. Techniques that are based on microarray slide or microfluidic cells often have the advantage of being able to analyze multiple glycan epitopes on multiple specific glycoproteins in a potentially high-throughput manner and using low amount (< 10  $\mu$ L) of serum sample [27, 29, 31]. However, when based on microtitre plates, this method may require large amounts of serum sample dependent on the abundance of the glycoprotein of interest [30]. Assays based on a microarray slides also require substantial knowledge and/or investment in conjugation chemistry, microarray printers and readers. In general, the accuracy of these techniques is highly dependent on the specificity of the antibodies that are used for capturing the glycoproteins of interest and the lectin or antibodies that are used for detecting their glycan epitopes. The sensitivity of the assay also depends on the binding affinity and/or steric hindrances of the detection lectin or antibody. Also, the necessity to chemically block the glycans of the capturing antibodies, so as to reduce background, may result in a loss of binding affinity probably due to the influences on their three-dimensional structure [32, 33].

Building on the need to achieve a simple and easily adoptable assay for glycan epitope quantification, we have developed a novel glycosidase plate-based assay for the quantification of galactosylation and sialylation on IgG glycoproteins. We aimed at overcoming some of the disadvantages of typical sandwich immuno/lectin-based assays such as adoptability and infrastructure investment, by designing our assay for a potential kit-like format that required a microtitre plate reader as the only instrumentation. IgG captured from human plasma was treated with exoglycosidases, and the released  $\beta$ ,D-Galactose monosaccharides were used in a coupled enzymatic redox reaction to produce a fluorescence signal which is directly proportional to the amount of released galactose residues. Furthermore, we made a comparison between the assay and an industry established hydrophilic interaction liquid chromatography (HILIC)-FLD-MS<sup>n</sup> method for the quantification of a galactosylation trait on IgG from 15 serum samples of patients suspected of having IBD.

## Materials and Methods

### Reagents and samples

Monosodium dihydrogen phosphate, disodium hydrogen phosphate, formic acid, resazurin, lyophilized diaphorase (Lipoamide Dehydrogenase from *Clostridium kluuyveri*) and Protein G beads (Protein G Sepharose®, Fast Flow) were purchased from Sigma Aldrich (UK). Nicotinamide adenine dinucleotide (NAD<sup>+</sup>) was purchased from Roche (Germany). Phosphate buffered saline (PBS) powder, pH 7.4 was purchased from Thermo Fisher Scientific (UK). Galactose monosaccharide standard (400 µg/mL), the buffer pH 8.5 that is used in the preparation of the reagent mix, and the solution of galactose dehydrogenase and mutarotase that is used in the preparation of the redox enzyme mix was obtained from a galactose quantification kit (K-Arga) which was purchased from Megazyme (Ireland). IMAC purified recombinant forms of the β(1-4) specific GH2 galactosidase Bt0461 [34] and α(2-3/6/8) specific GH33 sialidase Bt0455 [35] were provided by Newcastle University. The protein concentrations were determined to be 420 µM and 23 µM, respectively, by a protein quantification kit (Pierce BCA protein assay kit, Thermo Fisher Scientific, UK).

Sodium phosphate buffer solution (SPBS; 250 mM), pH 6 was prepared by dissolving 30.8 grams of monosodium dihydrogen phosphate and 4.95 grams of disodium hydrogen phosphate in a liter of deionized water and then filtering the solution through a 0.45 µm filter.

IgG glycoprotein standard from human serum was purchased from Sigma Aldrich (UK). The protein concentration was determined to be 6.13 mg/mL (~41µM) by a protein quantification kit (Pierce BCA protein assay kit) and which was in accordance with the manufacturer. Human plasma standard (containing 4% trisodium citrate as anticoagulant) was purchased from Sigma Aldrich (UK). Human patient serum samples (n = 15) were obtained from the IBD-BIOM cohort and which were collected as part of a biobank as was previously described [36]. The patient blood were collected in 3.5mL vacuette plastic SST II Advance tube with gel separator, clot activator, and BD Hemograd closure (BD, no 367956). The serum obtained from these preparations were used in the experiments of this paper.

### Glycosidase plate-based assay

Briefly, the assay involves affinity purification of IgG glycoprotein from human plasma and quantifying protein amount by measuring absorbance at 280 nm. These IgG glycoproteins are treated with exoglycosidases to release galactose and/or sialic acid residues from the glycans. The released galactose monosaccharides are then subjected to an enzymatic redox reaction to produce a fluorescence output that is measured in a microplate reader.

**Affinity purification of IgG glycoproteins from human plasma.** IgG purification is performed in a 96 well plate format using protein G conjugated Sepharose beads. 5 µL of protein G bead resin is transferred to each well of a 96 well plate (LC-PROC-96, Ludger, UK). The plate is fitted onto a vacuum manifold (LC-VAC-MANIFOLD-KIT, Ludger) and the beads are washed three times with 200 µL PBS by applying a vacuum of between 5 and 10 mbar. The plate is removed from the manifold, and 100 µL PBS is added to the wells containing the protein G beads followed by 10 µL of plasma samples. The plate is sealed and is incubated for 60 minutes at

room temperature on an orbital plate shaker at low speed. After incubation, the plate is once again fitted onto a vacuum manifold and the solution in the wells drained away. The protein G beads that have bound the IgG molecules are washed three times with 200  $\mu$ L PBS, followed by three washes with 200  $\mu$ L deionized water to wash away residual PBS. The plate is once again removed from the manifold, and the bottom is gently tapped dry on a paper towel. To elute the IgG from the protein G resin, the plate is stacked onto a 96 well collection plate (1.2 mL 96 Deepwell plate, 4titude, UK) and 30  $\mu$ L of an aqueous solution of 10 mM formic acid is added to the wells containing the resin. The top plate is covered with a lid and incubated for 15 minutes at room temperature on a plate shaker at low speed. After incubation, the stack plate setup is centrifuged at 1000 rpm for 5 minutes. The elution procedure is repeated a second time. The eluted IgG samples from the collection plate are then transferred to a conical bottom 96 well PCR plate (Framestar 96 non skirted, 4titude).

**Protein quantification and exoglycosidase treatment.** Each IgG sample was subjected to three treatment conditions which are **1)** galactosidase treatment for quantification of terminal galactosylation, **2)** combined galactosidase and sialidase treatment for quantification of total galactosylation and **3)** exoglycosidase untreated negative control for protein quantification and for sample fluorescence background and interference subtraction. All the following steps for assay preparation were performed on a robotic platform (Microlab starlet, Hamilton, Germany). The IgG sample plate and an empty 384 well plate (Ultravision plate 384, 4titude) were placed on the deck of the robotic platform along with 2 mL tubes (Sarstedt, UK) containing the galactosidase mix (2  $\mu$ M galactosidase Bt0461 in 125 mM SPBS, pH 6), galactosidase and sialidase mix (2  $\mu$ M galactosidase Bt0461 and 5  $\mu$ M sialidase Bt0455 in 125 mM SPBS, pH 6) and 125  $\mu$ M SPBS, pH 6. For each sample, 20  $\mu$ L of the respective mixes were transferred to separate wells in the 384 well plate (Supplementary figure S1). Additionally, 20  $\mu$ L of the 125 mM SPBS buffer was transferred to the blocks of wells of the galactose monosaccharide standards and of the IgG glycoprotein standards (Supplementary figure S1). 10  $\mu$ L of PBS was transferred to the IgG samples to prevent sticking of protein to the plasticware. 15  $\mu$ L of these IgG samples were transferred each as three sub-samples to the 384 well plate containing the corresponding treatment mixes (Supplementary figure S1). After this step, the IgG sample plate on the deck was swapped for another plate containing the IgG glycoprotein standards of known amounts (in 10 mM formic acid) for the preparation of the IgG glycoprotein absorbance standard curve. Again, 10  $\mu$ L of PBS was added to these standards and 15  $\mu$ L were transferred each to their respective block on the 384 well plate (Supplementary figure S1). The final amounts of the IgG glycoprotein standards making up the standard curve in the 384 well plate were 30.7  $\mu$ g, 61.3  $\mu$ g, 92  $\mu$ g, 122.6  $\mu$ g and 184  $\mu$ g.

The 384 well plate was removed from the deck and centrifuged for 2 minutes at 1000 g. The plate was placed on an orbital plate shaker for 3-5 minutes before centrifuging again. Protein amounts of the IgG samples were determined by measuring the absorbance at 280 nm for the exoglycosidase untreated sub-samples and equating it to the absorbance standard curve prepared from the IgG glycoprotein standards of known amounts. The absorbance measurements at 280 nm were taken on a microplate reader (Enspire 2300, Perkin Elmer Enspire, USA). Once the plate was inserted into the reader, it was shaken in an orbital motion for 60 seconds before the measurements were started, so as to ensure equilibration of its

temperature to the chamber which was 21 °C. The measurements were taken at a height of 2 mm. 10 flashes were taken for each well, of which 5 were integrated. The plate was measured three consecutive times and the measurements were averaged.

After measuring the absorbance, the plate was heat sealed (Peel heat seal, 4titude) and incubated overnight at 37 °C.

**Enzymatic redox reaction and fluorometric measurement of assay.** After the overnight exoglycosidase treatment, the samples were subjected to an enzymatic redox reaction which produces the detection mechanism of the assay. All the following steps were performed on the robotic platform. The 384 well plate containing the exoglycosidase treated samples, a fresh conical bottom 96 well PCR plate, and 2 mL and 0.5 mL tubes (Sarstedt) containing the reagent mix [5 µL of pH 8.5 buffer (K-Arga kit, Megazyme), 4 µL of 1 mM resazurin, 2.6 µL of 25 mM NAD<sup>+</sup> and 23.4 µL of deionized water] and the redox enzyme mix [4 µL of 4U/mL diaphorase and 1 µL solution of galactose dehydrogenase and mutarotase (K-Arga kit, Megazyme)] respectively, were placed on the deck of the robotic platform. The PCR plate was used for the preparation of a dilution series of galactose monosaccharide in water from a 400 µg/mL galactose stock solution (K-Arga kit, Megazyme, Ireland). This dilution series were prepared in triplicates and 15 µL of each was transferred to the 384 well plate. The final amounts of galactose monosaccharide standards in the 384 well plate were 8.3 pmol, 41.6 pmol, 83.3 pmol, 166.6 pmol, 333.2 pmol, 499.7 pmol and 666.3 pmol. Next, 35 µL of the reagent mix were transferred to the wells containing the samples and the diluted galactose monosaccharide standards. This was followed by transferring 5 µL of the redox enzyme mix to the same wells. The 384 well plate was sealed (qPCR Seal, 4titude) and centrifuged for 2 minutes at 1000 rpm. The plate was then placed on an orbital plate shaker for 5 minutes before centrifuging again and incubating in the dark at room temperature for 3 hours. After 3 hours, 25 µL of PBS was added to the wells of the plate. The plate was placed on an orbital plate shaker for 5 minutes before centrifuging for 2 minutes at 1000 rpm.

Fluorescence measurements were taken on the fluorescence microplate reader (Enspire 2300, Perkin Elmer Enspire, USA). The resorufin that was produced in the redox reaction was quantified by measuring its fluorescence (with an excitation wavelength ( $\lambda_{ex}$ ) = 571 nm and an emission wavelength ( $\lambda_{em}$ ) = 586 nm). The excitation illumination was from above the plate and the measurements were taken at a height of 10 mm. 100 flashes were taken for each well and were all integrated. The plate was measured three consecutive times and the measurements were averaged.

A standard curve was prepared from the fluorescence measurements of the galactose monosaccharide standard (K-Arga, Megayzme) solutions of known amounts. For each sample, the fluorescence measurement from the exoglycosidase untreated sub-sample was subtracted from the sub-samples treated with galactosidase and the combination of galactosidase and sialidase, respectively. This was done to exclude fluorescence background or interferences coming from the sample itself. The amount of galactosylation in the samples was calculated by equating their fluorescence measurements to the galactose standard curve.

## Method optimization and assay performance

For details pertaining to assay development and optimization please see section “Supplementary Methods”. The linear range of quantification of the assay for IgG glycoproteins was determined with human IgG glycoprotein standard (Sigma, UK). These standards were dried down using a vacuum centrifuge (Thermo-Savant, UK) and then reconstituted in PBS to make solutions of 30.7 µg/10 µL, 61.3 µg/10 µL, 92 µg/10 µL, 122.6 µg/10 µL and 245.2 µg/10 µL.

For demonstrating the assays ability to remove all possible galactose and/or sialic acid residues from the *N*-glycans on the intact IgG glycoproteins, IgG samples purified from human plasma were treated with the exoglycosidase regime as according to the workflow of the assay. However, after the overnight exoglycosidase treatment, these sub-samples were collected for HILIC-FLD-MS<sup>n</sup> analysis (see subsection 2.4).

The intermediate precision of the assay for sample preparation and measurement was performed in three independent experiments on three different days within a period of three weeks. For each of these experiments, the assay was performed with 16 replicates of IgG purified from a pooled human plasma (Sigma Aldrich).

## Hydrophilic interaction liquid chromatography analysis of IgG *N*-glycans

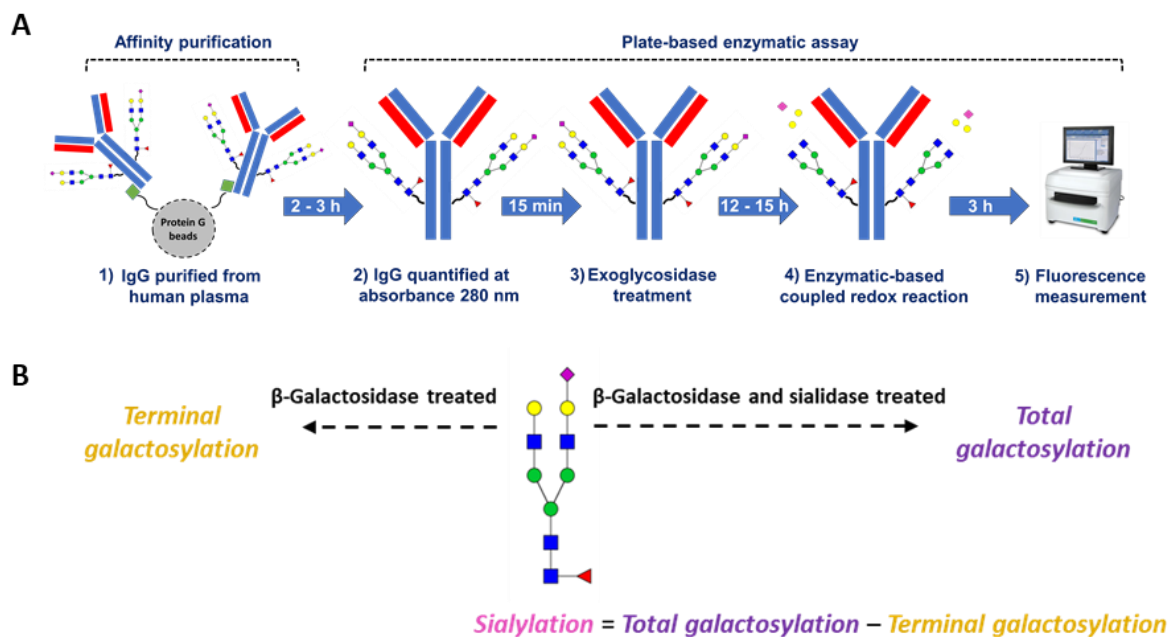
PNGaseF released *N*-glycans from human IgG were used for procainamide labelling and analysis on a HILIC-FLD-MS<sup>n</sup> platform as previously described [37]. IgG was affinity purified from human plasma/serum as described in section 2.2.1. 25 µL of the purified IgG samples were dried down in the 96 well PCR plate using a vacuumed centrifuge (manufacturer) before resuspending them in 9 µL water. The *N*-glycans were released using a PNGaseF kit (LZ-rPNGaseF-kit, Ludger). Briefly, the IgG glycoproteins were denatured by adding 1 µL of 10x denaturation buffer to the samples, heat sealing (Pierce heat seal, 4titude) the plate and incubating at 95°C for 10 minutes. After heat denaturation, the plate was allowed to cool to room temperature before adding 10 µL of a PNGaseF mix [1 µL PNGaseF, 2 µL 10x reaction buffer, 2 µL 10x NP-40 and 5 µL water]. The sample plate was heat sealed (Pierce heat seal, 4titude) before incubating it overnight at 37°C.

