Unraveling mucin type o-glycosylation signatures of colorectal cancer

Madunić. K.

Unraveling Mucin Type O-glycosylation Signatures of Colorectal Cancer

Katarina Madunić

UNRAVELING MUCIN TYPE *O*-GLYCOSYLATION SIGNATURES OF COLORECTAL CANCER

Katarina Madunić

ISBN: 978-94-6458-073-0

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

This work was supported by the European Commission's Horizon 2020 project "GlyCoCan", grant number 676421.

Cover Design: Wouter van Oldeneel tot Oldenzeel

Printed by: Ridderprint | www.ridderprint.nl.

UNRAVELING MUCIN TYPE *O*-GLYCOSYLATION SIGNATURES OF COLORECTAL CANCER

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 dinsdag 29 maart 2022
klokke 13:45 uur

door

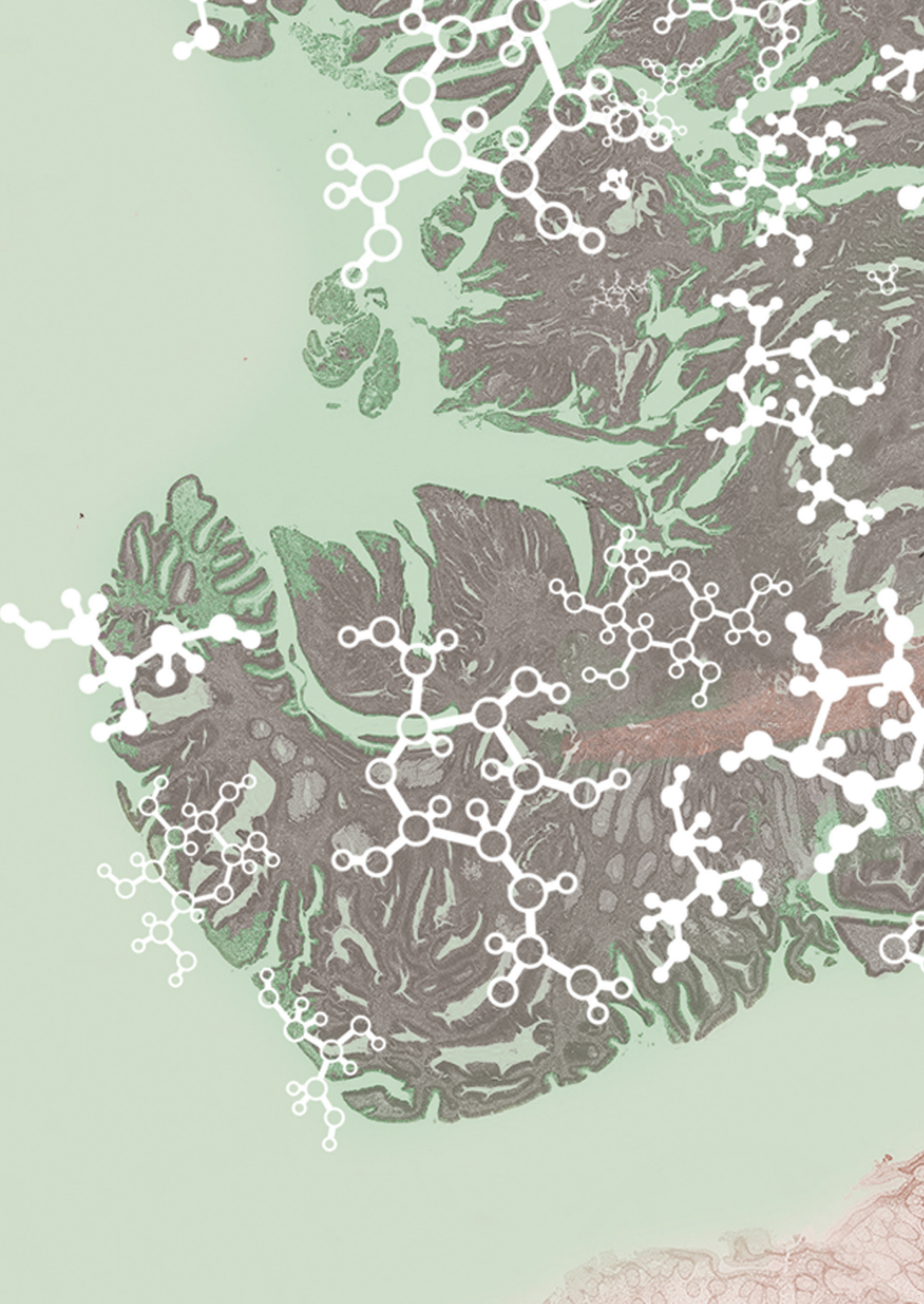
Katarina Madunić

geboren te Split (Kroatië)

in 1989

Promotores:	Prof. Dr. M. Wuhrer
	Prof. Dr. P. ten Dijke
Co-promotor:	Dr. G. S. M. Lageveen- Kammeijer
Leden van de promotiecommissie:	Prof. Dr. C. H. Hokke
	Dr. L. R. Ruhaak
	Prof. Dr. C. A. Reis
	University of Porto, Portugal
	Prof. Dr. H. H. Wandall
	University of Copenhagen, Denmark

TABLE OF CONTENTS



INTRODUCTION

Chapter 1

This introduction will cover colorectal cancer (CRC) pathogenesis and current treatment strategies for targeted pharmacotherapy of CRC. Moreover, it will introduce glycosylation and give a detailed overview of the biosynthesis of mucin type *O*-glycosylation, as well as changes in the expression observed in cancer. Finally, an overview will be provided regarding the methodologies being used for the analysis of mucin type *O*-glycans, with a specific focus on separation techniques as well as mass spectrometry as a detection technique.

COLORECTAL CANCER

PATHOGENESIS OF COLORECTAL CANCER

CRC is one of the most common types of cancers with nearly two million new cases in 2020 worldwide[1]. In addition, CRC is the leading cause of cancer related deaths[1], as patients are often asymptomatic until the late stages of the disease. CRC develops from precancerous adenoma, by the accumulation of somatic and germline mutations over a longer period of time. Among others, mechanisms such as chromosomal instability (CIN), microsatellite instability (MSI), and epigenetic instability also known as CpG island methylator phenotype (CIMP) have been linked to CRC pathogenesis[2]. According to The Cancer Genome Atlas (TCGA), which contains genomic data from a large sample set of CRC patients, a differentiation between hypermutated and non-hypermutated CRC can be made[3]. The hypermutated CRC subtype (15% of CRC patients) is characterized as MSI, with immune cell infiltration and better prognosis in the early stages. The prevalent non-hypermutated subtype is characterized by CIN, worse prognosis, and resistance to immunotherapy. Despite insights into molecular pathogenesis, this stratification provides limited clinical benefit for outcome prediction and therapeutic intervention. To provide a more accurate patient prognosis, stratification based upon the four consensus molecular subtypes (CMS) revealed promising results[4]. For this purpose, multiple molecular markers were combined which were derived from transcriptomic profiling of a large sample set of CRC patients (**Figure 1**). CMS 1 tumors were predominantly classified as MSI and showed high immune infiltration in the tumor microenvironment. This was associated with a better patient prognosis, however worse survival after relapse. Both CMS 2 and CMS 3 tumors showed strong epithelial differentiation signatures. CMS 2 is characterized by

CMS1 MSI immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
14%	37%	13%	23%
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
BRAF mutations		*KRAS* mutations	
Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGF-β activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival

Figure 1. Proposed classification of CRC reflecting the gene expression-based molecular subtypes. *Figure adapted from Guinney et al.*[4]

WNT and MYC activation, whereas the CMS 3 group revealed characteristic metabolic pathway dysregulation[4]. In contrast, the mesenchymal CMS 4 tumors were found to be characterized by stromal infiltration, TGF-β activation and upregulation of epithelial to mesenchymal transition (EMT) resulting in worse overall patient prognosis[4].

TARGETED THERAPY FOR CRC

Conventional therapies for the treatment of CRC include surgery, radiotherapy and chemotherapy and are associated with various side effects[5]. To increase the efficiency and to decrease the toxicity of CRC treatment, targeted therapies provide new perspectives as it intervenes with the function of specific molecules present on cancer cells or cells in the tumor microenvironment [6]. For example, promising results were achieved using antibodies for immuno-targeted therapies in cancer treatment[7]. Several targets have been approved for the antibody treatment of CRC, namely, the anti-VEGF (bevacizumab), anti-EGFR (cetuximab and panitumumab), anti-PD-1 immune checkpoint (pembrolizumab, nivolumab), and BRAF V600E inhibitors (several combinations of vemurafenib, irinotecan, and cetuximab or panitumumab)[6]. Many more are still undergoing clinical trials[6]. Due to the tumor heterogeneity and relative paucity of the mentioned protein targets

present on cancer cells, only a small part of patients clearly benefit from these treatments[6]. Therefore, better patient stratification and identification of new targets is a crucial aspect for the development of more effective therapies. Since aberrant glycosylation is one of the hallmarks of cancer[8], and cancer associated glycans are present in high abundance on the surface of cancer cells which are linked to many different proteins, it has an immense potential to be used as therapy target[9].

GLYCOSYLATION

The process where sugar molecules are covalently linked to a protein or lipid carrier by specific enzymes is called glycosylation and plays an important role in modulating their function. In fact, the surface of eukaryotic cells is composed of a very dense layer of oligosaccharides, linked to their carriers, which play an important role in cell-cell and cell-extracellular matrix interactions, leading to changes in immune response, differentiation and cell growth[10]. This layer, also known as the glycocalyx, will be extensively studied in this thesis in the context of CRC. Oligosaccharides can be present on various carriers and based upon the glycoconjugate a distinction is made between the different classes such as glycoproteins, glycosaminoglycans (GAGs), glycosylphosphatidylinositol (GPI) anchors, proteoglycans and glycosphingolipids (GSLs)[10]. In regard to glycoproteins, glycans can be bound to the protein via a nitrogen atom to the amino acid asparagine (Asn), with a specific sequence motif of three amino acids (Asn-X-Ser/Thr in which “X” is any amino acid except proline (Pro)) and are known as *N*-linked glycans whereas, in the case where the glycan is attached via an oxygen atom to a serine (Ser) or threonine (Thr) it is referred to as *O*-linked glycosylation. In eukaryotic organisms, the glycans are composed of a limited set of monosaccharide building blocks, which for glycoproteins consist of hexoses (galactose (Gal), mannoses (Man) and glucoses (Glc)), *N*-acetylhexosamines (*N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc)), deoxyhexose (fucose (Fuc)), pentose (xylose (Xyl)) and sialic acid (*N*-acetylneuraminic acid (Neu5Ac), and *N*-glycolylneuraminic acid (Neu5Gc)). Next to the various building blocks, the glycans become even more diverse as the monosaccharides can be differently linked to each other, namely, linkages can be formed on different hydroxyl groups in the sugar ring as well as in two anomeric configurations (α and

β). The glycan repertoire is further complicated by modifications such as the addition of acetyl groups, phosphate or sulfate.

O-LINKED GLYCOSYLATION

As stated before, *O*-linked glycosylation occurs on specific amino acids such as Ser and Thr, and in rare cases on tyrosine (Tyr)[11]. However, compared to *N*-glycosylation, no consensus sequence motif has been defined. The biosynthesis of *O*-GalNAc linked glycans is a process that takes place in the Golgi apparatus. The enzymes involved are type II transmembrane proteins, with catalytic domain in the lumen of the Golgi. The activated monosaccharide donors are transported from the cytosol. The oligosaccharide chains can be initiated by a GalNAc, GlcNAc, Fuc, Xyl or Man residues. Although it can also be present on other types of glycoproteins, the main glycosylation type of mucins is the *O*-GalNAc glycosylation, where GalNAc is the first monosaccharide attached to the Ser or Thr[11]. This glycosylation type occurs predominantly in central tandem repeat domains which are rich in Pro, Thr and Ser (PTS). Aberrant mucin expression and mucin type glycosylation have been associated with disease pathologies, such as cancer[11], therefore mucin type *O*-linked glycosylation will be the main focus of this thesis, and it will be described in further detail.

BIOSYNTHETIC PATHWAY OF MUCIN TYPE *O*-GLYCANS

BIOSYNTHESIS OF MAIN MUCIN TYPE GLYCAN CORES

The initial step of adding a GalNAc in α-linkage to Ser or Thr is catalyzed by polypeptide-*N*-acetylgalactosaminyltransferases (GALNT1-20), a family of 20 different enzymes[12] (**Figure 2**). This step results in the biosynthesis of the Tn antigen. When Neu5Ac is added in the α2-6 position of the core GalNAc, sialyl-Tn antigen is formed. Addition of β1-3-linked galactose by the action of *C1GALT1* forms the T antigen, a core 1 glycan, that can be further extended. The activity of T synthase is dependent on its molecular chaperone COSMC, and high expression of Tn and sialyl-Tn antigen in cancer are often associated with defects in the activity of T synthase or COSMC[13–15]. Next to the addition of a Neu5Ac, the Tn-antigen can also be extended differently, forming four well-known types of *O*-GalNAc glycan cores (1-4) **(Figure 2)**, as well as four less common ones (cores 5-8).

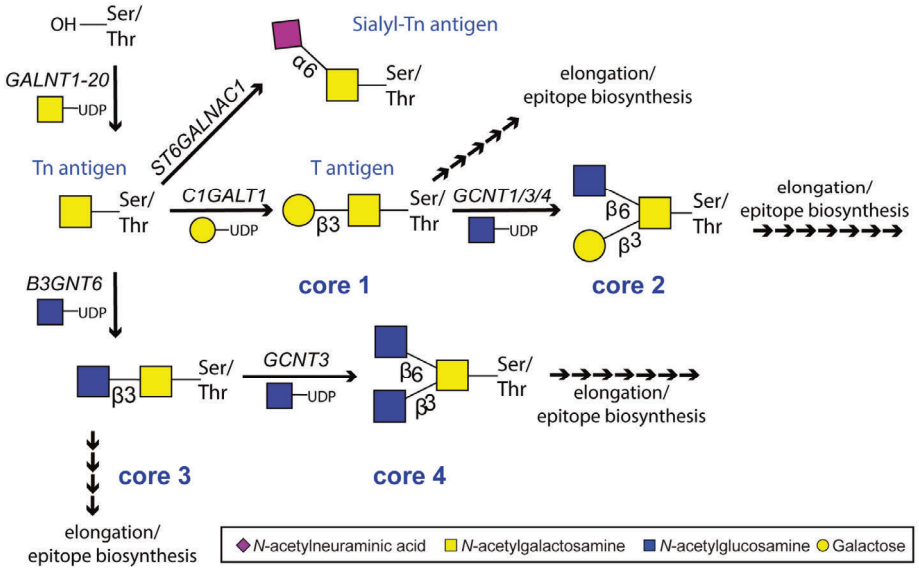


Figure 2. Biosynthesis of the most common mucin type glycan cores.

When the GlcNAc is added in β1-6-linkage to the GalNAc residue of the T antigen, a core 2 glycan is formed. This step is catalyzed by β1-6-*N*-acetylglucosaminyltransferases 1, 3 and 4 (GCNT1, GCNT3 and GCNT4)[14]. GCNT1 is the leukocyte type enzyme (L-type) which synthesizes only core 2 glycans, whereas the mucin(M)-type enzyme (GCNT3) can also synthesize core 4 glycans by adding the β1-6-GlcNAc to a core 3 glycan precursor[16]. The M-type enzyme is found predominately in high mucus secreting organ systems such as the gastrointestinal and respiratory tract. Core 2 and core 4 glycans can be further elongated on both branches by the activity of different glycosyltransferases (GSTs). Addition of β1-3-linked GlcNAc to the Tn structure by the action of β1-3-*N*-acetylglucosaminyltransferase (*B3GNT6*) forms core 3 glycans, that can be further extended into a linear glycan, or transformed into a branched core 4 glycan by the action of M-type β1-6 *N*-acetylglucosaminyltransferase (*GCNT3*)[10]. The expression of core 3 and core 4 glycans is limited to epithelial mucosa of the gastrointestinal and respiratory tract as well as salivary glands, and their expression is downregulated in cancer[17–19] due to downregulation of core 3 synthase (*B3GNT6)*[20]. Apart from the initiation of GalNAc type glycosylation, most GSTs involved in the core formation are pathway specific and do not have close paralogues, which allows to predict cell glycosylation based on the expression of

those enzymes[11]. Core structures are further extended by elongation with Type 1 and 2 chains, and/or capped with terminal motifs such as Lewis type and blood type antigens or by sialylation.

The four less common core types (cores 5-8) are all linear glycan types. Of which cores 5 and 7 contain a GalNAc linked to the core GalNAc in an α1-3- and α1-6-linkage, respectively. Core 6 glycans have a β1-6-linked GlcNAc to the core GalNAc whereas core 8 glycans contain a Gal which is α1-3-linked to the GalNAc.

CORE EXTENSION, ELONGATION OR BRANCHING

Extension of the pre-existing core GlcNAc in branched glycans can form two different types of chains; Type 1 is defined by the addition of the Gal by β1-3 GSTs (*B3GAL1/2/5*) and type 2 by β1-4 galactosyltransferases (*B4GALT1-4*) **(Figure 3)**. Additionally, a GlcNAc can be added to a terminal Gal to form a linear elongation by β1-3-*N*-acetylglucosaminyltransferases (*B3GNT2,3,4,7,8 and 9*) or otherwise a branching point by β1-6-*N*-acetylglucosaminyltransferase (*GCNT2, GCNT7* and *GCNT3*) forming the I-antigen[20,21]. Type 2 motifs form repeating units (polyLacNAc) or may be capped by terminal epitopes. Interestingly, the *GCNT2* is the main enzyme responsible for biosynthesis of the I-antigen on GSLs, whereas the M-type (*GCNT3)* plays a central role in the I-antigen next to core 2 and 4 biosynthesis of mucin type glycans in colon, small intestine, trachea and stomach[16]. A less common elongation is the formation of a LacdiNAc motif (GalNAcβ1-4GlcNAcβ1) by action of β1-4-GalNAc transferase (*B4GALNT3/4*) which can be further modified to a fucosylated LacDiNAc or terminated by sialylation[21,22]. GSTs involved in glycan elongation are considered non-pathway specific and, in combination with overlapping specificities, predictions of glycan expression become far more challenging. However, it is known that *B3GALT5* preferably acts on core 3 *O*-glycans, and *B3GNT3* prefers core 1 and 2 *O*-glycans as substrates[10].

BIOSYNTHESIS OF TERMINAL EPITOPES

GSTs responsible for the biosynthesis of terminal epitopes act on distinct types of glycans. The α1-2-fucosyltransferases *FUT1* and *FUT2* synthesize the blood group H antigens on *N*-glycans, *O*- glycans and GSLs by adding fucose to the terminal Type 1, 2, 3 and 4 linked galactose residues **(Figure 3).** Galactose is β1-3-linked to a GalNAc residue in Type 3 H antigen, either to reducing end GalNAc in mucin

type *O*-glycans, or to inner GalNAc residues in glycolipids. If those are globo-series glycolipids, it is referred to as Type 4[10]. The *H* allele encodes for the Type 2 and Type 4 specific *FUT1* transferase, forming blood group H in erythrocytes, whereas the *Se* allele encodes for *FUT2* transferase, forming the Type 1 and Type 3 H-antigen in epithelia of the gastrointestinal, respiratory and reproductive tract. Individuals with inactive *FUT2* gene, termed non-secretors, do not express soluble forms of blood group ABH epitopes[19]. When α1-3-linked GalNAc (by *A3GALNT*) or α1-3-linked Gal (by *A3GALT1*) is added to the galactose, A or B blood group antigens are formed, respectively. In secretor-positive individuals, formation of Lewis $(Le)^B$ antigen is possible by action of α1-4 fucosyltransferase *FUT3* on an already formed Type 1 H antigen. Le^A antigen is formed by the action of the same GSTs on a Type 1 LacNAc substrates. Similarly, Le^X and Le^Y antigens are formed

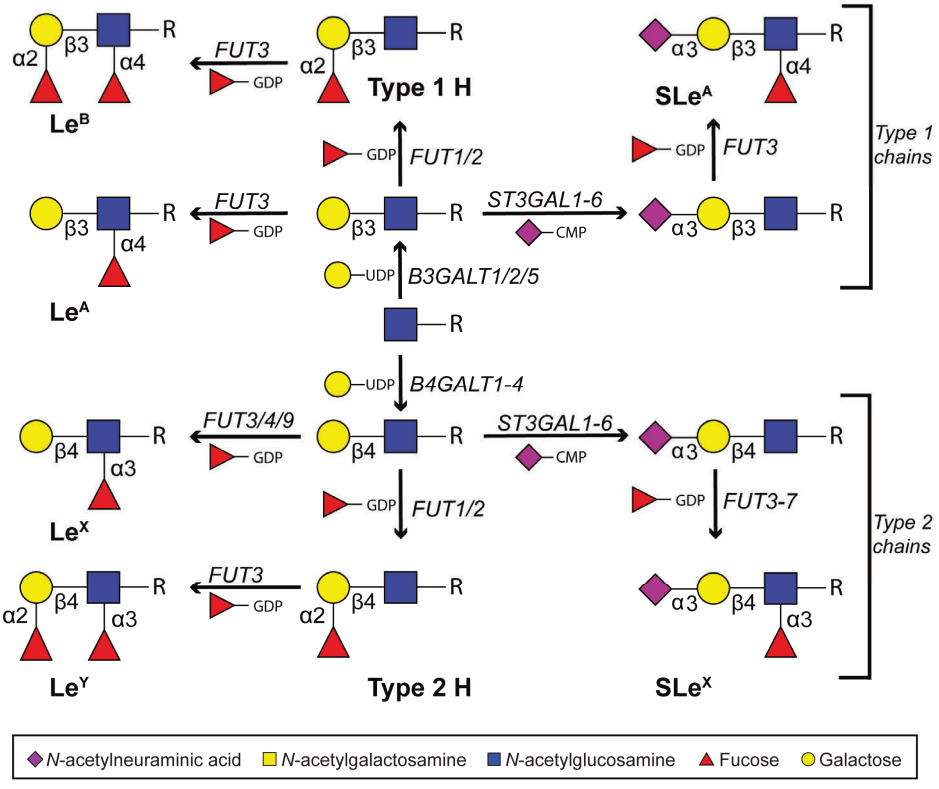


Figure 3. Common core extensions and terminal epitopes.

by action of α1-3-fucosyltransferases encoded by *FUT3, FUT4 or FUT9* on Type 2 chains and *FUT3* on Type 2 blood group H antigens, respectively[20].

A family of α2-3-sialyltransferases (*ST3GAL1-6)* is responsible for terminal sialylation of Gal, which is in competition with terminal fucosylation as described above (**Figure 3**). Fucosylation of the GlcNAc residue on sialylated substrates is still possible by the activity of *FUT3* or *FUT3/4/5/6/7* forming the sLe^A or sLe^X antigens, respectively. *ST3GAL1* is involved in the sialylation of core 1 substrates and, therefore, important for the biosynthesis of sialyl T and disialyl T antigens. On the other hand, α2-6 sialylation on *O*-glycans was described mostly in the context of the core GalNAc modification by the activity of a family of α2-6-sialyltransferases (*ST6GALNAC1-6*). *ST6GALNAC 1* and *2* are important for the biosynthesis of sialyl-6T, disialyl T and cancer associated sialyl-Tn antigen. Sialylation blocks further elongation of Gal by other monosaccharides other than the Neu5Ac, which can be added to form polysialic acids by the activity of a family of α2-8-sialyltransferases (*ST8SIA1-6*). Additionally, a specific epitope found in the colon named Sda antigen is formed by addition of a GalNAc to a Gal which is already substituted with an α2-3-NeuAc by β1-4-GalNAc transferase encoded by *B4GALNT2*[24]. This epitope is often found on core 3 glycans in the large intestine[24]. A similar antigen, named Cad, shares the same epitope sequence, however, it is carried by a core 1 *O*-glycan with a α2-6-linked Neu5Ac on the core GalNAc. Colonic glycans often show sulfation on the C3 position of Gal or at the C6 of the GlcNAc. The sulfation of GlcNAc is mediated by *CHST2,4-7* sulfotransferases, where *CHST2,4,5* and *7* are responsible for sulfation of the innermost GlcNAc in core 3 glycans. In addition, *CHST2* and *4* are involved in the sulfation of Lewis antigens[24]. The C3 sulfation of Gal is mediated by *GAL3ST2,3* and *4*, whereas the C6 sulfation of Gal is mediated by *CHST1* or *3*[25]. Sialic acid acetylation is also a common modification, which occurs on the level of the activated sugar donor CMP-Neu5Ac and is incorporated into the glycans by action of different sialyltransferases in the Golgi.

MUCIN TYPE GLYCOSYLATION IN CANCER

Mucins are a family of large secreted or membrane-bound glycoproteins, that play an important role in the epithelial cell homeostasis, acting as a protective shield[25].

The secreted mucins (MUC2, MUC5AC, MUC5B, MUC6, MUC7, and MUC19) play an important role in mucus rich tissues such as the gastrointestinal and respiratory tract. Of which, MUC1, MUC3, MUC4, MUC12, MUC16 and MUC17 are transmembrane mucins, which are also involved in cell signaling pathways. In CRC, low expression of MUC2 was reported in classical adenocarcinoma, while a high expression was reported in mucinous adenocarcinoma[26]. Additionally, higher MUC5AC expression was found in adenocarcinoma, while limited or no expression in normal colon mucosa[26]. The changes in the glycosylation that occur with cancer have shown to have an impact on the function of mucins, by changing their interaction with glycan binding proteins such as galectins[27], selectins[28] and siglecs[29]. These interactions can affect various cell processes including cell proliferation, adhesion, migration and immune surveillance which play an essential role in cancer pathogenesis[30–33]. Moreover, the very dense distribution of glycans in clustered PTS regions of mucins, as well as the possibility to target aberrant glycans present on different proteins on the cell surface, makes them attractive candidates for glycan-based immunotherapy as well as image guided surgery (IGS)[34].

The characteristic glycan alterations in cancer include specific aberrant expression of incomplete carbohydrate structures or de novo expression of carbohydrate antigens also known as tumor-associated carbohydrate antigens (TACAs). The expression of truncated structures such as Tn, sialyl-Tn, T and sialyl-T antigens was extensively described for various cancers including colon cancer[13,14]. Tn and Sialyl-Tn expression have been linked to somatic mutations in the chaperone essential for the function of core 1 synthase, COSMC (*C1GALTC1*) enabling the elongation of Tn antigen into core 1 structure, T antigen. However, this mechanism is applicable for a part of tumors expressing these antigens, and the fact that some tumors express both truncated and elongated glycans supports the hypothesis that there are other mechanisms involved such as silencing by promoter hypermethylation of COSMC[15,35]. Nevertheless, T synthase knock outs resulted in increased Tn antigen expression which enhanced cell proliferation, migration and invasiveness in colon cell cultures[36]. Additionally, the forced expression of truncated glycans directly induced oncogenic features in an organotypic epithelial tissue model[15].

The downregulation of core 3 and core 4 structures in cancer has been associated with both downregulation of core 3 synthase and upregulation of core 1 synthase. However, no study so far demonstrated the expression of those glycan types in patient derived cancer cells *vs* normal colon epithelia. The downregulation of core 3 synthase was described in colon and pancreatic adenocarcinoma and associated with poor patient prognosis and metastatic potential[17,18,37,38]. Moreover, the silencing of core 1 synthase induced increased expression of Tn, Sialyl-Tn and core 3 glycans in colon cancer cells[39]. On the other hand, the expression of the core 4 structures is dictated by the availability of core 3 precursors, and the activity of the C2GnT-M enzyme, encoded by the *GCNT3* gene, expressed in mucus secreting tissues[40]. *GCNT3* expression is downregulated in colon, ovarian and pancreatic cancer[41–43], and its transfection reduced cell proliferation, invasion and adhesion in CRC cells[40].

High expression of C2GnT-L type β1-6 GlcNAc transferase, which is responsible for biosynthesis of core 2 glycans, was described extensively in relation to cancer[44–47]. The formation of the 6-branch creates another scaffold for the activity of β1-3/4-galactosyltransferases creating Type 1 or Type 2 extensions, which have been shown to switch from primarily Type 1 in normal colon mucosa to Type 2 in CRC due to upregulation of B4GALT4[48]. Type 2 chains are precursors of sLe^X, and it has been shown that sLe^X expression is higher in CRC, whereas Le^A (Type 1) chains characteristic for normal mucosa are downregulated. The expression of β1-3-galactosyltransferase A Type 1 chain synthase, encoded by *B3GALT5,* has shown to be downregulated in colon cancer compared to normal colon mucosa[49]. It has also been previously described that the upregulation of $sLe^{X/A}$ antigens in CRC is due to the downregulation of *B4GALNT2* which encodes the β1-4-GalNAc transferase which synthesizes the Sda-antigen characteristic for the normal colon mucosa[24]. Additionally, the expression of sulfated (s)Le antigens was found to be high in normal colon epithelia, contrary to elevated sLe^X expression in CRC. Moreover, sLe^X antigens have been associated with promoting metastasis and showed correlation with poor patient prognosis[31,52]. Further research suggested that it is related to the interaction with selectins[53].

O-GLYCAN ANALYSIS

Dysregulation of mucin type *O*-glycosylation affects signaling pathways involved in cancer progression, therefore targeting mucin type glycosylation has a great clinical potential; however, it has remained an analytical challenge to enable an in-depth exploration as has been achieved for to *N*-glycosylation[53]. This is mainly due to the complexity and high heterogeneity of this type of glycosylation. Moreover, as it is not template driven, the cell glycosylation is a result of a complex interplay between different enzyme expression (GSTs and glycosidases), their competition, location within the endoplasmic reticulum and Golgi and the substrate availability. Additionally, compared to *N*-glycan analysis, *O*-glycan analysis brings along its own specific challenges as there is no universal enzyme that releases all types of *O*-glycans from glycoproteins. Therefore, the release is often performed with a chemical approach in alkaline conditions, called β-elimination, generally under reducing conditions to minimize glycan degradation (peeling). The non-reducing β-elimination is performed to enable reducing end glycan labeling, since the reduction results in non-modifiable glycan alditols. Additionally, hydrazinolysis can be used in the same manner, however, it requires special handling conditions due to the flammability of the chemicals used[54]. Additionally, *O*-glycans do not share a common core structure as *N*-glycans and can, therefore, express even more structural isomers for the same monosaccharide composition. This requires sophisticated analytical tools to be used for their analysis.

Glycan characterization from complex biological samples can be performed using a variety of methodologies that can be targeted or untargeted. Targeted approaches such as lectin and antibody binding assays evaluate the expression of known antigens whereas untargeted techniques such as mass spectrometry (MS) allow for discovery of novel glycan and glycoprotein targets.

GLYCOMICS USING MASS SPECTROMETRY

Glycan structure analysis can be performed in different ways, depending on the depth of the information needed. If the focus is on the in-depth structure sequencing of the glycan, techniques such as MS are often used. The most common ionization techniques applied for glycan analysis are matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), both can be

performed in positive and negative ionization mode. MALDI-MS is often used for analysis of released glycans; however, when measured in positive ionization mode, stabilization of sialic acids by permethylation or carboxyl group derivatization is necessary to avoid the partial loss of sialic acids due to in source and metastable decay. The obtained mass profiles provide insights into the possible monosaccharide compositions but is unable to provide information about structural isomers originating from positional or linkage variations. This can be partially overcome by tandem MS spectra (MS/MS), supplying more information about the glycan sequence. ESI can also generate positively or negatively charged ions while the most informative fragment ions for the elucidation of glycan structures are obtained by MS/MS collision induced dissociation (CID) of negatively charged precursors. In addition, the glycans do not undergo monosaccharide rearrangements, such as described in CID of positively charged precursors[55]. Nevertheless, sulfate modification rearrangements have been described in negative mode[56].