Following the overnight incubation, the reducing ends of the released glycans were deaminated to aldoses by acidifying the samples in 1 % formic acid (Sigma) at room temperature for 50-60 minutes. These acidified samples were then filtered through a protein binding plate (LC-PBM-96, Ludger). The wells of the plate were washed twice with 100 µL water. The washes collected along with the filtrate were dried down in a vacuum centrifuge. These *N*-glycans samples were labelled by reductive amination in 10 µL of water and 10 µL of procainamide labelling solution (LT-KPROC-24 containing NaCNBH<sub>3</sub>, Ludger), and incubated for 60 minutes at 65 °C. A HILIC-type clean-up plate (LC-PROC-96, Ludger) was used to remove unreacted procainamide dye. Procainamide labelled *N*-glycans were eluted twice in 100 µL water. These purified glycan samples were dried down in vacuum centrifuge and resuspended in 100 µL water.

Procainamide labelled samples were analyzed by HILIC-FLD-MS<sup>n</sup>. 12.5 µL of each sample was mixed with 37.5 µL acetonitrile, and 20 µL of this solution was injected into an ACQUITY BEH Glycan column (1.7 µm, 2.1 x 150 mm Waters Inc, USA) at 60°C on a Dionex Ultimate 3000 UHPLC instrument (Thermo, UK) with a fluorescence detector ( $\lambda_{ex} = 310 \text{ nm}$ ,  $\lambda_{em} = 370 \text{ nm}$ ) coupled in-line to an Amazon speed ETD (Bruker Daltonics, Bremen, Germany). The UHPLC gradient conditions were as follows: (solvent A – 50 mM ammonium formate, pH 4.4; solvent B – acetonitrile) 0 to 53.5 min, 76 to 51% B, 0.4 mL/min; 53.5 to 55.5 min, 51% to 0% B, 0.4 mL/min to 0.2 mL/min; 55.5 to 57.5 min, 0% B at a flow rate of 0.2 mL/min; 57.5 to 59.5 min, 0 to 76% B, 0.2 mL/min; 59.5 to 65.5 min, 76% B, 0.2 mL/min; 65.5 to 66.5 min, 76% B, 0.2 mL/min to 0.4 mL/min; 66.5 to 70.0 min, 76% B, 0.4 mL/min. The Amazon Speed settings used were as follows: source temperature, 250°C; gas flow, 10 L/min; capillary voltage, 4500 V; ICC target, 200,000; Max. accu. time (Maximum Accumulation Time), 50.00 ms; rolling average, 2; number of precursor ions selected, 3; release after 0.2 min; positive ion mode; scan mode, enhanced resolution; mass range scanned, 600 to 2000; target mass, 900.

## Results and discussion

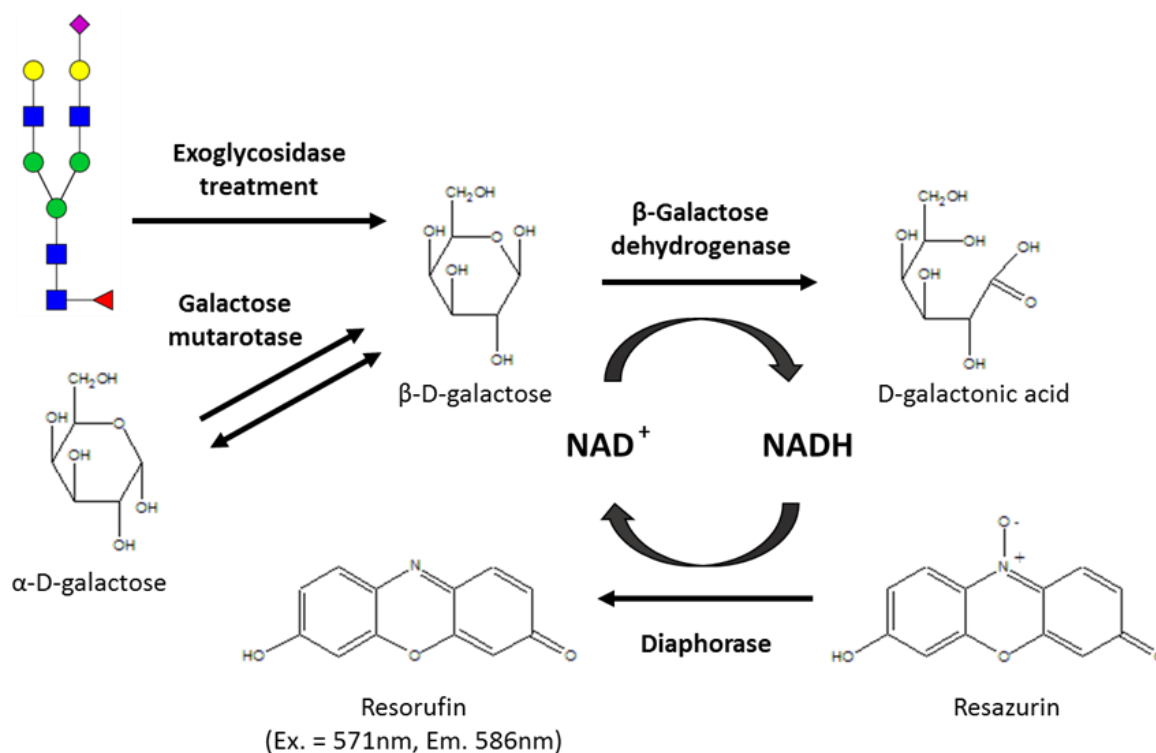
### Workflow and mechanism of assay



**Figure 1. Workflow and schematic description of the glycosidase plate-based assay. (A)** Workflow of the glycosidase plate-based assay for the quantification of galactosylation and sialylation on affinity purified IgG from human plasma. **(B)** Schematic description of exoglycosidase treatment regime for the quantification of total or terminal galactosylation. The difference between both quantitated galactosylation types is equivalent to the sialylation on the glycoproteins not typically exhibiting polysialylation. [Green squares: protein G; Blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid].

The workflow for the glycosidase plate-based assay is shown in Figure 1A. IgG purified from human plasma sample is transferred to a 384 well plate as a set of three sub-samples (Supplementary figure S1). The sub-samples without exoglycosidase treatment are used for quantification of IgG glycoprotein amounts by measuring absorbance at 280 nm. For the sub-samples treated with only the galactosidase, the exposed galactose residues (non-sialylated residues) are released from the *N*-glycans and will be referred to here as terminal galactosylation (Figure 1B). However, for the sub-samples treated with the combination of the galactosidase and a sialidase, all galactose residues on the *N*-glycan are released and will be referred to here as total galactosylation (Figure 1B). After the exoglycosidase treatments, the released galactose monosaccharides are utilized in an enzymatic redox reaction which forms the fluorometric detection mechanism of the assay (Figure 2). This fluorescence signal is used for calculating the molar amounts of terminal galactosylation and total galactosylation, and their difference is equivalent to the sialylation on the *N*-glycans (Figure 1B). However, these measurements are cumulative of the abundance of galactosylation and

sialylation on the IgG glycoproteins as well as the amount of IgG glycoprotein in samples. Hence, in-order to express only the changes in glycosylation traits, the measurements have to be normalized to account for the differences in IgG amounts. This was done as either **1)** an absolute quantification of moles of galactose or sialic acid per  $\mu\text{g}$  of IgG glycoprotein, when normalizing to the measured glycoprotein amounts in the sub-samples or as **2)** a ratio of terminal galactosylation to total galactosylation, when normalizing to the glycan traits itself. The latter will be referred to here as galactosylation index.



**Figure 2. Schematic of the enzymatic redox reaction resulting in fluorescent compound resorufin that forms the detection mechanism of the glycosidase plate-based assay.** Resorufin measured at excitation (Ex.) wavelength of 571nm and emission (Em.) wavelength of 586nm. [ $\text{NAD}^+$ : Nicotinamide adenine dinucleotide (oxidized);  $\text{NADH}$ : Nicotinamide adenine dinucleotide (reduced)]

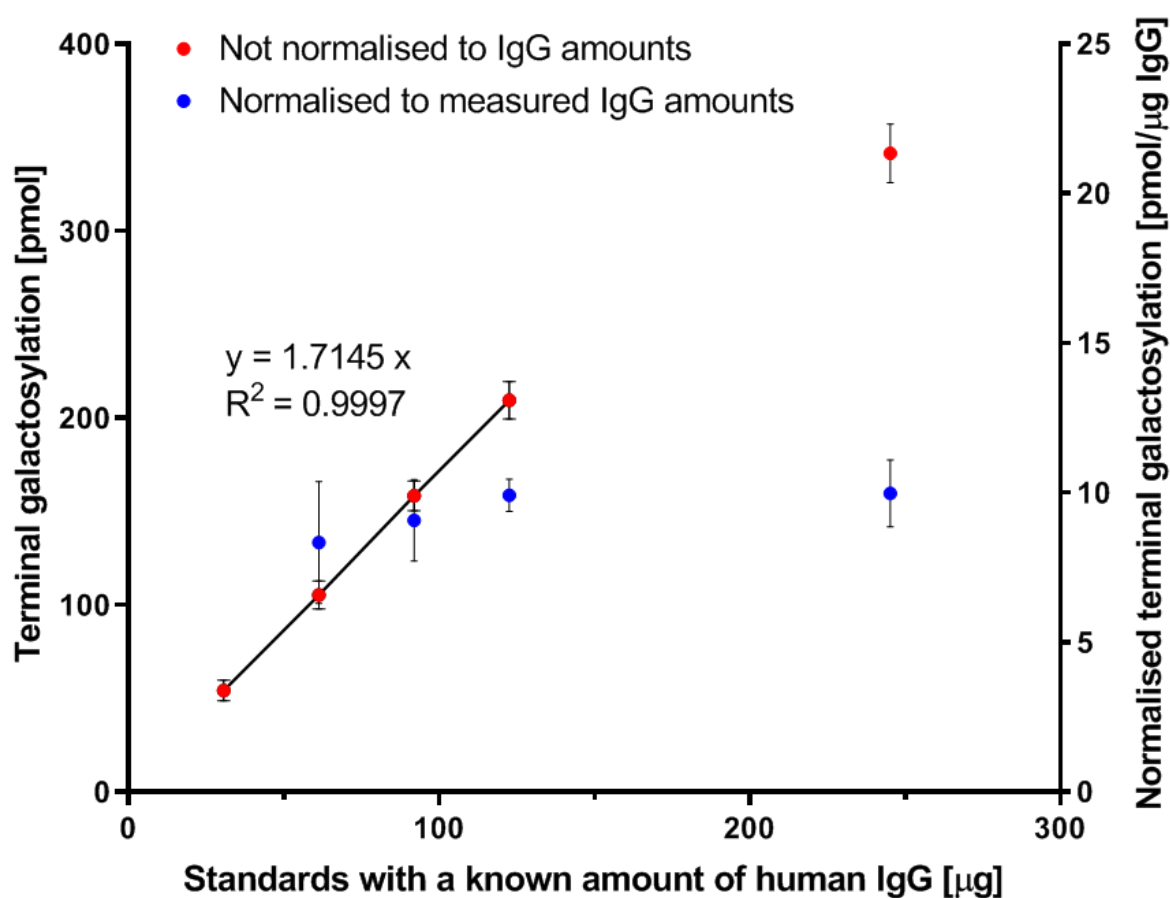
The principle for the enzymatic redox reaction is shown in Figure 2. After the exoglycosidase treatment of the IgG sub-samples, the released  $\beta$ -D-galactose monosaccharides are oxidized to D-galactonic acid by a  $\beta$ -galactose dehydrogenase which in turn reduces  $\text{NAD}^+$  to  $\text{NADH}$ . The  $\text{NADH}$  is then oxidised by a diaphorase which enables the reduction of resazurin to the fluorescent compound resorufin in molar proportions stoichiometric to the released galactose monosaccharides. The general disadvantage of such enzymatic assays is that their accuracy is often limited to the specificity/activity of the enzymes, which in this case are the galactosidase, sialidase and galactose dehydrogenase. However, since the assay functions on the combined activity of these enzymes, the cumulative effect of improved specificity to



galactose quantification could be achieved. Additionally, the incorporation of a mutarotase into the redox reaction helps maintain a  $\beta$  conformer majority of the released galactose which favors the  $\beta$ -galactose dehydrogenase activity and hence assay kinetics.

For details pertaining to assay development and optimization, please see section "Supplementary Methods".