NEGATIVE IONIZATION MODE COLLISION INDUCED DISSOCIATION

CID fragmentation predominantly induces glycosidic C-type ions derived from non-reducing end, as well as A-type cross ring non-reducing end fragments, compared to glycosidic reducing end Y- and non-reducing end B-type ions (**Figure 4**). The pioneering work from Harvey *et al.,* set the groundwork for patterns in fragmentation of deprotonated *N*-glycan alditols (reviewed in [57]) where specific D- and E-type ions reveal the composition of the α1-6- and α1-3-linked glycan antennae, respectively. Specific negative ion fragmentation of mucin type glycans was described extensively by the Karlsson group and is reviewed elsewhere[58–60]. Briefly, linear glycans, such as core 1 and core 3 yield a B-ion from the loss of reducing end GalNAcol ($[M - H]^{1-}$ - 223) which decreases in intensity with the size of the glycan. Additionally, linear structures show a characteristic $[M - H]^{1-}$ - 108 ion, from the partial loss of side groups from the reducing end GalNAcol ($C_3H_8O_4$) which is never present in the branched structures with GlcNAc β1-6-linked to the core (cores 2 and 4). Dominant ions in the MS/MS spectra of the branched core 2 and core 4 are the Z-fragment ions from the glycosidic cleavage in between the core GalNAc and C3 branch. Additionally, a characteristic cross ring $A^{0,4}$ cleavage of the GalNAcol gives information about the composition of the C6 branch of the

glycan. Both core 1 and core 2 glycan fragmentation gives Y- and Z-ions at *m/z* 384^{1-} and 366^{1-}, respectively, whereas the core 2 additionally generate YY-, YZ-

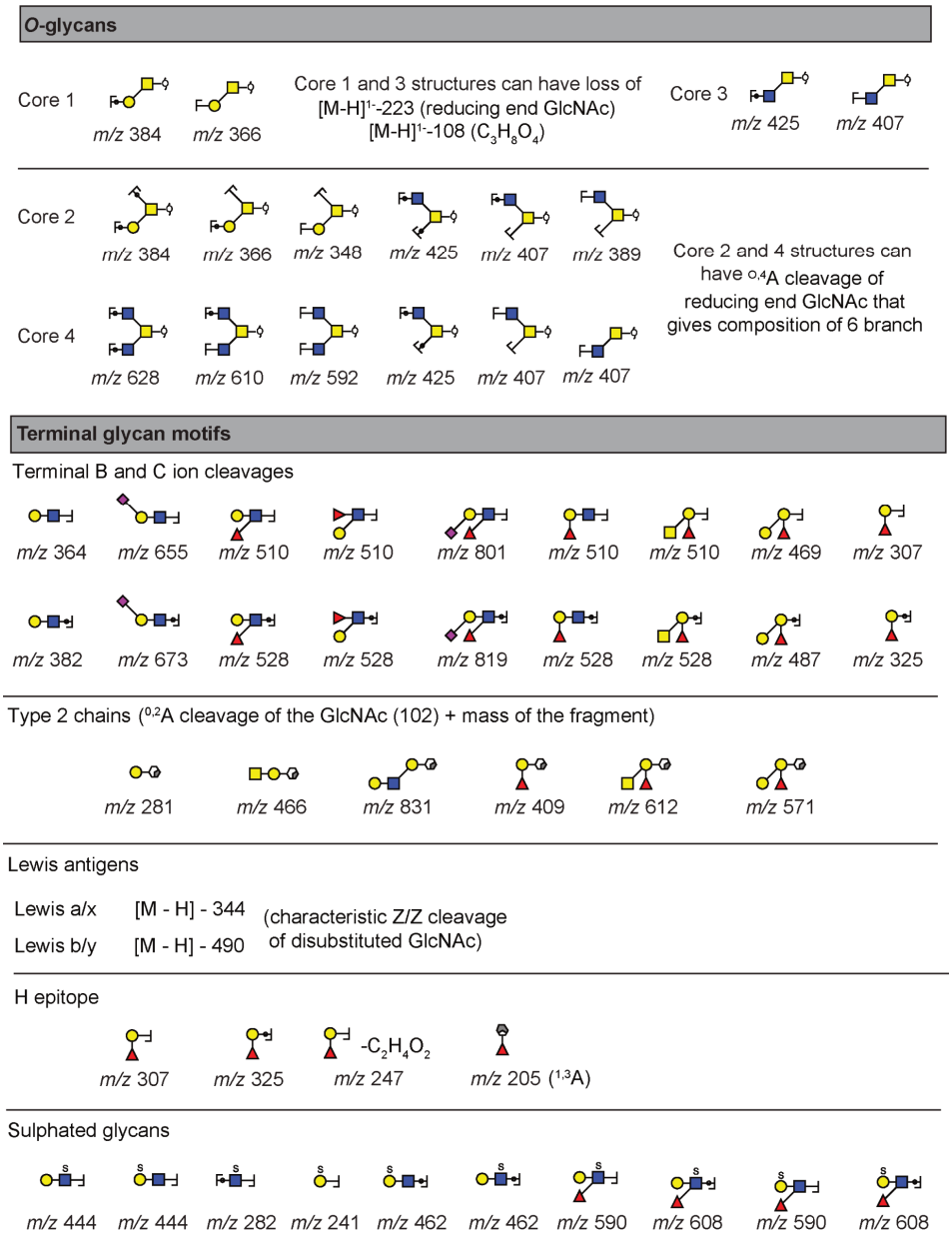


Figure 4. Characteristic fragment ions in negative mode tandem mass spectra analysis of reduced mucin type O-glycans. *Reprinted with permission from Everest-Dass et al.*[60]

and ZZ-ions ions at *m/z* 425^{1-}, $407^{1-,}$ and 389^{1-}. Type 2 LacNAc chains give specific cross ring $A^{0,2}$ cleavage, together with a concurrent ion arising from loss of water $A^{0,2} - H_2O$. While there are no diagnostic ions for Type 1 structures, the absence of the characteristic Type 2 cleavages can be used as an indication. Moreover, negative ion mode fragmentation spectra produce fragment ions enabling determination of the fucose position. Lewis type fucosylation gives rise to specific elimination of both C3 and C4 substituents of GlcNAc residue generating Z/Z and Z/Z- CH_2O ions. These ions are also present in the spectra of Le^B and Le^Y structures, although at lower intensity. These ions can also form an additional loss of acetyl and acetate groups from other acetylated monosaccharides (Z/Z-C_2H_2O and Z/Z-$C_2H_4O_2$). Unfortunately, no diagnostic ions are formed in MS/MS to distinguish between Le^X and Le^A, as well as Le^B and Le^Y. Blood group H determinants induce weak B- and C-type cleavages, at *m/z* 307^{1-}, *m/z* 325^{1-}, respectively. Additional loss of acetate (B2-$C_2H_4O_2$) at *m/z* 247^{1-}, and the cross ring fragment of galactose ($A^{1,3}$) at *m/z* 205^1 indicate an α1-2-linked fucose, however, those fragments can only be observed in MS/MS for small glycan structures. Moreover, $A^{0,2}$ cleavage with a loss of water from the GlcNAc at *m/z* 409^{1-}, is characteristic for Type 2 blood group H antigen. The Z- and Y-ions at *m/z* 495^{1-} and 513^{1-}, respectively, are diagnostic for α2-6-sialylation of the GalNAc and a cross ring $X^{0,2}$ can give additional indication of the presence of the α2-6-linked Neu5Ac. Determining the position of the sulfate modification on a particular monosaccharide is a very challenging task. The ions indicating the position of a sulfate on either Gal or GlcNAc are B-ions at *m/z* 444^{1-} which in case of fucosylation shifts to *m/z* 590^1. The location of the sulfate can be determined by MS/MS for very small structures, or by MS^3, where fragment ions at *m/z* 241^{1-} and 282^{1-} are indicative of its linkage to Gal and GlcNAc, respectively.

Of note, it has been shown that negative ion mode fragmentation performed by ion trap (IT) instruments produces reproducible spectra, even if instruments from different laboratories and manufacturers are used[61]. Therefore, a repository and a database Unicarb DR and DB, respectively, aims to collect a large number of annotated glycan spectra to facilitate spectral matching and automatic assignments of glycan structures[62].

However, ESI alone is unable to distinguish between glycan isomers, therefore coupling ESI with online separation techniques is often a preferred choice, particularly for the analysis of mucin type *O*-glycans.

ISOMERIC SEPARATION TECHNIQUES

Several liquid chromatography (LC) separation techniques are used for isomeric separation of glycans from complex samples including reverse phase (RP), hydrophilic interaction liquid chromatography (HILIC), porous graphitized carbon (PGC), and mesoporous graphitic carbon (MGC). RP-C18-MS was used recently for the analysis of both *N*- and *O*-linked permethylated glycans and demonstrated baseline separation of glycan isomers, together with mobile phase optimization by adding lithium which limited adduct heterogeneity and simplified data analysis[63]. New approaches include the use of micro pillar array (qPAC) columns for the analysis of permethylated glycans, demonstrating isomeric separation of *N*- and *O*-glycans, although the diversity of *O*-glycans tested was very limited[64]. Unlike RP, HILIC-LC can be used for the separation of non-permethylated, native or reducing end labeled glycans, and has recently been demonstrated as a valuable tool also for the isomeric separation of *N*-glycans[65,66]. Despite increasing efficiency of HILIC and RP stationary phases in separating glycan isomers, due to a specific combination of hydrophobic and polar interactions of carbohydrates with the planar surface of the graphite, PGC outstands with its superior results in separating both linkage and positional glycan isomers[67].

The usage of PGC chromatography was pioneered by the laboratory of prof. Packer for glycan desalting and purification prior to MS analysis using the material for solid phase extraction (SPE)[68]. It was thereafter employed for HPLC analysis or oligosaccharides where it showed excellent separation power for structural and linkage isomers[69]. Moreover, the glycans can be analyzed without prior derivatization, in their native form or after reduction thereby avoiding separation of reducing end α- and β-anomers. The advantages include shorter sample preparation, less sample loss due to additional purification steps and incomplete derivatization. Nevertheless, the PGC-LC was also used for the analysis of permethylated glycans, enabling separation at higher temperatures[70,71]. Coupling of PGC-LC with ESI in positive ion mode is used less frequently[72], although the signal intensities of neutral and negatively charged glycan species are more

comparable to UV detection LC of labeled glycans[73]. Despite the bias in ionization of negatively charged species, PGC is more often coupled with negative ion ESI, which results in MS/MS spectra that contain informative cross-ring cleavages useful for determining glycan structures[69]. Specific elution orders are observed for *N*-linked glycans such as bisected *N*-glycans elute prior to their branched isomers, core fucosylated glycans elute later than their antenna fucosylated isomer pairs, and finally α2-6-linked sialylated glycans elute earlier than their α2-3-linked sialylated isomers. However, in the case of specifically mucin type *O*-glycans, elution patterns are difficult to predict, as they do not retain based on the glycan size, as exemplified by the linear trisaccharide Fucα1-2Galβ1-3GalNAcol which elutes around 30 minutes after the branched trisaccharide Galβ1-3[GlcNAcβ1-6]GalNAcol (unpublished observation).

In summary, PGC-ESI-MS/MS in negative ion mode allows the analysis of mucin type glycosylation in sensitive and an in-depth manner, allowing chromatographic separation of different linkage and positional isomers, and their structural characterization via negative ion mode CID. In this way cancer specific glycan signatures can be revealed giving new insights into the cancer specific glycosylation pathways originating from the changes in the cell biosynthetic machinery.

SCOPE OF THE THESIS

The scope of this thesis was to investigate mucin type glycomic signatures of CRC in order to explore new potential therapeutic targets for immunotherapy.

Previous studies from Kolarich *et. al.*,[69] showed the potential of using polyvinylidene difluoride (PVDF) membrane-based protein immobilization strategy to facilitate sequential release of both *N*- and *O*-glycans from the same sample, followed by analysis with PGC-LC-MS/MS allowing isomeric separation and in-depth structural identification by negative ion mode CID. In **Chapter 2** we optimized their approach, allowing a higher throughput for sample preparation using 96-well plates enabling a robust and sensitive analysis of a greater number of samples at a time. In **Chapter 3** we explored the effects of different polar protic solvents for dopant enriched nitrogen gas in order to improve the sensitivity for the analysis of different *N*- and *O*-glycan species and to evaluate its effect on the charge state distribution.

In **Chapter 4** we applied the methodology developed in Chapter 1 and characterized the *O*-glycomes of 26 CRC cell lines. This work was a continuation of the previous work from our group characterizing the same set of cell lines for their *N*-glycomes[74]. Furthermore, the *O*-glycan phenotypes were associated with the GST expression relevant for their biosynthesis. Inspired by the associations of the cell line glycomes and cell differentiation, we evaluated in **Chapter 5** the changes in cell line glycome upon butyrate stimulation and spontaneous differentiation in culture. Moreover, we explored the association between the glycome changes with the changes in the cell proteome. **In Chapter 6** we optimized a previously developed protocol for *N*- and *O*-glycomics from formalin fixed paraffin embedded tissues and applied it for the analysis of CRC tissues and their patient matched colon mucosa controls. This allowed us to identify CRC specific glycan signatures, which showed association with the differential regulation of specific biosynthetic pathways in cancer previously described on the transcriptomic level.

Finally, **Chapter 7** gives a general discussion about the technical challenges encountered related to the analysis of *O*-glycans, evaluation of different models for studying glycomic changes in cancer, the association between transcriptomic signatures and glycomic profiles, as well as future perspectives on how our results could be used in clinical applications, as well as the remaining challenges to be addressed.

REFERENCES

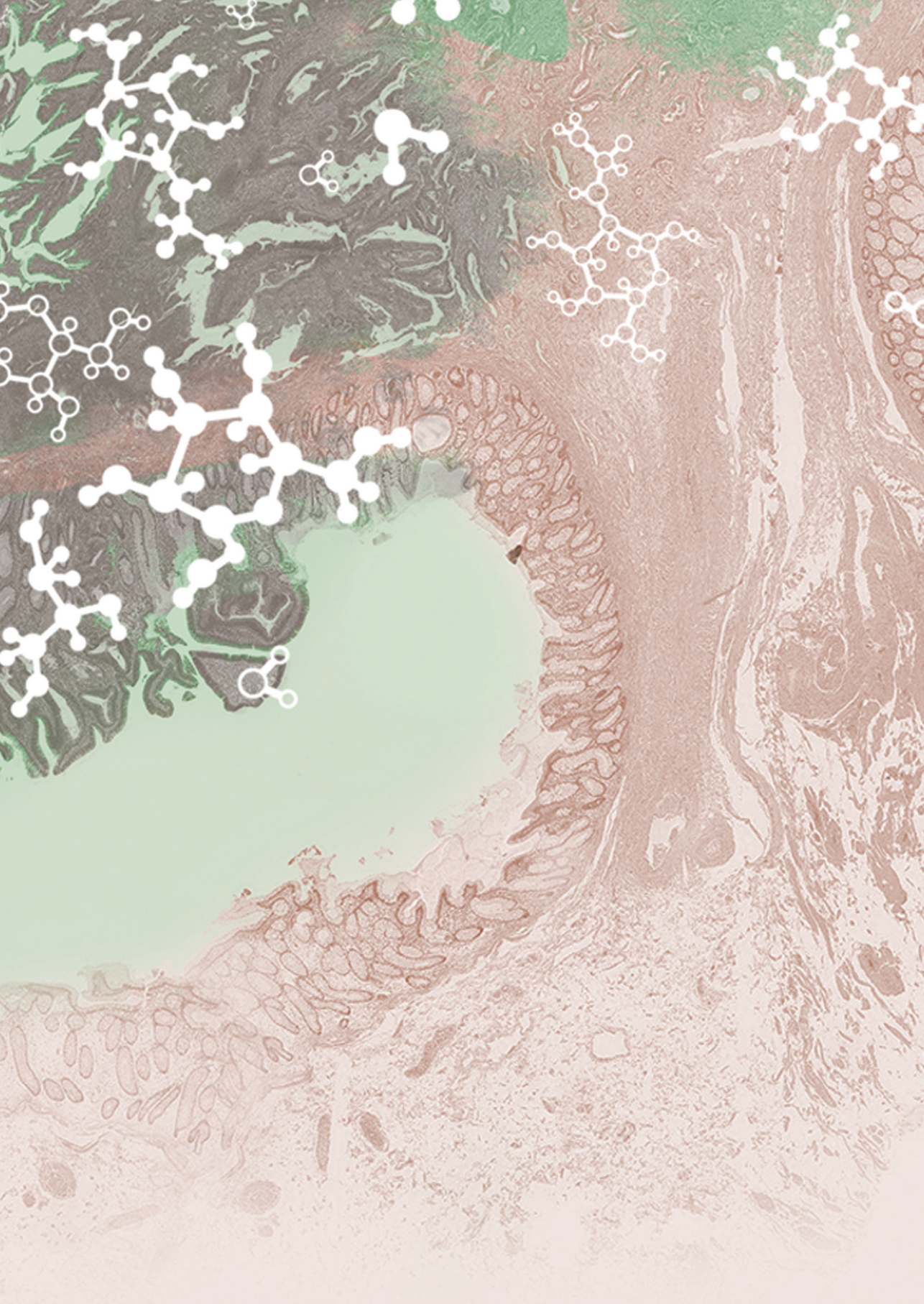
1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J. Clin. 71, 209–249 (2021).
2. Chung, D. C. The Genetic Basis of Colorectal Cancer: Insights Into Critical Pathways of Tumorigenesis. doi:10.1053/gast.2000.16507.
3. Angelova, M. et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biol. 16, 64 (2015).
4. Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. Nature Medicine (2015) doi:10.1038/nm.3967.
5. Ghani, S. et al. Recent developments in antibody derivatives against colorectal cancer; A review. Life Sci. 265, 118791 (2021).
6. Piawah, S. & Venook, A. P. Targeted therapy for colorectal cancer metastases: A review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer. *Cancer* **125**, 4139–4147 (2019).
7. Ayyar, B. V., Arora, S. & O'Kennedy, R. Coming-of-Age of Antibodies in Cancer Therapeutics. *Trends Pharmacol. Sci.* **37**, 1009–1028 (2016).
8. Holst, S., Wuhrer, M. & Rombouts, Y. *Glycosylation characteristics of colorectal cancer*. vol. 126 203–256 (Elsevier Inc., 2015).
9. Mereiter, S., Balmaña, M., Campos, D., Gomes, J. & Reis, C. A. Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? *Cancer Cell* vol. 36 6–16 (2019).
10. Stanley P, Cummings RD. Structures Common to Different Glycans. 2017. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 14.
11. Schjoldager, K. T., Narimatsu, Y., Joshi, H. J. & Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749 (2020).
12. Schjoldager, K. T.-B. G. & Clausen, H. Site-specific protein O-glycosylation modulates proprotein processing — Deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 2079–2094 (2012).
13. Cornelissen, L. A. M. *et al.* Tn Antigen Expression Contributes to an Immune Suppressive Microenvironment and Drives Tumor Growth in Colorectal Cancer. *Front. Oncol.* **10**, 1–15 (2020).
14. Matsumoto, T. *et al.* Tn antigen expression defines an immune cold subset of mismatch-repair deficient colorectal cancer. *Int. J. Mol. Sci.* **21**, 1–13 (2020).
15. Ju, T. *et al.* Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res.* **68**, 1636–1646 (2008).
16. Yeh, J. C., Ong, E. & Fukuda, M. Molecular cloning and expression of a novel β-1,6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I branches. *J. Biol. Chem.* **274**, 3215–3221 (1999).

17. Doi, N. *et al.* Clinicopathological significance of core 3 O-glycan synthetic enzyme, β1,3-N-acetylglucosaminyltransferase 6 in pancreatic ductal adenocarcinoma. *PLoS One* **15**, e0242851 (2020).
18. An, G. *et al.* Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived O-glycans. *J. Exp. Med.* **204**, 1417–1429 (2007).
19. Vavasseur, F., Yang, J. M., Dole, K., Paulsen, H. & Brockhausen, I. Synthesis of O-glycan core 3: Characterization of UDP-GlcNAc: GalNAc-R β3-N-acetyl-glucosaminyltransferase activity from colonic mucosal tissues and lack of the activity in human cancer cell lines. *Glycobiology* **5**, 351–357 (1995).
20. Lee, S. H. *et al.* Core3 O-glycan synthase suppresses tumor formation and metastasis of prostate carcinoma PC3 and LNCaP cells through down-regulation of α2β1 integrin complex. *J. Biol. Chem.* **284**, 17157–17169 (2009).
21. Saarinen, J., Welgus, H. G., Flizar, C. A., Kalkkinen, N. & Helin, J. N-Glycan structures of matrix metalloproteinase-1 derived from human fibroblasts and from HT-1080 fibrosarcoma cells. *Eur. J. Biochem.* **259**, 829–840 (1999).
22. Dell, A. *et al.* Structural Analysis of the Oligosaccharides Derived from Glycodelin, a Human Glycoprotein with Potent Immunosuppressive and Contraceptive Activities (∗). *Journal of Biological Chemistry* **270**, 24116–24126 (1995).
23. Dotz, V. & Wuhrer, M. Histo-blood group glycans in the context of personalized medicine. *Biochimica et Biophysica Acta - General Subjects* **1860**, 1596–1607 (2016).
24. Groux-Degroote, S. *et al.* B4GALNT2 gene expression controls the biosynthesis of Sda and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. *Int. J. Biochem. Cell Biol.* **53**, 442–449 (2014).
25. Corfield, A. P. Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochimica et Biophysica Acta - General Subjects* vol. 1850 (2015).
26. Krishn, S. R. *et al.* Mucins and associated glycan signatures in colon adenoma-carcinoma sequence: Prospective pathological implication(s) for early diagnosis of colon cancer. *Cancer Lett.* **374**, 304–314 (2016).
27. Belo, A. I., Van Der Sar, A. M., Tefsen, B. & Van Die, I. Galectin-4 Reduces Migration and Metastasis Formation of Pancreatic Cancer Cells. (2013) doi:10.1371/journal.pone.0065957.
28. Trinchera, M. *et al.* Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. *Biology* **6**, 16 (2017).
29. Rodriguez, E. *et al.* Sialic acids in pancreatic cancer cells drive tumour-associated macrophage differentiation via the Siglec receptors Siglec-7 and Siglec-9. *Nat. Commun.* **12**, 1–14 (2021).
30. Rabinovich, G. A. & Croci, D. O. Regulatory Circuits Mediated by Lectin-Glycan Interactions in Autoimmunity and Cancer. *Immunity* vol. 36 322–335 (2012).
31. Witz, I. P. The selectin-selectin ligand axis in tumor progression. *Cancer Metastasis Rev.* **27**, 19–30 (2008).

32. Bärenwaldt, A. & Läubli, H. The sialoglycan-Siglec glyco-immune checkpoint–a target for improving innate and adaptive anti-cancer immunity. *Expert Opinion on Therapeutic Targets* vol. 23 (2019).
33. Mantuano, N. R., Natoli, M., Zippelius, A. & Läubli, H. Tumor-associated carbohydrates and immunomodulatory lectins as targets for cancer immunotherapy. *Journal for ImmunoTherapy of Cancer* **8**, e001222 (2020).
34. Houvast, R. D. *et al.* Targeting glycans and heavily glycosylated proteins for tumor imaging. *Cancers* vol. 12 1–26 (2020).
35. Ju, T., Aryal, R. P., Kudelka, M. R., Wang, Y. & Cummings, R. D. The Cosmc connection to the Tn antigen in cancer. *Cancer Biomark.* **14**, 63–81 (2014).
36. Dong, X. *et al.* T-Synthase Deficiency Enhances Oncogenic Features in Human Colorectal Cancer Cells via Activation of Epithelial-Mesenchymal Transition. *Biomed Res. Int.* **2018**, (2018).
37. Iwai, T. *et al.* Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4572–4577 (2005).
38. Henion, T. R. & Schwarting, G. A. N-Linked polylactosamine glycan synthesis is regulated by co-expression of β3GnT2 and GCNT2. *J. Cell. Physiol.* **229**, 471–478 (2014).
39. Barrow, H., Tam, B., Duckworth, C. A., Rhodes, J. M. & Yu, L. G. Suppression of Core 1 Gal-Transferase Is Associated with Reduction of TF and Reciprocal Increase of Tn, sialyl-Tn and Core 3 Glycans in Human Colon Cancer Cells. *PLoS One* **8**, e59792 (2013).
40. Huang, M. C. *et al.* C2GnT-M is downregulated in colorectal cancer and its re-expression causes growth inhibition of colon cancer cells. *Oncogene* **25**, 3267–3276 (2006).
41. Fernández, L. P. *et al.* The role of glycosyltransferase enzyme GCNT3 in colon and ovarian cancer prognosis and chemoresistance. *Sci. Rep.* **8**, 8485 (2018).
42. González-Vallinas, M. *et al.* Clinical relevance of the differential expression of the glycosyltransferase gene GCNT3 in colon cancer. *Eur. J. Cancer* **51**, 1–8 (2015).
43. Gupta, R. *et al.* Global analysis of human glycosyltransferases reveals novel targets for pancreatic cancer pathogenesis. *Br. J. Cancer* **122**, 1661–1672 (2020).
44. Chen, Z., Gulzar, Z. G., St Hill, C. A., Walcheck, B. & Brooks, J. D. Increased expression of GCNT1 is associated with altered O-glycosylation of PSA, PAP, and MUC1 in human prostate cancers. *Prostate* **74**, 1059–1067 (2014).
45. Dalziel, M. *et al.* The Relative Activities of the C2GnT1 and ST3Gal-I Glycosyltransferases Determine O-Glycan Structure and Expression of a Tumor-associated Epitope on MUC1. *J. Biol. Chem.* **276**, 11007–11015 (2001).
46. St Hill, C. A., Baharo-Hassan, D. & Farooqui, M. C2-O-sLeX glycoproteins are E-selectin ligands that regulate invasion of human colon and hepatic carcinoma cells. *PLoS One* **6**, e16281 (2011).
47. St Hill, C. A., Bullard, K. M. & Walcheck, B. Expression of the high-affinity selectin glycan ligand C2-O-sLeX by colon carcinoma cells. *Cancer Lett.* **217**, 105–113 (2005).

48. Chen, W.-S., Chang, H.-Y., Li, C.-P., Liu, J. M. & Huang, T.-S. Tumor beta-1,4-galactosyltransferase IV overexpression is closely associated with colorectal cancer metastasis and poor prognosis. *Clin. Cancer Res.* **11**, 8615–8622 (2005).
49. Salvini, R., Bardoni, A., Valli, M. & Trinchera, M. β1,3-Galactosyltransferase β3Gal-T5 Acts on the GlcNAcβ1→3Galβ1→4GlcNAcβ1→R Sugar Chains of Carcinoembryonic Antigen and Other N-Linked Glycoproteins and Is Down-regulated in Colon Adenocarcinomas *. *J. Biol. Chem.* **276**, 3564–3573 (2001).
50. Nakamori, S. *et al.* Increased expression of sialyl Lewisx antigen correlates with poor survival in patients with colorectal carcinoma: clinicopathological and immunohistochemical study. *Cancer Res.* **53**, 3632–3637 (1993).
51. Yamada, N. *et al.* Increased expression of sialyl Lewis A and sialyl Lewis X in liver metastases of human colorectal carcinoma. *Invasion Metastasis* **15**, 95–102 (1995).
52. Paschos, K. A., Canovas, D. & Bird, N. C. The engagement of selectins and their ligands in colorectal cancer liver metastases. *J. Cell. Mol. Med.* **14**, 165–174 (2010).
53. Lageveen-Kammeijer, G. S. M., Kuster, B., Reusch, D. & Wuhrer, M. High sensitivity glycomics in biomedicine. *Mass Spectrom. Rev.* e21730 (2021).
54. Kozak, R. P., Royle, L., Gardner, R. A., Fernandes, D. L. & Wuhrer, M. Suppression of peeling during the release of O-glycans by hydrazinolysis. *Anal. Biochem.* **423**, 119–128 (2012).
55. Wuhrer, M., Koeleman, C. A. M., Hokke, C. H. & Deelder, A. M. Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. *Rapid Commun. Mass Spectrom.* **20**, 1747–1754 (2006).
56. Kenny, D. T., Issa, S. M. A. & Karlsson, N. G. Sulfate migration in oligosaccharides induced by negative ion mode ion trap collision-induced dissociation. *Rapid Commun. Mass Spectrom.* **25**, 2611–2618 (2011).
57. Harvey, D. J. Negative Ion Mass Spectrometry for the Analysis of N -linked Glycans. *Mass Spectrom. Rev.* **39**, 586–679 (2020).
58. Karlsson, N. G., Schulz, B. L. & Packer, N. H. Structural determination of neutral O-linked oligosaccharide alditols by negative ion LC-electrospray-MSn. *J. Am. Soc. Mass Spectrom.* **15**, 659–672 (2004).
59. Robbe, C., Capon, C., Coddeville, B. & Michalski, J. C. Diagnostic ions for the rapid analysis by nano-electrospray ionization quadrupole time-of-flight mass spectrometry of O-glycans from human mucins. *Rapid Commun. Mass Spectrom.* **18**, 412–420 (2004).
60. Everest-Dass, A. V., Abrahams, J. L., Kolarich, D., Packer, N. H. & Campbell, M. P. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J. Am. Soc. Mass Spectrom.* **24**, 895–906 (2013).
61. Campbell, M. P. *et al.* Validation of the curation pipeline of UniCarb-DB: Building a global glycan reference MS/MS repository. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1844**, 108–116 (2014).
62. Rojas-Macias, M. A. *et al.* Towards a standardized bioinformatics infrastructure for N- and O-glycomics. *Nat. Commun.* **10**, 3275 (2019).

63. Kurz, S., Sheikh, M. O., Lu, S., Wells, L. & Tiemeyer, M. Separation and Identification of Permethylated Glycan Isomers by Reversed Phase NanoLC-NSI-MSn. *Mol. Cell. Proteomics* **20**, 100045 (2021).
64. Cho, B. G., Jiang, P., Goli, M., Gautam, S. & Mechref, Y. Using micro pillar array columns (μPAC) for the analysis of permethylated glycans. *Analyst* **146**, 4374–4383 (2021).
65. Messina, A. *et al.* HILIC-UPLC-MS for high throughput and isomeric N-glycan separation and characterization in Congenital Disorders Glycosylation and human diseases. *Glycoconj. J.* **38**, 201–211 (2021).
66. Moravcová, D., Čmelík, R. & Křenková, J. Separation of labeled isomeric oligosaccharides by hydrophilic interaction liquid chromatography - the role of organic solvent in manipulating separation selectivity of the amide stationary phase. *J. Chromatogr. A* **1651**, 462303 (2021).
67. Pabst, M., Bondili, J. S., Stadlmann, J., Mach, L. & Altmann, F. Mass + Retention Time = Structure: A Strategy for the Analysis of N -Glycans by Carbon LC-ESI-MS and Its Application to Fibrin N -Glycans. *Anal. Chem.* **79**, 5051–5057 (2007).
68. Packer, N. H., Lawson, M. A., Jardine, D. R. & Redmond, J. W. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconj. J.* **15**, 737–747 (1998).
69. Jensen, P. H., Karlsson, N. G., Kolarich, D. & Packer, N. H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat. Protoc.* **7**, 1299–1310 (2012).
70. Zhou, S., Dong, X., Veillon, L., Huang, Y. & Mechref, Y. LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Anal. Bioanal. Chem.* **409**, 453–466 (2017).
71. Zhou, S., Hu, Y. & Mechref, Y. High-temperature LC-MS/MS of permethylated glycans derived from glycoproteins. *Electrophoresis* **37**, 1506–1513 (2016).
72. Hua, S. *et al.* Isomer-specific chromatographic profiling yields highly sensitive and specific potential N-glycan biomarkers for epithelial ovarian cancer. *J. Chromatogr. A* **1279**, 58–67 (2013).
73. Pabst, M. & Altmann, F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Anal. Chem.* **80**, 7534–7542 (2008).
74. Holst, S. *et al.* N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homebox 1 (CDX1)/Villin mRNA Expression. *Mol. Cell. Proteomics* **15**, 124–140 (2016).