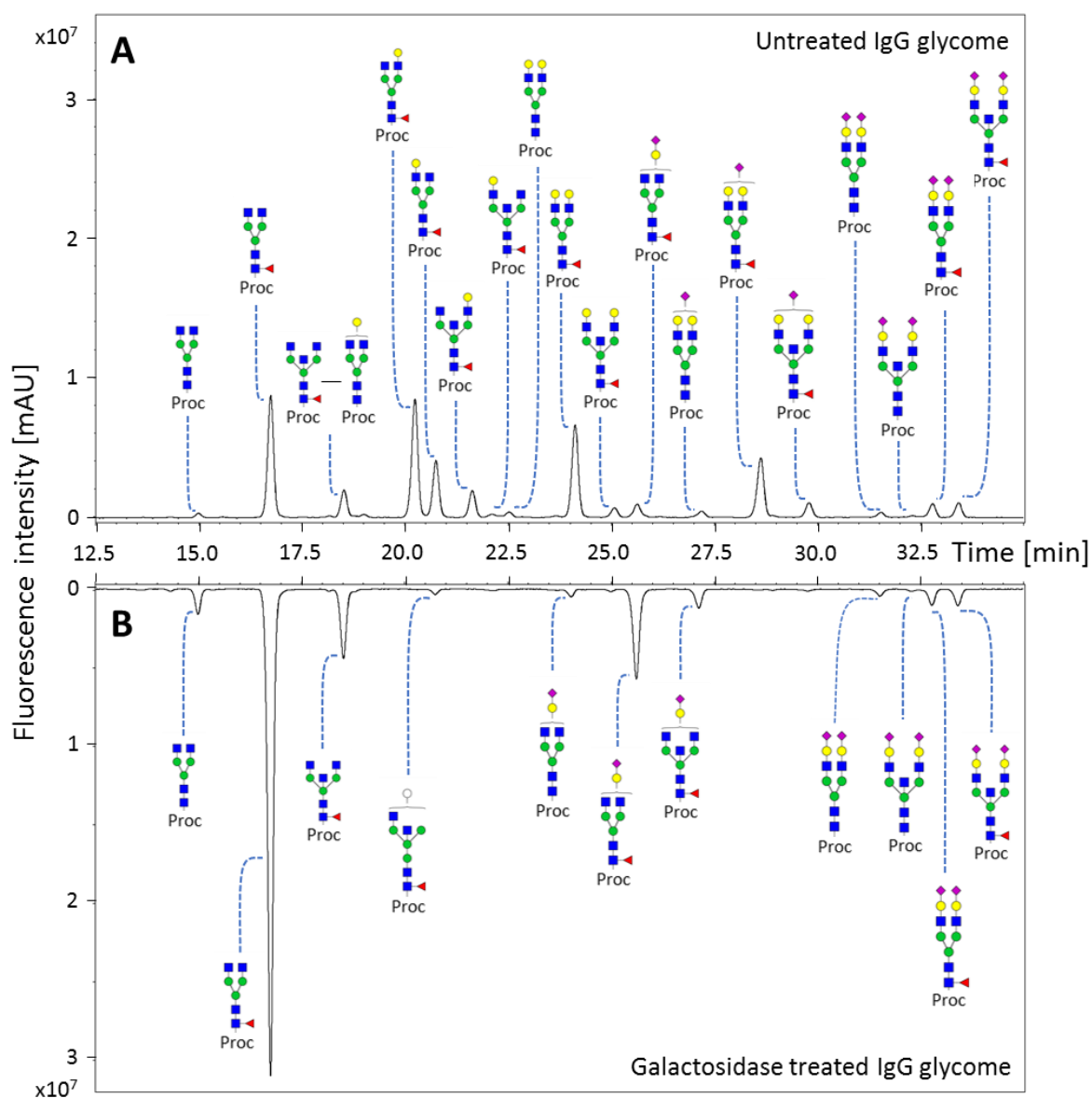
## Assay performance



**Figure 3. Linear range of quantification of the glycosidase plate-based assay for the quantification of terminal galactosylation on human IgG standards of known starting amounts.** The terminal galactosylation are expressed as not normalized values (Red) and as values normalized to the measured IgG amounts in the sub-samples of assay (Blue). The line formula is representative of the linear range (solid trend linear) of the not normalized terminal galactosylation values. The normalized values for the lowest amount of IgG tested, 30  $\mu\text{g}$ , was not calculable due to lack of measureable absorbance signals above background at 280nm.

The linear range of quantification for the assay, based on a galactose monosaccharides standard, is 42 pmol to 666 pmol (Supplementary figure S2). However, for IgG glycoprotein standards, the linear range of quantification of not normalized terminal galactosylation is  $54.1 \pm 5.6$  pmol to  $209.3 \pm 10.1$  pmol of galactose which corresponds to 30  $\mu\text{g}$  to 122  $\mu\text{g}$  of IgG glycoprotein standards respectively (Figure 3, Red). The loss of linearity of the not normalized values for amounts greater than 122  $\mu\text{g}$  of IgG is potentially due to over saturation of the protein G resin that is used for affinity purification of the IgG, thus resulting in incomplete capture of IgG from the samples. This is demonstrated by normalizing the terminal galactosylation to the measured IgG amounts in the sub-samples of the assay which will account for the losses of IgG during the processing steps of the assay. Similar amounts of normalized terminal galactosylation were obtained for the standards with different starting amounts of IgG (Figure 3, Blue). Furthermore, the normalized value of terminal galactosylation for the lowest amount of IgG tested, 30  $\mu\text{g}$ , was not calculable due to lack of measureable absorbance signals above background at 280nm (data not shown). Human serum contains 41  $\mu\text{g}$  – 217  $\mu\text{g}$  IgG per 10  $\mu\text{L}$  of serum depending on the age and/or disease state of the patient [38, 39]. Our assay has been developed to perform within this range of IgG amounts (Figure 3). However, to make a meaningful comparison between samples for the quantitated galactosylation and sialylation, these values will have to be normalized for the differences in IgG amounts between plasma samples as demonstrated in Figure 3.

When cumulatively comparing the normalized values of galactosylation for samples with different amounts of IgG, the combined relative standard deviations (RSD) are much larger for the absolute quantification of galactosylation (pmol of galactose per  $\mu\text{g}$  IgG) on IgG compared to the galactosylation index (Supplementary figure S3B). As mentioned above, the galactosylation index is the ratio of terminal galactosylation to total galactosylation in the IgG sample. On correlating terminal galactosylation to total galactosylation, a coefficient of 0.9947 was obtained (Supplementary figure S4). However, for absolute quantification, on correlating the measured IgG glycoprotein amounts to terminal galactosylation or total galactosylation, lower coefficients of 0.9597 or 0.9652 were obtained, respectively (Supplementary figure S5). The somewhat stronger correlations between terminal galactosylation and total galactosylation, makes the galactosylation index more robust at normalization of variations in IgG amounts, resulting in a lower combined RSD for the samples with different amounts of IgG glycoproteins (Supplementary figure S3). The weaker correlation of the measured IgG glycoprotein amounts to terminal galactosylation or total galactosylation may be explained by the high variation in protein quantification using absorbance at 280 nm in the assay. These variations may be attributed to the 384 well plate, interferences in the sample and/or variation in path lengths at the meniscus of the viscous sample solution.

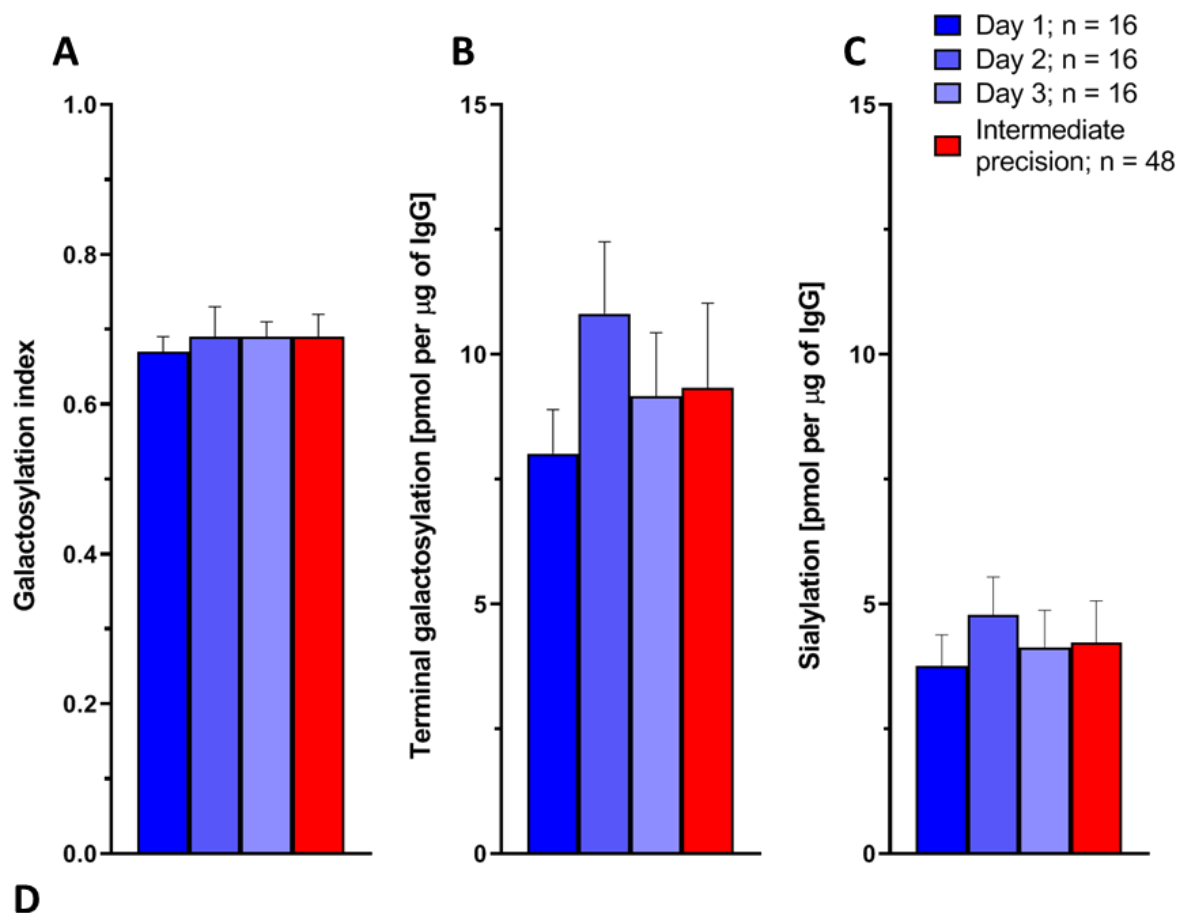


**Figure 4. Confirmation of complete de-galactosylation of terminal galactose residues of *N*-glycans on IgG glycoproteins by the exoglycosidase treatment in the glycosidase plate-based assay.** The released *N*-glycans from IgG glycoproteins (A) without galactosidase treatment was compared with (B) galactosidase treatment on a HILIC-FLD-MS<sup>n</sup> platform after labelling the reducing end with procainamide. For the confirmation of complete de-galactosylation and de-sialylation of *N*-glycans on IgG glycoproteins please see supplementary figure S6. [Proc: procainamide; Blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, white circle: ambiguous hexose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid].

The exoglycosidase treatment of the IgG sub-samples in the assay are performed on intact IgG *N*-glycans (PNGaseF untreated). Hence there is a possibility that the conformation and intra-IgG molecular interactions [10, 40] of the *N*-glycans and/or the protein domains could restrict the activity of the exoglycosidases. This is especially important since the Fc *N*-glycans

are largely located in the cavity formed by the C $\gamma$ 2 and C $\gamma$ 3 domains of the Fc homodimers and makes several interactions with the protein backbone [10]. However, in-order to contribute to the robustness of the assay, it is essential to ensure a complete de-sialylation and de-galactosylation of all *N*-glycans on IgG. To demonstrate this, the IgG sub-samples treated with and without exoglycosidases in the assay, were collected and their glycans released. The released glycans were labelled with procainamide and analyzed on a HILIC-FLD-MS<sup>n</sup> platform (Figure 4). For the exoglycosidase untreated IgG sub-samples, 70.9%  $\pm$ 0.8% abundance of the glycome consist of glycans having one or more exposed terminal galactose residues (Supplementary table S1). However, no such terminal galactosylated structures were detected after galactosidase treatment (Figure 4B and Supplementary table S2). Similarly, no detectable sialylated or galactosylated structures were observed after a combined sialidase and galactosidase treatment of IgG sub-samples, which was used for quantification of total galactosylation (Supplementary figure S6 and Supplementary table S3). Thus, the exoglycosidase treatment is not restricted to certain glycan structures or by the conformations and the molecular interactions of the glycan on the IgG molecule or by the IgG domains itself. The annotation list of the IgG glycans with and without exoglycosidase treatment are shown in Supplementary table S4, S5 and S6.

Intermediate precision of sample preparation and measurement of the assay was assessed by performing three independent experiments on different days each with 16 replicates of IgG purified from pooled human plasma (Figure 5). The galactosylation index was 0.69 with an RSD of 4.1% (Figure 5A and 5D). This was similar to the galactosylation index calculated from procainamide labelled IgG glycans measured on a HILIC-FLD-MS<sup>n</sup> platform which was 0.729 with an RSD of 0.5% (Supplementary table S7). These values signify that ~70% of all galactose residues on human IgG glycans are terminally exposed (terminal galactosylation) whilst the remaining 30% are capped by sialic acid residues. Alternatively, this can be stated as the amount of sialic acid residues are 30% of the total galactose residues. In a general comparison to some sandwich immuno/lectin-based glycomics assays, the intermediate precision was comparable [30], if not lower [29, 41]. Although the performance of such assays highly depends on the quantitated glycan traits, glycoprotein of interest, the matrix of the sample and/or the specificity of the lectins and antibodies used in assay. Furthermore, for our assay, the absolute quantification of terminal galactosylation was 9.3 pmol per  $\mu$ g IgG while the sialylation was 4.2 pmol per  $\mu$ g IgG, with RSDs of 18.2% and 19.9% respectively (Figure 5B, 5C, 5D). These higher RSD can be explained again by likely variations in quantification of glycoprotein amounts using absorbance at 280nm.



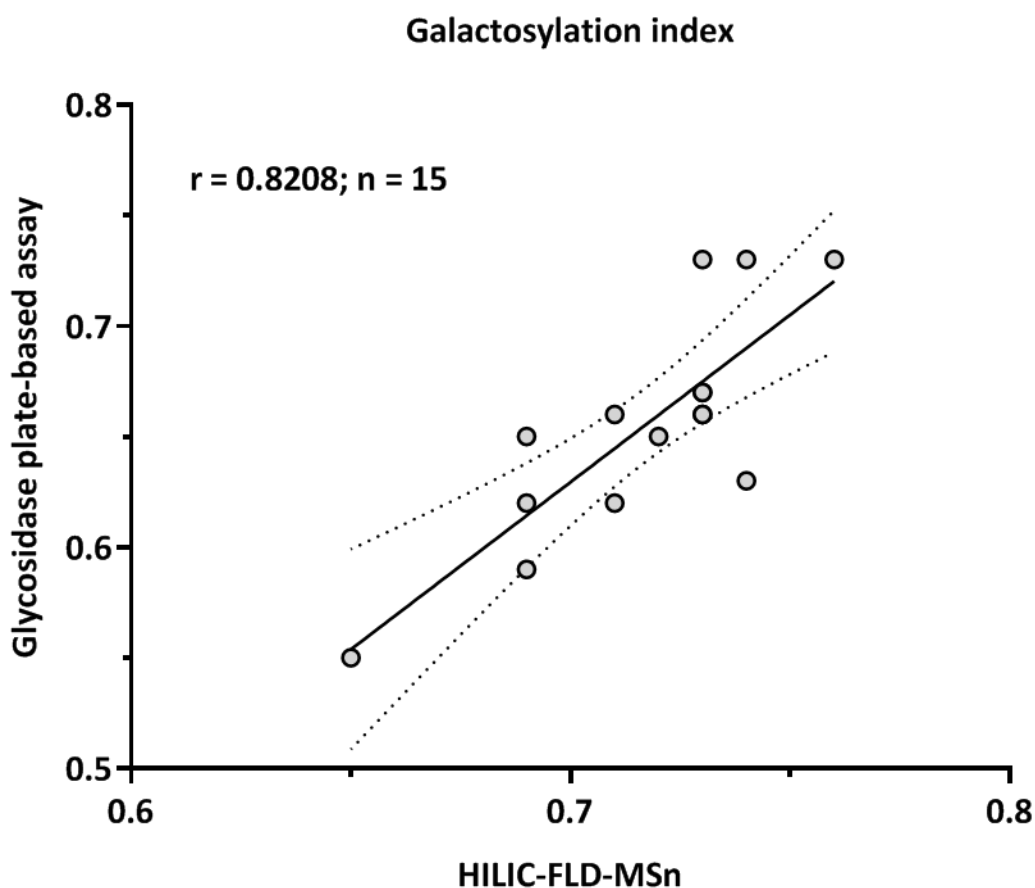
**Figure 5. Intermediate precision of sample preparation and measurements for the glycosidase plate-based assay on IgG from pooled human plasma.** The mean (A) galactosylation index, (B) terminal galactosylation [pmol] per µg IgG glycoprotein and (C) sialylation [pmol] per µg IgG glycoprotein are shown with error bars representing the standard deviation (n = 16). (D) The relative standard deviation are shown.  $[Galactosylation\ index = \frac{Terminal\ galactosylation}{Total\ galactosylation}]$

Both the absolute quantification of galactosylation or sialylation and the galactosylation index have their own advantages and disadvantages which may depend on the sample type and the purpose of the analysis. The low RSDs associated with the galactosylation index make it

suitable for the comparison of clinically relevant samples where the biological relevant variations in galactosylation and sialylation could be limited. Moreover, it is a single value that gives the relative abundance of both galactosylation and sialylation, which on IgG glycans, are associated with pro-inflammatory and anti-inflammatory states respectively. Although the absolute quantification is associated with high RSDs, it has significance in the characterization and/or routine quality control analysis of certain sample types such as bio-pharmaceutical recombinant monoclonal antibodies. This is because it provides information on the number of galactose or sialic acid residues per molecule of the pure glycoprotein. Furthermore, the absolute quantification entails the level of agalactosylation on IgG while the galactosylation index omits such estimates. This reflects the major drawback of the galactosylation index since agalactosylation has also been associated with pro-inflammatory states.

### **Comparison of glycosidase plate-based assay and HILIC-FLD-MS<sup>n</sup> for analysis of patient galactosylation index**

The purpose of this research is to introduce the glycosidase plate-based assay as a prototype method of an initial screening assay for IgG galactosylation in a patient population. Following this, patients which may have a likely change in galactosylation, sialylation or galactosylation index can selectively be used for a further in-depth IgG glycome analysis such as on a HILIC-FLD-MS<sup>n</sup> platform. HILIC-FLD-MS<sup>n</sup> is a well-established technique that is widely used in industry for the identification and quantification of glycosylation changes in IgG. In order to establish the potential of the glycosidase plate-based assay to be used as an initial screening assay, we wanted to quantitatively compare the galactosylation index of this technique with that of the HILIC-FLD-MS<sup>n</sup> method. We chose the galactosylation index as the appropriate value for making our comparison as it can be precisely calculated from the glycan profiles obtained by HILIC-FLD-MS<sup>n</sup> while neglecting the differences in IgG glycoprotein amounts. The analysis was performed on 15 patient plasma samples suspected of having IBD. The correlation of the methods resulted in a Pearson's *r* correlation coefficient of 0.8208 with *p* value 0.0002 (Figure 6 and Supplementary table S8). Although the assay demonstrated a good correlation to the HILIC-FLD-MS<sup>n</sup> method, the lack of a stronger correlation cannot simply be explained by the measurement variations of the assay as RSDs of < 5% were obtained for IgG from a pooled human plasma standard (Figure 4). However, presence of a quantification biasness between the methods could be explained by the *O*-glycans on IgG [42] and/or the presence of co-purified or contaminating proteins having *O*-glycosylation after the protein G purification of IgG from plasma. These undesirable glycoproteins could contribute to the quantified terminal and total galactosylation in the glycosidase plate-based assay and skew the galactosylation index of the patient samples. Notably the activity of the exoglycosidases used in the assay has not been tested on intact *O*-glycans of proteins. Additionally, the co-elution of multiple glycans under a fluorescence peak on the HILIC-FLD-MS<sup>n</sup> platform may also skew the calculated galactosylation index relative to the glycosidase plate-based assay.



**Figure 5. Correlation of the galactosylation index of IgG from human serum measured by the glycosidase plate-based assay, and by HILIC-FLD-MS<sup>n</sup> analysis of procainamide labelled released *N*-glycans.** The human serum sample (n=15) were obtained from an IBD cohort. A Pearson's *r* correlation coefficient of 0.8208 was calculated with a *p* value of 0.0002. The trendline drawn is a linear regression of the data points with the error lines as 95% confidence intervals.

The advantages of the glycosidase plate-based assay over HILIC-FLD-MS<sup>n</sup> are that it does not require extensive investments in high-end instrumentation techniques, software packages or technical skills. The only analytical instrument required is a fluorescence and absorbance microplate reader. We have demonstrated the assay as a semi-automated method on a robotic platform, but it can also easily be performed manually with the appropriate multichannel pipette. Once the samples have been processed in the assay, the measurement can be performed very rapidly in less than 5 minutes for measuring 96 samples, as opposed to several days on a HILIC-FLD-MS<sup>n</sup> platform. Furthermore, data processing is simple and quick as it simply provides the change in galactosylation and sialylation on IgG in the form of measured fluorescence intensities. However, a disadvantage of the assay is the lack of detailed information on glycosylation changes. It does not relay any information on the extent of galactosylation or sialylation on the different glycan structures such as bisecting or fucosylated structures. Hence, we do recognize the analytical benefits of the HILIC-FLD-MS<sup>n</sup>

technique. However, the purpose of this assay can be two-fold. As a simplistic initial screening assay to identify the samples that require further in-depth IgG glycomics analysis. Alternatively, where an established glycan related biomarker has already been demonstrated using high-end instrumentation, this assay could be used as a simpler analytical solution to focus on the 'diagnostic' glycan feature.

## Conclusion

We developed a prototype for a simple glycosidase plate-based assay for the quantification of galactosylation and sialylation traits on intact human IgG glycoproteins. The measured galactosylation and sialylation were normalised for the natural variation of IgG amounts in plasma by using the measured IgG glycoprotein amounts (absolute quantification) or by using the ratio of the glycan traits itself (galactosylation index). The latter has the advantage of lower intermediate RSDs (<5%) which has importance in comparing clinically relevant samples, whilst the former, which have much higher RSDs (~20%), has importance in characterization of glycoproteins such as biopharmaceuticals, since it provides information of the moles of galactose or sialic acid per mole of glycoprotein. Finally, we demonstrated the performance of the assay on clinically relevant human patient serum samples and obtained a Pearson's *r* correlation coefficient of 0.8208 ( $p = 0.0002$ ) on comparison of the measured galactosylation index to an industrially established HILIC-FLD-MS<sup>n</sup> platform. The next step would be to demonstrate the assay on a much large clinically relevant patient cohort. Furthermore, it will be interesting to investigate the versatility of this method for the quantification of galactosylation and sialylation on other immunoglobulins such as IgA and IgM and/or also non-human mammalian immunoglobulins that may contain different patterns of galactosylation and sialylation. Finally, it must be stated that although the optimized assay was performed with non-commercial exoglycosidases, there is no reason for the assay not be performed with other such exoglycosidases which may or may not be commercially sourced.

## Supplementary information

Supplementary tables and figures are available online free of charge via

<https://link.springer.com/article/10.1007/s10719-020-09953-9#Sec15>

## Acknowledgements

This research was supported by the European Union (GlySign, Grant No. 722095) and a BBSRC/Innovate UK IB catalyst award 'Glycoenzymes for Bioindustries' (BB/M029018/1) to DB. Human patient serum samples (IBD-BIOM) were provided Jack Satsangi (University of Oxford) and Alex Adams (University of Oxford). Ethical approval for the collection of IBD-BIOM patient samples was granted by the Tayside Committee on Medical Ethics B (UK, LREC 06/S1101/16, LREC 2000/4/192). The study has been funded by the following EU FP7 grant: European Commission IBD-BIOM (contract # 305479).



## References

1. Bondt, A., et al., Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res*, 2013. 12(10): p. 4522-31.
2. Sun, D., et al., Distribution of abnormal IgG glycosylation patterns from rheumatoid arthritis and osteoarthritis patients by MALDI-TOF-MS(n). *Analyst*, 2019. 144(6): p. 2042-2051.
3. Vuckovic, F., et al., Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. *Arthritis Rheumatol*, 2015. 67(11): p. 2978-89.
4. Lemmers, R.F.H., et al., IgG glycan patterns are associated with type 2 diabetes in independent European populations. *Biochim Biophys Acta Gen Subj*, 2017. 1861(9): p. 2240-2249.
5. Shinzaki, S., et al., Lectin-based immunoassay for aberrant IgG glycosylation as the biomarker for Crohn's disease. *Inflamm Bowel Dis*, 2013. 19(2): p. 321-31.
6. Nairn, A.V., et al., Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem*, 2008. 283(25): p. 17298-313.
7. Blomme, B., et al., Alteration of protein glycosylation in liver diseases. *J Hepatol*, 2009. 50(3): p. 592-603.
8. Jennewein, M.F. and G. Alter, The Immunoregulatory Roles of Antibody Glycosylation. *Trends Immunol*, 2017. 38(5): p. 358-372.
9. Lippold, S., et al., Glycoform-resolved FcRIIIa affinity chromatography-mass spectrometry. *MAbs*, 2019. 11(7): p. 1191-1196.
10. Krapp, S., et al., Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol*, 2003. 325(5): p. 979-89.
11. Ferrara, C., et al., Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proceedings of the National Academy of Sciences of the United States of America*, 2011. 108(31): p. 12669-12674.
12. Karsten, C.M., et al., Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of FcγRIIB and dectin-1. *Nat Med*, 2012. 18(9): p. 1401-6.
13. Thomann, M., et al., Fc-galactosylation modulates antibody-dependent cellular cytotoxicity of therapeutic antibodies. *Mol Immunol*, 2016. 73: p. 69-75.
14. Kaneko, Y., F. Nimmerjahn, and J.V. Ravetch, Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science*, 2006. 313(5787): p. 670-3.
15. Anthony, R.M., et al., Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science*, 2008. 320(5874): p. 373-6.
16. Malhotra, R., et al., Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med*, 1995. 1(3): p. 237-43.
17. Lageveen-Kammeijer, G.S.M., et al., Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. *Nature Communications*, 2019. 10(1): p. 2137.

18. Royle, L., et al., HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Analytical Biochemistry*, 2008. 376(1): p. 1-12.
19. Stumpo, K.A. and V.N. Reinhold, The N-Glycome of Human Plasma. *Journal of Proteome Research*, 2010. 9(9): p. 4823-4830.
20. Trbojevic Akmacic, I., et al., Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm Bowel Dis*, 2015. 21(6): p. 1237-47.
21. Freidin, M.B., et al., The Association Between Low Back Pain and Composition of IgG Glycome. *Sci Rep*, 2016. 6: p. 26815.
22. Mahan, A.E., et al., A method for high-throughput, sensitive analysis of IgG Fc and Fab glycosylation by capillary electrophoresis. *J Immunol Methods*, 2015. 417: p. 34-44.
23. Vanderschaeghe, D., et al., Endoglycosidase S Enables a Highly Simplified Clinical Chemistry Procedure for Direct Assessment of Serum IgG Undergalactosylation in Chronic Inflammatory Disease. *Mol Cell Proteomics*, 2018. 17(12): p. 2508-2517.
24. Bondt, A., et al., Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Molecular & cellular proteomics : MCP*, 2014. 13(11): p. 3029-3039.
25. Gebrehiwot, A.G., et al., Exploring serum and immunoglobulin G N-glycome as diagnostic biomarkers for early detection of breast cancer in Ethiopian women. *BMC Cancer*, 2019. 19(1): p. 588.
26. Huhn, C., et al., IgG glycosylation analysis. *Proteomics*, 2009. 9(4): p. 882-913.
27. Haab, B.B., et al., Glycosylation variants of mucins and CEACAMs as candidate biomarkers for the diagnosis of pancreatic cystic neoplasms. *Ann Surg*, 2010. 251(5): p. 937-45.
28. Tang, H., et al., Glycan motif profiling reveals plasma sialyl-lewis x elevations in pancreatic cancers that are negative for sialyl-lewis A. *Mol Cell Proteomics*, 2015. 14(5): p. 1323-33.
29. Yue, T., et al., The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol Cell Proteomics*, 2009. 8(7): p. 1697-707.
30. Gornik, O. and G. Lauc, Enzyme linked lectin assay (ELLA) for direct analysis of transferrin sialylation in serum samples. *Clinical Biochemistry*, 2007. 40(9): p. 718-723.
31. Shang, Y., Y. Zeng, and Y. Zeng, Integrated Microfluidic Lectin Barcode Platform for High-Performance Focused Glycomic Profiling. *Sci Rep*, 2016. 6: p. 20297.
32. Chen, S., et al., Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nat Methods*, 2007. 4(5): p. 437-44.
33. Keusch, J., et al., Analysis of different glycosylation states in IgG subclasses. *Clinica Chimica Acta*, 1996. 252(2): p. 147-158.
34. Briliūtė, J., et al., Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology*, 2019. 4(9): p. 1571-1581.
35. Park, K.-H., et al., Structural and biochemical characterization of the broad substrate specificity of *Bacteroides thetaiotaomicron* commensal sialidase. *Biochimica et biophysica acta*, 2013. 1834(8): p. 1510-1519.

36. 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*, 2015. 10(4): p. e0123028.
37. Kozak, R.P., et al., Comparison of procainamide and 2-aminobenzamide labeling for profiling and identification of glycans by liquid chromatography with fluorescence detection coupled to electrospray ionization–mass spectrometry. *Analytical Biochemistry*, 2015. 486: p. 38-40.
38. Gonzalez-Quintela, A., et al., Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. *Clinical and experimental immunology*, 2008. 151(1): p. 42-50.
39. French, M., Serum IgG subclasses in normal adults. *Monogr Allergy*, 1986. 19: p. 100-7.
40. Nagae, M. and Y. Yamaguchi, Function and 3D structure of the N-glycans on glycoproteins. *International journal of molecular sciences*, 2012. 13(7): p. 8398-8429.
41. Srinivasan, K., et al., A Quantitative Microtiter Assay for Sialylated Glycoform Analyses Using Lectin Complexes. *Journal of biomolecular screening*, 2015. 20(6): p. 768-778.
42. Plomp, R., et al., Hinge-Region O-Glycosylation of Human Immunoglobulin G3 (IgG3). *Molecular & cellular proteomics : MCP*, 2015. 14(5): p. 1373-1384.





# **Chapter 6**

## **Discussion and perspectives**



Glycomics has proven to be valuable in clinical and medical research. Associations with disease states complement and occasionally supersede the diagnostic potential of proteomics, metabolomics or genomics [1-4]. Sometimes, glycomics is also being developed as an alternative to invasive examinations, such as in the case of prostate [5, 6] and colorectal [7, 8] cancers. However, the analysis is often complicated by the complexity of the glycan structures which show variations in both composition and linkages of the constituent monosaccharide residues [9, 10]. The major human plasma proteins are estimated to contain at least 167 *N*-glycan structures which complicates meaningful associations of disease states to glycomics data [11, 12]. The identification and quantification of these glycans are usually performed using chromatographic and spectrometric techniques [12-15]. The compilation of glycan motifs and epitope features into so-called glycan traits has enabled an improved comparison of variations in glycosylation between patients [13, 16, 17]. These academic research successes have created an interest in glycomics as a clinical prognostic assay. Accordingly, contract research organizations as well as assay developers have entered the field of glycomics [18-21]. This thesis addresses this commercial interest by developing assays for routine industrial applications with commercial potential for the development into off-the-shelf, on-site products.