DEVELOPMENT OF A 96-WELL PLATE SAMPLE PREPARATION METHOD FOR INTEGRATED *N*- AND *O*-GLYCOMICS USING POROUS GRAPHITIZED CARBON LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY

Chapter 2

Tao Zhang[‡,a], Katarina Madunić[‡,a], Stephanie Holst[a], Jing Zhang[b], Chunsheng Jin[c], Peter ten Dijke[b], Niclas G. Karlsson[c], Kathrin Stavenhagen[a] and Manfred Wuhrer[*,a]

[‡] *These authors contributed equally to this paper.*

[a] *Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands.*

[b] *Department of Cell and Chemical Biology and Oncode Institute, Leiden University Medical Center, Leiden, The Netherlands*

[c] *Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Sweden*

Reproduced from Mol. Omics, 2020, 16, 355-363, DOI: 10.1039/C9MO00180H with permission from the Royal Society of Chemistry.

ABSTRACT

Changes in glycosylation signatures of cells have been associated with pathological processes in cancer as well as infectious and autoimmune diseases. The current protocols for comprehensive analysis of *N*-glycomics and *O*-glycomics derived from cells and tissues often require a large amount of biological material. They also only allow the processing of very limited numbers of samples at a time. Here we established a workflow for sequential release of *N*-glycans and *O*-glycans based on PVDF membrane immobilization in 96-well format from 5 x 10^5 cells. Released glycans are reduced, desalted, purified, and reconstituted all in 96-well format plates, without additional staining or derivatization. Glycans are then analyzed on porous graphitized carbon nano-liquid chromatography coupled to tandem mass spectrometry using negative-mode electrospray ionization, enabling the chromatographic resolution and structural elucidation of glycan species including many compositional isomers. The approach was demonstrated using glycoprotein standards and further applied to analyze the glycosylation of the murine mammary gland NMuMG cell line. The developed protocol allows the analysis of *N*- and *O*-glycans from relatively large numbers of samples in a less time consuming way with high repeatability. Inter- and intraday repeatability of the fetuin *N*-glycan analysis showed two median intraday coefficients of variations (CVs) of 7.6% and 8.0%, and a median interday CV of 9.8%. Median CVs of 7.9% and 8.7% for the main peaks of *N*- and *O*-glycans released from NMuMG cell line indicate a very good repeatability. The method is applicable to purified glycoproteins as well as to biofluids and cell- or tissue-based samples.

INTRODUCTION

Protein glycosylation is involved in many biological processes such as cellular signaling, eliciting of immune responses and cancer progression[1,2]. Glycosylation studies focus on defining the diversity of glycan structures carried by specific glycoproteins or whole cells facilitating the understanding of the contribution of glycans in biological processes and the development of diseases. Due to the structural complexity and heterogeneity of glycans in complex biological samples, the available methodologies often have difficulties to achieve a comprehensive characterization, and require large amounts of biological material[3].

Mass spectrometry (MS)-based glycomics has become one of the most powerful methods for analyzing glycans released from glycoproteins and has vastly benefitted from rapid advances in sample preparation, chromatographic separation, MS methodology and data processing[4]. Importantly, MS structural analysis enhanced by powerful separation may provide detailed information and deeper understanding of glycan expression at high sensitivity. However, the low ionization efficiency, sialic acid decay and fucose migration in MS-based glycomics often pose problems in glycan characterization[5-8]. To address these challenges, derivatization and/or labeling followed by additional purification steps are employed. However, these methodologies often lack sufficient chromatographic separation resulting in failure to differentiate isomeric glycan species.

Another challenge in the field is the increasing need for high throughput sample preparation for glycan analysis, due to a vast increase of sample numbers and complexity required in functional glycomics, systems biology, and clinical applications. Currently, there are several high throughput approaches for *N*-glycan analysis (mainly glycoprotein)[9-11], while less methods are available for *O*-glycomics[12]. Therefore, there is a high demand for a high throughput, reproducible and robust analytical methods for integrated *N*- and *O*-glycomics. Scaling up the conventional glycomics analysis to higher-throughput approaches especially with respect to the sample preparation workflow is of

great importance for ensuring sufficient sample size providing more reliable results.

Here, we present an integrated *N*- and *O*-glycomics approach for sequential release of *N*- and *O*-glycans from biological samples based on protein immobilization on polyvinylidene fluoride (PVDF) membrane filter plates in 96-well format. The approach is based on methodology developed by Packer and coworkers[13, 14], who have pioneered the analysis of glycan alditols by porous graphitized carbon nano-liquid chromatography coupled to a tandem mass spectrometer (PGC nano-LC-ESI-MS/MS) using negative electrospray ionization[13]. Powerful separation capacity enabling discrimination between glycan isomers is a major advantage of PGC chromatography[15,16]. Taking their methodology as a starting point[13], considerable modifications were introduced with respect to sample preparation in order to allow a repeatable and less time consuming analysis in a higher-throughput manner. The method was demonstrated by analysis of glycans released from glycoprotein standards, and finally applied to the characterization of a murine breast cancer cell line. The method is applicable to a range of complex biological samples, including biofluids, cell lines and tissues.

MATERIALS AND METHODS

MATERIALS

Ammonium bicarbonate (ABC), Dowex cation-exchange resin (50W-X8), trifluoroacetic acid (TFA), fetuin from fetal bovine serum (FBS), hydrochloric acid (HCl), DL-dithiothreitol (DTT), ammonium acetate, sodium chloride (NaCl), and sodium borohydride ($NaBH_4$) were obtained from Sigma-Aldrich (Steinheim, Germany). 8 M guanidine hydrochloride (GuHCl), Dulbecco's modified Eagle's medium (DMEM) and FBS were purchased from Thermo Fisher Scientific. Peptide *N*-glycosidase F (PNGase F) was purchased from Roche Diagnostics (Mannheim, Germany). Glacial acetic acid (HAc) and potassium hydroxide (KOH) were purchased from Honeywell Fluka. Solid phase extraction (SPE) bulk sorbent Carbograph was obtained from Grace Discovery sciences (Columbia, USA). HPLC SupraGradient acetonitrile (MeCN) was obtained from Biosolve (Valkenswaard, The Netherlands) and other reagents and solvents

such as methanol (MeOH), ethanol, and 2-propanol were purchased from Merck (Darmstadt, Germany). MultiScreen® HTS 96 multiwell plates (pore size 0.45 µm) with high protein-binding membrane (hydrophobic Immobilon-P PVDF membrane) and 96-well PP Microplate were purchased from Millipore (Amsterdam, The Netherlands), conical 96-well Nunc plates from Thermo Scientific (Roskilde, Denmark). 96-well PP filter plate from Orochem Technologies (Naperville, IL, USA). Ultrapure water was used for the all preparations and washes, generated from a Q-Gard 2 system (Millipore).

CELL CULTURE AND CELL LYSIS

Mouse mammary gland (NMuMG) epithelial cell lines were obtained from American Type Culture Collection (Manassas, VA, USA) and cultured as previously described[17]. In brief, the cells were cultured in DMEM, supplied with 100 U/mL penicillin-streptomycin and 10% FBS at 37°C in 5% CO_2 and harvested after reaching 80% confluency. Pre-warmed trypsin was added to detach the cells. After detaching more than 90% of the cells, the equivalent of 2 volumes of pre-warmed complete DMEM was added to quench the enzyme activity. The cells were collected from the flask by washing with pre-warmed phosphate buffered saline (1x PBS) for 3 times. Thereafter, the cells were transferred to a 15 mL conical tube and centrifuged at 1000 rpm for 5 min. Cells were washed with PBS to remove the medium for 3 times by resuspension and centrifugation. The cells were counted using the Countess® Automated Cell Counter. The cells were washed twice with 5 mL of 1 x PBS, aliquoted to 2.0 x 10^6 cells per mL of 1 x PBS and pelleted by centrifuging 3 min at 1500 *g*. Finally, the supernatant was removed, and the cell pellets were stored in the freezer at -20 °C.

Mechanical cell lysis was performed in water as described previously[18]. Briefly, cell pellets with approximately 2 x 10^6 cells were resuspended in 100 µL of water by pipetting for 30 sec and vortexed for 30 sec followed by sonication for 60 min in an ultrasound bath at room temperature (RT).

GLYCOPROTEIN AND CELL LYSATE IMMOBILIZATION, DENATURATION AND REDUCTION

The 96-well plates with hydrophobic Immobilon-P PVDF membrane were preconditioned with 200 µL 70 % ethanol followed by equilibration 3 times with

200 µL water. Purified glycoproteins, biofluids and cell lysates (derived from cell lines, organoids and tissue samples) can be applied to the PVDF-membrane directly. In this work, either 2 µL (20 µg in water) fetuin aliquot or 25 µL cell lysate (approximately 5 x 10^5 cells) were applied per well followed by horizontal shaking for 15 min. Protein denaturation was achieved by applying 75 µL denaturation mix (72.5 µL 8 M GuHCl and 2.5 µL 200 mM DTT) in each well, followed by shaking for 15 min and incubating at 60 °C in a moisture box for 30 min. Subsequently the unbound material was removed by washing three times with water each time followed by centrifugation at 500 *g* for 1 min. Any residual liquid was subsequently removed from the membrane by pipetting and discarded.

ENZYMATIC *N*-GLYCAN RELEASE, REDUCTION AND PURIFICATION

The *N*-glycan release was performed as previously described[18] with small modifications. PNGase F (2 U of enzyme diluted with water to 15 µL) was added to each well followed by 15 min shaking at RT. The plate was incubated in a moisture box for 15 min at 37 ˚C. Subsequently, another 15 µL of water was added in each well. The plate was incubated overnight in a humidified plastic box at 37 ˚C to avoid evaporation of the digestion solution. Released *N*-glycans were recovered from the PVDF plate by centrifugation and washing three times with 30 µL of water. Subsequently the glycosylamine forms of the released *N*-glycans were hydrolyzed by adding 20 µL of 100 mM ammonium acetate (pH 5), incubated at RT for 1 h, and dried in a SpeedVac concentrator 5301 (Eppendorf, Hamburg, Germany) at 35 ˚C. Collected *N*-glycans were then reduced, desalted, followed by PGC cleanup using a 96-well plate based protocol adapted from Jensen *et al.*[13] For the reduction, 20 µL of 1 M $NaBH_4$ in 50 mM KOH was added to each well followed by incubation for 3 h at 50˚C in a humidified plastic box. To neutralize and quench the reaction, 3 µL of HAc was added to each sample. Desalting of the samples was performed using a strongly acidic cation exchange resin Dowex 50W-X8 which was self-packed into 96-well filter plates. Briefly, 100 µL of the resin slurry in MeOH (50/50; *v/v*) was added to each well in the filter plate. The columns were preconditioned by 3 x 100 µL of 1 M HCl, followed by 3 x 100 µL MeOH and 3 x 100 µL water each time removed by centrifuging at 500 *g*. The samples containing *N*-glycans were

loaded onto the columns and eluted two times with 40 µL of water followed by centrifugation at 500 *g*. The combined flow-through and wash were collected and dried in a SpeedVac concentrator at 35 ˚C. The remaining borate was removed by several rounds of co-evaporation using 100 µL MeOH in the SpeedVac concentrator at 35 ˚C. SPE cleaning step was performed by packing 60 µL (approximately 6 mg) of bulk sorbent carbograph slurry in MeOH into a 96-well filter plate. The columns were preconditioned by 3 x 100 µL of 80% MeCN in water containing 0.1% TFA, and 3 x 100 µL water with 0.1% TFA. After loading the sample, the columns were washed three times with 80 µL of 0.1% TFA, followed by *N*-glycan elution by 3x 40 µL of 60% MeCN in water containing 0.1% TFA. Samples were dried in a SpeedVac concentrator directly in polymerase chain reaction (PCR) plates for injection and re-dissolved in 10 µL of water prior to PGC nano-LC-ESI-MS/MS analysis.

O-GLYCAN RELEASE AND PURIFICATION

After removal of *N*-glycans, the *O*-glycans were released from the same PVDF membrane immobilized sample via reductive β-elimination. Briefly, 50 µL of 0.5 M $NaBH_4$ in 50 mM KOH was applied onto each PVDF membrane well after rewetting with 3 µL of MeOH. Plates were shaken for 15 min on a horizontal shaker and incubated in a humidified plastic box for 16 h at 50 °C. After incubation and cooling to RT, released *O*-glycans were recovered by centrifugation at 1000 *g* for 2 min into 96-well collection plates. The wells were rewetted by 3 µL of MeOH and washed three times with 50 µL of water with 10 min incubation steps on a horizontal shaker prior to centrifugation at 500 *g* for 2 min. Prior to desalting, the collected samples were concentrated to approximately 30 µL under vacuum in a SpeedVac concentrator at 35 °C for 2 h. Subsequently, 3 µL of HAc was added to neutralize the reaction followed by brief centrifugation to collect the sample at the bottom of the well. The following high throughput desalting and PGC SPE purification were performed as described in *the N-glycan preparation section*. The purified *O*-glycan alditols were re-suspended in 10 µL of water prior to PGC nano-LC-ESI-MS/MS analysis.

ANALYSIS OF RELEASED *N*- AND *O*-GLYCAN ALDITOLS USING PGC NANO-LC-ESI-MS/MS

The analysis of *N*- and *O*-glycans was performed on an Ultimate 3000 UHPLC system (Dionex/Thermo) equipped with a Hypercarb PGC trap column (5 µm Hypercarb Kappa, 32 µm x 30 mm, Thermo Fisher Scientific, Waltham, MA) and a Hypercarb PGC nano-column (3 µm Hypercarb 75 µm x 100 mm, Thermo Fisher Scientific, Waltham, MA) coupled to an amaZon ETD speed ion trap (Bruker, Bremen, Germany). Mobile phase A consisted of 10 mM ABC, while mobile phase B was 60 % acetonitrile/10 mM ABC (*v/v*). To analyze glycans, 1 µL of sample prepared from glycoprotein fetuin standard (31.25 picomole) and 3 µL of each sample derived from cell line NMUMG (1.5 x 10^5 cells) was injected and trapped on the trap column using a 6 µL/min loading flow in 2 % buffer B for *N*-glycan and 1 % buffer B for *O*-glycans for 5 min. Separation was achieved with a multi-step gradient of B: 2-9 % in 1 min and 9-49 % in 80 min for *N*-glycan and 1- 52 % over 72 min for *O*-glycans followed by a 10 min wash step using 95 % of B at a flow of rate of 0.6 µL/min. The column was held at a constant temperature of 45˚C. Ionization was achieved using the CaptiveSpray nanoBooster source (Bruker) with a capillary voltage of 1000 V applied and a dry gas temperature of 280˚C at 5 L/min and nebulizer at 3 pound per square inch (psi). Acetonitrile enriched dopant nitrogen was used. MS spectra were acquired within an *m/z* range of 500-1850 for *N*-glycans and 380-1850 for *O*-glycans in enhanced mode, smart parameter setting (SPS) was set to *m/z* 1200 for *N*-glycans and *m/z* 900 for *O*-glycans; ion charge control (ICC) to 4 x 10^3 and maximum acquisition time to 200 ms. MS/MS spectra were generated using collision-induced dissociation over an *m/z* range from 100 to 2500 on the top three most abundant precursors, applying an isolation width of 3 Thomson. The fragmentation cut-off was set to 27 % with 100 % fragmentation amplitude using the Enhanced SmartFrag option from 30-120 % in 32 ms and ICC was set to 150,000.

DATA PROCESSING

Glycan structures were assigned based on glycan composition obtained from accurate mass, relative PGC elution position, MS/MS fragmentation pattern in negative-ion mode and general glycobiological knowledge[15, 19], with help of GlycoWorkbench[20] and Glycomod[21] software tools. Extracted ion chromatograms were used to integrate area under the curve (AUC) for each

individual glycan isomer using Compass Data Analysis software v.5.0. The most abundant peaks in the glycan profile were manually picked and integrated. Relative quantitation of individual glycans was performed on the total area of all included glycans within one sample normalizing it to 100 %. The raw data is available on the GlycoPost repository: https://glycopost.glycosmos.org/preview/7996273655e60ec00bb52c.

RESULTS AND DISCUSSION

INTEGRATED *N*- AND *O*-GLYCAN SAMPLE PREPARATION IN 96-WELL PLATE FORMAT

A high throughput sample preparation in 96-well plate format was established for integrated *N*- and *O*-glycomics combined with an in-depth characterization of the *N*- and *O*-glycans using PGC nano-LC-ESI-MS/MS in negative ion mode, which can be applied for the analysis of purified glycoproteins and biological samples containing complex mixtures of glycoproteins (**Figure 1**). This approach is based on the sequential structural analysis of *N*- and *O*-glycans released from glycoproteins developed by Jensen *et al.,*[13] and a high throughput *N*-glycan preparation workflow that was recently reported by our group[18]. The highlights of the adapted workflow for combined *N*- and *O*-glycomics include high throughput approach in 96-well plate format for the entire sample preparation, integration of *N*- and *O*-glycan analysis, and powerful isomer separation on PGC for in-depth characterization of glycans. Importantly, large numbers of samples can be analyzed in the same batch with potential for method transfer to a liquid-handling robotic workstation prior to glycan analysis by PGC nano-LC-ESI-MS/MS.

The current protocols for glycomics analysis of cells often require a large quantity of biological material (4-10 x 10^6 cells)[11, 22, 23], and therefore are of limited use to decipher the glycosylation of cells that are available in minor amounts. Upon considerable modifications, with this approach we were able to analyze *N*- and *O*-glycosylation derived from 5 x 10^5cells. Our aim was to provide a high throughput workflow suitable for profiling cells ensuring sufficient material for glycomics analysis assuming different glycosylation characteristics of different biological samples. Therefore, we have not attempted to test the lowest cell amount necessary for the analysis of glycans derived from NMuMG cell line, as this would be specific for this cell line. However, PGC-MS based glycomics workflows have recently undergone major improvements with

respect to sensitivity[24-26]. Using post-column make-up flow (PCMF) for enhancing sensitivity has been shown to allow glycomics analyses from minor amounts of biological material such as rare cell populations as well as patient derived materials. As shown in **Figure 1**, (glyco)proteins from complex biological samples such as cell pellets from cultured cells, organoids or tissue should be first extracted by cell lysis. For tough or fibrous tissues, homogenization is often needed to assist extraction of glycoproteins prior to cell lysis. Taking into consideration the compatibility with MS, cells were lysed mechanically in water or lysis buffer without any detergents. Next, glycoproteins in solution or in cell lysates were then applied and immobilized on 96-well PVDF membrane filter plates followed by denaturation and reduction to achieve efficient and complete release of glycans. To avoid nonspecific binding of PNGase F to the PVDF membrane, the effect of blocking agent PVP-40 was tested, however it showed no benefit on the overall *N*-glycan signal when cell lysates from 5 x 10^5 cells were used, likely due to high amounts of protein in the samples which presumably already largely occupied the PVDF membrane pores (data not shown). After removing

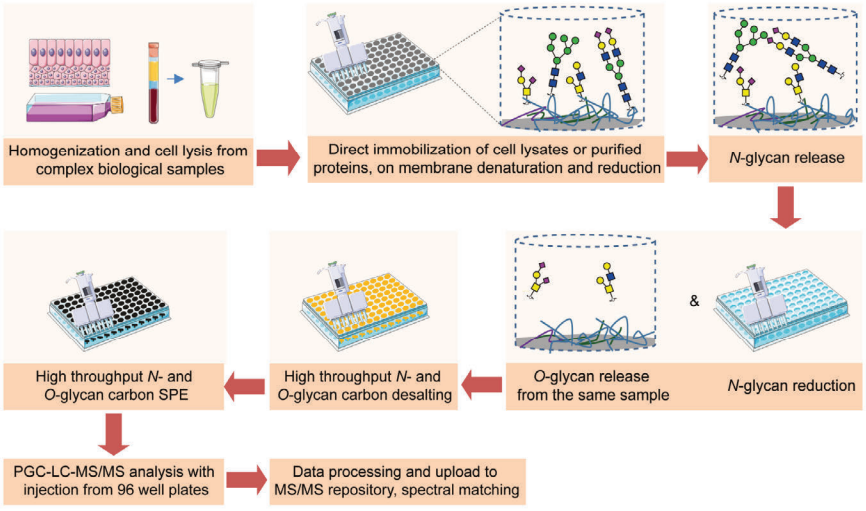


Figure 1. Workflow for integrated N-and O-glycomics based on a 96-wellplate sample preparation and analysis on PGC nano-LC-ESI-MS/MS. *Glycoproteins were extracted from biological samples by cell lysis and immobilized on PVDF membrane. Proteins were denatured and reduced by guanidine HCl and DTT without detergent followed by enzymatic N-glycan release by PNGase F. The released N-glycans were collected and reduced. O-Glycans were then released from the same sample by reductive β-elimination. Reduced glycans were further desalted, purified, reconstituted and injected all in 96-well format plates, and subjected to PGC nano-LC-ESI-MS/MS analysis in negative ion mode.*

the denaturation and reduction agents, *N*-glycans were released from immobilized glycoproteins on PVDF membrane using 2 U PNGase F, as illustrated in **Figure 1**. To ensure the efficient generation of reducing end glycans and subsequent quantitative reduction, glycosylamines resulting from the release deamidation were hydrolyzed using acidic buffer (pH 5). The released *N*-glycans were collected and further subjected to reduction, desalting, and PGC SPE clean-up all in 96-well plate format to ensure efficient sample-throughput. Subsequently, *O*-glycans were released by reductive β-elimination from the same wells and purified in a high throughput manner following the same procedure together with the released *N*-glycans. Thereafter, the released *N*-glycan and *O*-glycan alditols were separated on PGC chromatographic column and analyzed with an ion trap mass spectrometer in negative ionization mode.

Most methodologies employ sialic acid derivatization for their stabilization and chemical labeling to increase ionization efficiency[6, 18]. In addition, the fucose migration in positive mode often poses a problem for structural analysis[8, 27]. Working in negative ionization mode has key advantages to address these challenges due to its high ionization efficiency especially for sialylated glycans, stability of sialic acids and no migration of fucose. More importantly, negative ionization mode fragmentation allows in-depth structural identification of glycan structures[14, 16, 28-30], since additional structural information can be obtained from cross-ring fragments providing diagnostic ions for the characterization of glycan linkages[28], MS/MS spectra were publicly available by exporting from the DataAnalysis software using GlycoWorkbench workspace and uploading to Unicarb DR repository following a standardized bioinformatics infrastructure[31], serving as an open resource for automated *N*- and *O*-glycan identification via spectral matching.

APPLICATION OF THE WORKFLOW FOR THE RELEASE OF *N*- AND *O*-GLYCANS FROM BOVINE FETUIN STANDARD

The established approach was evaluated using bovine fetuin standard, which contains highly sialylated *N*-glycans[32, 33] and *O*-glycans[34, 35] differing in the number of sialic acids as well as their linkages. To assess the technical variation of glycan profiles, aliquots of 20 µg fetuin were applied to PVDF membrane wells in three technical replicates. *N*-glycans and *O*-glycans were prepared following this approach and analyzed by PGC nano-LC-ESI-MS/MS in negative ion mode. Glycan structures and linkages were assigned following known MS/MS fragmentation patterns in negative-

ion mode[15, 19], known retention behavior of different sialic acid isomers on PGC column[32] and general glycobiological knowledge. Within this analysis, 30 different *N*-glycan structures were detected of which 24 passed the quantification criteria (signal-to-noise ratio (S/N) ≥ 9). As expected, sialylated *N*-glycans were detected as dominant glycan types, with different number of sialic acids and their linkages. The profiles obtained by our analysis (**Figure 2A**) are very similar to other published characterizations of bovine fetuin *N*-glycans[32, 33, 36]. Relative quantitation was performed on the top 13 most abundant *N*-glycans within one sample normalized to 100 %. Details on glycan structures and relative abundance are given in **Supporting Information, Table S1.** Inter- and intraday repeatability of the fetuin *N*-glycan analysis showed two median intraday coefficient of variations (CVs) of 7.6 % and 8.0 %, and a median interday CV of 9.8 % within six technical replicates distributed into two plates over one month. The highest intraday CV for the relative quantification of *N*-glycans released from fetuin was 15.7 %, indicating the good repeatability of the workflow (**Figure 2B**). Our results showed good reduction efficiency, without detecting the unreduced species for the top 10 most abundant *N*-glycans released from fetuin and with less than 2 % of unreduced species in the top 10 of most abundant *N*-glycans released from NMuMG cells (data not shown).

The powerful separation capacity of PGC chromatography in *N*- and *O*-glycans has been widely demonstrated by previous work, enabling separation of glycan isomers based on glycan structure and linkage specificity[28, 32, 37-39]. **Figure 2A** shows four glycan isomers with the same composition containing five hexoses, four *N*-acetylhexosamines and two *N*-acetylneuraminic acids (H5N4S2; blue trace) which are separated due to different sialic acid linkage combinations. First the glycan with two α2-6-linked *N*-acetylneuraminic acids elutes at 27.2 min, followed by the two coeluting isomers at 32.9 min, which have both one α2-6- and one α2-3-linked *N*-acetylneuraminic acids, carried on different arms. Lastly, the same glycan with two α2-3-linked *N*-acetylneuraminic acids elutes at 37.3 min. From this result, we can confirm the previous reports that glycans containing α2-6-linked sialic acid residues elute significantly earlier than the α2-3-linked counterparts[32]. While the order of isomer elution is in line with previous reports, we observe slightly different relative retention of *e.g.* later eluting monosialylated compared to disialylated diantennary *N*-glycans.

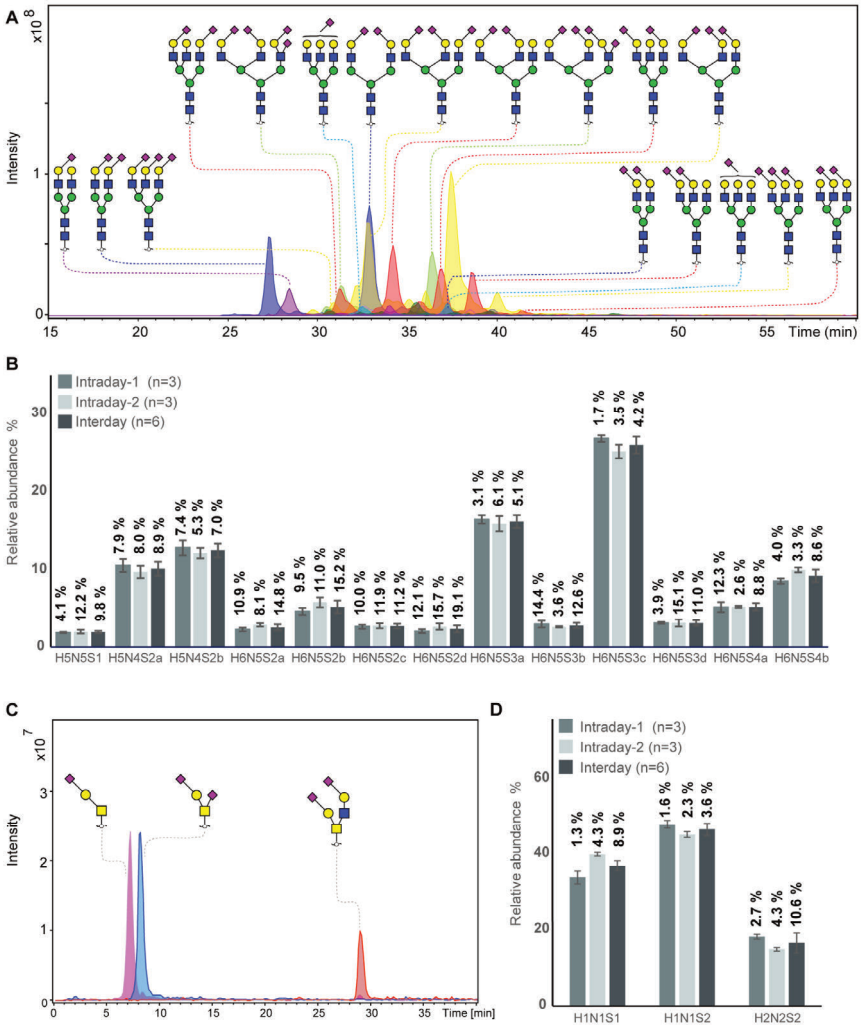


Figure 2. Analysis of N-glycans and O-glycans derived from bovine fetuin standard. ***(A)*** *Combined extracted ion chromatograms (EIC) of N-glycans released from bovine fetuin standard.* ***(B)*** *Inter- and intraday repeatability of the fetuin N-glycan analysis based on relative quantification of top 13 most abundant N-glycans. Inter- and intraday repeatability of the fetuin N-glycan analysis showed two median coefficients of variation (CV) of 7.6 % and 8.0 % (n = 3; same plate) and a median CV of 9.8 % (n = 6; two plates over one month), mean relative abundance are shown plus standard deviation. More details see* ***Supporting information, Table S1 C)*** *Combined EICs of 5 O-glycans released from bovine fetuin standard, in which the top three most abundant O-glycans account for 98 % of the relative abundance.* ***(D)*** *Inter- and intraday repeatability of the top three most abundant O-glycans released from fetuin after removing N-glycans. Mean relative abundance displayed plus standard deviation, for more details see* ***Supporting information, Table S2****. CVs of each glycans were listed on the top of the bar; intraday n = 3, interday n = 6, derived from two different plates performed over one month)*

Next to *N*-glycans, three *O*-glycans were also detected which were released and prepared from fetuin after *N*-glycan removal (**Figure 2C**), showing the feasibility of integrated *N*- and *O*-glycan sample preparation in a 96-well plate format. **Figure 2D** shows the relative quantification and repeatability data for the three glycans which are sialylated core 1 and core 2 structures. The median CVs for two intra- and interday repeatability of all quantified structures were 1.9 %, 2.3 % and 8.9 %, respectively (**Supporting Information**, **Table S2**).

APPLICATION OF THE WORKFLOW FOR THE RELEASE N- AND O-GLYCANS FROM MURINE NMUMG CELLS

To demonstrate that this established integrated *N*- and *O*-glycomics approach is applicable for complex biological samples, *N*- and *O*-glycans were prepared from murine NMuMG cell line following the method described in **Figure 1**. For complex biological samples, cell lysis was required to extract glycoproteins from the cell membrane. In this study, we utilized mechanical cell lysis in water via sonication for 60 min in an ultrasound bath. In addition, lysis buffer without detergent can be used as an alternative.

N-glycan (**Figure 3A**) and *O*-glycan (**Figure 3B**) alditols were prepared from 5 x 10^5 cells in a high throughput manner following our approach. *N*-glycan alditols were analyzed using PGC nano-LC-ESI-MS/MS platform which revealed 94 different *N*-glycan structures in a single analysis, including 3 paucimannose, 16 oligommanose, 11 hybrid and 64 complex glycans (**Supporting Information, Table S3**). Additional low abundant glycans may be present but were not considered further. The composition and proposed structures of *N*-glycans were assigned on the basis of general knowledge on *N*-glycan biosynthesis in mice, previous publications on *N*-glycan fragmentation in negative ion mode[40-44] as well as using automated spectral matching tool available online via Unicarb DB. **Figure 3A** shows the combined extracted ion chromatograms of the 25 most abundant *N*-glycan structures. The *N*-glycan profile obtained from 2 mg of total proteins extracted from NMuMG cell lines has been reported previously, which revealed only 27 *N*-glycan structures by MALDI-TOF/TOF-MS analysis after acetohydrazide derivatizaton of sialic acids[22]. With this approach we were able to confirm all of the *N*-glycans observed in the previous work[22], but also revealed a large number of additional glycans including paucimannose, phosphorylated oligomannose and more sialylated *N*-glycans (**Figure 3A**).

Phosphorylated oligomannose *N*-glycans were characterized by a decreased retention time on PGC compared to unmodified ones (**Supporting Information,**

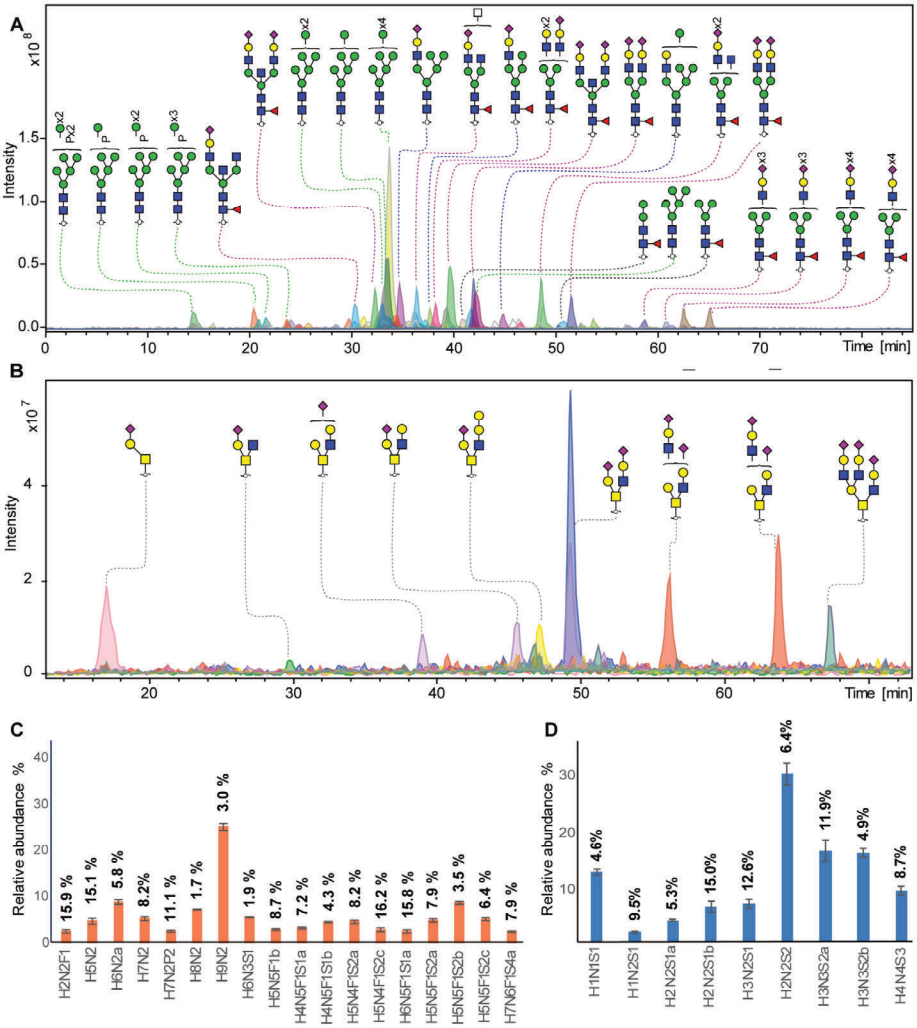


Figure 3. Integrated N-and O-glycans analysis derived from NMuMG cells following the 96-well plate sample preparation method and analyzed using PGC nano-LC-ESI-MS/MS. (A) *Combined EICs of N-glycans released from 5 105 NMuMG cells. (Black line: paucimannose; green line: oligomannose; blue line: hydride; red line; complex type).* ***(B)*** *Combined EICs of O-glycans released from NMuMG cells after removing N-glycans.* ***(C)*** *Relative quantification of top 18 most abundant N-glycans. (displayed as mean relative abundance plus standard deviation; CVs of each glycans were listed on the top of the bar; n = 3). More details displayed in* ***Supporting Information, Table S4. D)*** *Relative quantification of top 9 most abundant O-glycans. (displayed as mean relative abundance plus standard deviation; CVs of each glycans were listed on the top of the bar; n = 3). More details displayed in* ***Supporting Information, Table S5****.*

Figure S1). In contrast, addition of a sulfate group often results in an increased retention on PGC (unpublished data). Notably, powerful isomer separation on PGC vastly contributed to the higher number of identified glycans, compared to MALDI-TOF/TOF-MS analysis. In agreement with literature[32], we confirmed overall the earlier elution of oligomannose *N*-glycans, and an increasing retention of *N*-glycans with increasing numbers of sialic acids and antennae.

Relative quantitation of the top 18 most abundant glycans was performed on the total area of the selected glycans within one sample normalized to 100 %. The relative abundance and corresponding standard deviation are shown in **Figure 3C** and **Supporting Information, Table S4**. The robustness of the method was confirmed by a median CV of 7.9 % as well as the highest CV of 16.2 % within the top 18 glycan species. Specifically, 13 out of 18 the main peaks showed a CV lower than 10 %, indicating a very good repeatability. In contrast to *N*-glycans, mucin-type *O*-glycans do not share a common core beyond the innermost GalNAc, resulting in a vast structural diversity as well as the presence of multiple isomers. Released *O*-glycan analysis from the same 5 x 10^5 NMuMG cells revealed 11 *O*-glycans (**Figure 3B**). All *O*-glycans in NMuMG were mono-and disialylated core 1 and core 2 structures. The power of isomer separation of PGC can be displayed by the two glycan isomers with composition H2N2S1 at *m/z* 1040.4 (**Figure 3B**; purple trace), which carry the same H2N2 neutral core structure but with different linkage of sialic acid or sialic acid linked to different galactoses, as well as two different core 2 glycan isomers with composition H3N3S2 at *m/z* 1696.5 (**Figure 3B**; orange trace). The distribution of the *O*-glycans was determined by relative quantification, revealing the dominance of disialylated core 2 *O*-glycans (**Figure 3D** and **Supporting Information, Table S5**). The technical variability of *O*-glycans derived from cell lines was determined in three technical replicates (**Figure 3D**) with CVs less than 14.9 %.

CONCLUSIONS

In summary, we present an integrated approach for the analysis of *N*- and *O*-glycomics from purified proteins and complex biological samples combining sequential release of *N*- and *O*-glycans based on protein immobilization on PVDF membrane filter plates in 96-well format and structural elucidation based on PGC nano-LC-ESI-MS/MS using negative electrospray ionization, inspired

by the methodology developed by Packer and coworkers[13, 14]. We here provide a method with high repeatability for combined *N*- and *O*-glycomics of biological samples, in which glycans were prepared in a higher-throughput and less time-consuming manner as compared to previous workflows[13, 14]. Still, this method requires multiple manual handling steps and several sample transfers between different 96-well plates. The transfer of this sample preparation method to a liquid-handling robotic workstation should be feasible. Here, the method was successfully applied for *N*- and *O*-glycomics of NMuMG cells as an example of complex biological samples. In the future, the method may be used to handle larger numbers of samples for a comprehensive glycomics analysis in an in-depth manner in the context of functional glycomics, systems biology, and clinical applications.

SUPPORTING INFORMATION

Supporting information is available free of charge via DOI: 10.1039/c9mo00180h

ACKNOWLEDGMENTS

We thank A.L. Hipgrave Ederveen, C.A.M. Koeleman and L. de Neef for technical support, and Y. Mohammed for bioinformatics support. This work was supported by the European Commission's Horizon 2020 programme "GlyCoCan" project, grant number 676421.

REFERENCES

1. S. S. Pinho and C. A. Reis, Nat Rev Cancer, 2015, 15, 540-555.
2. J. G. Rodrigues, M. Balmana, J. A. Macedo, J. Pocas, A. Fernandes, J. C. M. de-Freitas-Junior, S. S. Pinho, J. Gomes, A. Magalhaes, C. Gomes, S. Mereiter and C. A. Reis, Cell Immunol, 2018, 333, 46-57.
3. T. Song, D. Aldredge and C. B. Lebrilla, Anal Chem, 2015, 87, 7754-7762.
4. S. Gaunitz, G. Nagy, N. L. Pohl and M. V. Novotny, Anal Chem, 2017, 89, 389-413.
5. S. Holst, B. Heijs, N. de Haan, R. J. van Zeijl, I. H. Briaire-de Bruijn, G. W. van Pelt, A. S. Mehta, P. M. Angel, W. E. Mesker, R. A. Tollenaar, R. R. Drake, J. V. Bovee, L. A. McDonnell and M. Wuhrer, Anal Chem, 2016, 88, 5904-5913.
6. G. S. M. Kammeijer, I. Kohler, B. C. Jansen, P. J. Hensbergen, O. A. Mayboroda, D. Falck and M. Wuhrer, Anal Chem, 2016, 88, 5849-5856.
7. C. W. Sutton, J. A. O'Neill and J. S. Cottrell, Anal Biochem, 1994, 218, 34-46.

8. M. Wuhrer, C. A. Koeleman, C. H. Hokke and A. M. Deelder, Rapid Commun Mass Spectrom, 2006, 20, 1747-1754.
9. I. Trbojevic-Akmacic, M. Vilaj and G. Lauc, Expert Rev Proteomics, 2016, 13, 523-534.
10. A. Shubhakar, K. R. Reiding, R. A. Gardner, D. I. Spencer, D. L. Fernandes and M. Wuhrer, Chromatographia, 2015, 78, 321-333.
11. S. Holst, G. W. van Pelt, W. E. Mesker, R. A. Tollenaar, A. I. Belo, I. van Die, Y. Rombouts and M. Wuhrer, Methods Mol Biol, 2017, 1503, 185-196.
12. M. Kotsias, R. P. Kozak, R. A. Gardner, M. Wuhrer and D. I. R. Spencer, Plos One, 2019, 14, e0210759.
13. P. H. Jensen, N. G. Karlsson, D. Kolarich and N. H. Packer, Nat Protoc, 2012, 7, 1299-1310.
14. C. Ashwood, B. Pratt, B. X. MacLean, R. L. Gundry and N. H. Packer, Analyst, 2019, 144, 3601-3612.
15. N. G. Karlsson, N. L. Wilson, H. J. Wirth, P. Dawes, H. Joshi and N. H. Packer, Rapid Commun Mass Spectrom, 2004, 18, 2282-2292.
16. A. V. Everest-Dass, J. L. Abrahams, D. Kolarich, N. H. Packer and M. P. Campbell, J Am Soc Mass Spectrom, 2013, 24, 895-906.
17. E. Piek, A. Moustakas, A. Kurisaki, C. H. Heldin and P. ten Dijke, J Cell Sci, 1999, 112 (Pt 24), 4557-4568.
18. S. Holst, A. I. Belo, E. Giovannetti, I. van Die and M. Wuhrer, Sci Rep, 2017, 7, 16623.
19. N. G. Karlsson, B. L. Schulz and N. H. Packer, J Am Soc Mass Spectrom, 2004, 15, 659-672.
20. A. Ceroni, K. Maass, H. Geyer, R. Geyer, A. Dell and S. M. Haslam, J Proteome Res, 2008, 7, 1650-1659.
21. C. A. Cooper, E. Gasteiger and N. H. Packer, Proteomics, 2001, 1, 340-349.
22. Z. Tan, W. Lu, X. Li, G. Yang, J. Guo, H. Yu, Z. Li and F. Guan, J Proteome Res, 2014, 13, 2783-2795.
23. H. Hamouda, M. Kaup, M. Ullah, M. Berger, V. Sandig, R. Tauber and V. Blanchard, J Proteome Res, 2014, 13, 6144-6151.
24. H. Hinneburg, S. Chatterjee, F. Schirmeister, T. Nguyen-Khuong, N. H. Packer, E. Rapp and M. Thaysen-Andersen, Anal Chem, 2019, 91, 4559-4567.
25. H. Hinneburg, P. Korac, F. Schirmeister, S. Gasparov, P. H. Seeberger, V. Zoldos and D. Kolarich, Mol Cell Proteomics, 2017, 16, 524-536.
26. T. Nguyen-Khuong, A. Pralow, U. Reichl and E. Rapp, Glycoconj J, 2018, 35, 499-509.
27. E. Mucha, M. Lettow, M. Marianski, D. A. Thomas, W. B. Struwe, D. J. Harvey, G. Meijer, P. H. Seeberger, G. von Helden and K. Pagel, Angew Chem Int Ed Engl, 2018, 57, 7440-7443.
28. C. Ashwood, C. H. Lin, M. Thaysen-Andersen and N. H. Packer, J Am Soc Mass Spectrom, 2018, 29, 1194-1209.
29. D. J. Harvey and J. L. Abrahams, Rapid Commun Mass Spectrom, 2016, 30, 627-634.
30. J. L. Abrahams, M. P. Campbell and N. H. Packer, Glycoconj J, 2018, 35, 15-29.
31. M. A. Rojas-Macias, J. Mariethoz, P. Andersson, C. Jin, V. Venkatakrishnan, N. P. Aoki, D. Shinmachi, C. Ashwood, K. Madunic, T. Zhang, R. L. Miller, O. Horlacher, W. B. Struwe, Y.

Watanabe, S. Okuda, F. Levander, D. Kolarich, P. M. Rudd, M. Wuhrer, C. Kettner, N. H. Packer, K. F. Aoki-Kinoshita, F. Lisacek and N. G. Karlsson, Nat Commun, 2019, 10, 3275.

32. G. Palmisano, M. R. Larsen, N. H. Packer and M. Thaysen-Andersen, Rsc Adv, 2013, 3, 22706-22726.
33. E. D. Green, G. Adelt, J. U. Baenziger, S. Wilson and H. Van Halbeek, J Biol Chem, 1988, 263, 18253-18268.
34. M. Windwarder and F. Altmann, J Proteomics, 2014, 108, 258-268.
35. Y. Goso, T. Sugaya, K. Ishihara and M. Kurihara, Anal Chem, 2017, 89, 8870-8876.
36. M. Nakano, R. Saldanha, A. Gobel, M. Kavallaris and N. H. Packer, Mol Cell Proteomics, 2011, 10, M111 009001.
37. K. Stavenhagen, D. Kolarich and M. Wuhrer, Chromatographia, 2015, 78, 307-320.
38. L. R. Ruhaak, A. M. Deelder and M. Wuhrer, Anal Bioanal Chem, 2009, 394, 163-174.
39. M. Pabst, J. S. Bondili, J. Stadlmann, L. Mach and F. Altmann, Anal Chem, 2007, 79, 5051-5057.
40. D. J. Harvey, J Am Soc Mass Spectr, 2005, 16, 631-646.
41. D. J. Harvey, J Am Soc Mass Spectr, 2005, 16, 622-630.
42. D. J. Harvey, J Am Soc Mass Spectr, 2005, 16, 647-659.
43. D. J. Harvey, M. Edgeworth, B. A. Krishna, C. Bonomelli, S. A. Allman, M. Crispin and J. H. Scrivens, Rapid Commun Mass Sp, 2014, 28, 2008-2018.
44. J. H. Chik, J. Zhou, E. S. Moh, R. Christopherson, S. J. Clarke, M. P. Molloy and N. H. Packer, J Proteomics, 2014, 108, 146-162.

DOPANT-ENRICHED NITROGEN GAS FOR ENHANCED ELECTROSPRAY IONIZATION OF RELEASED GLYCANS IN NEGATIVE ION MODE

Chapter 3

Katarina Madunić[1], Sander Wagt[1], Tao Zhang[1], Manfred Wuhrer[1], Guinevere S.M. Lageveen-Kammeijer[1] *

[1] *Center for Proteomics and Metabolomics, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands*

Reprinted with permission from Analytical Chemistry 2021, 93, 18, 6919–6923
DOI: 10.1021/acs.analchem.1c00023

ABSTRACT

The desolvation and ionization process of analytes can significantly be improved by enriching the nebulizing gas with a dopant (dopant enriched nitrogen (DEN)-gas) in the electrospray source. However, for the analysis of released glycans in negative ion mode, the usage of DEN gas remains largely unexplored. For this purpose, we investigated the effect of different polar protic solvents (methanol, ethanol and isopropanol) as well as using solely the nebulizing gas or ambient air on the ionization and charge state distribution of released *N*- and *O*-glycans. Compared to the standard acetonitrile enriched nitrogen gas, isopropanol showed the highest increase in regards to peak areas. Moreover, it showed large benefits for the identification of glycan structures at high sensitivity as the increased precursor intensities subsequently resulted in higher intensities in tandem MS mode. While similar effects are noted for both neutral and sialylated species, the most significant effect was observed for early eluting glycans where very low acetonitrile concentrations were present in the eluent. The best results in terms of S/N ratios were obtained with methanol, with less effect on the MS/MS signal enhancement. Therefore, the use of this dopant would be particularly beneficial for high sensitivity MS-mode applications. In conclusion, isopropanol enriched DEN gas greatly improves the detection of both *N*-and *O*-glycan species, their tandem mass spectra, particularly for the early eluting species whose ionization in the absence of DEN gas is hindered by low organic concentrations.

INTRODUCTION

Liquid chromatography coupled to mass spectrometry (LC-MS) has become a leading analytical platform for the identification and (relative) quantification of a wide range of compounds at high sensitivity. For the latter, ionization efficiency is of key importance. Because this is largely regulated by the electrospray ionization (ESI) process at the interface of the LC-MS, optimization of the mobile phase can already significantly improve the ionization of specific analytes[1]. With the addition of solvent vapors at the ion source, the ionization efficiency can be further improved[2]. To this end, the nebulizing gas (nitrogen; N_2) will be enriched with a solvent vapor that is guided around the spray emitter and spray plume to enhance the nebulization and desolvation process[3]. This approach is of specific interest for the glycomic field as carbohydrates usually ionize less efficiently than various other analytes due to their hydrophilic nature. Specifically, applications using acetonitrile (MeCN) as a dopant in positive ionization mode, revealed a 25-fold increase in sensitivity for the analysis of glycopeptides[4,5].

Porous graphitized carbon (PGC) nano-nanoLC-MS, where the ESI is operated in negative ion mode, is a very powerful platform for the analysis of *N*- and *O*-glycan alditols featuring glycan isomer separation[6–8]. However, the usage of a dopant enriched nitrogen (DEN) gas remains largely unexplored for these applications. When compared to positive-ion mode, apart from the often higher ionization efficiency of sialylated and sulphated glycan species, the main benefit of negative mode is its ability for in-depth characterization of linkage isomers as the deprotonated glycan fragments provide not only glycosidic but also cross-ring fragments in tandem MS[9]. However, it has been observed that the early eluting species have less ionization efficiency in PGC-LC-MS due to low MeCN concentrations in this elution range[10]. This particularly affects early eluting *N*-linked glycans as well as *O*-linked glycans such as T and sialylTn antigens. Thereby, these very important tumor antigens might be underrepresented or completely missed in the analysis.

Previous studies have attempted to account for this issue using post column make-up flow (PCMF) for capillary flow PGC-LC where MeCN and isopropanol (IPA)-based PCMF increased signal response, decreased variation and improved MS/MS spectral quality compared to reference approach without PCM[11]. A recent study has employed

the usage of DEN gas for the analysis of *N*-glycans with capillary zone electrophoresis (CZE), coupled to electrospray ionization mass spectrometry in negative ion mode and observed a significant increase in MS signals[12]. However, little is known about the effect of different dopants supplemented directly into the MS source for the PGC-nanoLC analysis of both *N*- and *O*-glycans in negative ion mode.

Here, we explore the effect of various DEN gas conditions on the ionization and charge state distribution of glycans in negative ion mode. For this purpose, different dopant solvents (methanol (MeOH), ethanol (EtOH) and IPA) as well as using the nebulizing gas unaccompanied or using solely ambient air were compared to the standard MeCN enriched nitrogen gas. Therefore, the aim of this study was to provide a systematic overview of the different dopant choices based upon the needs and the analytes present in the sample.

MATERIALS AND METHODS

CHEMICALS

Glacial acetic acid, IPA (> 99.9% Chromasol LC-MS, cat: 34965-2.5L) and potassium hydroxide (KOH) were purchased from Honeywell Fluka. EtOH (Reag. Ph. Eur, Prod. Nr: 1.00983.1000) and mucin from bovine submaxillary glands (BSM), type I-S was purchased from Merck (Darmstadt, Germany). MeCN (Ultra LC-MS, Art. Nr: 801023802) and MeOH (Ultra LC-MS, Art. Nr: 813013802) were purchased from Actu-All Chemicals (Oss, the Netherlands). Peptide N-glycosidase F (PNGase F) was obtained from Roche Diagnostics (Mannheim, Germany). Ammonium bicarbonate (ABC), trifluoroacetic acid (TFA), fetuin from fetal bovine serum (FBS), cation-exchange resin Dowex (50W-X8), hydrochloric acid (HCl), sodium borohydride ($NaBH_4$), DL-dithiothreitol (DTT) and ammonium acetate and were obtained from Sigma-Aldrich (Steinheim, Germany). Bulk sorbent Carbograph was obtained from Grace Discovery sciences (Columbia). Ultrapure water was generated from a Q-Gard 2 system, MultiScreen HTS 96-well plates (hydrophobic Immobilon-P PVDF membrane), 96-well PP Microplate were obtained from Millipore (Amsterdam, the Netherlands). 96-well PP filter plate from Orochem Technologies (Naperville, IL). Immunoglobulin G from human plasma was purchased from Athens Research & Technology (Georgia, USA). Sialylated fetuin *N*-linked alditols standard was obtained from Dionex (Thermo Fisher, CA, USA)

N- AND *O*-GLYCAN RELEASE FROM PURIFIED GLYCOPROTEINS

Twenty µg of purified glycoproteins IgG, bovine fetuin and bovine submaxillary mucin were blotted onto separate wells of preconditioned hydrophobic Immobilon-P PVDF membrane. Both *N*- and *O*-glycans were released from the glycoproteins as described before[13]. Detailed information about *N*- and *O*-glycan release is provided in **Supporting Information 1, S-1**. Abundances of glycans per replicate per condition are available in **Supporting Information 2, Table S-1**. Averaged relative abundances of glycans per condition are listed in in **Supporting Information 2, Table S-2** (*N*-glycans) and **Table S-3** (*O*-glycans). Raw data is available online at https://glycopost.glycosmos.org/preview/17497542156058d85d3a5c2.

MEASUREMENTS WITH PGC-LC-MS/MS USING DIFFERENT DOPANT SOLVENTS

Analysis was performed using a PGC nano-LC Ultimate 3000 UHPLC system (Dionex/Thermo, Sunnyvale, CA) coupled to an amaZon ETD speed ion trap (Bruker Daltonics, Bremen, Germany). The samples were re-dissolved in 20 µL of water prior to analysis. The released *N*-glycans from IgG and fetuin were mixed together in ratio 1:1 and 1 µL of the mixture was injected into the system. An estimate of the injected amount is shown in **Supporting Information 2, Table S-4**. Furthermore, 1 µL of the released *O*-glycans from BSM was injected for analysis. The glycans were separated on a custom-made PGC column. The LC system was coupled to an amaZon ETD speed ESI ion trap MS using the CaptiveSpray™ source (Bruker Daltonics) in negative-ionization mode. The drying gas (N_2) temperature was set at 280 ˚C and the flow to 3 L/min. The nebulizer gas pressure was kept at 3 psi (for all conditions except for ambient air). The nanoBoosterTM bottle (Bruker Daltonics) was filled with different dopant solvents, namely MeOH, EtOH, IPA and MeCN. The measurements were performed by filling the bottle with 800 mL solvent, and the level of the dopant inside the bottle was never below 500 mL. Detailed information about the LC-MS/MS settings is provided in **Supporting Information 1, S-1.**

RESULTS AND DISCUSSION

EFFECT OF DOPANT SOLVENTS ON THE IONIZATION OF *N*-GLYCANS

Compared to the standard MeCN enriched nitrogen gas, all the polar protic solvents tested (IPA, followed by EtOH and MeOH) gave an approximately 8-fold increase in peak areas for the selected 15 most abundant *N*-glycans released from IgG and

bovine fetuin glycoproteins (**Figure 1A**). Similar effects were detected for the neutral and sialylated glycans. However, significantly higher effect on the peak area is seen for the tetrasialylated early eluting glycan with composition H6N4S4a. Our data also shows that the increase in peak areas comes with a slight decrease in signal to noise (S/N) ratios for EtOH and IPA compared to the standard MeCN approach (**Figure 1B**). From the polar protic solvents, the S/N ratios are the highest when MeOH is used as dopant. The relative abundance remains rather consistent between the different conditions (**Supporting Information 1, Figure S-1A**). While, limited research has been performed about the usage and composition of DEN gas for the analysis of *N*-glycans in negative mode, previous research showed an increased response of glucose in negative ion mode when MeOH was employed as organic co-solvent compared to MeCN[14]. This effect has been attributed to different physical properties of these solvents, relevant for ESI, such as surface tension, vapor pressure and gas phase thermochemistry. Of note, only a single monosaccharide has been used to study these effects, in contrast to our study which studies the effect of MeOH addition on a broad range of *N*- and *O*-glycans. Moreover, a study investigating the effect of polar protic (MeOH and water) and polar aprotic (MeCN and acetone) solvents on ionization of small acidic molecules revealed that the compounds ionized better in polar protic solvents, leading to increased sensitivity and lower limits of detection[15]. The effect was consistent in flow injection experiments, when pure solvents were used. However, analytes eluting at low organic concentrations showed less pronounced increase in response and sensitivity when MeOH was used as the organic mobile phase in gradient chromatographic separations. Similarly, Henriksen *et al.,* have shown that most of the tested acidic analytes ionize better in negative ion mode when infused in pure MeOH than MeCN, which was attributed to the protic nature of MeOH, a stronger solvation of negative ions and ion pairs results in more deprotonated molecules[16]. Intriguingly, they have also observed that the benefits of MeOH on the ionization are gone in aqueous mixtures in LC separations. This might explain why in a similar investigation for capillary-flow PGC-LC-MS/MS, the authors observed very similar signal intensities for the *N*- and *O*- glycans when using MeCN or polar protic solvent (MeOH and IPA) using PCMF[11]. This finding emphasizes the advantage of our approach, where there is a clear benefit of an alcohol enriched nebulizing gas even for the aqueous based separations. On the other hand, compared to the reference MeCN approach, the setup using solely N_2 gas or ambient air, shows similar results

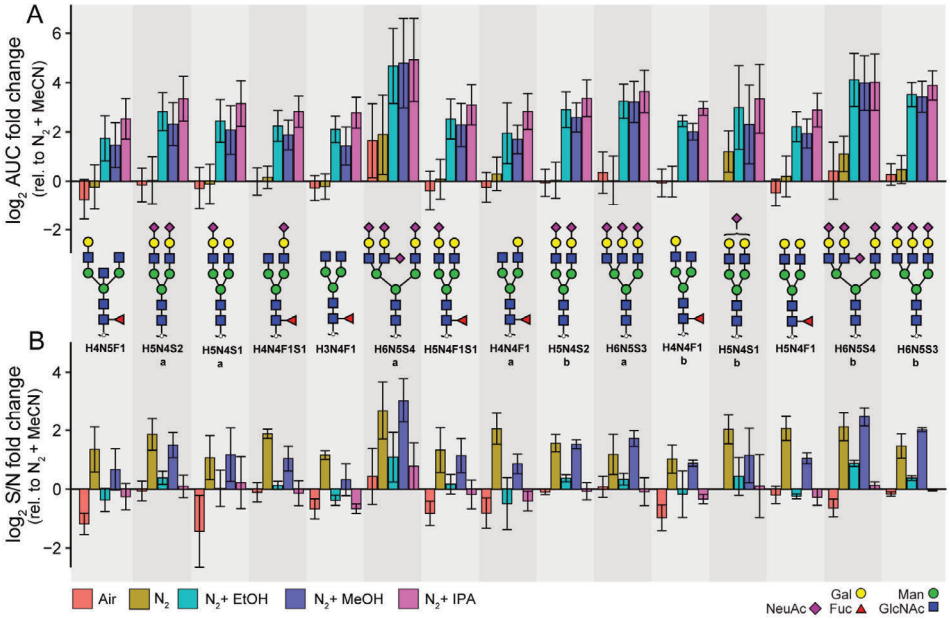


Figure 1. Effect of different dopant solvents on N-glycans. (A) *The fold change of the area under the curve relative to the reference approach (MeCN enriched N_2) and* ***(B)*** *Fold change of S/N ratios relative to the reference approach (MeCN enriched N_2) for the 15 most abundant N-glycans derived from IgG and bovine fetuin (N = 3). The error bars represent the standard deviation from the mean of the technical replicates (N = 3). The glycans are ordered based upon their elution time. H-Hexose, N- N-acetylhexosamine, F-deoxyhexose, S- N-acetylneuraminic acid, Sg- N-glycolylneuraminic acid.*

or even a decrease in peak areas for most of the glycans studied. Recently, a study employing DEN gas in CZE coupled to MS in negative ion mode demonstrated that IPA based DEN gas outperforms MeCN, leading to up to 40-fold increase in peak areas for larger neutral glycans[12]. An effect on the charge state distribution of glycans was also observed as illustrated in **Figure 2**. The polar protic solvents (IPA, MeOH and EtOH) show the main boosting effect on the doubly charged species of the analytes (**Figure 1A, Supporting Information 2, Table S-5**), making them the most dominant species per analyte (**Figure 2B**). Nonetheless, increased peak areas were also observed for the triply charged species, compared to reference approach. Consequently, S/N of the fragment ions were enhanced as well (**Supporting Information 2, Table S-6**). Interestingly, the triply charged species are relatively more abundant in the approaches that do not use dopant solvents (N_2 and ambient air). While the polar protic solvents showed this charge state effect in our DEN gas setup, a similar effect has not been observed in a previous PCMF setup[11]. Moreover, another

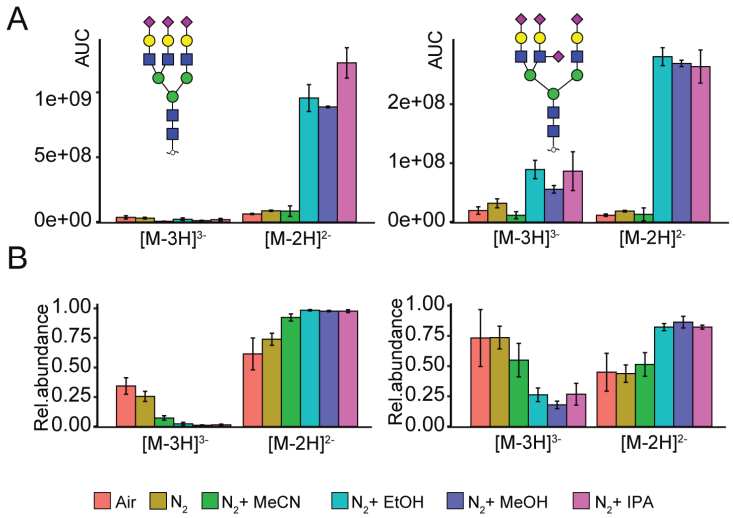


***Figure 2. Effect of different dopant solvents on the charge state distribution of N-glycans**. **(A)** Area under the curve and **(B)** relative abundance of doubly and triply charged species of (1) H6N5S3 and (2)H6N5S4 observed during PGC-LC-MS (N= 3).*

study revealed that IPA based DEN gas showed an increase in ratios of abundances for triply and doubly charged species compared to conventional CZE-MS[12]. The differences between our study and these could be explained by the different solvent compositions during ionization as compared to the post column solvent supplementation in the PCMF study[11], and no gradient elution in the CZE study[12].