## Analytical techniques in glycomics

In clinical glycomics, development of analytical techniques and medical knowledge is co-dependent. Advances in separation and identification technologies have allowed for understanding and confirming the complex mechanism of glycosylation in a living entity [4, 22, 23]. The more we understand the biology behind glycosylation, the more demand in advanced analytical techniques is developing to further push the boundaries of our understanding. Early glycomics assays were based on lectin blots which utilise the structure-specific binding activities of lectins to identify certain glycan motifs [24-26]. Around the same time, a move into performing glycomics assays on liquid chromatography (LC) - fluorescence detector (FLD) and pulsed amperometric detection (PAD) platforms gained momentum. In fact, LC-based assays still remain the main workhorses of industrial and academic glycomics. They are applied, for example, for the glycosylation characterization of individual proteins and complex biological samples, such as TPNG [13, 27, 28]. Analyte identification using this technique is based on retention times which are normalised to a glucose polymer ladder or other glycan standards which enables data set comparisons [29]. For the elucidation of glycan structures, a sequential process of exoglycosidase digestion is performed before the LC analysis [29]. By correlating the shifts in peaks between the digestion regimes, the structure of a glycan can be built up [29, 30]. Although these chromatographic techniques have been exceptionally successful in glycomics, they can sometimes be limited by their molecular resolution. This holds true for hydrophilic interaction liquid chromatography (HILIC) which is one of the most prominent LC methods used in glycomics assays. The separation performance of HILIC methods depends on the analytes partitioning between the mobile phase layers created on the polar stationary phase [31]. However, *N*-glycans have a characteristic pentasaccharide core structure, with variations mainly in the antennary branching and decorations of galactosylation, sialylation and fucosylation [9, 32]. Thus, minor differences



between glycans, such as a different linkage of a monosaccharide, may result in only minute differences of physicochemical properties making their HILIC separation challenging. This may result in several glycan structures co-eluting under a single chromatographic peak. Despite this limitation, HILIC analysis has time and time again proven to be successful in drawing associations to certain disease states in patients [15, 33]. Many of the analyses rely on comparison of the relatively abundance of the chromatographic peaks rather than the individual glycans that form a peak [34].

However, the analytical questions posed in cutting-edge medical research and biopharmaceutical development, require detailed analytical answers regarding e.g. the structure of individual glycans and their quantities [35-39]. Thus, separation techniques with improved resolutions such as multimodal LC [40], capillary electrophoresis (CE) [11] or capillary gel electrophoresis (CGE) [41] have been explored. LC and CE can be coupled to mass spectrometry (MS) devices, mainly via electrospray ionisation (ESI) – MS, which provides a second analytical dimension namely  $m/z$  identification. Thus, even if a chosen separation technique is unable to differentiate between glycan structures, a mass spectrum obtained may allow to identify and quantify these structures. Furthermore, improvements in ion optics over the past decades have allowed for improvements in selective gas phase fragmentations that enable structural sequencing of the analytes.

## **Existing glycomics assays**

The importance of glycomics is growing along with our understanding of the correlations of glycomics changes to disease progression. As a result, a growing interest for glycomics is created in industry. There is only a handful of analytical chemistry companies that specialise in glycomics analysis of clinical and biopharmaceutical samples. Many of these industrial applications rely on LC and MS-based methods that screen an entire glycome. However, due to the complexity of glycomes, such as TPNG, often structures that have clinical or medical importance are unresolved. The work of this thesis primarily addresses this issue by developing novel glycomics assays that are designed to specifically quantify targeted epitopes in TPNG or IgG glycome samples. These assays thus enhance quantification of certain important glycan epitopes such as antennary fucosylation which are normally tedious to quantify by the commonly used LC-FLD/MS methods in industry. Besides, the industrial glycomics services, some of these organisations are also involved in developing kits for glycomics assays. However, many of these analytical kits are made for sample preparation such as reducing end labelling [42, 43] and methylation [44] of glycans for analysis on LC/CE/CGE-FLD/MS platforms [45-48]. The work of this thesis addresses the demand for innovative kit development by introducing an assay which does not require the latter analytical platforms but rather a spectrophotometric plate-reader. These assays could hence meet demand in laboratories that are not specialised in routine glycomics analysis. Especially since now-a-days many laboratories embark on cross-disciplinary research topics, such analytical glycomics kits could serve as a quick and efficient solution for attempting to address a glycomics research question.

## Simple glycomics assays

Glycomics is heavily dependent on chromatography, electrophoresis and MS. However, unlike proteomics and genomics, not much investigation into alternative non-chromatographic assays has been done [49, 50]. It is true that detailed analytical questions in proteomics and genomics require some form of separation. However, many assays have been created as an alternative to answer certain niche and targeted analytical questions without the need for chromatography. For instance, enzyme linked immuno sorbent assay (ELISA) are often used in proteomics and quantitative polymerase chain reactions (qPCR) assays are used in genomics to quantify the abundance of certain proteins or DNA epitopes in a sample, respectively. Such plate-based assays have a plethora of advantages which will be discussed along the span of this section. An established alternative to chromatographic /electrophoretic/mass spectrometric glycomics assays for complex samples, such as TPNG, is the lectin microarray assay [51, 52]. Similar to an ELISA format, an analyte of interest is immobilised and quantified based on the binding specificity of a detection probe.

However, these microarrays are not widely adopted due to the technical skills required for assay establishment as well as multiple sample processing steps [53, 54]. Moreover, performance of the assay is heavily dependent on specificity and binding affinity of the lectin probes. A few research groups have developed another sort of a plate-based assay that is designed around redox reactions of monosaccharides units to produce a colorimetric signal [55, 56]. Building from this research, this thesis introduces the development of an assay using enzyme-based redox reactions for the quantification of glycan epitopes. In general, analytes and reactants are mixed within a microtitre plate and the colorimetric signal is measured on a spectrophotometric plate-reader [55, 56]. Thus, these assays do not require the tedious process of immobilising antibodies/lectins to the microarray. The immobilisation is a common source of variability arising from variations in immobilised lectin density and stoichiometry [54].

## Critical aspects for assay development targeting industrial application

### One-pot reactions

Analytical chemistry assays used in industrial applications should preferably have a streamlined workflow with minimal hands-on time. There are several reasons for these requirements, but the most important one is the operator's ability to perform an assay with minimal training. Therefore, the assays consist of a minimal number of steps; hence one/two-pot reactions are desirable. This also reduces the necessary laboratory cost on consumables and waste. Such one pot assays usually also allow for an easy transfer of the assay to a fluidic handling robotic platforms which aids in throughput. Furthermore, such assays usually allow for concise comprehensive standard operating protocols. In chemistry, compiling several reaction steps into a one-pot reaction procedure is seldom easy, due to possible interferences between the constituents required for the different reaction steps. However, many biochemical assays such as ELISA and qPCR are based on one-pot biocatalysed enzyme-based

reactions [57, 58]. Due to the specificity of enzymes to catalyse certain reaction steps, several steps can be compiled into a one-pot reaction with little to no interference of intermediate reaction steps. However, since enzymes are functional proteins, a deviation from their suitable reaction conditions can limit their activity [59]. Thus, when designing such one-pot reactions with multiple enzyme catalysed steps, the physio-chemical conditions of the reaction should strike a balance between the efficiencies of the constituent enzyme activities so as not to limit the overall efficiency of the assay [60]. In case of drastic differences in the required physico-chemical conditions of the different constituent reaction steps, the individual reactions of the one-pot reaction can be performed in a sequential manner. Chapter 5 demonstrates such a sequential one-pot assay. Exoglycosidases, dehydrogenases and an epimerase were the enzyme biocatalysts used in this multi-step one-pot reaction spectrophotometric glycomics assay. By performing several reaction steps sequentially with increasing reaction volumes, tolerable biochemical conditions for enzymes were achievable whilst obvious interferences between reactions steps were limited.

## **Instrumentation and data**

Assays based on spectrophotometric plate-readers are advantageous when it comes to the overall infrastructure and costs of an assay. These plate-readers usually cost much less than LC, CE and MS platforms, and often do not require extensive regular maintenance or skilled technicians. Modern plate-readers are also compact benchtop instruments that require much less laboratory space and no gas lines or vents as compared to LC, CE or MS instruments. Thus, a dedicated analytical service laboratory can furnish many such plate-readers easily, if demand increases. Additionally, the machine time for data analysis of a sample in a spectrophotometric plate-reader is only a fraction of the time required by separation-based platforms, which increases throughput. Their ease-of-use is highly desired for industrial analytical application, e.g enabling quick training of technical staff.

The identification and quantification ability of LC and CE techniques is based on their ability to separate analytes either in time or space irrespective of the detection system used. This is usually in contrast to assays based on spectrophotometric plate-readers. A sample containing one or more analytes of interest is placed in a single well of a microtitre plate, and these analytes are exclusively selected based on their spectrophotometric profile of the absorbed and/or emitted light. Thus, the identification and quantification ability of a spectrophotometric plate-reader assay is limited to the differences in the spectrophotometric profile of the analytes in the sample, i.e. analytes with similar or overlapping spectrophotometric profiles cannot be differentiated [61, 62]. However, physical chemical approaches using quantum dots and Förster resonance energy transfer (FRET) are being investigated to improve the multiplexing of spectrophotometric assays based on time dependence and narrow bandwidths of spectrophotometric profiles [63, 64].

LC, CE and MS quantification often rely on relative area comparison of the integrated peaks in a chromatogram, electropherogram and mass spectrum, respectively. In contrast, for spectrophotometric plate-reader assays, an absolute quantification is performed based on comparison to a standard curve which is a concentration series of the analyte of interest. The

latter can be limiting when it comes to the analysis of some biological molecules, due to the availability of standards. For example, the quantification of a glycan epitope on a specific glycoprotein is limited by the fact that pure protein glycoform standards hardly exist. When only the relative biological variation of the glycan epitopes is clinically relevant, then measuring absolute glycan epitope amounts would also incorporate the biological variation of the glycoprotein expression which may not necessarily have clinical relevance. In this case, the quantified glycan epitope would need to be normalised for the clinically irrelevant variations of the glycoprotein. One approach is to take the ratio of the glycan epitope amounts to the protein amounts, wherein the glycan epitopes and protein amounts are firstly quantified against their respective molecular standard curves. Alternatively, a relative quantification approach can also be performed by taking a ratio of two glycan epitopes that may be structurally inter-dependent in the parent biomolecule. The added benefit of a normalisation approach using the ratio of epitopes quantified by similar chemical and detection mechanisms, is the elimination of the need for a standard curve to compare between samples. For example, taking the ratio of galactosylation and sialylation for a glycoprotein can eliminate variations coming from the glycoprotein amounts of different samples. This was exemplified in Chapter 5.

### **Other research aspects**

The assays introduced in this thesis were developed with the intention of easily transforming them into analytical kits. Besides being potential products for use in established industrial and academic laboratories, a major advantage of these kit-based assays is the portability for in-field research applications. Such kits and spectrophotometric plate-readers can be easily transported and do not require laboratory infrastructure such as gas lines and vents. Furthermore, the use of light emitting diodes (LEDs) with wavelengths of interest rather than white-light sources, has allowed for a scaling down of plate readers to the size of booklets (Absorbance 96, Byonoy GmbH, Germany) and hence has further improving portability [65].

In-field research has its advantages as it saves time for answering an analytical question but more importantly it removes the need to transport temperature-sensitive biological samples to an analytical laboratory which is not always located near a site of the field research. It must be mentioned that the assay in Chapter 5 was performed almost entirely on a Hamilton robotic system, but there is no reason why it cannot be performed similarly with multichannel pipettes, which was the case during its development.

## Identification options in (LC-)MS assays

### Fragmentation and glycan epitope identification

The complexity of glycomics partially derives from the linkage variants of residues. However, obtaining information on the exact loci of these residues within a molecule can further increase the complexity of glycomics analysis. Considering only linkage variations, often two or more linkage variants of a residue can occur on the same molecule. Sometimes, these linkage iterations cannot be resolved with a separation method. Thus, a coupled MS dimension helps to derive the structural features by gas phase fragmentation which, in glycomics, is usually collision induced dissociation (CID) [66]. Most MS based glycomics assays are performed in positive ion mode due to the higher (compared to negative ion mode) ionisation efficiency [66]. However, CID in positive ion mode results in glycosidic bond breaks which may not always be sufficiently informative, especially in the case of co-eluting linkage variants [67]. Furthermore, complete coverage of molecular fragmentation patterns is not always obtainable. However, if we consider *N*-glycosylation in humans, the linkage variants of clinical interest are mostly of fucoses (Fuc), *N*-acetylglucosamines (GlcNAc) and sialic acids (Sia), which are almost entirely *N*-acetylneuraminic acid. Linkage variants of galactose (Gal) and mannose can also be found. Each of these residues has its own generally accepted manner of identification. In case of Fuc, diagnostic fragment ions representing parts of the glycan core structure are used for identifying its  $\alpha(1-6)$  or  $\alpha(1-3/4)$  variants which are exclusively core or antennary located, respectively [68]. Similar diagnostic ions representing part of the core can be used for differentiating between  $\beta(1-4)$  bisecting GlcNAc and  $\beta(1-2/4/6)$  antennary GlcNAcs [69]. However, differentiating between the linkage variants of antennary GlcNAcs by CID fragmentation is particularly difficult.

In the previous paragraph, we only considered the identification of linkage variants without considering the loci of these variants within a molecule. It is worth mentioning that glycans having the same composition including their constitutional isomers, may not necessarily have the same structure because the residues of antennary GlcNAc, Gal, antennary Fuc and Sia can be assorted onto different antennary arms. The possibility of the presence of such regioisomers becomes especially concerning in glycans with multiple antennae. CID in positive ion mode is unable to provide sufficient information to elucidate the position of such residues. However, CID in negative ion mode results in sequential cross-ring fragmentation patterns from which such information can be derived [67]. This makes this method a very powerful tool in glycomics analysis. However, the drawback is that ESI of glycans in negative ion mode suffers from low ionisation efficiencies. Furthermore, fragmentation data is not always easily comprehensible and hence negative ion mode analysis is not widely practiced in glycomics.

Of special relevance to this thesis are the diagnostic fragment ions that are used for differentiating the linkage variants of Fuc residues due to their core or antennary localisation on the glycan structure. However, these diagnostic ions are not conclusive due to the phenomenon of Fuc rearrangement. The  $\alpha(1-3/4)$  antennary Fuc residues have a tendency to undergo a gas phase molecular rearrangement to form isomeric ions, mimicking core fragment ions [70]. Considering that antennary fucosylation in TPNG has been associated with several cancers [27, 71], a demand for its accurate quantification has been created. In aid of

improving the accuracy of antennary Fuc quantification, Chapter 4 takes a different approach to add selectivity. The novelty of this assay lies in its ability to accurately quantify antennary fucosylation in complex TPNG samples without the use of MS fragmentation. The assay relies on an exoglycosidase depletion of core Fuc residues so as to obtain information on the antennary Fuc residues by matrix assisted laser desorption ionisation (MALDI)-time of flight (TOF)-MS. This allows for the calculation of antennary Fuc traits without the need for complex and sometimes inconclusive MS/MS fragmentation experiments. Thus by taking advantage of the selectivity of exoglycosidase, the need for technical and experimental complexity can be reduced in glycomics. This approach has potential in differentially branched antennary GlcNAc isomer quantification where MS/MS experiments required for their identification are very complex. This is why a portion of the work of this thesis was dedicated to characterising the substrate specificities of exoglycosidase, which could have potential not only for glycan profiling but also for similar aforementioned exoglycosidase-based analytical assays. More of this will be discussed in the later section.