EFFECT OF DOPANT SOLVENTS ON THE IONIZATION OF *O*-GLYCANS

Overall, the *O*-glycan analysis benefitted a lot from the use of DEN gas. The effects of the dopant solvents were further investigated on a wider range of analyte sizes and elution times. For this purpose, the absolute peak areas of the five most abundant *O*-glycan species in BSM were investigated over three retention time (RT) windows. Similar to the *N*-glycans, polar protic solvents (IPA, MeOH and EtOH) provide approximately a 5-fold increase in peak areas compared to MeCN for all *O*-glycans in RT windows 2 and 3 (**Figure 3A**). While similar S/N ratios were obtained for the *O*-glycans with all dopants (**Figure 3B**), MeOH shows approximately a 4-fold increase compared to the reference MeCN condition. Therefore, this dopant should be considered for applications that require high sensitivity. While IPA and EtOH provide higher relative areas of early eluting *O*-glycans, the non-enriched N_2 and ambient air

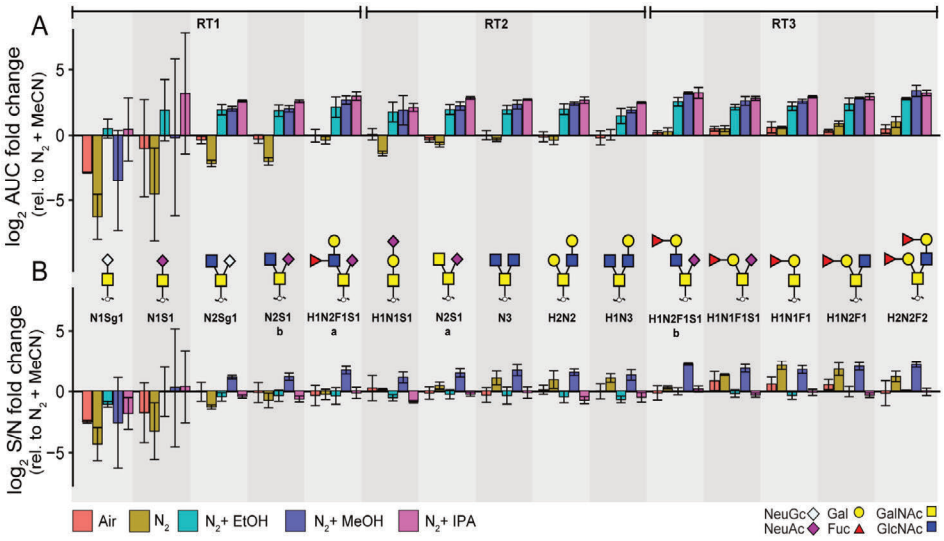


Figure 3. Effect of different dopant solvents on O-glycans. (A) *The fold change of the area under the curve relative to the reference approach (MeCN enriched N_2) and* ***(B)*** *Fold change of S/N ratios relative to the reference approach (MeCN enriched N_2) for the BSM O-glycans (N = 3) in three retention time windows corresponding to different gradient solvent composition RT1: 2-10% MeCN, RT2: 10- 15% MeCN, RT3: 15-26% MeCN. The error bars represent the standard deviation from the mean of the technical replicates (N = 3). The glycans are ordered based upon their elution time. H- Hexose, N- N-acetylhexosamine, F- deoxyhexose, S- N-acetylneuraminic acid, Sg- N-glycolylneuraminic acid.*

showed higher relative areas of later eluting glycans (**Supporting Information 1, Figure S-1B**). This clearly illustrates that the boosting effect of the dopants is dependent on the elution time of the glycans as shown in **Figure 4**. It has been reported previously that early eluting species in the PGC-LC-MS setup ionize less due to the low organic content in the gradient[10]. Therefore, it is important to note that the most significant effect of all dopants can be seen on the glycans eluting in the first 10 min of the gradient, *e.g.* sialyl Tn occupied with an *N*-acetylneuraminic acid or *N*-glycolylneuraminic acid. Those glycans elute at very low concentrations of MeCN and therefore the ionization of these compounds is expected to be significantly lower compared to the later eluting species[10]. It should be noted that the retention of these glycans to the PGC stationary phase is limited, therefore high coefficient of variations (CVs) in this region are expected. Overall, the peak areas are significantly lower if no dopants are used (solely N_2 or ambient air). Therefore, for high-sensitivity detection of these compounds, the use a dopant setup is advisable. IPA as dopant solvent gives the most pronounced boosting effect, increasing the precursor intensities which is

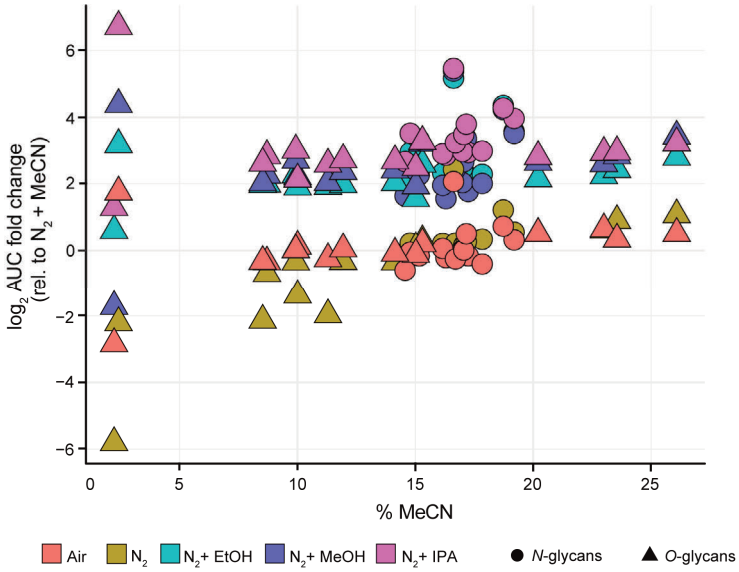


Figure 4. Early eluting glycans show the highest boost with DEN gas. *The fold change of the area under the curve with different dopant solvents for both N-and O-glycans as a function of the percent of MeCN in the elution gradient.*

reflected in the quality of the fragmentation spectra as exemplified for the early eluting *O*-glycans in **Supporting Information 1, Figure S-2 and S-3**. The MS/MS spectra measured with IPA showed many fragments useful for structural determination of the glycan and in addition their S/N were significantly improved compared to both MeOH and MeCN (**Supporting Information 2, Table S-7**). While similar fragments were observed in the spectra with MeOH as a dopant, the improved MS2 data obtained with IPA provided higher confidence in the identification of the glycan isomers due to the higher number and quality of informative fragment ions. While the proposed adaptations give considerable benefit for the analysis of released glycan alditols in negative ionization mode, these benefits may in in part depend on the specific electrospray ionization source. In recent years DEN gas has been successfully adapted to different Bruker electrospray sources and sprayers, as well as to other types of mass spectrometers[12,17]. Although alcohol-based DEN gas significantly improves the signal for both *N*- and *O*-glycans in negative ion mode, analysis of limited sample amounts are still challenging and further optimization is needed. Moreover, future studies should focus on improving the charging of highly sialylated species, to obtain informative MS/MS spectra.

CONCLUSION

In this study, we demonstrated that the usage of polar protic solvents as dopants, such as IPA, EtOH and MeOH, have a beneficial effect on the ionization of both *N*- and *O*-glycans in negative ion mode compared to the reference MeCN enriched nitrogen gas. While, IPA shows the highest effect on peak areas, increasing the precursor intensities, resulting in higher intensities of fragment ions, therefore showing large benefits for the identification of glycan structures. MeOH provides the best results in terms of S/N ratios, therefore this dopant would be particularly beneficial for high sensitivity MS-mode applications.

SUPPORTING INFORMATION

Supporting information is available free of charge via doi.org/10.1021/acs.analchem.1c00023

AUTHOR CONTRIBUTIONS

KM performed the experiments. KM, SW, TZ, GL and MW conceptually designed the work. KM, GL, and MW wrote the manuscript. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENTS

This work was supported by the European Commission's Horizon 2020 programme ''GlyCoCan'' project, grant number 676421.

REFERENCES

(1) Mysling, S.; Palmisano, G.; Hojrup, P.; Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Anal. Chem.* 2010, *82* (13), 5598–5609. https://doi.org/10.1021/ac100530w.

(2) Nguyen, S.; Fenn, J. B. Gas-Phase Ions of Solute Species from Charged Droplets of Solutions. *Proc. Natl. Acad. Sci. U. S. A.* 2007, *104* (4), 1111–1117. https://doi.org/10.1073/pnas.0609969104.

(3) Stephanie Kaspar, Thorsten Ledertheil, Ralf Hartmer, Thomas Hagedorn, Carsten Baessmann, Kerry Nugent: Increasing Peptide Identification Rates for Proteomics Samples by Controlling Peptide Charge States Using CaptiveSpray NanoBooster-Technical Note # TN-4.

(4) Alagesan, K.; Kolarich, D. To Enrich or Not to Enrich: Enhancing (Glyco)Peptide Ionization Using the CaptiveSpray NanoBooster™. *bioRxiv*. 2019. https://doi.org/10.1101/597922.

(5) Kammeijer, G. S. M.; Kohler, I.; Jansen, B. C.; Hensbergen, P. J.; Mayboroda, O. A.; Falck, D.; Wuhrer, M. Dopant Enriched Nitrogen Gas Combined with Sheathless Capillary Electrophoresis–Electrospray Ionization-Mass Spectrometry for Improved Sensitivity and Repeatability in Glycopeptide Analysis. 2016. https://doi.org/10.1021/acs.analchem.6b00479.

(6) Möginger, U.; Grunewald, S.; Hennig, R.; Kuo, C.-W.; Schirmeister, F.; Voth, H.; Rapp, E.; Khoo, K.-H.; Seeberger, P. H.; Simon, J. C.; Kolarich, D. Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and Squamous Cell Carcinoma. *Front. Oncol.* 2018, *8*, 70. https://doi.org/10.3389/fonc.2018.00070.

(7) Hinneburg, H.; Korać, P.; Schirmeister, F.; Gasparov, S.; Seeberger, P. H.; Zoldoš, V.; Kolarich, D. Unlocking Cancer Glycomes from Histopathological Formalin-Fixed and Paraffin-Embedded (FFPE) Tissue Microdissections. *Mol. Cell. Proteomics* 2017, *16* (4), 524–536. https://doi.org/10.1074/mcp.M116.062414.

(8) Thomsson, K. A.; Karlsson, N. G.; Hansson, G. C. Liquid Chromatography-Electrospray Mass Spectrometry as a Tool for the Analysis of Sulfated Oligosaccharides from Mucin Glycoproteins. *J. Chromatogr. A* 1999, *854* (1–2), 131–139. https://doi.org/10.1016/S0021-9673(99)00625-1.

(9) Harvey, D. J. Negative Ion Mass Spectrometry for the Analysis of N -linked Glycans. *Mass Spectrom. Rev.* 2020, *39* (5–6), 586–679. https://doi.org/10.1002/mas.21622.

(10) Pabst, M.; Altmann, F. Influence of Electrosorption, Solvent, Temperature, and Ion Polarity on the Performance of LC-ESI-MS Using Graphitic Carbon for Acidic Oligosaccharides. *Anal. Chem.* 2008, *80* (19), 7534–7542. https://doi.org/10.1021/ac801024r.

(11) Hinneburg, H.; Chatterjee, S.; Schirmeister, F.; Nguyen-Khuong, T.; Packer, N. H.; Rapp, E.; Thaysen-Andersen, M. Post-Column Make-Up Flow (PCMF) Enhances the Performance of Capillary-Flow PGC-LC-MS/MS-Based Glycomics. *Anal. Chem.* 2019, *91* (7), 4559–4567. https://doi.org/10.1021/acs.analchem.8b05720.

(12) Marie, A.-L.; Ray, S.; Lu, S.; Jones, J.; Ghiran, I.; Ivanov, A. R. High-Sensitivity Glycan Profiling of Blood-Derived Immunoglobulin G, Plasma, and Extracellular Vesicle Isolates with Capillary Zone Electrophoresis-Mass Spectrometry. 2021. https://doi.org/10.1021/acs.analchem.0c03102.

(13) Zhang, T.; Madunić, K.; Holst, S.; Zhang, J.; Jin, C.; Ten Dijke, P.; Karlsson, N. G.; Stavenhagen, K.; Wuhrer, M. Development of a 96-Well Plate Sample Preparation Method for Integrated: N - And O -Glycomics Using Porous Graphitized Carbon Liquid Chromatography-Mass Spectrometry. *Mol. Omi.* 2020, *16* (4), 355–363. https://doi.org/10.1039/c9mo00180h.

(14) Thacker, J. B.; Schug, K. A. Effects of Solvent Parameters on the Electrospray Ionization Tandem Mass Spectrometry Response of Glucose. *Rapid Commun. Mass Spectrom.* 2018, *32* (15), 1191–1198. https://doi.org/10.1002/rcm.8155.

(15) Huffman, B. A.; Poltash, M. L.; Hughey, C. A. Effect of Polar Protic and Polar Aprotic Solvents on Negative-Ion Electrospray Ionization and Chromatographic Separation of Small Acidic Molecules. *Anal. Chem.* 2012, *84* (22), 9942–9950. https://doi.org/10.1021/ac302397b.

(16) Henriksen, T.; Juhler, R. K.; Svensmark, B.; Cech, N. B. The Relative Influences of Acidity and Polarity on Responsiveness of Small Organic Molecules to Analysis with Negative Ion Electrospray Ionization Mass Spectrometry (ESI-MS). *J. Am. Soc. Mass Spectrom.* 2005, *16* (4), 446–455. https://doi.org/10.1016/j.jasms.2004.11.021.

(17) Lageveen-Kammeijer, G. S. M.; de Haan, N.; Mohaupt, P.; Wagt, S.; Filius, M.; Nouta, J.; Falck, D.; Wuhrer, M. Highly Sensitive CE-ESI-MS Analysis of N-Glycans from Complex Biological Samples. *Nat. Commun.* 2019, *10* (1), 2137. https://doi.org/10.1038/s41467-019-09910-7.

COLORECTAL CANCER CELL LINES SHOW STRIKING DIVERSITY OF THEIR *O*-GLYCOME REFLECTING THE CELLULAR DIFFERENTIATION PHENOTYPE

Chapter 4

Katarina Madunić[1], Tao Zhang[1], Oleg A. Mayboroda[1], Stephanie Holst[1], Kathrin Stavenhagen[1], Chunsheng Jin[2], Niclas G. Karlsson[2], Guinevere S.M. Lageveen-Kammeijer[1], Manfred Wuhrer[1]

[1] *Leiden University Medical Center, Center for Proteomics and Metabolomics, Postbus 9600, 2300 RC Leiden, The Netherlands*

[2] *Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*

Reprinted with permission from Cell. Mol. Life Sci., 2021, 78, 337–350

ABSTRACT

Alterations in protein glycosylation in colorectal cancer (CRC) have been extensively studied using cell lines as models. However, little is known about their *O*-glycome and the differences in glycan biosynthesis in different cell types. To provide a better understanding of the variation in *O*-glycosylation phenotypes and their association with other molecular features, an in-depth *O*-glycosylation analysis of 26 different CRC cell lines was performed. The released *O*-glycans were analyzed on porous graphitized carbon nano-liquid chromatography system coupled to a mass spectrometer via electrospray ionization (PGC-nano-LC-ESI-MS/MS) allowing isomeric separation as well as in-depth structural characterization. Associations between the observed glycan phenotypes with previously reported cell line transcriptome signatures were examined by canonical correlation analysis. Striking differences are observed between the *O*-glycomes of 26 CRC cell lines. Unsupervised principal component analysis reveals a separation between well differentiated colon-like and undifferentiated cell lines. Colon-like cell lines are characterized by a prevalence of I-branched and sialyl Lewis X/A epitope carrying glycans while most undifferentiated cell lines show absence of Lewis epitope expression resulting in dominance of truncated α2-6-core sialylated glycans. Moreover, the expression of glycan signatures associates with the expression of glycosyltransferases that are involved in their biosynthesis, providing a deeper insight into the regulation of glycan biosynthesis in different cell types. This untargeted in-depth screening of cell line *O*-glycomes paves the way for future studies exploring the role of glycosylation in CRC development and drug response leading to discovery of novel targets for the development of anti-cancer antibodies.

INTRODUCTION

With over 18 million new cases worldwide in 2018, colorectal cancer (CRC) is the third most common cancer in the world[1]. The disease is very heterogeneous with a high variability in patient prognosis and treatment response[2]. There have been various attempts to classify CRC patients into clinically relevant groups[2,3] using single genomic markers such as microsatellite instability (MSI), BRAF and KRAS mutations. While these markers can give insights into disease processes, they cannot provide a full understanding of the molecular pathology and prediction of patient outcome[4]. Therefore, the CRC subtyping consortium (CRCSC) has recently made a systematic comparison of gene mutation and gene expression of primary tumors from a large set of samples and defined four subtypes of CRC with different clinical and molecular markers[4]. Consensus molecular subtype (CMS)1 tumors show prevalence of MSI together with high immune infiltration in the tumor microenvironment, associated with a better patient prognosis[3]. In contrast, the mesenchymal CMS4 tumors are characterized by infiltration with cancer associated fibroblasts and upregulation of epithelial to mesenchymal transition (EMT) resulting in worse overall patient prognosis[3]. Both CMS2 and CMS3 tumors show strong epithelial differentiation signatures, with characteristic metabolic pathways dysregulation in the CMS3 group[4]. Although the proposed classification provides a deeper understanding of CRC and its differential molecular signatures, it is not yet clear which features will be relevant for accurate patient stratification. In order to design subtype specific therapeutic strategies, translation of the CMS classification to preclinical models is needed, to enable large-scale drug screenings.

Various genetic studies have confirmed that cancer cell lines recapitulate the molecular features of the tumors[5,6] and the same has been confirmed for CMS[7] . Recently, profiling of 34 CRC cell lines revealed consistency at the gene, microRNA and protein levels, dominated by two distinct clusters. The colon-like cluster has high expression of gastrointestinal specific markers, while the second cluster contains undifferentiated cell lines showing upregulation of transforming growth factor (TGF)-β induced genes and EMT signatures[8]. These two groups significantly associate with CMS groups, where CMS1 and CMS4 cluster as undifferentiated, and CMS2 and CMS3 as colon-like[8].

Next to genetic, metabolic and proteomic signatures, protein glycosylation is a major factor in colon differentiation and CRC development[9]. It has been shown that malignant transformation changes the glycosylation machinery of the cells, affecting the function of the oncogenic receptors that are involved in the control of cell proliferation and differentiation[10]. Moreover, glycan binding proteins, expressed by immune cells in the tumor microenvironment, respond to these changes, often resulting in an immunosuppressive response[11]. Therefore, unravelling glycan-based interactions in cancer is instrumental for disclosure of molecular mechanisms underlying cancer biology.

Due to their continuous availability, cell lines are often used as models for studying glycosylation changes in cancer. Recently, the *N*-glycosylation of a set of CRC cell lines has been characterized, revealing association of antennary fucosylation with differentiation and Caudal Type Homeobox 1 (CDX1) expression[12,13]. Another major class of colon glycans are mucin-type *O*-glycans, mainly carried by heavily glycosylated mucin proteins, which are the major components of mucus layer in the gastrointestinal tract. Mucin type *O*-linked glycosylation is initiated by the transfer of *N*-acetylgalactosamine (GalNAc) to Ser/Thr of both mucin and non-mucin glycoproteins which are shown to be altered in various cancers including CRC[14,15]. Unfortunately, little is known about *O*-glycosylation of cell lines due to its complexity, the presence of multiple isomeric structures as well as the lack of enzymatic release methods, making it overall a challenging task[14].

Here, we present an in depth structural analysis of *O*-glycosylation phenotypes of 26 CRC cell lines derived from both primary tumors and metastatic sites. We optimized a 96-well plate PVDF membrane based method[12] for preparation of released *O*-glycans from 500,000 cells via reductive β-elimination[16]. Released *O*-glycans were analyzed on a sensitive analytical platform, namely, porous graphitized carbon nano-liquid chromatography coupled to a tandem mass spectrometer (PGC nano-LC-ESI-MS/MS) using negative electrospray ionization. Major differences are observed between 26 analyzed CRC cell lines, revealing the diversity of the CRC cell line *O*-glycome. Moreover, associations are found between the observed glycome phenotypes and cell line gene expression as well as their differentiation.

MATERIALS AND METHODS

CELLS AND CELL CULTURE

Human CRC cell lines were obtained from the Department of Surgery of the Leiden University Medical Center (LUMC), Leiden, The Netherlands, as well as the Department of Pathology of the VU University Medical Center (VUmc), Amsterdam, The Netherlands. Further details are provided in **Supporting Information, S-1**.

O-GLYCAN RELEASE AND ANALYSIS

Lysed cell pellets containing 5 x 10^5 cells were loaded to the preconditioned PVDF membrane plate wells and denatured with guanidine hydrochloride and dithiothreitol (DTT) at 60 °C. After removing denaturation agent, *N*-glycans were released by PNGase F digestion overnight at 37 °C. Upon removal of *N*-glycans, 50 µL of 0.5 M sodium borohydride ($NaBH_4$) in 50 mM potassium hydroxide (KOH) was added to each well and incubated for 16 hrs at 50 ˚C for the release of *O*-glycans via reductive β-elimination. Desalting of the samples was performed using a cation exchange resin Dowex 50W X8 which was self-packed into 96-well filter plates. Desalted *O*-glycans were further purified via solid phase extraction by packing bulk sorbent carbograph slurry into 96-well filter plates. Analysis was performed using a PGC nano-LC-ESI-MS/MS platform. More details are provided in **Supporting Information, S-1**.

GLYCAN STRUCTURE ANALYSIS AND RELATIVE QUANTIFICATION

Identification of glycans was performed based on PGC retention time, known biosynthetic pathways, and manual inspection of fragmentation spectra following known MS/MS fragmentation pathways of *O*-glycan alditols in negative-ion mode[17,18]. Glycan sequences and linkages were confirmed by the analysis of glycans upon α2-3 neuraminidase, α1-3/4 fucosidase, and β1-4 galactosidase digestion. Relative quantitation was performed on the total area of all *O*-glycans within one sample normalizing it to 100 %. MS/MS mass lists were exported from the DataAnalysis software for upload to Unicarb DR repository[19]. A more detailed description is provided in **Supporting Information, S-1**.

STATISTICAL ANALYSIS

An imputation of the minimum positive number (0.0001) was performed to enable use of the statistical tools sensitive to the missing values such as principal component analysis (PCA). Regularized canonical correlation analysis was performed using rcc

function as it is implemented in the "mixOmics" package[20]. Data analysis and visualization was performed in ''R'' software.

RESULTS

HIGH THROUGHPUT AND ROBUST *O*-GLYCAN RELEASE FROM CELLS

In order to analyze the *O*-glycosylation of 26 CRC cell lines, we established a high throughput sample preparation in 96-well format (**Supporting Information, Figure S1**). The combination and optimization of two previously established protocols[12,16] allowed sequential release of both *N*- and *O*-glycans from cell lysates in a higher throughput manner using 96-well plates. *O*-glycans were analyszd on PGC nano-LC-ESI-MS/MS platform which revealed 178 different *O*-glycan structures. Of these, 153 passed the quality control criteria and were included in the analysis (**Supporting Information, Figure S3** and **Supporting Information, Table 1: S3- S31**). To assess the technical and biological variation of *O*-glycan profiles for each cell line, the complexity of each sample was reduced by compiling the relative peak areas for all glycans to single mass spectrometry average compositions (MSAC)[21] (**Supporting Information, Table 1: S1**) representing the normalized number of sugar residues and modifications per glycan molecule. Low technical variability of our workflow is illustrated by the close clustering of scores in the PCA model from two technical replicates of each cell line as well as the triplicate of *O*-glycans released from bovine fetuin (standard) (**Supporting Information, Figure S2**). In addition, the close clustering scores from the cell lines which were cultured and analyzed in three biological replicates (HT29, HCT116, SW480, SW620, and HCT8) revealed a low biological variability (marked as A, B and C in the respective PCA plot). The highest variability in glycosylation profiles was observed in cell lines SW480, HT29 and HCT116, which may be attributed to the fact that the replicates for these cell lines were prepared at different sites (VUmc and LUMC) as well as using different media (detailed description in **Supporting Information, S-1**).

HIGH DIVERSITY OF CRC CELL LINES *O*-GLYCOSYLATION PROFILES

The comparison of the 26 CRC cell lines revealed striking differences between the *O*-glycomes (**Supporting Information, Figure S3**). Many, yet undescribed, *O*-GalNAc linked glycans were detected varying in size from two up to 14 monosaccharide residues. An illustration of the diversity between cell lines is shown in **Figure 1**. The

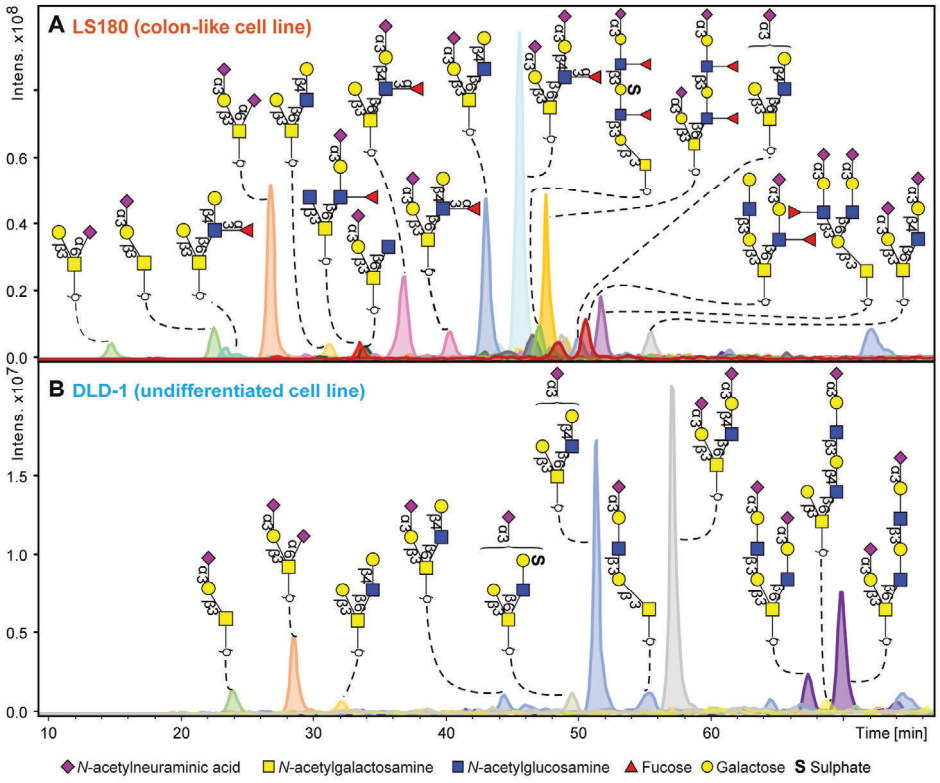


Figure 1. O-glycan profiles from two exemplary CRC cell lines. (A) *O-glycan profile of the mucin secreting well differentiated (colon-like) cell line LS180 is characterised by high expression of sialyl Lewis X/A and Lewis X/A antigens. These antigens were found to be present on core 1, core 2 and core 4 glycan structures, both in linear and I-branched form. In contrast, panel* ***(B)*** *displays poorly differentiated human colon adenocarcinoma cell line DLD-1, showing an O-glycan profile which is dominated by core 2 sialylated glycans lacking any fucosylated antigens.*

top panel shows the glycan profile of the mucin secreting cell line LS180 (human colon adenocarcinoma) which is characterized by high expression of sialyl Lewis X/A and Lewis X/A antigens. These antigens were found to be present on core 1, core 2 and core 4 glycan structures, both in linear and I-branched form. In contrast, DLD-1 (bottom panel), a poorly differentiated human colon adenocarcinoma cell line, showed an *O*-glycan profile which is dominated by core 2 sialylated glycans lacking any fucosylated antigens.

Only three glycans were present in all samples, albeit at highly varying abundances: disialyl-T antigen (NeuAcα2-3Galβ1-3(NeuAcα2-6)GalNAcol), the α2-3 sialyl-T antigen (NeuAcα2-3Galβ1-3GalNAcol), and the di-sialylated core 2 glycan (NeuAcα2-

3Galβ1-3(NeuAcα2-3Galβ1-4GlcNAcβ1-6)GalNAcol). The main core structures in most cell lines were core 2 followed by elongated structures of core 1. Core 3 structures were only detected in low amounts, while core 4 structures were expressed the most in HCT-15, LOVO and SW1116 cell lines (**Supporting Information, Figure S4**). Interestingly, all cell lines were dominated by sialylated glycan species (**Supporting Information, Figure S4**). *N*-acetylneuraminic acids (Neu5Ac) on *O*-glycans were mostly α2-3-linked to a galactose residue. Moreover, α2-6-linked sialylation was observed on the innermost GalNAc predominantly in the context of sialyl-T and disialyl-T antigens. Cell lines HT29 and WiDr showed the highest expression of α2-3-sialylated *O*-glycans (**Supporting Information, Table 2: S2**). In addition, 42 *O*-glycans were found to contain sulphate modifications mainly expressed by SW948 and LS411N cell lines (**Supporting Information, Table 2: S2**).

A major advantage of PGC chromatography is the high separation power that enables to discriminate between glycan linkages and positional isomers[17,18]. Identification of *O*-glycans was performed based on PGC retention time, described biosynthetic pathways and manual inspection of fragmentation spectra following known MS/MS fragmentation patterns of *O*-glycan alditols in negative-ion mode [17,22]. All annotated

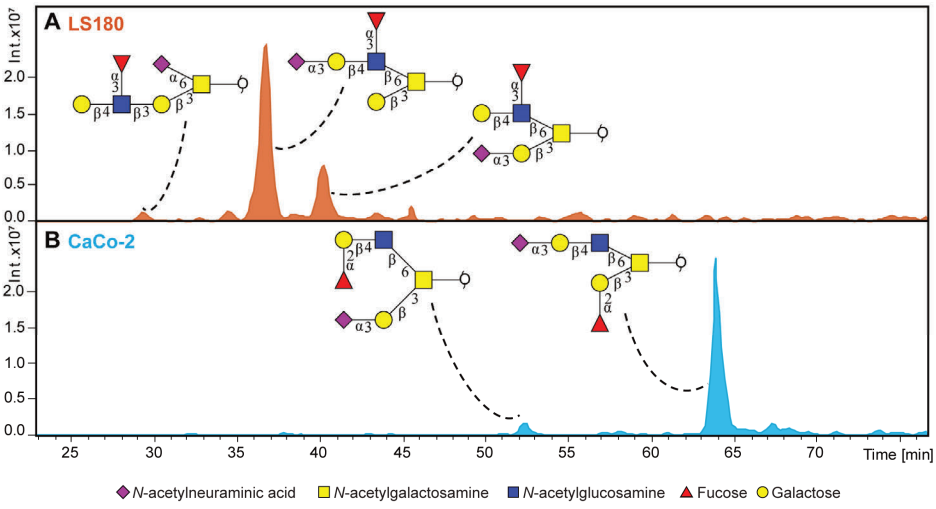


Figure 2. Chromatographic separation of O-glycan isomers at m/z 1186.40 [M-H]⁻ carrying different antigens*. The extracted ion chromatograms show a different retention behaviour of Lewis antigen isomers in the* ***(A)*** *LS180 cell line. Namely, core 1 carrying Lewis X antigen with α2-6-linked NeuAc linked to the core GalNAc (RT= 30.2 min), core 2 glycan carrying sialyl-Lewis X antigen on the 6 arm (RT= 37.2 min), and core 2 glycan carrying Lewis x antigen on the 6 arm (RT = 40.7 min). Additionally, the* ***(B)*** *CaCo-2 cell line illustrates the separation of core 2 blood group antigen H type 2 (RT = 52.4 min), and a core 2 mucin blood group antigen H type 3 (RT = 63.8 min).*

structures are listed in **Supporting Information, Table 2: S1**. MS/MS peak lists with glycan annotations per cell line are available via an online repository Unicarb DR[19] (http://unicarb-dr.biomedicine.gu.se/). **Figure 2** shows the powerful chromatographic separation of five glycan isomers with the same composition H2N2F1S1 in both LS180 and CaCo-2 cell lines. Glycan sequences and linkages were confirmed by the analysis of glycans upon α2-3 neuraminidase digestion, as well as additional combined α1-3/4 fucosidase and β1-4 galactosidase digestion as demonstrated by **Supporting Information, Figure S5-7**. With this approach we were able to identify the most abundant glycans representing more than 95% of the relative intensity for 13 cell lines, and over 90 % of the relative intensity for the remaining 13 cell lines.