It is also worth mentioning, that when it comes to analytical speed, MALDI-MS methods are faster than LC-ESI-MS methods on which most antennary Fuc assays are based. This makes the assay even more attractive for an industrial application as throughput is often vital.

## Derivatization

Few years ago, identification of Sia linkages by fragmentation was very challenging due to its terminal location at the non-reducing end and lack of differentiating diagnostic ions. However, derivatisation methods such as methyl esterification [72] and ethyl esterification, and its adaptations [73-75], have greatly enhanced identification of the most prominent  $\alpha(2-3/6)$  Sia variants. During these esterification reactions, a methyl/ethyl ester is preferably formed on the C1 -COOH of  $\alpha(2-6)$  Sia rather than  $\alpha(2-3)$  Sia, due to their differential relative proximity to the C2 -OH of the penultimate galactose (Gal) residue [73]. This reaction selectivity causes a differential mass shift between both linkage variants. As a result the composition of the Sia variants within a structure can be identified even without the need for fragmentation [76]. Such chemical derivatisation approaches have not been investigated for Fuc and bisecting/antennary GlcNAc residues, and could possibly be very tedious due to the lack of unique and suitable chemically active functional groups in these residues. Thus, exoglycosidase-based glycan profiling still remains the industrial gold-standard for identifying Fuc and GlcNAc residue linkages. Hence, obtaining fucosidases and GlcNAc-ases of isomer specificity is important.

## **Exoglycosidases for analytical use**

The performance of the analytical assays developed in this thesis as well as other glycoprofiling techniques are highly dependent on the specificity of the exoglycosidases used. Hence, it is important that these glycosidases are well characterised in terms of substrate specificity so as to ensure high accuracy in quantification of isomeric analytes. Glycosidases have an ever growing demand in bioindustry and academic research applications. Hence, many of these glycosidases are readily obtainable from commercial vendors as recombinant or naturally sourced products. Furthermore, molecular biologist with experience in gene expression can produce these recombinant glycosidases in-house. However, the source of the glycosidases is highly dependent on the user preferences and reliability of the enzymes available. This can be exemplified by the choice of the sources of glycosidases used in the assays developed in this thesis. These glycosidases are either commercially available or can be produced readily by molecular biologist in-house as recombinant proteins. Commercially available enzymes are usually well characterised for specificity and kinetic activity and hence ideal for assay development as it removes the need for enzyme characterisation. In an academic sense, availability of commercially sourced enzymes also allows the assay to be reproduced readily by other research groups that are interested in the topic. This was one of the underlying motives in Chapter 4, where a commercial core fucosidase was used rather than the non-commercially sourced recombinant core fucosidases which were also available. However, commercial enzymes can sometimes be quite costly. Since the assay was intended for cost effective high-throughput analysis, the use of in-house sourced recombinant exoglycosidases was ideal rather than the more expensive commercial options. Many of these exoglycosidases can also be inexpensively commercially available as natural sourced enrichments. These naturally sourced enzymes usually have one or more purification steps for enriching the enzyme of interest. Since these purification steps are often limited in purification efficiency, glycan and/or other exoglycosidase contaminants could introduce variations in the assay when using different batches of the enzymes. Hence to avoid such variations, recombinant enzymes of high purity should be used.

## **Exoglycosidases of potential analytical interest**

Although there is a large repertoire of available glycosidases, there is still a demand for certain glycosidase activities which can be based on specificity to a certain linkage or on a certain glycan structure. There will always be a demand for obtaining glycosidases of certain specificity which depends on the needs of a glycomics assay or a glycobiological application. Enzymology is a large research field and so to convey the importance of glycosidases in glycomics, I shall exemplify from the work done in this thesis on an antennary fucosidase (Chapter 2) and a bisecting GlcNAc specific GlcNAcase ([77], Chapter 3). These enzymes were chosen for screening due to difficulties in identifying their substrate glycan isomers by analytical glycoprofiling techniques. As mentioned above, identification of isomeric fucose residues on *N*-glycans is tedious due to unreliable diagnostic fragment ions. Similar diagnostic fragments ions of bisecting GlcNAc isomers are usually present in low abundance along with the low occurrence of these bisecting glycans in TPNG [78, 79].

The antennary fucosidase here specifically removes  $\alpha(1-3/4)$ -linked antennary Fuc residues irrespective of sialylation on the Lewis X epitope. This activity is unique, as all reported antennary fucosidases are hindered by a sialylation on these epitopes. This fucosidase has immense potential in LC-based glycomics assay for antennary fucose identification. The mechanism of this activity was elucidated relying on structural characterisation of the enzyme. The presence of an open void in the vicinity of the active site allowed for the unhindered accommodation of a  $\alpha(2-3)$ -linked Sia residue which was not the case for other previously reported antennary fucosidases. The information obtained from this chapter will go on to aid efforts in genetic engineering of other antennary fucosidases with hope of introducing such a favourable trait. Such approaches of engineering desirable traits into enzymes have repeatedly proven valuable. A relevant example, the incorporation of a predominant transfucosidase activity into the  $\alpha$ -1,3/4-L-fucosidase BbAfcB from *Bifidobacterium bifidum* was achieved by swapping an  $\alpha$ -helical loop with the transfucosidase CpAfc2 from *Clostridium perfringens* [80]. This has important interest in industrial production of human milk oligosaccharides since the desirable activities from a pathogen derived enzyme was incorporated into a probiotic enzyme, thus avoiding possible controversy.

A GlcNAcase was identified that is specific to bisecting GlcNAc residues which gives it a novelty among known GlcNAcases [77]. However, this enzyme was highly limited in its activity as it gets hindered by the presence of galactosylation and sialylation. Chapter 3 is an attempt to understand the mechanism of the enzyme in more detail. Although the mechanism of antennary GlcNAcase catalysis [81] is well known, the mechanism behind its specificity to bisecting GlcNAc residues remains a mystery due to the lack of co-crystallised glycans within its active site. Often these GlcNAcases rely on associated residues encompassing the entry to the active site to regulate the positioning of glycan's antennary arms or bisection [77, 82]. Such previously described regulatory residues [77, 82] could not be identified within this structure. Furthermore, attempts at computer-assisted docking of the glycan into the active site were not conclusive. Follow-up research would be required to understand the regulatory mechanism of the enzyme specificity.

## On the importance of interdisciplinary research

Finally, I would like to end this chapter by saying that glycomics like so many other sectors of analytical chemistry, heavily relies on interdisciplinary research. Analytical chemistry projects are heavily dependent on instrumentation and assay development, and thus we require a constant input from other sciences to develop and advance our methods. This brings me to the importance of interdisciplinary collaboration in the research of this thesis as it would not be possible without a consortium-like investigation by analytical chemists, biochemists, structural biologists, and process engineers for commercialisation of potential future products.



## References

1. Zoldoš, V., T. Horvat, and G. Lauc, Glycomics meets genomics, epigenomics and other high throughput omics for system biology studies. *Current Opinion in Chemical Biology*, 2013. 17(1): p. 34-40.
2. Perakakis, N., et al., Non-invasive diagnosis of non-alcoholic steatohepatitis and fibrosis with the use of omics and supervised learning: A proof of concept study. *Metabolism*, 2019. 101: p. 154005.
3. Taniguchi, N., et al., The Second Golden Age of Glycomics: From Functional Glycomics to Clinical Applications. *Journal of Proteome Research*, 2009. 8(2): p. 425-426.
4. Durand, G.v. and N. Seta, Protein Glycosylation and Diseases: Blood and Urinary Oligosaccharides as Markers for Diagnosis and Therapeutic Monitoring. *Clinical Chemistry*, 2000. 46(6): p. 795-805.
5. Tkac, J., et al., Glycomics of prostate cancer: updates. *Expert Review of Proteomics*, 2019. 16(1): p. 65-76.
6. Kyselova, Z., et al., Alterations in the Serum Glycome Due to Metastatic Prostate Cancer. *Journal of Proteome Research*, 2007. 6(5): p. 1822-1832.
7. Snyder, C.M., et al., Complementary Glycomic Analyses of Sera Derived from Colorectal Cancer Patients by MALDI-TOF-MS and Microchip Electrophoresis. *Analytical Chemistry*, 2016. 88(19): p. 9597-9605.
8. Sethi, M.K., et al., In-depth N-glycome profiling of paired colorectal cancer and non-tumorigenic tissues reveals cancer-, stage- and EGFR-specific protein N-glycosylation. *Glycobiology*, 2015. 25(10): p. 1064-1078.
9. Maverakis, E., et al., Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *Journal of Autoimmunity*, 2015. 57: p. 1-13.
10. Peng, W., et al., Clinical application of quantitative glycomics. *Expert Review of Proteomics*, 2018. 15(12): p. 1007-1031.
11. Lageveen-Kammeijer, G.S.M., et al., Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. *Nature Communications*, 2019. 10(1): p. 2137.
12. Adamczyk, B., et al., High-Throughput Analysis of the Plasma N-Glycome by UHPLC. *Methods Mol Biol*, 2017. 1503: p. 97-108.
13. Trbojević Akmačić, I., et al., Inflammatory Bowel Disease Associates with Proinflammatory Potential of the Immunoglobulin G Glycome. *Inflammatory Bowel Diseases*, 2015. 21(6): p. 1237-1247.
14. Saldova, R., et al., Association of Medication with the Human Plasma N-Glycome. *Journal of Proteome Research*, 2012. 11(3): p. 1821-1831.
15. Dotz, V. and M. Wuhler, N-glycome signatures in human plasma: associations with physiology and major diseases. *FEBS Letters*, 2019. 593(21): p. 2966-2976.
16. Bladergroen, M.R., et al., Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *Journal of Proteome Research*, 2015. 14(9): p. 4080-4086.

17. 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 <sup>\*</sup>. *Molecular & Cellular Proteomics*, 2017. 16(2): p. 228-242.
18. Ruhaak, L.R., et al., Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *Journal of Proteome Research*, 2011. 10(4): p. 1667-1674.
19. Gudelj, I., et al., Estimation of human age using N-glycan profiles from bloodstains. *International Journal of Legal Medicine*, 2015. 129(5): p. 955-961.
20. Rebello, O.D., et al., A novel glycosidase plate-based assay for the quantification of galactosylation and sialylation on human IgG. *Glycoconjugate Journal*, 2020. 37(6): p. 691-702.
21. Gudelj, I., G. Lauc, and M. Pezer, Immunoglobulin G glycosylation in aging and diseases. *Cellular Immunology*, 2018. 333: p. 65-79.
22. Ohtsubo, K. and J.D. Marth, Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*, 2006. 126(5): p. 855-867.
23. Hennet, T., Diseases of glycosylation beyond classical congenital disorders of glycosylation. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 2012. 1820(9): p. 1306-1317.
24. Zhao, J., et al., Glycoprotein Microarrays with Multi-Lectin Detection: Unique Lectin Binding Patterns as a Tool for Classifying Normal, Chronic Pancreatitis and Pancreatic Cancer Sera. *Journal of Proteome Research*, 2007. 6(5): p. 1864-1874.
25. Etxebarria, J., et al., Lectin-Array Blotting: Profiling Protein Glycosylation in Complex Mixtures. *ACS Chemical Biology*, 2012. 7(10): p. 1729-1737.
26. Qiu, Y., et al., Plasma Glycoprotein Profiling for Colorectal Cancer Biomarker Identification by Lectin Glycoarray and Lectin Blot. *Journal of Proteome Research*, 2008. 7(4): p. 1693-1703.
27. Doherty, M., et al., Plasma N-glycans in colorectal cancer risk. *Scientific Reports*, 2018. 8(1): p. 8655.
28. Ventham, N.T., et al., Changes to Serum Sample Tube and Processing Methodology Does Not Cause Inter-Individual Variation in Automated Whole Serum N-Glycan Profiling in Health and Disease. *PLOS ONE*, 2015. 10(4): p. e0123028.
29. Royle, L., et al., HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Analytical Biochemistry*, 2008. 376(1): p. 1-12.
30. Gotz, L., et al., GlycoDigest: a tool for the targeted use of exoglycosidase digestions in glycan structure determination. *Bioinformatics*, 2014. 30(21): p. 3131-3133.
31. Hao, Z., B. Xiao, and N. Weng, Impact of column temperature and mobile phase components on selectivity of hydrophilic interaction chromatography (HILIC). *Journal of Separation Science*, 2008. 31(9): p. 1449-1464.
32. Lannoo, N. and E.J.M. Van Damme, Review/N-glycans: The making of a varied toolbox. *Plant Science*, 2015. 239: p. 67-83.

33. Saldova, R., et al., Core fucosylation and  $\alpha$ 2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology*, 2011. 21(2): p. 195-205.
34. Novokmet, M., et al., Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Scientific Reports*, 2014. 4(1): p. 4347.
35. Harvey, D.J., et al., Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Analytical Biochemistry*, 2008. 376(1): p. 44-60.
36. Pagel, K. and D.J. Harvey, Ion Mobility–Mass Spectrometry of Complex Carbohydrates: Collision Cross Sections of Sodiated N-linked Glycans. *Analytical Chemistry*, 2013. 85(10): p. 5138-5145.
37. Hua, S., et al., Technologies for glycomic characterization of biopharmaceutical erythropoietins. *TrAC Trends in Analytical Chemistry*, 2015. 68: p. 18-27.
38. Szabo, Z., et al., High Performance Anion Exchange and Hydrophilic Interaction Liquid Chromatography Approaches for Comprehensive Mass Spectrometry-Based Characterization of the N-Glycome of a Recombinant Human Erythropoietin. *Journal of Proteome Research*, 2018. 17(4): p. 1559-1574.
39. Lageveen-Kammeijer, G.S.M., et al., High sensitivity glycomics in biomedicine. *Mass Spectrometry Reviews*, 2021. n/a(n/a): p. e21730.
40. Stavenhagen, K., et al., Site-Specific N- and O-Glycopeptide Analysis Using an Integrated C18-PGC-LC-ESI-QTOF-MS/MS Approach, in *High-Throughput Glycomics and Glycoproteomics: Methods and Protocols*, G. Lauc and M. Wuhrer, Editors. 2017, Springer New York: New York, NY. p. 109-119.
41. 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. *Journal of Proteome Research*, 2010. 9(12): p. 6655-6664.
42. Kozak, R.P., et al., Comparison of procainamide and 2-aminobenzamide labeling for profiling and identification of glycans by liquid chromatography with fluorescence detection coupled to electrospray ionization–mass spectrometry. *Analytical Biochemistry*, 2015. 486: p. 38-40.
43. Ruhaak, L.R., et al., Glycan labeling strategies and their use in identification and quantification. *Analytical and Bioanalytical Chemistry*, 2010. 397(8): p. 3457-3481.
44. Zauner, G., et al., Protein O-glycosylation analysis. *Biological Chemistry*, 2012. 393(8): p. 687-708.
45. Lim, M.S., et al., Validation of Rapi-Fluor method for glycan profiling and application to commercial antibody drugs. *Talanta*, 2019. 198: p. 105-110.
46. Turnbull, J.E. and R.A. Field, Emerging glycomics technologies. *Nature Chemical Biology*, 2007. 3(2): p. 74-77.
47. Hilliard, M., et al., Glycan characterization of the NIST RM monoclonal antibody using a total analytical solution: From sample preparation to data analysis. *mAbs*, 2017. 9(8): p. 1349-1359.

48. Keser, T., et al., Comparison of 2-Aminobenzamide, Procainamide and RapiFluor-MS as Derivatizing Agents for High-Throughput HILIC-UPLC-FLR-MS N-glycan Analysis. *Frontiers in Chemistry*, 2018. 6(324).
49. Cao, W.-Q., et al., Novel methods in glycomics: a 2019 update. *Expert Review of Proteomics*, 2020. 17(1): p. 11-25.
50. Kailemia, M.J., et al., Recent Advances in the Mass Spectrometry Methods for Glycomics and Cancer. *Analytical chemistry*, 2018. 90(1): p. 208-224.
51. Katrlík, J., et al., Glycan and lectin microarrays for glycomics and medicinal applications. *Medicinal Research Reviews*, 2010. 30(2): p. 394-418.
52. Pilobello, K.T. and L.K. Mahal, Deciphering the glycode: the complexity and analytical challenge of glycomics. *Current Opinion in Chemical Biology*, 2007. 11(3): p. 300-305.
53. Durbin, S.V., W.S. Wright, and J.C. Gildersleeve, Development of a Multiplex Glycan Microarray Assay and Comparative Analysis of Human Serum Anti-Glycan IgA, IgG, and IgM Repertoires. *ACS Omega*, 2018. 3(12): p. 16882-16891.
54. Temme, J.S., C.T. Campbell, and J.C. Gildersleeve, Factors contributing to variability of glycan microarray binding profiles. *Faraday Discussions*, 2019. 219(0): p. 90-111.
55. Kilcoyne, M., et al., Periodic acid–Schiff's reagent assay for carbohydrates in a microtiter plate format. *Analytical Biochemistry*, 2011. 416(1): p. 18-26.
56. Motilva, M.-J., et al., Adaptation of the standard enzymatic protocol (Megazyme method) to microplaque format for  $\beta$ -(1,3)(1,4)-d-glucan determination in cereal based samples with a wide range of  $\beta$ -glucan content. *Journal of Cereal Science*, 2014. 59(2): p. 224-227.
57. Ma, L.-n., et al., An overview on ELISA techniques for FMD. *Virology Journal*, 2011. 8(1): p. 419.
58. Postollec, F., et al., Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food Microbiology*, 2011. 28(5): p. 848-861.
59. Daniel, R.M., et al., The Role of Dynamics in Enzyme Activity. *Annual Review of Biophysics and Biomolecular Structure*, 2003. 32(1): p. 69-92.
60. Baruch, A., D.A. Jeffery, and M. Bogyo, Enzyme activity – it's all about image. *Trends in Cell Biology*, 2004. 14(1): p. 29-35.
61. Jarvis, C.E. and J.R.L. Walker, Simultaneous, rapid, spectrophotometric determination of total starch, amylose and amylopectin. *Journal of the Science of Food and Agriculture*, 1993. 63(1): p. 53-57.
62. Li, C., et al., A universal multi-wavelength fluorescence polarization immunoassay for multiplexed detection of mycotoxins in maize. *Biosensors and Bioelectronics*, 2016. 79: p. 258-265.
63. Geißler, D., et al., Quantum Dot Biosensors for Ultrasensitive Multiplexed Diagnostics. *Angewandte Chemie International Edition*, 2010. 49(8): p. 1396-1401.
64. Petryayeva, E. and W.R. Algar, Multiplexed Homogeneous Assays of Proteolytic Activity Using a Smartphone and Quantum Dots. *Analytical Chemistry*, 2014. 86(6): p. 3195-3202.
65. Szymula, K.P., et al., An Open-Source Plate Reader. *Biochemistry*, 2019. 58(6): p. 468-473.

66. Zaia, J., Mass Spectrometry and Glycomics. *OMICS: A Journal of Integrative Biology*, 2010. 14(4): p. 401-418.
67. Harvey, D.J., NEGATIVE ION MASS SPECTROMETRY FOR THE ANALYSIS OF N-LINKED GLYCANS. *Mass Spectrometry Reviews*, 2020. 39(5-6): p. 586-679.
68. Rebello, O.D., et al., A Matrix-Assisted Laser Desorption/Ionization—Mass Spectrometry Assay for the Relative Quantitation of Antennary Fucosylated N-Glycans in Human Plasma. *Frontiers in Chemistry*, 2020. 8: p. 138.
69. Chen, Q., et al., The Essential Functions and Detection of Bisecting GlcNAc in Cell Biology. *Frontiers in Chemistry*, 2020. 8(511).
70. Wuhrer, M., et al., Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. *Rapid Communications in Mass Spectrometry*, 2006. 20(11): p. 1747-1754.
71. Benicky, J., et al., Quantification of Fucosylated Hemopexin and Complement Factor H in Plasma of Patients with Liver Disease. *Analytical Chemistry*, 2014. 86(21): p. 10716-10723.
72. Wheeler, S.F., P. Domann, and D.J. Harvey, Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of  $\alpha(2 \rightarrow 3)$ - and  $\alpha(2 \rightarrow 6)$ -isomers. *Rapid Communications in Mass Spectrometry*, 2009. 23(2): p. 303-312.
73. de Haan, N., et al., Glycomics studies using sialic acid derivatization and mass spectrometry. *Nature Reviews Chemistry*, 2020. 4(5): p. 229-242.
74. Pongracz, T., M. Wuhrer, and N. de Haan, Expanding the Reaction Space of Linkage-Specific Sialic Acid Derivatization. *Molecules*, 2019. 24(19).
75. Nishikaze, T., Sialic acid derivatization for glycan analysis by mass spectrometry. *Proceedings of the Japan Academy, Series B*, 2019. 95(9): p. 523-537.
76. de Haan, N., et al., Linkage-Specific Sialic Acid Derivatization for MALDI-TOF-MS Profiling of IgG Glycopeptides. *Analytical Chemistry*, 2015. 87(16): p. 8284-8291.
77. Briliūtė, J., et al., Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology*, 2019. 4(9): p. 1571-1581.
78. 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*, 2017. 16(2): p. 228-242.
79. Stumpo, K.A. and V.N. Reinhold, The N-Glycome of Human Plasma. *Journal of Proteome Research*, 2010. 9(9): p. 4823-4830.
80. Zeuner, B., et al., Loop engineering of an  $\alpha$ -1,3/4-l-fucosidase for improved synthesis of human milk oligosaccharides. *Enzyme and Microbial Technology*, 2018. 115: p. 37-44.
81. Mark, B.L., et al., Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. *J Biol Chem*, 2001. 276(13): p. 10330-7.

82. Langley, D.B., et al., Structure of N-acetyl-beta-D-glucosaminidase (GcnA) from the endocarditis pathogen *Streptococcus gordonii* and its complex with the mechanism-based inhibitor NAG-thiazoline. *J Mol Biol*, 2008. 377(1): p. 104-16.



# **Addendum**



## Summary

A large repertoire of medical and clinical research has extensively demonstrated the importance of glycosylation in human health. This has created a demand for the development of precision glycomics assays. These assays are not only valuable in academic research, but have immense potential for high-throughput industrial applications of quality control, quality assurance and contract-research glycan profiling. Precision glycomics assays are often based on chromatographic and mass spectrometric techniques which require instrumentation infrastructure and technical skills that are not always readily available. This creates a demand in industry for the development of glycomics assays that have a low infrastructure cost and that are user-friendly with minimal training requirements.

Chapter 1 of this thesis provides an introduction of disease-associated changes in protein *N*-glycosylation observed in human plasma and the various analytical techniques used for their analysis. A description is given of benefits and challenges of chromatography and mass spectrometry techniques widely used in industrial glycan profiling. These applications often use glycoenzymes, and the roles of these enzymes as an analytical tool in glycomics assays are described. Chapter 1 is concluded by defining the scope of this thesis which, in a nutshell, is the development of novel high-throughput glycomics assays via a re-fashioned approach to exoglycosidase-based glycan profiling.

The assays developed in this thesis are highly reliant on the exoglycosidases for their accuracy and precision. Thus, great importance was given to the sourcing of exoglycosidases of certain desired traits. Chapter 2 outlines the workflow for the screening of exoglycosidases from natural sources. A metagenomic study of a gut bacteriome highlighted certain genes that coded for antennary fucosidases. These putative enzymes were recombinantly expressed and shown to indeed have an antennary fucosidase activity by a commonly used industrial liquid chromatography-mass spectrometry technique. Remarkably, one of the enzymes was capable of removing the antennary  $\alpha(1-3/4)$  fucosyl linkage from sialylated arms of *N*-glycans which makes it an interesting tool in glycan profiling. This was a novel and previously unreported activity for an antennary fucosidase. An in-depth structural study of the enzyme revealed the mechanism for its novel activity.

Sourcing novel enzymes for desired traits and understanding their specificity is not always as straightforward as exemplified in chapter 2. Chapter 3 outlines such a scenario in the sourcing of an *N*-acetylglucosaminidase that has potential for industrial applications. This enzyme was reported to specifically remove the bisecting *N*-acetylglucosamine residues from *N*-glycans which makes it novel in its activity. Such an activity has importance in glycan profiling. A structural study of the *N*-acetylglucosaminidase was performed to understand its mechanism and specificity. However, this understanding remained inconclusive and hence further research into the mechanism is required.

The predominant goal of this thesis is the development of exoglycosidase-based glycomics assays which have potential for industrial applications and commercialization. These assays were developed with a stern backbone i.e. they should be high-throughput, user friendly, cost-effective and have commercialization potential as analytical kits. Furthermore, these assays should provide a unique approach to an important analytical challenge.

The first analytical challenge identified was the quantification of antennary fucosylation in human plasma *N*-glycosylation. This glycosylation trait is often associated with inflammatory diseases such as autoimmune diseases, diabetes and cancer, thus having relevance in medical sciences. The industrial techniques used for the quantification of antennary fucosylation are based on HILIC(LC)-FLD/MS<sup>n</sup>. Often these chromatographic techniques take several minutes up to an hour for analysis of a single sample. In this respect, MALDI-TOF-MS is the superior techniques when it comes to high-throughput analysis. Chapter 4 introduces a novel MALDI-TOF-MS based assay for the quantification of antennary fucosylation on human plasma *N*-glycosylation. The core fucosylation of *N*-glycans was depleted by exoglycosidases and the remaining antennary fucosylation was quantified. By combination of this workflow with sialic acid derivatization, sialyl Lewis X/A epitopes could be quantified. This MALDI-based assay not only outperformed an industrial LC-FLD-MS<sup>n</sup> technique, but also significantly reduced the required time for analysis which is an important desirable trait in busy industrial analytical laboratories.

The second analytical challenge was the development of spectrophotometric assays instead of the previous combinations of chromatography and mass spectrometry. This type of assays has commercial interest as they are easily producible as kits and can cater to a large clientele of laboratories that are not equipped for routine high-end glycomics analysis. Chapter 5 introduces a novel plate-based spectrophotometric assay for the quantification of galactosylation and sialylation on human antibodies. Exoglycosidases are used for releasing sialic acids and galactose residues from the glycoproteins. These released residues are subjected to a redox reaction that produces a fluorescence signal which is quantified. This assay was shown to perform equally well as an industrial LC-FLD-MS<sup>n</sup> technique for glycan profiling of antibodies, whilst not requiring the infrastructure or skills of an MS-based glycomics laboratory.

Finally, the analytical challenges for quantifying *N*-glycosylation are discussed in chapter 6 with respect to assay development. Firstly, the widely used techniques of chromatography, electrophoresis and mass spectrometry are discussed along with the less common techniques of glycan/lectin microarrays. The general pros and cons for these techniques are outlined and evaluated. Next, the most important and critical characteristics of glycomics assays for industrial applications are discussed in detail. Many of these assays are based on LC-FLD/-MS<sup>n</sup> techniques and their molecular identification options are discussed in-depth. This includes collision induced dissociation, derivatisation and exoglycosidase-based profiling. Finally, the importance of glycosidases in industry is discussed, especially the need for sourcing and/or engineering exoglycosidase as analytical tools in industrial glycan profiling applications.

## Nederlandse samenvatting

Het belang van glycosylering met betrekking tot gezondheid is uitvoerig aangetoond in medisch en klinisch onderzoek met als gevolg een nieuwe behoefte aan nauwkeurige *glycomics assays*. De ontwikkeling van deze testen is niet alleen waardevol voor het uitvoeren van academisch onderzoek, maar is ook gewenst voor industriële toepassingen met hoge doorvoer van kwaliteitscontrole, kwaliteitsborging en glycaanprofilering bij contractonderzoek. Precisie *glycomics assays* zijn vaak gebaseerd op chromatografische en massaspectrometrische technieken die een infrastructuur en technische vaardigheden vereisen die niet eenvoudig voorhanden zijn. Dit creëert een vraag in de industrie naar de ontwikkeling van *glycomics assays* die lage infrastructuurkosten hebben en die gebruiksvriendelijk zijn met minimale trainingsvereisten.

Hoofdstuk 1 van dit proefschrift geeft een introductie van ziekte-geassocieerde *N*-glycosyleringsveranderingen die zijn waargenomen in humaan plasma en de verschillende analytische technieken die voor hun analyse worden gebruikt. Er wordt een beschrijving gegeven van de voordelen en uitdagingen van chromatografie- en massaspectrometrietechnieken die veel worden gebruikt bij industriële glycaanprofilering. Deze toepassingen maken vaak gebruik van glyco-enzymen en de rol van deze enzymen als analytisch hulpmiddel bij *glycomics assays* wordt beschreven. Hoofdstuk 1 wordt afgesloten met het kaderen van dit proefschrift, in het kort is dit de ontwikkeling van nieuwe *glycomics assays* met hoge doorvoer via een vernieuwde benadering van op exoglycosidase gebaseerde glycaanprofilering.

De testen die in dit proefschrift zijn ontwikkeld, zijn in hoge mate afhankelijk van de exoglycosidasen voor hun nauwkeurigheid en precisie. Het was daarom van groot belang exoglycosidasen te verkrijgen met bepaalde gewenste eigenschappen. Hoofdstuk 2 schetst de strategie voor het evalueren van exoglycosidasen uit natuurlijke bronnen. Een metagenomische studie van een darmbacterioom bracht bepaalde genen aan het licht die codeerden voor antennefucosidasen. Deze hypothetisch verwachte enzymen werden op recombinante wijze tot expressie gebracht en er werd aangetoond dat ze inderdaad een antennefucosidase-activiteit hebben door een algemeen gebruikte industriële vloeistofchromatografie-massaspectrometrietechniek. Opmerkelijk was dat één van de enzymen in staat was om van *N*-glycaan antennes een fucoseresidu dat  $\alpha(1-3/4)$  gebonden is aan gesialyleerde armen te verwijderen, wat het een interessant hulpmiddel maakt bij glycanprofilering. Dit is een nieuwe en voorheen niet-gerapporteerde activiteit voor een antennefucosidase. Een diepgaande structurele studie van het enzym onthulde het mechanisme voor zijn nieuwe activiteit.

Het vinden van nieuwe enzymen voor gewenste eigenschappen en het begrijpen van hun specificiteit is niet altijd zo eenvoudig als wordt geïllustreerd in hoofdstuk 2. Hoofdstuk 3 schetst een dergelijk scenario bij het verkrijgen van een *N*-acetylglucosaminidase dat potentieel heeft voor industriële toepassingen. Van dit enzym werd gemeld dat het specifiek de zogenaamde *bisecting N*-acetylglucosamine uit *N*-glycanen kan verwijderen, waarmee het nieuw zou zijn in zijn activiteit. Een dergelijke activiteit is van belang bij glycaanprofilering. Een structurele studie van de *N*-acetylglucosaminidase werd uitgevoerd om het mechanisme

en de specificiteit ervan te begrijpen. De resultaten waren echter niet eenduidig en daarom is verder onderzoek naar het mechanisme vereist.