GLYCAN TRAITS ARE ASSOCIATED WITH CELL LINE DIFFERENTIATION

To explore the specific *O*-glycan phenotypes in CRC more in-depth, the assigned glycans were relatively quantified and grouped based on glycan structural features such as core (1, 2, 3, or 4), I-branch (GlcNAcβ1-6Gal-R), α2-3- or α2-6-sialylation, Lewis X/A (Galβ1-4/3(Fucα1-3/4)GlcNAc-R, sialyl Lewis x/a (NeuAcα2-3Galβ1-4/3(Fucα1-3/4)GlcNAc-R, (sialyl) dimeric Lewis X/A (NeuAcα2-3Galβ1-4/3(Fucα1-3/4)GlcNAcβ1-3Galβ1-4/3(Fucα1-3/4)GalNAc-R, blood group A GalNAcα1-3(Fucα1-2)Galβ1-3/4-R, blood group B (Galα1-3(Fucα1-2)Galβ1-3/4-R, blood group H type 3 (Fucα1-2Galβ1-3(R-)GalNAcol) and blood group H type 2 (Fucα1-2Galβ1-4GlcNAc-R) (**Supporting Information, Table 2: S2**). The structures that could not be unambiguously assigned were not included in the calculations of the structural features. The obtained CRC *O*-glycosylation signatures were further explored by PCA (**Figure 3**). Analysis of the score and loadings plot (**Figure 3A** and **B,** respectively) of the model shows that HT29 and its derivative WiDr cell line are positioned close to the center of the Hoteling circle and as such could be viewed as the examples of the average glycosylation profile. The cell lines derived from the same patient (HCT15, HCT8 and DLD-1) could also be considered as the examples of the average profile. Yet, a distance between them indicates the differences in their *O*-glycomes. This is supported by the **Figure 4** where DLD-1 cell line shows higher expression of I-branched glycans, and no expression of blood group antigen H, while the HCT8 and HCT15 express glycans carrying blood group H-antigens. Moreover, another two cell lines derived from the same patient, SW480 from primary tumor and SW620 from lymph node metastasis, do not cluster together on the score plot. Here, differences

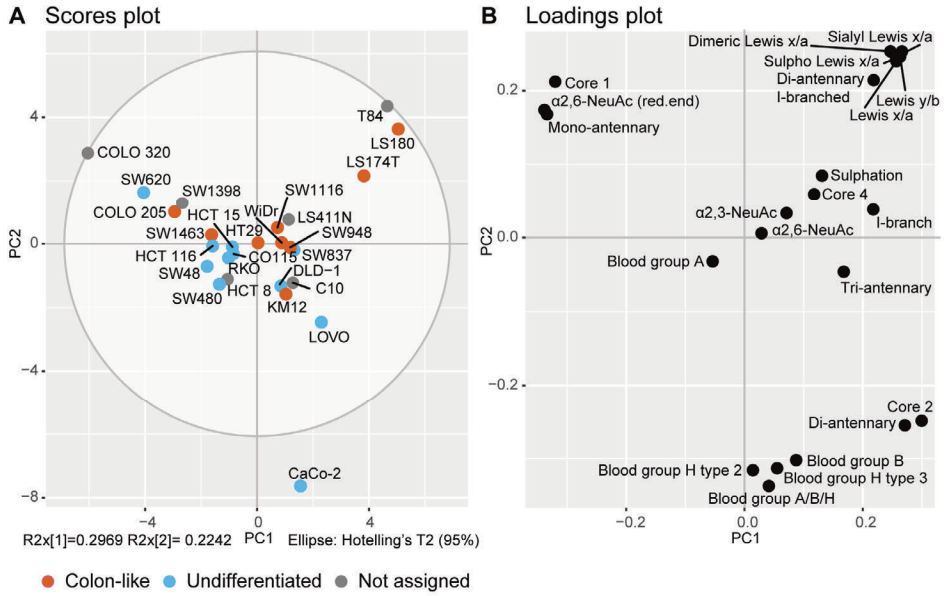


Figure 3. PCA based on relative abundance (%) of calculated structural glycan features*. A separation between colon-like and undifferentiated cell lines is illustrated in the* ***(A)*** *PCA score plot of PC1 against PC2.* ***(B)*** *The PCA loading plot displays the variables that drive the separation in the PCA model. The top three of principal components explain 68.58 % of the variation within the data. Biological and technical replicates were averaged per cell line.*

are found in the expression of blood group antigen H carrying glycans, which are present in higher levels in SW480 cell line, while SW620 does not express any fucosylated epitopes. Finally, the model reveals a similarity in the glycosylation profiles of the closely positioned cell line variants LS180 and LS174T on the outskirts of the score plot, as a result of very high expression of Lewis x/a antigens.

Differentiation phenotype of the cells (colon-like and undifferentiated)[8] appears as a visible trend separating the first two components of the score plot (**Figure 3A**), while no association could be found with other cancer cell characteristics (e.g. MSI, oncogene mutations, tumor stages; data not shown). In order to decipher which structural glycan features drove the separation in the PCA model, we explored the variables in the PCA loadings plot (**Figure 3B**). Cell lines clustering in the left part of the score plot in **Figure 3A** (Colo320, SW620, SW1398, SW48, SW480, HCT116, RKO, Co115, and HCT15) have higher expression of core 1 and α2-6-linked Neu5Ac linked to the core (**Figure 3B**) which is also displayed in **Supporting Information, Figure S4**. Most of the cell lines in this cluster have been previously classified as

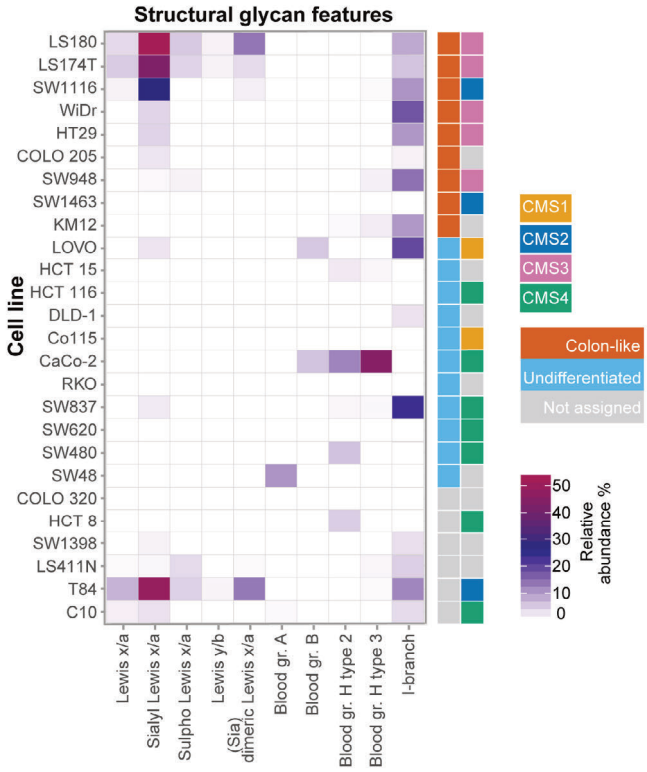


***Figure 4. Relative abundance of structural glycan features per cell type**. Geometrical tile of the relative abundance (%) of the calculated structural glycan features (x-axis) and cell line type (y-axis). Different classifications of the cell lines are displayed with colour codes based on gene expression (colon-like in red and undifferentiated in light blue) and consensus molecular subtypes (CMS1 in yellow, CMS2 in dark blue, CMS3 in pink, and CMS4 in green). Non-assigned cell lines were marked in grey for the gene expression as well as for the CMS status.*

undifferentiated based on low expression of gastrointestinal specific genes[8]. On the other hand, mucin secreting cell lines LS180, LS174T[23] and T84[24] are characterized by abundant and diverse glycosylation with very high expression of Lewis antigens clustering in the upper right part of the PCA score plot. In contrast, CaCo-2 cell line displays a different phenotype, rich in blood group H type 2 and type 3 antigens, as well as Sd[a] (GalNAcβ1-4(NeuAcα2-3)Galβ-) and Cad (GalNAcβ1-4(NeuAcα2-3)Galβ1-3(NeuAcα2-6)-) antigens. Colon-like cell lines show higher expression of Lewis-like antigens, predominantly sialyl Lewis X/A epitopes (**Figure 4**) which could be a trait of the CMS3 metabolic subtype (represented by LS180, LS174T, HT29, WiDr and SW948). The I-antigen branching is found more often in the colon-like cell lines

(**Figure 5**) such as WiDr, SW948, SW1116 and HT29. However, one of the undifferentiated cell lines, SW837, shows the highest expression of I-branched glycans. Most of CMS4 cell lines are characterized by a high level of overall α2-3-sialylation, no expression of Lewis antigens (with exception of SW837 and C10 cell lines), together with relatively higher expression of blood group antigen type 2.

GLYCOMIC SIGNATURES ASSOCIATE WITH CELL LINE TRANSCRIPTOMICS

In order to understand which genes are underlying the glycosylation phenotypes observed in this study, we performed a canonical correlation analysis using mRNA expression data from an online available dataset[8]. Primarily, we selected genes involved in the biosynthesis of *O*-glycans which showed the largest fold change in expression ($\log_e > 0.5$) when comparing colon-like and undifferentiated cell lines with high statistical significance (Bonferroni corrected *p*-value < 0.05), displayed in the volcano plot (**Supporting Information, Figure S8**). Colon-like cells show higher expression of the following genes encoding for glycosyltransferases (GTs): *B3GNT3, FUT2, FUT3, FUT6, GALNT3, GCNT3,* and *ST6GALNAC1.* In contrast, the undifferentiated cell lines show a slightly higher expression of *ST3GAL3.* The correlations of glycan structural features with expression of the selected genes are illustrated in the clustered image map of the canonical model (**Figure 5**). A moderate correlation between the *FUT3* and Lewis x/a (r = 0.40) as well as between *FUT6* and sialyl Lewis x/a epitopes (r = 0.36) is observed. *ST6GALNAC1* gene expression also correlates with the expression of sialyl Lewis X/A (r = 0.38) and Lewis x/a epitopes (r = 0.40). Additionally, *GCNT3* shows a strong correlation with α2-3-sialylation (r = 0.70) and I-branching (r = 0.66) reflecting the expression of sialylated I-branched diantennary and triantennary *O*-glycans. *GALNT3* encoding for *O*-glycosylation initiating enzyme shows positive correlation with core 2 (r = 0.43), I-branching (r = 0.44), and α2-3-sialylation (r = 0.38), as well as a negative correlation with α2-6-core sialylation (r = -0.50), and core 1 glycans (r = -0.45). *B3GNT3*, a gene encoding for β-1-3-*N*-acetylglucosaminyltransferase 3 and involved in the biosynthesis of poly-*N*-acetyllactosamine chains, is found to be negatively correlated with the expression of blood group A, B or H (blood group A/B/H) antigens (r = -0.57), core 2 glycans (r = -0.45) and shows a positive correlation with core 1 glycans (r = 0.40). *ST3GAL3* is the only gene involved in the *O*-glycan biosynthesis which shows higher expression in the

undifferentiated cell lines, displaying correlation with core 1 (r = 0.45) and α2-6-sialylated glycan expression (r = 0.46).

TRANSCRIPTIONAL REGULATION OF GLYCOSYLTRANSFERASE EXPRESSION

To gain more insight into the regulation of *O*-glycosylation, we also examined the associations of glycan epitopes with transcription factors which showed the highest fold change in expression ($\log_e > 0.5$) when comparing colon-like and undifferentiated cell lines[8] with high statistical significance (Bonferroni corrected *p*-value < 0.05) (**Supplementary Figure S8**). The colon-like cells show significantly higher expression of the following transcription factors: *CDX1, ETS2, HNF1A, HNF4A, MECOM* and *MYB.* In contrast, the undifferentiated cell lines showed elevated expression levels of the following transcription factors: *MLLT10, MSX1, SIX4, ZNF286A* and *ZNF286B*. The clustered image map of the canonical model (**Figure 5**) displays the moderate positive correlation between the expression of Lewis antigens and transcription factors *CDX1* and *ETS2*. Relatively strong correlation is seen for *ETS2* with α2-3-sialylation (r = 0.57) and I-branching (r = 0.51). The *MYB* gene, which is highly expressed in colon-like CRC cells, showed a strong positive correlation with the expression of core 1 (r = 0.59) and α2-6-sialylated glycans (r = 0.57) and a strong negative correlation with blood group A, B or H (Blood group A/B/H) carrying structures (r = -0.62). On the other hand, two transcription factors *ZNF286A* and *ZNF286B,* expressed in undifferentiated cell lines, show high negative correlation with α2-3-sialylation (r=-0.63, and r= -0.60 respectively) and I-branching (r = -0.55 and r = -0.53 respectively), together with a positive correlation with blood group antigen A, B or H carrying glycans (r= 0.58 and r= 0.53, respectively). To identify the possible transcriptional regulation of specific GTs, we tested the associations between transcription factor expression and GT expression (**Supplementary Figure S9**)**.** In contrast with the rather weak correlations of *MYB* expression and Lewis antigen expression, *MYB* expression does show correlation with *FUT3* and *B3GNT3* genes. We also observe strong negative correlations between expression of *ZNF286A* and *ZNF286B* genes with the expression of *GCNT3* and *B3GNT3* genes involved in the elongation and branching of *O*-glycan molecules next to a positive correlation with the expression of *ST3GAL3*.

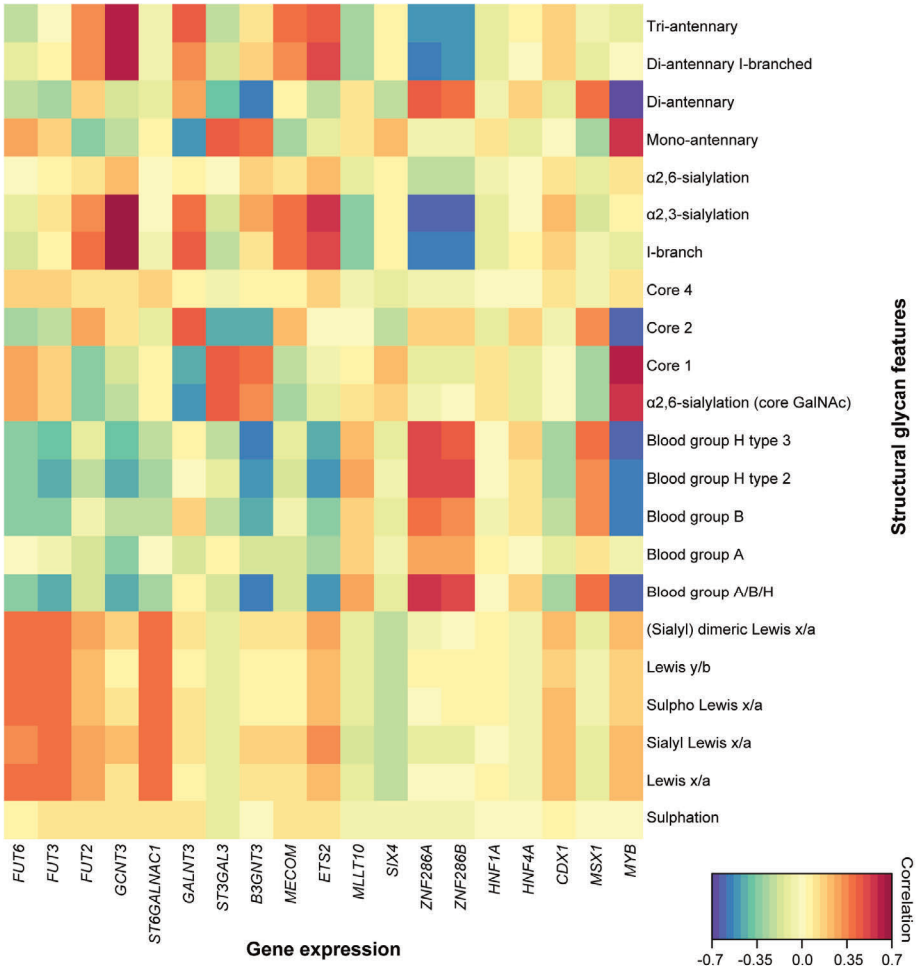


Figure 5. Associations of structural glycan features with gene expression. *The clustered image map of the canonical model displaying associations of glycan structural features with gene expression of relevant GTs and transcription factors differentially expressed in colon-like vs undifferentiated cell lines*

DISCUSSION

In the present study, we investigated the *O*-glycosylation phenotypes of 26 CRC cell lines derived from both primary tumors and metastatic sites, revealing pronounced differences between the cell lines. The PGC nano-LC-MS/MS platform allowed separation of isomeric *O*-glycan species which were structurally elucidated relying on negative mode tandem mass spectra and exoglycosidase treatment. Using this approach, we were able to create a detailed CRC cell line *O*-glycan MS/MS spectral

library, which will be available via https://unicarb-dr.biomedicine.gu.se/, serving as an important resource leading towards automated *O*-glycan identification via spectral matching tools.

An exploratory, qualitative analysis of the entire pool of our data clearly shows a difference in *O*-glycome profiles between the colon-like well differentiated cell lines and undifferentiated ones. A proposed model explaining the differences in glycosylation and biosynthesis in different cell types is displayed in **Figure 6**. For instance, the well-differentiated cell lines show an overall higher expression of Lewis antigens and I-branched glycans, while the undifferentiated cell lines show a higher abundance of glycans carrying an α2-6-linked NeuAc to the core GalNAc. Few exceptions could also be observed. For instance, Colo205 a colon-like but also a metastatic cell line[25], clusters together with the undifferentiated cell lines expressing higher relative amounts of core 1 α2-6-sialylated glycan species. Similarly, the rectal adenocarcinoma SW1463 cell line[23] does not show expression of Lewis antigens characteristic for other mucus producing cell lines. Such observations can be valuable on their own, but offer only a limited space for a functional interpretation of the data. Taking advantage of the existing pool of published transcriptomics data on the selected cell lines, we attempted to generate regulational hypotheses using the correlations between the *O*-glycan profiles and the transcripts most strongly associated with the differentiation status of the studied cell lines. Associations were found between gene expression and cell line differentiation based on the *O*-glycome profiles of 26 CRC cell lines. Here, Lewis antigen expression was found to be the most abundant in the mucin secreting cell lines LS174T [24], LS180 and T84[26] correlating with the expression of genes *FUT3* and *FUT6* encoding for the GTs involved in the biosynthesis of the Lewis antigens. Expression of Lewis antigens was also associated with *ST6GALNAC1* gene expression encoding for an α2-6-sialyltransferase which acts on the core GalNAc residue. Although we have detected sialyl-Tn antigen in the mucin secreting cell lines (LS180 and LS174T), we could not quantify mono- and disaccharide *O*-glycan alditols in a reliable manner due to limitations of the method. Previously, sialyl-Tn antigen was found to be highly expressed in LS174T cell line, and also showed correlation with upregulation of *ST6GALNAC1*[27]. In concordance with previous studies, demonstrating an association of CDX-1 transcription factor with the expression of GTs *FUT3* and *FUT6* and expression of multifucosylated *N*-glycans in

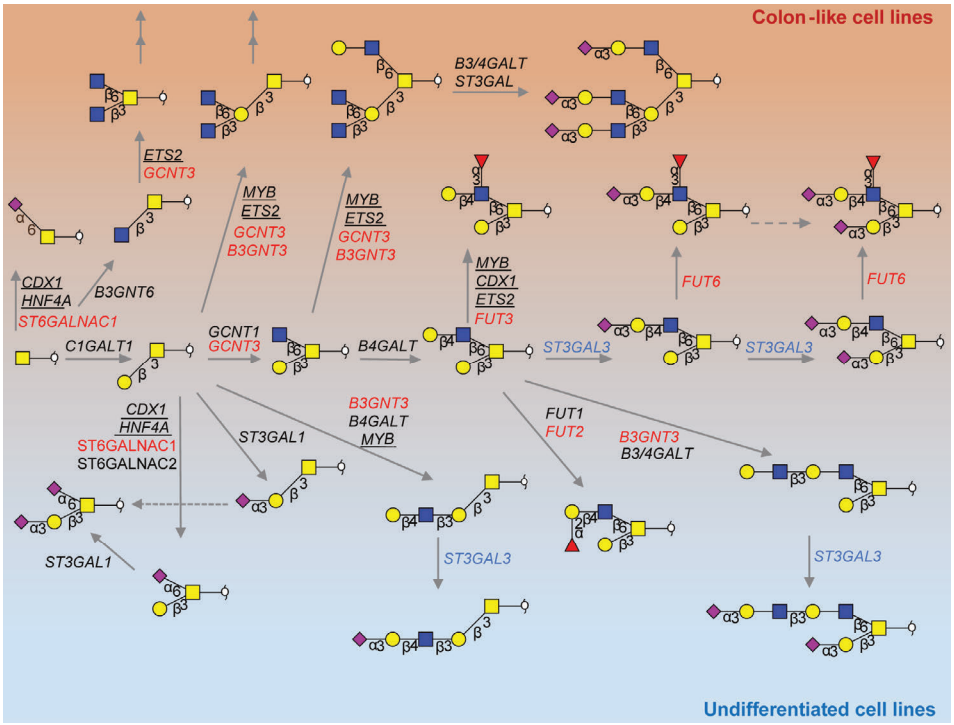


Figure 6. Proposed model explaining the differences in glycosylation between colon-like and undifferentiated cell lines. *The most abundant structures in colon-like and undifferentiated cell lines and their biosynthesis pathways with genes encoding for the GTs involved in the biosynthesis. GTs up-regulated in colon-like cell lines are marked in red, whereas genes upregulated in undifferentiated cell lines are marked in blue. Doubled arrows indicate structures that can be elongated further by action of different GTs. Transcription factors showing correlation with the expression of GTs are underlined. Please note that other GTs could also be involved in the biosynthesis of the displayed glycans*

cell lines[12,13], a correlation between Lewis antigen expressing cell lines and *CDX1* expression was observed. Moreover, some cell lines with high expression of α2-3-sialylated and multifucosylated *N*-glycans[12] (HT29, WiDr, T84, and LOVO) also revealed a high expression of *O*-glycans carrying sialyl Lewis X/A antigens. Strikingly, the cell lines with the highest expression of sialylated Lewis X/A type antigens on *O*-glycans (LS180, LS174T and SW1116) express mainly non-sialylated Lewis antigens on *N*-glycans[12]. These findings emphasize the importance of studying glycosylation in a glycan type-specific manner, as they can exhibit different glycan motifs and may convey different functions.

In contrast to high Lewis antigen and I-branching signatures of colon-like cell lines, the undifferentiated cell lines show a high abundance of core 1 glycans carrying an α2-6-

linked Neu5Ac to the core GalNAc. This signature is mainly reflecting the high relative abundance of the ubiquitous disialyl-T antigen compared to a lower abundance of elongated and branched *O*-glycans. The addition of Neu5Ac in the α2-6-position of the Tn or T antigen will prevent formation of different core structures and further elongation of cores with different structural epitopes[28]. This may explain the absence of core 3 structures in cell lines, which similarly to inflamed colon tissue[21], show low abundance of core 3 structures compared to normal colon tissue. Some reports have shown that upregulation of *ST3GAL3*, *ST6GALNAC1* and *ST6GALNAC2* could be associated with invasion to lymph nodes and poor patient prognosis [29]. Furthermore, enhanced sialylation of tumour cells can engage inhibitory sialic acid binding lectin-Siglec receptors on natural killer (NK) cells, providing an immune evasion mechanism[11]. The competition between different GTs for the core GalNAc residue can explain the strong inverse correlations between core 2 and α2-6-sialylation observed in our data (**Figure 6**). Competition involved in core 2 glycan biosynthesis such as *GCNT1*, *GCNT3* and *ST6GALNAC1* have been described extensively in the literature[30,31]. *GCNT3* gene encodes a mucin type β1,6-*N*-acetylglucosamine transferase, which adds a *N*-acetylglucosamine (GlcNAc) to the core GalNAc forming core 2 or core 4 structures. Importantly, *GCNT3* also has I-antigen biosynthetic activity, by adding the residue to the terminal galactose forming another branching point. *GCNT3* expression has been associated with reduction of proliferation and invasion as well as with higher sensitivity to the chemotherapeutic 5-Fluorouracil[32]. Low *GCNT3* expression has been proposed as prognostic marker that could be used to identify early-stage colon cancer patients at high risk of relapse[33]. It has been demonstrated that the expression of *GCNT3* is higher in the colon-like cell lines[8], showing a strong correlation with the expression of I-branched glycans as well as with α2-3 sialylation in our data. However, one of the undifferentiated cell lines SW837, derived from the rectum[34], shows the highest expression of I-branched glycans.

Our results also show expression of histo-blood group ABH antigens carried by *O*-glycans mainly from undifferentiated cell lines. Apart from the erythrocytes, the ABH antigens can be expressed also in the gastrointestinal tract and secretions, where *Se*-FUT enzyme (*FUT2*) is responsible for their biosynthesis[35]. However, several reports have demonstrated the expression of blood group ABH antigens in colon cancers from non-secretors with inactive enzyme[36,37]. Cell lines showing expression of blood group

H antigens such as CaCo-2, HCT15, HCT8, SW948, SW1116, and SW837 were derived from tumors of blood group O secretor positive individuals[38]. We have also observed expression of blood group B antigens by CaCo-2 cell line, and this incompatible expression in the gut of blood group O individuals has been previously observed in colon cancer[39]. LOVO cell line, derived from a B blood type individual, expressed the blood group B carrying glycans, however the SW48 (AB blood group individual) shows expression of only blood group A glycans. Cell lines SW480, SW620 and HT29 derived from blood group A individuals do not show expression of the A antigen carried by *O*-glycans although these cell lines have shown a substantial A transferase activity[40]. HCT8 cell line derived from the ileocecal colon shows higher expression of blood group antigens H type 2, compared to HCT15 derived from sigmoid colon of the same individual, which is in concordance with the reports demonstrating a decreasing gradient of ABH antigen expression towards the distal part of the colon[41]. However, more research is needed to see if CRC cell line glycosylation is comparable with blood group antigen distribution in the healthy colon. We have also observed that blood group antigen H types 2 and 3 were more abundant than H type 1 antigen in CRC cell lines. Similarly, antibody staining studies showed region specific expression of mainly H type 1 blood group antigen H in healthy colon tissue, in contrast to colon cancer tissues aberrantly expressing blood group antigens H type 2 and 3 associated with tumor progression in the distal colon and rectum[36].

The majority of the cell line *O*-glycans had α2-3-linked Neu5Ac at the terminal end of the glycan. Overall sulphation was found to be much less abundant than sialylation, showing highest expression in SW948 and LS411N cell lines. As compared with normal colon tissue[41], cancer associated *O*-glycans often show increased sialylation and less sulphation[42]. Furthermore, we detected structures carrying Sd^a/Cad epitopes only in one cell line, namely, CaCo-2, supporting other reports from the literature [43]. Structures terminated by Sd^a/Cad epitopes have been described before as characteristic for normal colon mucin tissue[44,45]. Our results also show a dominance of core 2 and core 1 *O*-linked glycans, and a very low abundance of core 3 and core 4 structures. Core 3 and core 4 are known to be characteristic for the healthy colon mucins[46] and most of the cell lines did not show high expression of those structures. Moreover, core 3 synthase has been reported to be downregulated in colon cancer[47], with a loss of activity in many CRC cell lines[48]. Additionally, a decrease in core 3

structures together with a concomitant increase in core 1 structures has been observed when moving from healthy tissue to tumour regions [15]. Previously, no core 3 structures could be detected in the glycan profiles of five CRC cell lines, in contrast to substantial expression in CRC tissues[27]. We can hypothesize that due to the lack of precursor core 3 structures, *GCNT3*, which is highly expressed in many cell lines, exhibits the core 2 and I-branching activity, resulting in predominance of core 2 and I-branched structures in the cell lines.

To provide a better understanding of regulation of *O*-glycosylation, we also examined the associations of glycan epitopes with transcription factor expression, and generated hypotheses by association analysis. *ETS2* transcription factor expression shows associations with I-branched and α2-3-sialylated glycan expression (**Figure 5**) supported by its correlation with the expression of the *GCNT3* gene (**Supplementary Figure S9**). *ETS2* is a Wnt pathway target gene, whose inactivation leads to increased crypt cell proliferation[49]. It has previously been associated with *MGAT2* promoter activation for the biosynthesis of complex type *N*-glycans, however its potential regulation of *O*-glycan biosynthesis is still unknown. Oncogenic transcription factor c-Myb, important for both cell proliferation and cell differentiation[50], is also involved in the Wnt pathway activation, and shows a positive correlation with the expression of core 1 and α2-6-core sialylated glycans (**Figure 5**). Activated β-catenin and *MYB* induce upregulation of *MYC* promoter, and c-Myc transcription factor expression[51]. It has been found that *ST3GAL1, 2 and 4* are transcriptionally upregulated by c-Myc[52], which might contribute to the higher expression of disialyl and sialyl-T antigens, together with a relatively lower expression of core 2 structures (**Figure 5**). The very strong negative correlation with blood group antigen H expression in the samples could be explained by the strong correlation of *MYB* gene with *B3GNT3* gene and is in direct competition with the biosynthesis of blood group antigens[31] (**Figure 6**). Another strong positive correlation was found between the expression of *ZNF286A* and *ZNF286B* genes and the expression of blood group antigens. This might be a consequence of the strong negative correlations with both the expression of *B3GNT3* and *GCNT3* genes, which are involved in the elongation and branching of *O*-glycan structures. Additionally, a positive correlation with the expression of *ST3GAL3,* encoding the α2-3-sialyltransferase, terminating the elongation of glycan structures

with Neu5Ac residues was found. Further studies need to be performed to confirm these hypotheses.

Although it has been demonstrated before that cell lines can be good representatives of the tumors based on gene mutations, gene expression, and protein expression[6–8], a large scale study comparing glycosylation of cell lines and tumor tissues is necessary to evaluate whether the cell lines can be used as glycobiological model systems. A small scale study, which compared *N*- and *O*-glycomes derived from CRC cell lines and tumors, revealed that some cell lines are not representing the *O*-glycosylation of the tumor tissues, showing lack of expression of sialyl- Tn and core 3 disialyl Lewis X hexasacharide[27], previously associated with malignant transformation[53]. Additionally, we found notable differences between the *O*-glycomes of cell lines cultured at different sites in different culturing conditions, therefore the impact of these variables also needs to be considered.

The presented mass spectrometric approach revealing the complete repertoire of glycans expressed by a specific cell type is a valuable resource for understanding glycosylation related changes in cancer. While the CRC cell lines have shown enormous diversity of individual glycan structures, their structural features could be associated with the cell differentiation. Moreover, glyco-phenotypes were found to be associated with the expression of specific glycosyltransferases involved in their biosynthesis, providing more insight into the regulation of glycan biosynthesis in different cell types. Finally, this untargeted in-depth screening of cell line glycome phenotypes will provide an important resource for future studies exploring the role of cell glycosylation in CRC progression and drug response leading to discovery of novel targets for the development of anti-cancer antibodies.

SUPPORTING INFORMATION

Supporting information is available free of charge via doi.org/10.1007/s00018-020-03504-z

DATA AVAILABILITY

The raw mass spectrometric data files that support the findings of this study are available in GlycoPOST in mzXML format, with the identifier GPST000035, accessible via the following link:

https://glycopost.glycosmos.org/preview/863090025d7bddcbd8bcf. The MS/MS spectra of glycan structures are available in UnicarbDR repository, accessible via the following link https://unicarb-dr.biomedicine.gu.se/references.

AUTHOR CONTRIBUTIONS

K.M. performed the experiments. K.M., S.H., T.Z., K.S. and M.W. conceptually designed the work. K.M., G.L., O.M. and M.W. wrote the manuscript. N.C. and C.J. assisted with data interpretation. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank G. van Pelt for support with cell culture, H. Duarte, and M. Pirro for reading the manuscript. We also thank C.A.M. Koeleman, A.L. Hipgrave Ederveen, and L. de Neef for technical support. This work was supported by the European Commission's Horizon 2020 programme ''GlyCoCan'' project, grant number 676421.

REFERENCES

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
2. De Sousa E Melo, F. *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat. Med.* **19**, 614–618 (2013).
3. Becht, E. *et al.* Immune and stromal classification of Colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clin. Cancer Res.* **22**, 4057–4066 (2016).
4. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
5. Linnekamp, J. F. *et al.* Consensus molecular subtypes of colorectal cancer are recapitulated in in vitro and in vivo models. *Cell Death Differ.* **25**, 616–633 (2018).
6. Mouradov, D. *et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res.* **74**, 3238–3247 (2014).
7. Sveen, A. *et al.* Colorectal cancer consensus molecular subtypes translated to preclinical models uncover potentially targetable cancer cell dependencies. *Clin. Cancer Res.* **24**, 794–806 (2018).
8. Berg, K. C. G. *et al.* Multi-omics of 34 colorectal cancer cell lines - a resource for biomedical studies. *Mol. Cancer* **16**, 116 (2017).
9. Holst, S., Wuhrer, M. & Rombouts, Y. *Glycosylation characteristics of colorectal cancer. Advances in Cancer Research* **126**, (Elsevier Inc., 2015).

10. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
11. Rodríguez, E., Schetters, S. T. T. & Van Kooyk, Y. The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nature Reviews Immunology* **18**, 204–211 (2018).
12. Holst, S. *et al.* N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homebox 1 (CDX1)/Villin mRNA Expression. *Mol. Cell. Proteomics* **15**, 124–140 (2016).
13. Holst, S. *et al.* N-Glycomic and Transcriptomic Changes Associated with CDX1 mRNA Expression in Colorectal Cancer Cell Lines. *Cells* **8**, 273 (2019).
14. Ju, T., Aryal, R. P., Kudelka, M. R., Wang, Y. & Cummings, R. D. The Cosmc connection to the Tn antigen in cancer. *Cancer Biomarkers* **14**, 63–81 (2014).
15. Mihalache, A. *et al.* Structural Characterization of Mucin O-Glycosylation May Provide Important Information to Help Prevent Colorectal Tumor Recurrence. *Front. Oncol.* **5**, 217 (2015).
16. Jensen, P. H., Karlsson, N. G., Kolarich, D. & Packer, N. H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat. Protoc.* **7**, 1299–1310 (2012).
17. Karlsson, N. G., Schulz, B. L. & Packer, N. H. Structural determination of neutral O-linked oligosaccharide alditols by negative ion LC-electrospray-MSn. *J. Am. Soc. Mass Spectrom.* **15**, 659–672 (2004).
18. Karlsson, N. G. *et al.* Negative ion graphitised carbon nano-liquid chromatography/mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid Commun. Mass Spectrom.* **18**, 2282–2292 (2004).
19. Rojas-Macias, M. A. *et al.* Towards a standardized bioinformatics infrastructure for N- and O-glycomics. *Nat. Commun.* **10**, 3275 (2019).
20. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
21. Hayes, C. A., Nemes, S. & Karlsson, N. G. Statistical analysis of glycosylation profiles to compare tissue type and inflammatory disease state. *Bioinformatics* **28**, 1669–1676 (2012).
22. Seymour, J. L., Costello, C. E. & Zaia, J. The Influence of Sialylation on Glycan Negative Ion Dissociation and Energetics. *J. Am. Soc. Mass Spectrom.* **17**, 844–854 (2006).
23. Kerbel, R. S., Pross, H. F. & Leibovitz, A. Analysis of established human carcinoma cell lines for lymphoreticular-associated membrane receptors. *Int. J. Cancer* **20**, 673–679 (1977).
24. Van Klinken, B. J. W. *et al.* The human intestinal cell lines Caco-2 and LS174T as models to study cel-type specific mucin expression. in *Glycoconjugate Journal* **13**, 757–768 (1996).
25. Semple, T. U., Quinn, L. A., Woods, L. K. & Moore, G. E. Tumor and lymphoid cell lines from a patient with carcinoma of the colon for a cytotoxicity model. *Cancer Res.* **38**, 1345–55 (1978).
26. McCool, D. J., Marcon, M. A., Forstner, J. F. & Forstner, G. G. The T84 human colonic adenocarcinoma cell line produces mucin in culture and releases it in response to various secretagogues. *Biochem. J.* **267**, 491–500 (1990).

27. Chik, J. H. L. *et al.* Comprehensive glycomics comparison between colon cancer cell cultures and tumours: Implications for biomarker studies. *J. Proteomics* **108**, 146–162 (2014).
28. Brockhausen, I., Yang, J., Lehotay, M., Ogata, S. & Itzkowitz, S. Pathways of mucin O-glycosylation in normal and malignant rat colonic epithelial cells reveal a mechanism for cancer-associated Sialyl-Tn antigen expression. *Biol. Chem.* **382**, 219–232 (2001).
29. Schneider, F. *et al.* Overexpression of sialyltransferase CMP-sialic acid:Galbeta1,3GalNAc-R alpha6-Sialyltransferase is related to poor patient survival in human colorectal carcinomas. *Cancer Res.* **61**, 4605–11 (2001).
30. Lo, C. Y. *et al.* Competition between core-2 GlcNAc-transferase and ST6GalNAc-transferase regulates the synthesis of the leukocyte selectin ligand on human P-selectin glycoprotein ligand-1. *J. Biol. Chem.* **288**, 13974–13987 (2013).
31. Brockhausen, I. Pathways of O-glycan biosynthesis in cancer cells. *Biochim. Biophys. Acta - Gen. Subj.* **1473**, 67–95 (1999).
32. Fernández, L. P. *et al.* The role of glycosyltransferase enzyme GCNT3 in colon and ovarian cancer prognosis and chemoresistance. *Sci. Rep.* **8**, 8485 (2018).
33. González-Vallinas, M. *et al.* Clinical relevance of the differential expression of the glycosyltransferase gene GCNT3 in colon cancer. *Eur. J. Cancer* **51**, 1–8 (2015).
34. Leibovitz, A. *et al.* Human Colorectal Adenocarcinoma Cell Lines. *Cancer Res.* **36**, 4562–4569 (1976).
35. Dotz, V. & Wuhrer, M. Histo-blood group glycans in the context of personalized medicine. *Biochim. Biophys. Acta - Gen. Subj.* **1860**, 1596–1607 (2016).
36. Fujitani, N. *et al.* Expression of H type 1 antigen of ABO histo-blood group in normal colon and aberrant expressions of H type 2 and H type 3/4 antigens in colon cancer. *Glycoconj. J.* **17**, 331–338 (2000).
37. Orntoft, T. F., Greenwell, P., Clausen, H. & Watkins, W. M. Regulation of the oncodevelopmental expression of type 1 chain ABH and Lewisb blood group antigens in human colon by α-2-L-fucosylation. *Gut* **32**, 287–293 (1991).
38. Dahiya, R., Itzkowitz, S. H., Byrd, J. C. & Kim, Y. S. ABH Blood Group Antigen Expression, Synthesis, and Degradation in Human Colonic Adenocarcinoma Cell Lines. *Cancer Res.* **49**, 4550–4556 (1989).
39. Clausen, H., Hakomori, S., Graem, N. & Dabelsteen, E. Incompatible A antigen expressed in tumors of blood group O individuals: immunochemical, immunohistologic, and enzymatic characterization. *J. Immunol.* **136**, 326–30 (1986).
40. Dahiya, R., Itzkowitz, S. H., Byrd, J. C. & Kim, Y. S. ABH blood group antigen expression, synthesis, and degradation in human colonic adenocarcinoma cell lines. *Cancer Res.* **49**, 4550–6 (1989).
41. Robbe, C. *et al.* Evidence of regio-specific glycosylation in human intestinal mucins: Presence of an acidic gradient along the intestinal tract. *J. Biol. Chem.* **278**, 46337–46348 (2003).
42. Brockhausen, I. Mucin-type O-glycans in human colon and breast cancer: Glycodynamics and functions. *EMBO Rep.* **7**, 599–604 (2006).

43. Malagolini, N., Dall'Olio, F. & Serafini-Cessi, F. UDP-Ga1NAc:NeuAcα2,3Galβ-R (GalNAc to Gal) β1,4-N-acetyl-galactosaminyltransferase responsible for the Sda specificity in human colon carcinoma CaCo-2 cell line. *Biochem. Biophys. Res. Commun.* **180**, 681–686 (1991).
44. Capon, C., Maes, E., Michalski, J. C., Leffler, H. & Kim, Y. S. Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochem J* **358**, 657–664 (2001).
45. Groux-Degroote, S. *et al.* B4GALNT2 gene expression controls the biosynthesis of Sda and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. *Int. J. Biochem. Cell Biol.* **53**, 442–449 (2014).
46. Robbe, C., Capon, C., Coddeville, B., & Michalski, J. C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J.* **384**, 307–316 (2004).
47. Iwai, T. *et al.* Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc. Natl. Acad. Sci.* **102**, 4572–4577 (2005).
48. Vavasseur, F., Yang, J. M., Dole, K., Paulsen, H. & Brockhausen, I. Synthesis of O-glycan core 3: Characterization of UDP-GlcNAc: GalNAc-R β3-N-acetyl-glucosaminyltransferase activity from colonic mucosal tissues and lack of the activity in human cancer cell lines. *Glycobiology* **5**, 351–357 (1995).
49. Múnera, J., Ceceña, G., Jedlicka, P., Wankell, M. & Oshima, R. G. Ets2 regulates colonic stem cells and sensitivity to tumorigenesis. *Stem Cells* **29**, 430–439 (2011).
50. Alitalo, K. *et al.* Aberrant expression of an amplified c-myb oncogene in two cell lines from a colon carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4534–4538 (1984).
51. Ciznadija, D. *et al.* Intestinal adenoma formation and MYC activation are regulated by cooperation between MYB and Wnt signaling. *Cell Death Differ.* **16**, 1530–1538 (2009).
52. Sakuma, K., Aoki, M. & Kannagi, R. Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF/bFGF-induced epithelial-mesenchymal transition. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7776–7781 (2012).
53. Robbe-Masselot, C. *et al.* Expression of a core 3 disialyl-Lex hexasaccharide in human colorectal cancers: A potential marker of malignant transformation in colon. *J. Proteome Res.* **8**, 702–711 (2009).

INTEGRATED GLYCOMIC AND PROTEOMIC SIGNATURES OF BUTYRATE-STIMULATED COLORECTAL CANCER CELL LINE DIFFERENTIATION

Chapter 5

K. Madunić[a], Y.M.C.A. Luijkx[b,c], O.A. Mayboroda[a], G.M.C. Janssen[a], P.A. van Veelen[a], K. Strijbis[c], T. Wennekes[b], G.S.M. Lageveen-Kammeijer[a], M. Wuhrer[a]

[a] *Center for Proteomics and Metabolomics, Leiden University, The Netherlands.*

[b] *Department Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands.*

[c] *Department Biomolecular Health Sciences, Utrecht University, Utrecht, The Netherlands*

Manuscript submitted

ABSTRACT

Gut microbiota of the gastrointestinal tract provide health benefits to the human host *via* bacterial metabolites. Bacterial butyrate has beneficial effects on intestinal homeostasis and is the preferred energy source of intestinal epithelial cells, capable of inducing differentiation. It was previously observed that changes in the expression of specific proteins as well as protein glycosylation occur with differentiation. In this study, specific mucin *O*-glycans were identified that mark butyrate-induced epithelial differentiation of the intestinal cell line CaCo-2, by applying porous graphitized carbon nano-liquid chromatography with electrospray ionization tandem mass spectrometry (PGC-nanoLC-MS/MS). Moreover, a quantitative proteomic approach was used to decipher changes in the cell proteome. It was found that the fully differentiated butyrate-stimulated cells are characterized by a higher expression of sialylated *O*-glycan structures, whereas fucosylation is downregulated with differentiation. By performing an integrative approach, we generated hypotheses about glycosylation signatures of specific cell adhesion proteins. These insights pave the way for future endeavors to study the dynamic *O*-glycosylation patterns in the gut, either produced *via* cellular biosynthesis or through the action of bacterial glycosidases as well as the functional role of these patterns in homeostasis and dysbiosis at the gut-microbiota interface.

INTRODUCTION

The human gut microbiota is a complex ecology of a variety of different microorganisms. Among viruses, prokaryotes, and eukaryotes, bacteria are the most abundant inhabitants of the human gastrointestinal (GI) tract. By co-evolution with the host, a symbiotic relationship has been formed between the GI tract and its bacteria[1,2]. The maintenance of homeostasis in the GI tract depends on the complex process of epithelial cell differentiation[3]. Bacterial metabolites also play a functional role in the maintenance of this homeostasis[4]. Butyrate, a short chain fatty acid, is a bacterial metabolite produced by bacterial fermentation of dietary fibers and is known to have a beneficial effect on the intestinal homeostasis. Physiological concentrations of butyrate inhibit cell proliferation, induce differentiation and increase the rate of apoptosis for a number of tumor cell types *in vivo* and *in vitro*[5–8]. Furthermore, butyrate is the main energy source of intestinal epithelial cells and capable of upregulating the gene expression of both secreted and membrane-linked mucins[9–13]. Notably, bacterial short chain fatty acids, such as butyrate, are suspected of also extending their effect outside of the gut as they have been implicated as possible mediators of phenomena observed in the gut-brain axis[14].

The mucus layer covers the epithelial cells to prevent them from being in direct contact with the microbiota. The function of the mucosal layer as a barrier is largely maintained by the gel-forming mucins, which are large, highly *O*-glycosylated proteins secreted by the epithelial cells[15]. Furthermore, epithelial cells express membrane-linked mucins that have both a barrier and a signaling function[16]. The intracellular tails of the membrane-linked mucins can be phosphorylated and mediate signaling, thereby regulating cell-cell interactions, differentiation, and apoptosis[16]. Additionally, the mucus layer and especially its *O*-glycans represent an important nutrient source for the surrounding microbiota and thereby contribute to bacterial colonization in the human gut. Typically, the majority of the mucin dry weight is made up of *O*-glycans and is the main modification that largely determines the properties and function of the mucins[17]. A wide variety of oligosaccharide structures can be attached to the mucins as well as other *O*-glycoproteins and the composition of these structures can vary within cell types due to differential expression of glycosidases and glycosyltransferases[18].

Furthermore, membrane-linked *O*-glycoproteins are seen as important oncogenic proteins as their intracellular domain and their glycosylated extracellular domain link to pathways involved in cell differentiation and apoptosis[19]. Notably, altered glycosylation can be correlated to the levels of the bacterial metabolite butyrate that regulates specific glycogenes including galectin-1 and β-galactoside-α-2-6-sialyltransferase (ST6GAL1)[20–23].

In this study the intestinal cell line, CaCo-2 (*Ca*ncer *co*li-*2*), derived from a human colorectal carcinoma (CRC) in 1977, is investigated and became a well-established model since then to study cellular differentiation. The cell line can differentiate spontaneously or by exposure to butyrate into polarized cells with morphological and biochemical features of the mature colonic epithelium[24]. Spontaneous differentiation of CaCo-2 cells has been studied previously, including transcriptomic, proteomic and glycomic analysis[25–32].The changes in *O*-glycosylation upon butyrate differentiation have yet to be characterized and insights into these changes is crucial for a better understanding of the role of glycans in gut homeostasis. The aim of the present study was to identify specific *O*-glycan features that define butyrate-induced epithelial differentiation in relation to spontaneous differentiation to gain insights into *O*-glycan signatures of colon cancer cell line differentiation derived from the gut epithelium.

MATERIALS AND METHODS

MATERIALS

Sodium borohydride ($NaBH_4$), sodium chloride (NaCl), Dowex cation-exchange resin (50W-X8), ammonium bicarbonate (ABC), trifluoroacetic acid (TFA) Dulbecco's Phosphate Buffered Saline (DPBS), Hydrochloric acid (HCl), DL-dithiothreitol (DTT), and were purchased from Sigma-Aldrich (Steinheim, Germany). Ethanol (Reag. Ph. Eur) and Mucin from bovine submaxillary glands, type I-S were purchased from Merck (BSM; Darmstadt, Germany). TMTpro Label Reagents, 8 M guanidine hydrochloride (GuHCl), Dulbecco's modified Eagle's medium (DMEM), 0.25% trypsin/EDTA, and fetal calf serum (FCS) were obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Potassium hydroxide (KOH) was obtained from Honeywell Fluka. Solid phase extraction (SPE) bulk sorbent Carbograph was obtained from Grace Discovery sciences (Columbia, USA). HPLC SupraGradient acetonitrile (MeCN) was obtained from Biosolve (Valkenswaard, The Netherlands). Peptide N-

glycosidase F (PNGase F) and complete EDTA free protease inhibitor cocktail tablets were purchased from Roche Diagnostics (Mannheim, Germany). 96-well PP filter plate was purchased from Orochem Technologies (Naperville, IL, USA). MultiScreen HTS 96 multiwell plates (hydrophobic Immobilon-P PVDF membrane) and 96-well PP Microplate were obtained from Millipore (Amsterdam, the Netherlands).

CELL CULTURE

Human colorectal adenocarcinoma Caco-2 cells were grown in Dulbecco's Modified Eagle medium (Gibco, Thermo Fisher Scientific) with 10 % FCS. Cells were subcultured at 80% confluency and maintained at 37 °C in a humidified incubator with 5 % CO_2. At day 0, when cells reached full confluency, 2 mM butyrate was added to the appropriate cells. On days 5, 7, 14, 21 and 24 cells were collected in biological triplicates. For harvesting of the cells, medium was removed and adherent cells were washed twice with DPBS and trypsinized using 0.25 % trypsin- 1 mM EDTA. To stop the trypsin activity, medium (without FCS) in a ratio of 2:5 (trypsin:medium; *v/v*) was added and cells were pelleted at 300 x g for 5 min. Cells were resuspended in DPBS and counted and aliquoted to ~2.0 x 10^6 cells per replicate and washed twice with 1 mL DPBS for 3 minutes at 100 x g. The supernatant was removed and cell pellets stored at -20 °C until further use.

PHASE-CONTRAST MICROSCOPY

Phase images were acquired using phase contrast microscopy with 40 times objective. Images were acquired through an CMEX 5000 Microscope camera.

ALKALINE PHOSPHATASE ACTIVITY

Alkaline phosphatase (ALP) activity was measured in cell lysates using a colorimetric assay according to the manufacturer's instructions (Abcam). The cells were harvested with 0.25 % trypsin- 1 mM EDTA, washed with PBS and subjected to ultra-sonification in the supplied assay buffer. The obtained cell homogenates were stored at -80 °C until assayed. Cells lysed at different growth days were incubated with 1.7 mM *p*-nitrophenylphosphate (*p*-NPP) for 60 min at 25 °C. Absorbance readings were taken at 405 nm with a microplate reader (Fluorimeter Fluorstar Omega, BMG Labtech). Results are expressed as milli units/mg protein. One unit is defined as the activity that hydrolyzes 1 micromole of *p*-NPP/min at 25 °C.

CELL LYSIS AND *O*-GLYCAN RELEASE

Three biological replicates from each time point were analyzed. Frozen cell pellets containing ~2.0 x 10^6 cells were resuspended in 100 µL of lysis buffer containing Tris HCl, EDTA, NaCl and protease inhibitor cocktail. The cells were lysed using a Branson sonicator rod at 1.5, and 25 µL of the suspension (containing ~ ~5 x 10^5 cells) were loaded onto the preconditioned PVDF membrane plate wells. BSM (10 µg) was blotted in three different wells of the same PVDF membrane plate. The denaturation as well as the *N*- and *O*-glycan release were performed as described previously[31,33]. Briefly, the proteins were denatured on membrane using guanidine hydrochloride and DTT at 60 °C. Upon removal of denaturing agents, the *N*-glycans were released by PNGase F digestion overnight at 37 °C, and recovered in MQ water. Upon recovery of *N*-glycans, the *O*-glycans were released from the same wells by reductive beta elimination, using sodium borohydride in potassium hydroxide incubating at 50 °C for 16 h. Sample desalting by cation exchange resin Dowex 50 W X8 and graphitized carbon solid phase extraction were performed in self-packed 96-well filter plates. The samples were dried after cleaning and stored at –20 °C until analysis.

PGC-LC-MS/MS ANALYSIS

The *O*-glycan samples were then reconstituted in 20 µL of MQ water, and 2 µL were injected for analysis. Analysis was performed using a PGC nano-LC Ultimate 3000 UHPLC system (Thermo Fisher Scientific, Sunnyvale, CA) coupled to an amaZon ETD speed ion trap (Bruker Daltonics, Bremen, Germany). The samples were loaded using 100 % buffer A (10 mM ammonium bicarbonate) at a loading flow of 6 µL/min on a custom-made trap column (size 30 × 0.32 mm) packed with 5 µm particle size PGC stationary phase from Hypercarb PGC analytical column (size 100 × 4.6 mm, 5 µm particle size, Thermo Fisher Scientific. Afterwards, the *O*-glycans were separated at a 0.6 µL/min flow rate on a custom-made PGC column (100 mm × 100 µm, 3 µm particle size obtained from Thermo Fisher Scientific) by applying a linear gradient from 1 % to 50 % buffer B (MeCN, 10 mM ammonium bicarbonate) over 73 min. During the procedures a constant column temperature of 45˚C was maintained. To continue, the LC system was coupled to an amaZon ETD speed ESI ion trap MS using the CaptiveSpray™ source (Bruker Daltonics) with an applied capillary voltage of 1000 V in negative-ionization mode. The flow rate was set to 3 L/min and the drying gas (N_2) temperature was set at 280 ˚C. The nebulizer gas pressure was kept at 3 psi. The

nanoBooster™ bottle (Bruker Daltonics) was filled with methanol, as a dopant solvent[34]. MS spectra were acquired in enhanced mode within a mass to charge ratio (*m/z*) range of 380-1850. The maximum acquisition time was set to 200 ms mass, the ion charge control (ICC) to 40,000, and the target of smart parameter setting was set to *m/z* 900. MS/MS spectra were generated by collision-induced dissociation on the three most abundant precursors, applying an isolation width of 3 Thomson. Additionally, ICC was set to 150,000 and the fragmentation cut-off was set to 27 % with a 100 % fragmentation amplitude using the Enhanced SmartFrag option (30 – 120 % in 32 ms). To integrate area under the curve (AUC) for each individual glycan isomer, extracted ion chromatograms of the first three isotopes were used in Compass DataAnalysis software (v.5.0). Peaks were manually picked and integrated. Total area normalization was employed for relative quantification of *O*-glycan species. Identification of *O*-glycan species was performed by comparison with PGC retention time, MS/MS spectra, and a BSM standard.

QUANTITATIVE PROTEOMICS USING TMT LABELLING

Cell lysis, digestion and TMT labeling was performed as described previously[35]. In short, ~1.0 x 10^6 Caco-2 cells were lysed in SDS lysis buffer (SDS [5 %], Tris-HCl [100 mM, pH 7.6]) at 95°C for 4 min. Protein concentration was determined by Pierce BCA protein assay (Thermo Fisher Scientific). Around 100 µg of protein was used for subsequent reduction with 5 mM TCEP, alkylation with 15 mM iodoacetamide and quenching with 10 mM DTT. Protein lysates were purified by methanol-chloroform precipitation. The resulting protein pellets were resuspended in 40 mM HEPES pH 8.4 and digested with trypsin (10 µg) for 16 hrs at 37 °C. Peptide concentration was measured with Pierce BCA assay.

The 10 different conditions together with common reference samples were arranged in triplicate in three TMTpro 13-plex experiments. Of each peptide preparations, 10 µg was dissolved in 25 µl of HEPES buffer (40 mM, pH 8.4) and incubated with 40 µg of one of 13 amino reactive TMTpro Label for 1h at ambient temperature. Excess TMT label was quenched by addition of 6 µL 5 % hydroxylamine and incubation for 15 min at ambient temperature. Labeled peptide samples were then mixed and freeze-dried. TMT-labeled mixtures were dissolved in 1 mL 10 mM ammonium bicarbonate and fractionated on 1cc C18 SPE cartridges (Oasis HLB, Waters) using 5 %, 10 %, 15 %, 20 %, 25 % and 35 % MeCN in 10 mM ammonium bicarbonate.

TMT-labeled peptides were dissolved in water/formic acid (100/0.1 *v/v*) and subsequently analyzed twice by on-line C18 nanoHPLC MS/MS with a system consisting of an Ultimate3000nano gradient HPLC system (Thermo, Bremen, Germany), and an Exploris480 mass spectrometer (Thermo). Fractions were injected onto a cartridge precolumn (300 µm × 5 mm, C18 PepMap, 5 um, 100 A, and eluted via a homemade analytical nano-HPLC column (50 cm × 75 µm; Reprosil-Pur C18-AQ 1.9 um, 120 A (Dr. Maisch, Ammerbuch, Germany). The gradient was run from 2 % to 40 % solvent B (20/80/0.1 water/acetonitrile/formic acid (FA); *v/v/v*) in 120 min. The nano-HPLC column was drawn to a tip of ~10 µm and acted as the electrospray needle of the MS source. The mass spectrometer was operated in data-dependent MS/MS mode for a cycle time of 3 s, with a HCD collision energy at 30 V and recording of the MS2 spectrum in the orbitrap, with a quadrupole isolation width of 1.2 Da. In the master scan (MS1) the resolution was 120,000, the scan range 400-1500, at standard AGC target at maximum fill time of 50 ms. A lock mass correction on the background ion *m/z* at 445.12 was used. Precursors were dynamically excluded after n = 1 with an exclusion duration of 45 s, and with a precursor range of 20 ppm. Charge states 2-5 were included. For MS2 the first mass was set to 120 Da, and the MS/MS scan resolution was 45,000 at an AGC target of 200 % at maximum fill time of 60 ms.

In a post-analysis process, raw data were first converted to peak lists using Proteome Discoverer version 2.2 (Thermo Electron), and submitted to the Uniprot database (Homo sapiens, 20596 entries), using Mascot v.2.2.07 (www.matrixscience.com) for protein identification. Mascot searches were with 10 ppm and 0.02 Da deviation for precursor and fragment mass, respectively, and trypsin as enzyme. Up to two missed cleavages were allowed. Methionine oxidation and acetyl on protein N-terminus were set as a variable modification; carbamidomethyl on Cys, TMTpro on N-terminus and Lys were set as a fixed modification. Protein FDR was set to 1 %. Normalization was performed on total peptide amount.

STATISTICAL ANALYSIS

Data analysis and visualization was performed using in-house developed “R” scripts. To enable use of principal component analysis, imputation of minimum positive number (0.0001) was performed. Differences between groups were tested using two-way ANOVA analysis of variance both for glycomics and proteomics datasets.

RESULTS

CELL DIFFERENTIATION ASSAY

O-glycosylation of CaCo-2 cells was analyzed at different time points during the process of spontaneous and butyrate-induced differentiation. Cells were proliferated to confluency, which is marked as day 0, after which cells start their differentiation. Differentiation of CaCo-2 cells was evaluated by a colorimetric assay measuring alkaline phosphatase (AP) activity, which is a marker for late stage differentiation, indicative for the presence of an established brush border[36]. AP activity levels of both spontaneous differentiation and induced differentiation showed a continuing increase

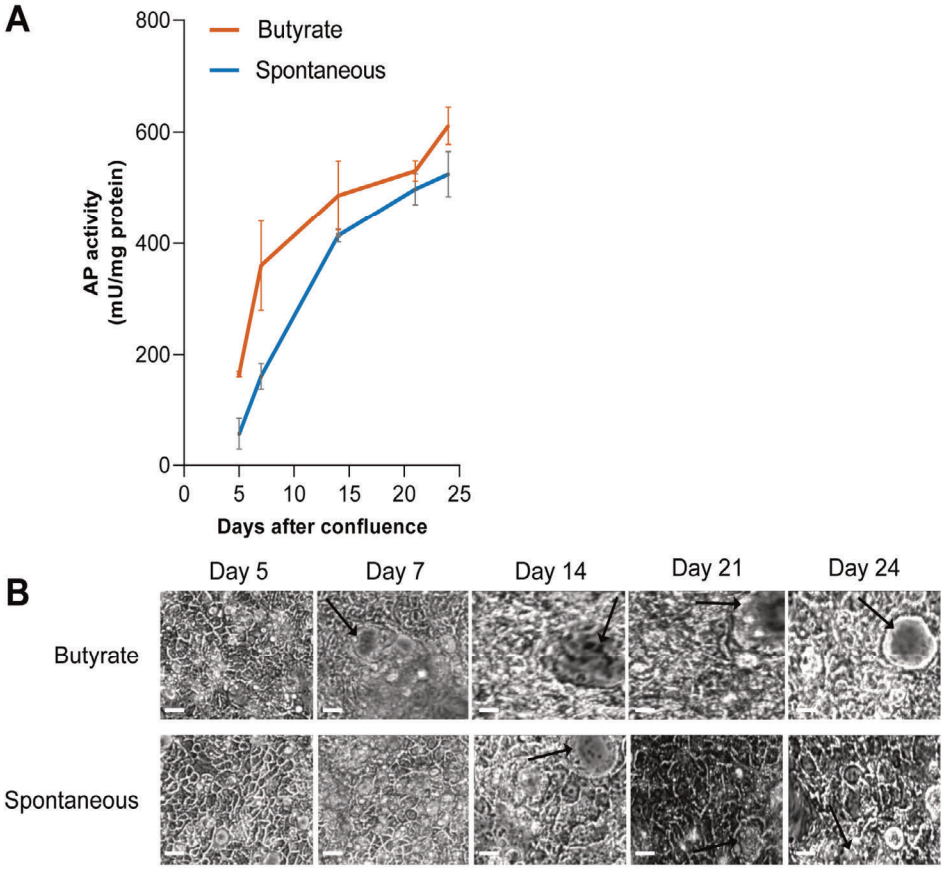


Figure 1. Biochemical and morphological validation of CaCo-2 differentiation. a) *Alkaline phosphatase (AP) activity during CaCo-2 growth. Stimulated by 2 mM butyrate (orange line) or spontaneous differentiation (blue line). Standard deviation is indicated by error bars (n = 3).* ***b)*** *Phase contrast images of CaCo-2 cells on culture dishes at different stages of growth. Arrow; dome-like structure. Original magnification; 40x. White scale bar = 50 µm.*

(**Figure 1a**). A significantly higher level of AP activity for the induced differentiation was observed for days 5 and 7, compared to the spontaneous differentiation.

In addition to AP activity levels as a validation for differentiation, phase images were taken of the apical surface to observe differentiation induced "dome" formation (**Figure 1b**). CaCo-2 cells imaged at different growth times started displaying dome formation at day 7 for the butyrate-treated cells, whereas the spontaneously differentiated cells showed dome formation much later (day 14). After having established that CaCo-2 cells were differentiated, our next aim was to assess the effect of butyrate on the *O*-glycomic profile of these CaCo-2 cells.

GLYCOMIC ANALYSIS

The analysis of *O*-glycans can be challenging in comparison to *N*-glycans, due to the lack of specific enzymes that can release the intact *O*-glycans from glycosylated proteins. In this study, we used a previously established protocol that allowed the chemical release of *O*-glycans from cell lysates via reductive β-elimination in 96-well plate format[31]. *O*-glycans of the differentiating CaCo-2 cells were analyzed at days 5, 7, 14, 21 and 24 post confluency. To further validate the robustness of this approach, full technical triplicates of the cell lysate (5 x 10^5 cells) from the same biological replicate (5 days, spontaneous differentiation) were processed and analyzed independently by porous graphitized carbon nano-liquid chromatography coupled to mass spectrometry (PGC-nanoLC–ESI-MS/MS). The mean relative area of the 11 most abundant glycan species and corresponding standard deviations are shown in **Supplementary Information 1, Figure S1a**, demonstrating the low technical variability of the workflow. Parallel to this, the PGC-nanoLC–ESI-MS performance was assessed by releasing *O*-glycans from 10 µg of bovine submaxillary mucin (BSM) standard and measuring them across five days, as illustrated in **Supplementary Information, Figure S1b**. Overall, the method showed very good precision.