Het belangrijkste doel van dit proefschrift is de ontwikkeling van op exoglycosidase gebaseerde *glycomics assays* die potentieel hebben voor industriële toepassingen en commercialisering. Deze assays moeten voldoen aan een hoge doorvoer, dienen gebruiksvriendelijk en kosteneffectief te zijn en verder een commercialiseringspotentieel te hebben als analytische kits. Bovendien moeten deze testen een unieke benadering bieden voor een belangrijke analytische uitdaging.

De eerste analytische uitdaging die werd geïdentificeerd, was de kwantificering van antenne-fucosylering in *N*-glycosylering van humaan plasma. Deze glycosyleringseigenschap wordt vaak geassocieerd met ontstekingsziekten zoals auto-immuunziekten, diabetes en kanker, en is dus relevant in de medische wetenschappen. De industriële technieken die gebruikt worden voor de kwantificering van antennefucosylering zijn gebaseerd op HILIC(LC)-FLD/MS<sup>n</sup>. Vaak nemen deze chromatografietechnieken enkele minuten tot een uur in beslag voor de analyse van een enkel monster. In dit opzicht is MALDI-TOF-MS de superieure techniek als het gaat om *high-throughput analyse*. Hoofdstuk 4 introduceert een nieuwe op MALDI-TOF-MS gebaseerde test voor de kwantificering van antenne-fucosylering op *N*-glycosylering in humaan plasma. De zogenaamde core-fucosylering van *N*-glycanen werd verwijderd door exoglycosidasen en de resterende antennefucosylering werd gekwantificeerd. Door deze workflow te combineren met derivatisering van sialzuur, konden sialyl Lewis X/A-epitopen worden gekwantificeerd. Deze op MALDI gebaseerde test presteerde niet alleen beter dan een industriële LC-FLD-MS<sup>n</sup>-techniek, maar verminderde ook aanzienlijk de benodigde tijd voor analyse, wat een belangrijke wenselijke eigenschap is in industriële analytische laboratoria.

De tweede analytische uitdaging was de ontwikkeling van spectrofotometrische testen die geen chromatografie en massaspectrometrie toepassen. Dit type assays heeft commercieel belang omdat ze gemakkelijk als kits kunnen worden geproduceerd en geschikt zijn voor een grote klantenkring van laboratoria die niet zijn uitgerust voor routinematige *high-end glycomics-analyse*. Hoofdstuk 5 introduceert een nieuwe op platen gebaseerde spectrofotometrische test voor de kwantificering van galactosylering en sialylering op humane antilichamen. Exoglycosidasen worden gebruikt voor het verwijderen van sialzuren en galactoseresiduen uit de glycoproteïnen. Deze vrijgekomen residuen worden onderworpen aan een redoxreactie die een fluorescentiesignaal produceert dat wordt gekwantificeerd. Deze test bleek even goed te presteren als een industriële LC-FLD-MS<sup>n</sup>-techniek voor glycanprofilering van antilichamen, terwijl de infrastructuur of vaardigheden van een *glycomics* laboratorium met massaspectrometrie niet nodig waren.

Tenslotte worden de analytische uitdagingen voor het kwantificeren van *N*-glycosylering besproken in hoofdstuk 6 met betrekking tot de ontwikkeling van assays. Ten eerste worden de huidige veelgebruikte technieken van chromatografie, elektroforese en massaspectrometrie besproken, samen met de minder populaire technieken van glycan/lectine *microarrays*. De algemene voor- en nadelen van deze technieken worden geschetst en geëvalueerd. Vervolgens worden de belangrijkste en meest kritische kenmerken

van *glycomics assays* voor industriële toepassingen in detail besproken. Veel van deze testen zijn gebaseerd op LC-FLD/-MS<sup>n</sup>-technieken en hun moleculaire identificatie-opties worden uitgebreid besproken. Dit omvat door botsingen geïnduceerde dissociatie, derivatisering en op exoglycosidase gebaseerde profilering. Tenslotte wordt het belang van glycosidasen in de industrie besproken, met name de noodzaak om exoglycosidase te sourcen en/of te engineeren als analytische hulpmiddelen in industriële glycanprofileringstoepassingen.

## **Acknowledgements**

Firstly, I want to thank my promoter Prof. dr. Manfred Wuhrer for offering me this PhD position, mentoring me in my scientific research and making me believe in my capabilities.

I would like to thank my co-promotor Dr. David Falck for mentoring me in my scientific research, indulging my scientific curiosity and instilling upon me an analytical mindset. I will greatly miss our long fury filled scientific discussions / debates.

I would also like to acknowledge and thank Dr. Daniel Spencer for recruiting me into Ludger Ltd. which made obtaining this PhD possible. I am glad to be mentored under him because I learnt something which academia would never teach me i.e. professionalism and work attitude in a commercial and industrial scientific organisation.

Collaborations with other institutes and research groups were a vital part of my thesis, and so I want to acknowledge and thank my collaborators Dr. Lucy Crouch, Dr. Arnaud Basil, Dr. Haiyang Wu, Prof. dr. Nathalie Juge and Dr. David Bolam. These collaborations introduced me into such an exciting world of interdisciplinary research that tickled my interest, and which made my thesis possible.

Of course, I would like to acknowledge my colleagues at the CPM, Steffen, Katarina, Di, Sander, Tamas, Alan and Iwona for their friendship and creating a great work environment for my time at the LUMC.

Importantly vital, I acknowledge the financial grant organisations that made my PhD possible, 1) The European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement GlySign No 722095 and 2) The Innovate UK Biocatalyst grant Glycoenzymes for Bioindustries (BB/M029042/).

Lastly but most importantly, I want to acknowledge and thank all my colleagues at Spectrometry Vision B.V. (MSVision) for their love and encouragement during the completion of my PhD.

## Abbreviations

<b>2'FL</b>	2-Fucosyllactose
<b><math>\alpha</math>Gal-LeX</b>	$\alpha(1-3)$ Gal-Lewis X
<b>BBSRC</b>	Biotechnology and Biological Sciences Research Council
<b>CAZymes</b>	Carbohydrate-Active enZymes
<b>CBM</b>	Carbohydrate binding module
<b>CE</b>	Capillary electrophoresis
<b>CFG</b>	Consortium for Functional Glycomics
<b>CGE</b>	Capillary gel electrophoresis
<b>CID</b>	Collision induced dissociation
<b>CQA</b>	Critical quality attribute
<b>CZE</b>	Capillary zone electrophoresis
<b>ECD</b>	Electron capture dissociation
<b>EDC</b>	1-Ethyl-3-(3-(dimethylamino)-propyl)carbodiimide hydrochloride
<b>ELISA</b>	Enzyme linked immune sorbent assay
<b>ESI</b>	Electrospray ionisation
<b>FLD</b>	Fluorescence detection
<b>FRET</b>	Förster resonance energy transfer
<b>FT-ICR</b>	Fourier transform-ion cyclotron resonance
<b>Fuc</b>	Fucose
<b>Gal</b>	Galactose
<b>GalNac</b>	<i>N</i> -acetylgalactosamine
<b>GH</b>	Glycoside hydrolase
<b>GiRP</b>	Girard's Reagent P
<b>Glc</b>	Glucose
<b>GlcNac</b>	<i>N</i> -acetylglucosamine
<b>HILIC</b>	Hydrophilic interaction liquid chromatography
<b>HMOs</b>	Human milk oligosaccharides
<b>HOBt</b>	1-hydroxybenzotriazole
<b>HRP</b>	Horseradish peroxidase
<b>IBD</b>	Inflammatory bowel disease
<b>ICR</b>	Ion cyclotron resonance
<b>IMAC</b>	Immobilized metal affinity chromatography
<b>IMS</b>	Ion mobility mass spectrometry
<b>IPTG</b>	Isopropyl B-D-thiogalactopyranoside
<b>ITC</b>	Isothermal titration calorimetry
<b>LacNac</b>	<i>N</i> -acetyl-D-lactosamine
<b>LC</b>	Liquid chromatography
<b>LeA</b>	Lewis A
<b>LeB</b>	Lewis B
<b>LED</b>	Light emitting diodes
<b>LeX</b>	Lewis X
<b>LeY</b>	Lewis Y
<b>LIF</b>	Laser induced fluorescence
<b>MALDI</b>	Matrix assisted laser desorption ionisation
<b>Man</b>	Mannose
<b>MD</b>	Molecular dynamics

<b>MS</b>	Mass spectrometry
<b>NAD<sup>+</sup></b>	Nicotinamide adenine dinucleotide
<b>NADP<sup>+</sup></b>	Nicotinamide adenine dinucleotide phosphate
<b>Neu5Ac</b>	5- <i>N</i> -acetylneuraminic acid
<b>NHAc</b>	Acetamido group
<b>NMR</b>	Nuclear magnetic resonance
<b>NP-40</b>	Nonidet P-40
<b>PAD</b>	Pulsed amperometric detection
<b>PBS</b>	Phosphate buffered saline solution
<b>PNGase F</b>	Peptide <i>N</i> -glycosidase F
<b>pNP-Fuc</b>	4-nitrophenyl $\alpha$ -L-fucopyranoside
<b>QA</b>	Quality assurance
<b>QC</b>	Quality control
<b>qPCR</b>	Quantitative polymerase chain reaction
<b>RMSD</b>	Root mean square deviation
<b>RP</b>	Reverse phase
<b>RSD</b>	Relative standard deviations
<b>S/N</b>	Signal to noise
<b>SAD</b>	Single anomalous diffraction
<b>SDS</b>	Sodium dodecyl sulfate
<b>Sia</b>	Sialic acid
<b>sLeX</b>	Sialyl Lewis X
<b>SPBS</b>	Sodium phosphate buffer solution
<b>SPE</b>	Solid-phase extraction
<b>SPR</b>	Surface plasmon resonance
<b>SSN</b>	Sequence similarity networks
<b>STD</b>	Saturation transfer difference
<b>TOF</b>	Time of flight
<b>TPNG</b>	Total plasma <i>N</i> -glycome
<b>TSNG</b>	Total serum <i>N</i> -glycome



## **Curriculum vitae**

Osmond D. Rebello was born on the 21<sup>th</sup> of July, 1992 in Dubai, U.A.E. In 2013, he obtained a bachelor's degree in Chemistry and Industrial microbiology from Pune University, India. and in 2016, he obtained a master's degree in Biochemistry and Molecular Biology from Bremen University, Germany. During his master's studies, Osmond worked as an assistant laboratory technician at the Nano-SIMS facility of the Max Planck Institute, Bremen. During this period, Osmond gained an immense interest in analytical chemistry and instrumentation and so focused his master internships and thesis towards an analytical chemistry track. For his master's thesis, Osmond developed assays for quantifying the binding affinity of lectin domains from trans-sialidases in the laboratory of Prof. dr. Sorge Kelm, Bremen University.

In 2017, Osmond obtained a three-year Marie Curie fellowship as an early-stage researcher for the GLYSIGN consortium. During this project, Osmond was primary based at Ludger Ltd., United Kingdom, as a R&D scientist developing glycomics techniques. In collaboration with Prof. dr. Nathalie Juge of Quadram Institute, Dr. Lucy Crouch of University of Birmingham and Dr. David Bolam of NewCastle University, Osmond was involved in the identification of several novel exoglycosidases by LC-MS<sup>n</sup> techniques. Additionally, Osmond also developed a plate-based spectrophotometric glycomics technique. Osmond spent a portion of this period working on research projects at the Center for Proteomics and Metabolomics (CPM) in Leiden University Medical Center (LUMC) where he developed mass spectrometry-based and capillary electrophoresis-based glycomics assays and techniques under the guidance of Dr. David Falck and Prof. dr. Manfred Wührer. Osmond also contributed to automating a IgG capture protocol onto a Hamilton liquid handling robotic platform. His contribution to the GLYSIGN consortium is compiled into this PhD. thesis.

Since 2020, Osmond is working as a R&D scientist / mass spectrometry engineer at Spectrometry Vision B.V. (MSVision), Netherlands. In this role, Osmond pursues a career in mass spectrometry and other analytical instrumentation development.

## List of publications

### This thesis

- 1. A novel glycosidase plate-based assay for the quantification of galactosylation and sialylation on human IgG**  
**Osmond D. Rebello**, Richard A. Gardner, Paulina A. Urbanowicz, David N. Bolam, Lucy I. Crouch, David Falck, Daniel I. R. Spencer.  
Glycoconjugate Journal 37, 691–702, 2020
- 2. A Matrix-Assisted Laser Desorption/Ionization—Mass Spectrometry Assay for the Relative Quantitation of Antennary Fucosylated N-Glycans in Human Plasma**  
**Osmond D. Rebello**, Simone Nicolardi, Guinevere S. M. Lageveen-Kammeijer, Jan Nouta, Richard A. Gardner, Wilma E. Mesker, Rob A. E. M. Tollenaar, Daniel I. R. Spencer, Manfred Wuhrer and David Falck  
Frontiers in Chemistry 8, 138, 2020
- 3. Fucosidases from the human gut symbiont Ruminococcus gnavus**  
Haiyang Wu, **Osmond Rebello**, Emmanuelle H. Crost, C. David Owen, Samuel Walpole, Chloe Bennati Granier, Didier Ndeh, Serena Monaco, Thomas Hicks, Anna Colvile, Paulina A. Urbanowicz, Martin A. Walsh, Jesus Angulo, Daniel I. R. Spencer, Nathalie Juge  
Cellular and Molecular Life Sciences, 78:675–693, 2021

### Other

- 4. Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci**  
Justina Briliūtė, Paulina A. Urbanowicz, Ana S. Luis, Arnaud Baslé, Neil Paterson, **Osmond Rebello**, Jenifer Hendel, Didier A. Ndeh, Elisabeth C. Lowe, Eric C. Martens, Daniel I. R. Spencer, David N. Bolam & Lucy I. Crouch  
Nature Microbiology 4, 1571–1581, 2019
- 5. Carbon recycling efficiency and phosphate turnover by marine nitrifying archaea**  
Travis B. Meador, Niels Schoffelen, Timothy G. Ferdelman, **Osmond Rebello**, Alexander Khachikyan, Martin Könneke  
Science Advances, 6(19):eaba1799, 2020
- 6. GLP-2 Is Locally Produced From Human Islets and Balances Inflammation Through an Inter-Islet-Immune Cell Crosstalk**  
Wei He, **Osmond D. Rebello**, Antonia Henne, Fabian Nikolka, Thomas Klein and Kathrin Maedler  
Frontiers in Endocrinology, 12:69712005, 2021

**7. TLR4 triggered complex inflammation in human pancreatic islets**

Wei He, **Osmond Rebello**, Rocco Savino, Rosa Terracciano, Carole Schuster-Klein, Beatrice Guardiola, Kathrin Maedler

Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1865(1), 86-97, 2018

**8. Two new isoforms of the human hepatoma-derived growth factor interact with components of the cytoskeleton**

Jessica Nüße, Ursula Mirastschijski, Mario Waespy, Janina Oetjen, Nadine Brandes, **Osmond Rebello**, Federico Paroni, Sørge Kelm and Frank Dietz

Biological Chemistry 397(5), 417-436, 2016