To explore the *O*-glycomic profiles a principal component analysis (PCA) was performed on relative abundances of individual glycans detected in the glycomic profiles (**Supplementary Information 2, Table S1).** As illustrated in **Supplementary Information 1**, **Figure S2a**, the model showed narrow clustering of most biological replicates, although substantial differences were observed within replicates for time points 24 days (butyrate) and 5 days (spontaneous). Different time points of

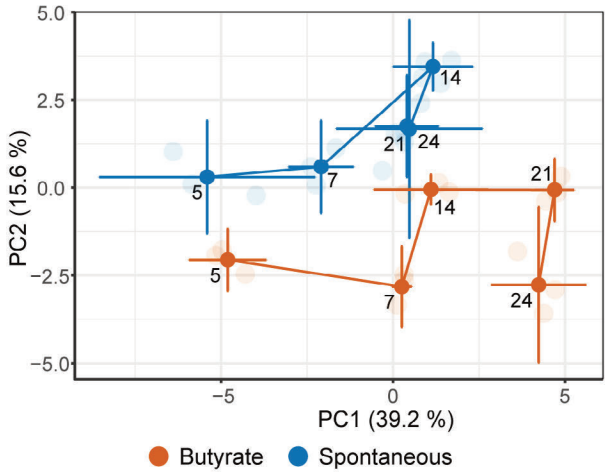


Figure 2. The geometric trajectory visualization of glycomic changes with differentiation. *A distinction between the butyrate-stimulated group (orange) and spontaneous differentiation (blue) is observed in the PCA model based on relative abundance (%) of different O-glycans. The separation between different time points in the two groups is illustrated as a trajectory. The top two principal components (PC) explain 54.8% of the variation within the data. The PCA scores from different biological replicates (faded color) were averaged to create the trajectory.*

differentiation were separated along the PC1 (39.2 %). A clear separation was apparent between the butyrate-stimulated samples and spontaneously differentiated samples along PC2 (15.6 %). A relative higher abundance of fucose containing *O*-glycans was found for the samples clustering in the upper left corner of the loadings plot, whereas sialylated *O*-glycans were enriched in the lower right part of the plot (**Supplementary Information, Figure S2b**). Further analysis of the scores plot (**Figure 2**) showed that for days 5 and 7 both groups are different from one another as well as from days 14, 21 and 24, as no overlap is observed of the representations. Additionally, when focused on the difference overtime within the butyrate and non-stimulated group (**Figure 2**), the scores plot indicates that the CaCo-2 *O*-glycome undergoes progressive changes when cells are grown from confluency to 7 days post-confluence, at which point these cells are not fully differentiated yet. Upon further differentiation, a distinct difference between the butyrate-stimulated group trajectory and the spontaneous differentiation trajectory can be observed. For the spontaneous differentiated CaCo-2 cells the representations of days 14, 21 and 24 are close to one another. This indicates that the *O*-glycome appears to stabilize in the later differentiation phases. However, the butyrate-stimulated cells are different from each

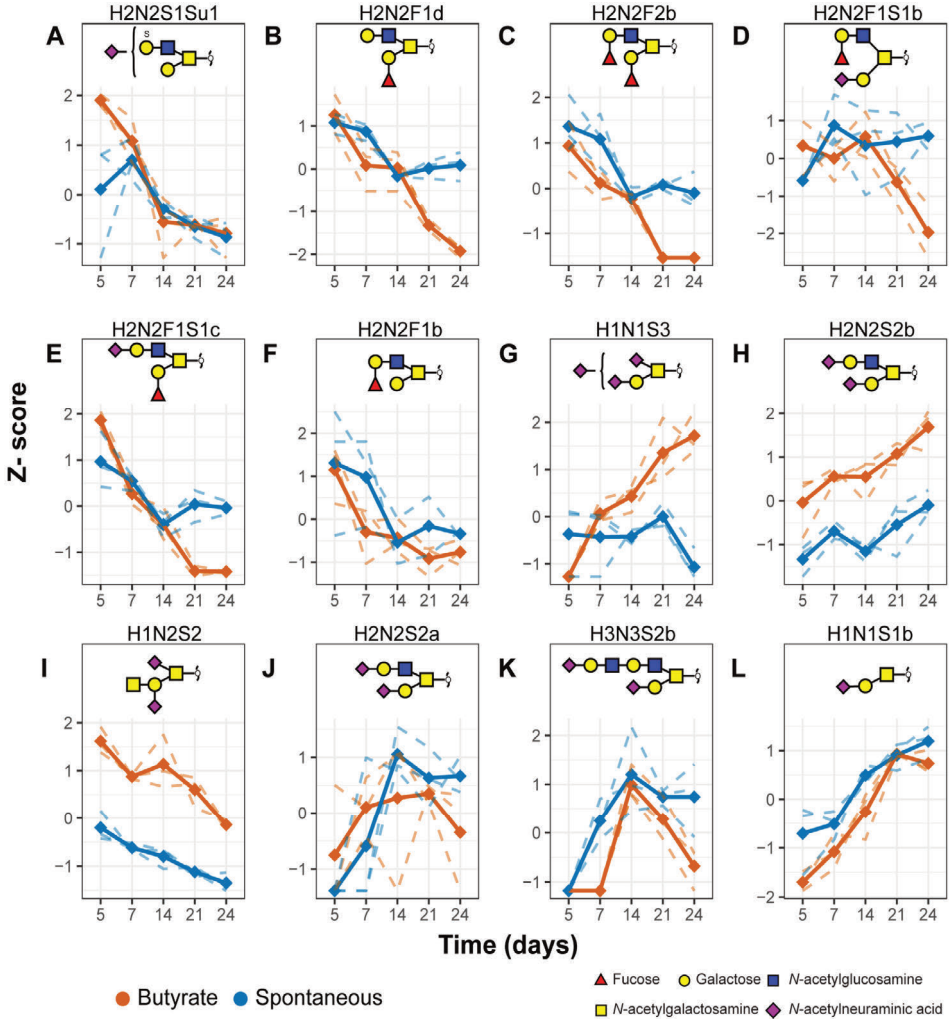


Figure 3. Differentiation induces significant downregulation (a-f; i) or upregulation (g, h, j, k. l) of CaCo-2 O-glycans as well as some butyrate-specific changes. *Spaghetti plots of the selected glycan abundances that show significant difference between time points selected from the analysis of variance (ANOVA). The dashed lines represent the scaled z-scores of the measured values of each biological replicate, whilst the continuous lines represent the z-scores of the mean values per biological replicate. Su: sulfation.*

other between days 14, 21 and 24, which indicates that the butyrate-stimulated cells still undergo progressive changes in their *O*-glycome in the late differentiation phases. To support this visual interpretation and identify the glycosylation signatures changing with differentiation, two-way ANOVA analysis of variance was performed on relative abundances of individual *O*-glycans (**Supplementary Information 2, Table S2).**

Specific *O*-glycans that show statistically significant changes with time (butyrate-stimulated *versus* spontaneous differentiation) are summarized in **Supplementary Information 1, Figure S3** and **Supplementary Information 2, Table S3.** The *O*-glycans that show a difference between the groups (butyrate-stimulated and spontaneous differentiation) are listed in **Supplementary Information 2, Table S4**. Moreover, the glycans that show a significant change with both butyrate stimulation and time are listed in **Supplementary Information 2, Table S5.**

O-glycans that show a clear consecutive pattern with differentiation were selected for visualization (**Figure 3**). **Figure 3a-f** show the *O*-glycans that are significantly downregulated with differentiation and are carriers of terminal blood group antigens (H2N2F1b, H2N2F1d, H2N2F2b, H2N2F1S1b, H2N2F1S1c), as well as sulfation (H2N2S1Su1). All *O*-glycans that show an increase overtime carry terminal *N*-acetylneuraminic acids and no fucosylation (H2N2S2b, H2N2S2a, H3N3S2b, H1N1S1a, H1N2S2, H1N1S3). A glycan with a disialyl motif, with composition H1N1S3 (**Figure 3g**) is upregulated in the butyrate stimulated cells. Notably, the Cad antigen (GalNAcβ1-4(Neu5Acα2-3)Galβ1-3[Neu5Acα2-6]GalNAc) present on *O*-glycans with composition H1N2S2 (**Figure 3i**) shows downregulation with differentiation. However, it is significantly higher in the butyrate-stimulated group compared to the spontaneous differentiation. Overall, we observed a downregulation of blood group antigen H fucosylation, and upregulation of terminal sialylation with differentiation.

STRUCTURAL IDENTIFICATION OF GLYCAN SPECIES

Due to reports from the literature that provide contradictory *O*-glycan assignments, we made additional efforts to support the assignment of the specific *O*-glycan structures. As the *O*-glycomic profile of BSM has been very well characterized by both NMR and tandem MS, we used this as a reference standard[37–39]. At *m/z* 895.34, two isomers with composition H2N2F1 were observed in our Caco-2 samples (**Supplementary Information 1, Figure S4a;** isomer *1* and *2a*). The major isomer expressed in Caco-2 cell line (2a), matched with the major isomer in BSM standard (**Supplementary Information 1, Figure S4b;** *2b*) by retention time and MS/MS spectrum which allowed us to confirm that the major *O*-glycan isomer (*2a*) in the Caco-2 samples carried the terminal blood group H antigen[38]. The MS/MS spectra of this isomer (**Supplementary Information 1, Figure S4d** and **e**) shows the presence of a highly abundant Z-ion (*m/z* 569.20 (H1N2) indicating the occupancy of the 6'-antenna of the core 2 *O*-glycan.

On the other hand, isomer *1* was found solely in Caco-2 cells, and the MS/MS spectra (**Supplementary Information 1, Figure S4c)** revealed the presence of terminal fucose linked to a type 2 *N*-acetyllactosamine on the 6' antenna, indicated by the characteristic cross-ring fragment of the β1-4-galactosylated GlcNAc (*m/z* 409.02). Previous studies have reported an α1-2-linked terminal L-fucose, as part of blood group antigen H[40]. In contrast, a recent study has reported the presence of an unconventional terminal α1-6-linked fucose[30]. Since this glycan was not present in the BSM standard, it was not possible to determine the linkage of the terminal fucose residue and further experiments will be needed. Additionally, two isomers with composition H2N2F2 at *m/z* 1041.40 were observed in our Caco-2 samples (**Supplementary Information 1, Figure S5a,** isomer *1* and *2a*). Isomer *2a* matched by retention time and MS/MS spectra with the major isomer of the BSM standard (**Supplementary Information 1, Figure S5b** and **d;** isomer *2b*), indicating a blood group antigen H related terminal α1-2-fucose[30]. Isomer *1* could not be identified by MS/MS and did not provide insightful differences in the fragmentation patterns due to low abundance. Interestingly, a recent study reported the presence of two α1-6- linked terminal fucosides carried by a major isomer with composition H2N2F2 in Caco-2[30]. The structural identification of an *O*-glycan with composition H1N1F1S1 at *m/z* 821.30 is illustrated in **Supplementary Information 1, Figure S6.** Based upon fragments *m/z* 530.23 and 512.14 this glycan carries a blood group antigen type 3. The signal at *m/z* 495.25 indicated an α2-6-linked sialic acid on the innermost GalNAc, which is in contrast to a previously reported annotation where the fucose was stated to be linked to the innermost GalNAc, whereas the sialic acid was linked to terminal galactose in α2-6 linkage[30]. Structural elucidation of the *O*-glycan with composition H1N1S3 at *m/z* 1257.44 is shown in **Supplementary Information 1, Figure S7**. Two sialic acids appeared to be linked to each other, indicated by the presence of a characteristic fragment ion at *m/z* 581.18. While our data did not provide sufficient evidence to deduce the location of this disialyl motif, it has been previously described that this occurs on the α2-6-linked sialic acid linked to the innermost GalNAc[41]. Moreover, structural elucidation of a glycan with Cad antigen, upregulated in the butyrate stimulated cells, was depicted in **Supplementary Information 1**, **Figure S8**. Overall, we observed a downregulation of terminal blood group H antigen expression and an upregulation of both galactose α2-3-sialylation and core GalNAc α2-6-sialylation with differentiation.

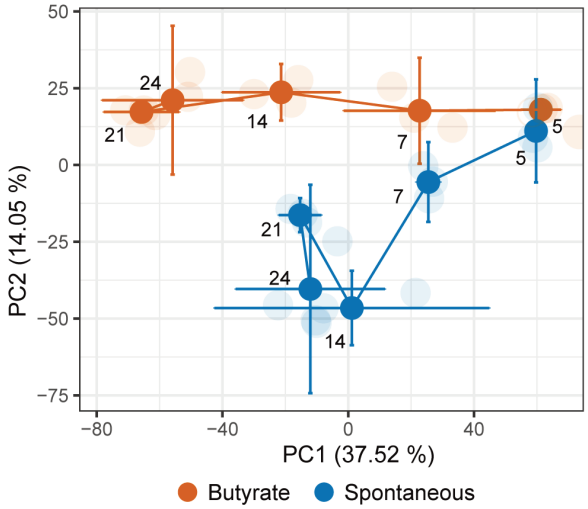


Figure 4. The geometric trajectory visualization of proteomic changes with differentiation. *PCA model based on abundance of different proteins shows a separation between the butyrate-stimulated group (orange) and spontaneous differentiation (blue). The separation between different time points in the two groups is illustrated as a trajectory. The top two principal components (PC) explain 51.57 % of the variation within the data. The PCA scores from different biological replicates (faded color) were averaged to create the trajectory.*

Proteomic analysis

Butyrate has been extensively used for stimulation of differentiation. However, to our knowledge a quantitative proteomics approach has not been employed to study the effect as well as differences between spontaneous differentiation and butyrate-stimulated differentiation in the Caco-2 cell line. Therefore, we performed quantitative bottom-up proteomics on the very same set of samples analyzed for the profiling of the *O*-glycome. The samples were digested with trypsin, isotopically labeled with a tandem mass tag (TMT), and analyzed by LC-MS/MS. With this approach a total of 5050 proteins could be confidently quantified (mascot score >21, unique peptides >= 1; **Supplementary Information 3, Table S2**).

To explore specific proteome variations that correlate to the differentiation of butyrate-stimulated samples, principal component analysis (PCA) was used (**Supplementary Information 1, Figure S9**). The geometric trajectory in **Figure 4** (based on the PCA model) illustrates changes in the cell proteome with differentiation as well as the differences in butyrate stimulated cells. At day 5, both groups cluster closely together, whereas the changes start to be more prominent at day 7 culminating at day 14. In the

non-stimulated cells the proteome appears to stabilize after day 14. In contrast, butyrate stimulation induced significant changes in the proteome until day 21. Proteins that show statistically significant changes (Two-way ANOVA, Bonferroni corrected p-value < 0.05) with time, with butyrate stimulation, as well as with both time and with butyrate stimulation are listed in **Supplementary Information 3, Tables S4, S5** and **S6**, respectively. Interestingly, significant upregulation of alkaline phosphatases (ALPP, and ALPG) and cytokeratin 20 (KRT20) are present only in the butyrate stimulated cells (**Supplementary Information 1, Figure S10**), Specific transcription factors known for their role in regulation of colon differentiation also showed a change (**Supplementary Information 1, Figure S11)**. Hepatocyte nuclear factor 1α (HNF1A) decreased in abundance over time in both groups, however, in the butyrate stimulated cells it continued to decline after 14 days. Hepatocyte nuclear factor 4α (HNF4A) showed an upregulation after 5 days, with a decline after 7 days in spontaneous differentiation, whereas in the butyrate stimulated cells in continued to rise until 14 days, followed by a decline. Transcription factors GATA6 and FOXA1 showed a continuous decrease in abundance both in spontaneous and butyrate stimulated cells. Although not statistically significant, transcription factor CDX2 decreased over time in both groups. All transcription factors which show a change with differentiation are listed in **Supplementary Information 3, Table S9.** Moreover, proteins that show the most prominent upregulation with butyrate stimulation are protocadherin Fat 1 (FAT1), galectin-1 (LGALS1), and phosphoenolpyruvate carboxykinase, (PCK1) (**Supplementary Information 3, Table S5**). The majority of the proteins changing specifically with butyrate stimulation are related to cell metabolism (hsa01100-Kegg pathways).

In order to relate the *O*-glycomic changes to changes in the abundance of specific *O*-glycoproteins, we continued our analysis of the proteomic data by selecting the proteins involved in the GO ontologies-Glycosylation, GO-Monosaccharide transport, as well as *O*-glycoproteins from different cell lines discovered by the Simple cell approach[42]. From this selection, 120 proteins showed a statistically significant change with time, whereas only 36 proteins showed a significant change with both time and butyrate stimulation (**Supplementary Information 3, Table S7** and **S8**, respectively). Interestingly, the only mucin that showed an upregulation over time was MUC13 (**Supplementary Information 3, Table S7)**. Additionally, changes were detected in

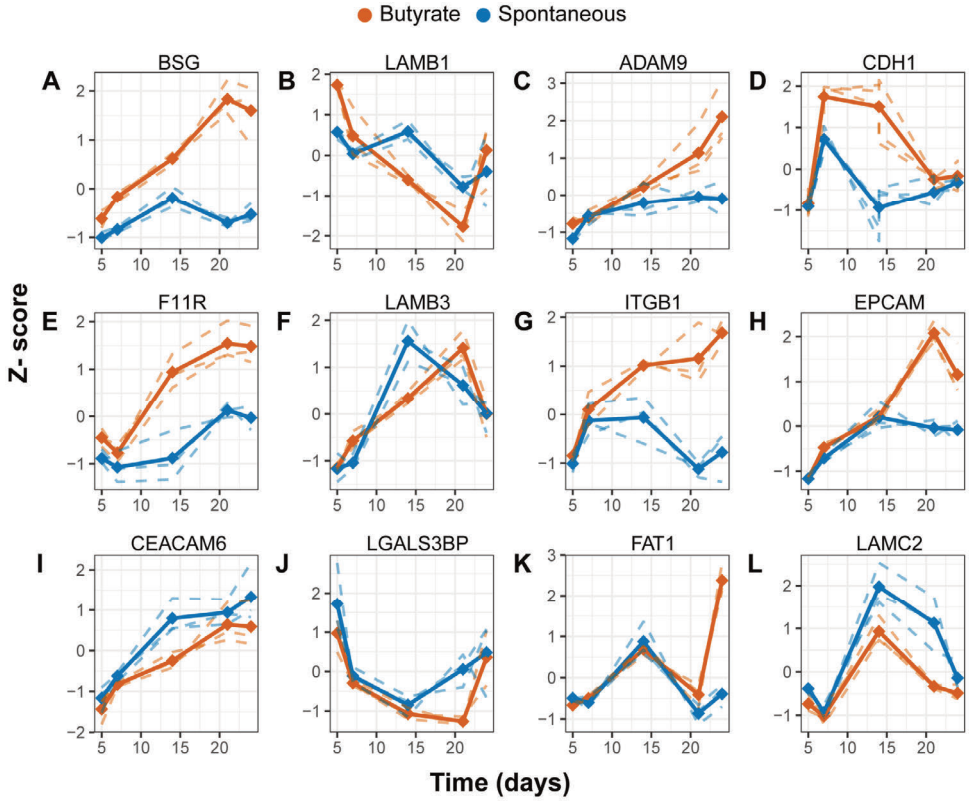


Figure 5. Differentiation induces significant changes in the abundance of cell adhesion O-glycoproteins. *Spaghetti plots of the protein abundances that show significant difference between groups selected from the analysis of variance (ANOVA). The dashed lines represent the scaled Z-scores of the measured values of each biological replicate, whilst the continuous lines represent the Z-scores of the mean values per biological replicate.*

monosaccharide transport related proteins including solute carrier family facilitated monosaccharide transporters (SLC2A1, SLC2A3, SLC2A5) as shown in **Supplementary Information 1, Figure S12**. Also, the abundances of polypeptide GalNAc transferases 2 and 3 (GALNT2, GALNT3 respectively) were increased over time (**Supplementary Information 1, Figure S12**). However, other glycosyltransferases involved in the biosynthesis of *O*-glycans did not show a significant change or were not detected with our proteomics workflow. Interestingly, the abundance of GDP-mannose 4,6-dehydratase (GMDS) was downregulated with differentiation, however, the change was not statistically significant.

In regard to time, significant changes were observed for the abundances of different *O*-glycoproteins involved in cell adhesion such as laminin (LAMB3, LAMC2, LAMB1), galectin 3 binding protein (LGALS3BP), cadherin 1 (CDH1), epithelial cell adhesion molecule (EPCAM), intracellular adhesion molecule1 (ICAM1), carcinoembryonic antigen related cell adhesion molecule 6 (CEACAM6), integrin beta 1 (ITGB1), disintegrin and metalloproteinase domain containing protein 9 (ADAM9), protocadherin (FAT1) and junctional adhesion molecule A (F11R) (**Figure 5**). Interestingly, butyrate-stimulated cells induce higher abundances of cell adhesion proteins such as protocadherin (FAT 1), cadherin 1 (CDH1) and EPCAM than spontaneously differentiated cells (**Figure 5** and **Supplementary Information 3, Table S8**).

DATA INTEGRATION

To identify associations between the proteomic and glycomic changes with differentiation we used a data integration approach (MixOmics)[43]. For this purpose, only the above-mentioned proteins of interest were included in the analysis. Due to the complexity of the model, the proteome and *O*-glycome differentiation signatures over time were integrated separately for the spontaneous and butyrate-stimulated cells. The clustered image map (CIM) indicates that changes on the glycome level highly associates with the changes on the proteome level in spontaneous differentiation, revealing potential glycosylation changes of specific proteins or relationship with specific enzymes and monosaccharide transporters (**Figure 6**). Among others, cathepsin D (CTSD) was significantly downregulated with differentiation, which associated with fucosylated glycomic signatures H2N2F1d, H2N2F2b and H2N2F1S1c. Interestingly cell adhesion glycoprotein cadherin 1 (CDH1) showed strong upregulation at day 7 together with *O*-glycan signatures such as H3N3, and sialylated glycans H3N3S2a, H1N2S1, H2N2F1S1b, H2N2S2b. At day 14 prominent upregulation was seen in the abundance of other cell adhesion proteins such as EPCAM, ICAM1, CEACAM6, which showed a similar expression pattern as sialylated glycans H1N1S1b and H2N2S2a. Moreover, an upregulation of SLC2A1, a monosaccharide transporter of a wide range of aldoses is seen at this stage of differentiation, as well as glutamine-fructose-6-phosphate aminotransferase 1 (GFPT1) and phosphoacetylglucosamine mutase (PGM3) which catalyzes the conversion of GlcNAc-6-P into GlcNAc-1-P during the synthesis of uridine UDP-

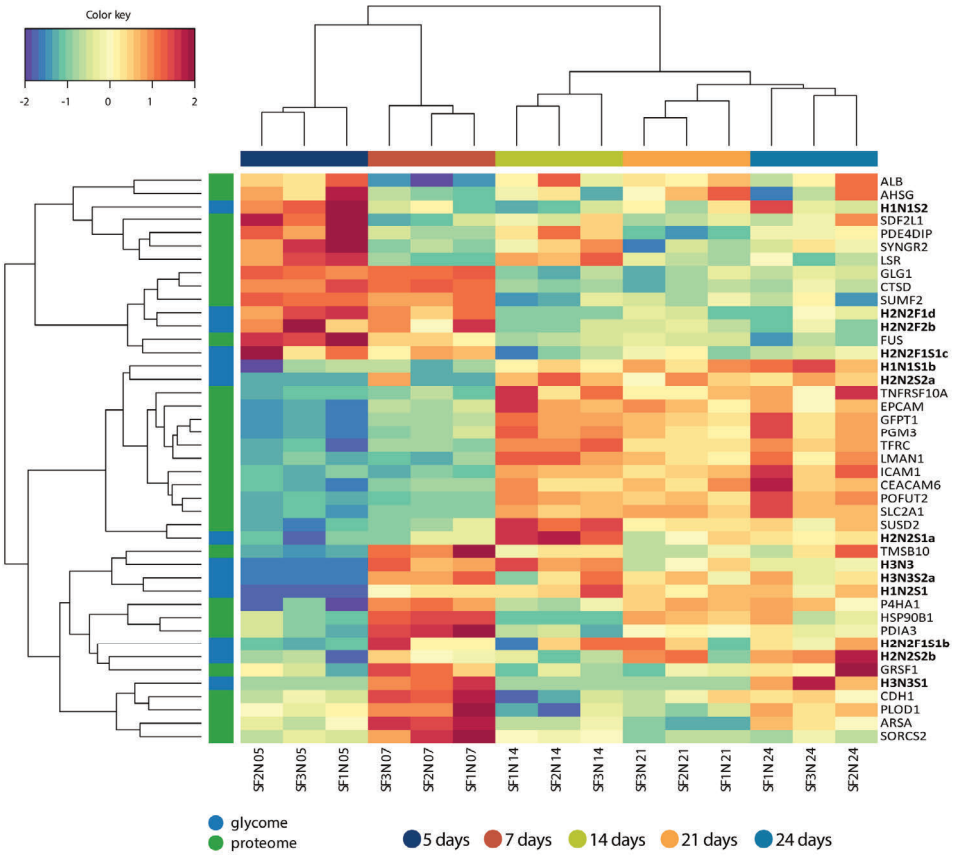


Figure 6. Multi-omics signature of spontaneous differentiation. *Clustered Image Map illustrating changes in both proteome and glycome with differentiation based on a combination of the PLS (Partial Least Square Regression) generalized for the multiple matched datasets and LASSO (least absolute shrinkage and selection operator) based variable selection. The model was tuned for the maximal correlation.*

GlcNAc. The changes in the later time points are very similar as for day 14 due to the stabilization that was observed in the proteome and glycome of spontaneously differentiated cells. On the other hand, butyrate stimulated differentiation induces significant changes on the proteome level even until day 21, and less prominent changes until day 24 (**Figure 7**). Downregulation of, among others, apolipoproteins E and A1, alpha-2-HS-glycoprotein (AHSG), albumin (ALB), lipolysis-stimulated lipoprotein receptor (LSR), sodium/ascorbate transporter (SLC23A1), chitobiosyldiphosphodolichol beta-mannosyltransferase (ALG1), beta-hexosaminidase is observed in spontaneously differentiated cells. On the other hand,

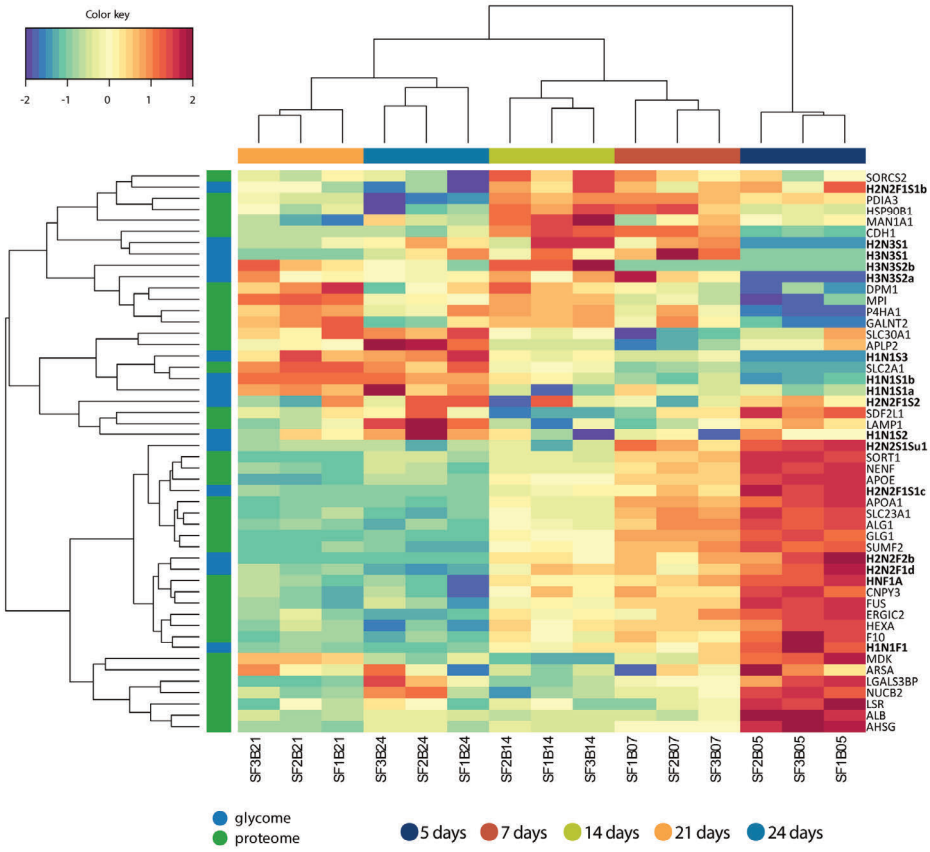


Figure 7. Multi-omics signature of butyrate stimulated differentiation. *Clustered Image Map illustrating changes in both proteome and glycome with differentiation based on a combination of the PLS (Partial Least Square Regression) generalized for the multiple matched datasets and LASSO (least absolute shrinkage and selection operator) based variable selection. The model was tuned for the maximal correlation.*

butyrate stimulated differentiation induces subunit alpha (EXA) and hepatocyte nuclear factor 1-alpha (HNF1A) is observed, which are correlated with fucosylated *O*-glycomic signatures such as H2N2F1S1c, H2N2F2b, H2N2F1d and H1N1F1. Furthermore, it also induces downregulation of galectin-3-binding protein (LGALS3BP) which promotes integrin-mediated cell adhesion. Remarkably, at the same time, these cells showed upregulation of another cell adhesion protein cadherin 1 (CDH1), which is associated with the upregulation of glycans H2N2F1S1b, H2N3S1, H3N3S1, H3N3S2b, H3N3S2a. Moreover, butyrate stimulation induced upregulation of proteins involved in the monosaccharide metabolism, such as mannose-6-phosphate isomerase (MPI) involved in the synthesis of the GDP-mannose as well as solute

carrier family 2, facilitated glucose transporter member 1 (SLC2A1). Upregulation of proteins involved in *N*-glycan biosynthesis was also induced by butyrate stimulation such as mannosyl-oligosaccharide 1,2-alpha-mannosidase IA (MAN1A1) and dolichol-phosphate mannosyltransferase subunit 1 (DPM1). Upregulation of polypeptide GalNAc transferase 2 (GALNT2) shows a similar pattern to the upregulation of the following sialylated glycans H1N1S3, H1N1S1b, H1N1S1a, H2N2F1S2 and H1N1S2 all showing increasing abundances with cell differentiation.

The associations found with HNF1A triggered us to explore if other transcription factors associate with the *O*-glycomic changes. Therefore, we performed the integrative analysis selecting only the *O*-glycans and the transcription factors that show a significant change with differentiation (**Supplementary Information 3, Table S9**). Downregulation of fucosylation in spontaneous differentiation showed associations with FOXA1, FOXP4, GATA6, STAT1 and HNF1A transcription factors. Specifically, HNF4A expression was associated more with the expression of H2N2S1Su1 and H3N3S1 glycomic signatures (**Supplementary Information 1, Figure S13)**. Similar to HNF1A, that showed associations with glycomic signatures carrying terminal blood group H (H2N2F2b, H2N2F1S1c) (**Figure 7**), transcription factor HNF4A associated with H2N2F1S1b in the butyrate stimulated cells (**Supplementary Information 1, Figure S14)**. On the other hand, upregulation of sialylated species associated with NF-Kappa-B Transcription Factor P65 (RELA) and MYC promoter-binding protein 1 (ENO1) in butyrate stimulated cells.

DISCUSSION

In this study we observed changes in the Caco-2 *O*-glycome correlated to differentiation over time. Both butyrate-stimulated and spontaneous differentiation were accompanied by the elevation of galactose α2-3-sialylation, and a downregulation of the terminal blood group H antigen over time. However, it should be noted that there are contradictory reports in literature regarding the correlation of the terminal blood group H-antigen expression with differentiation. Namely, Murakami *et al.* reported no detection of the H type blood group antigen, while Vincente *et al.* reported diffused expression of H type antigen, Amano *et al.* reported higher H type 1 antigen expression with differentiation and Xu *et al.* reported no changes in the *O*-linked fucosylation[30,44–46]. It is important to note that there were considerable

differences in the experimental design regarding induction of differentiation, days in culture as well as detection methods. Moreover, it was previously postulated that Caco-2 cells can form different subpopulations during growth and differentiation, therefore considerable differences between laboratories have previously been reported for identical experimental conditions[47,48]. Therefore, we first validated that the cells showed changes in the morphology related to cell differentiation (**Figure 1b**), as well as changes in the activity of alkaline phosphatase (**Figure 1a**), which is a well-established marker of epithelial differentiation. Thereafter we aimed at deciphering differences between spontaneous and butyrate stimulated changes with cell differentiation focusing on the *O*-glycome and proteome by MS.

Our study on the relative abundances of individual *O*-glycans showed that the Caco-2 *O*-glycome for spontaneous differentiation stabilizes after day 14, this is in agreement with a proteomics analysis on the spontaneous differentiation of Caco-2 cells previously published by Stierum *et al.*[25]. The *O*-glycome for the butyrate-stimulated group behaved significantly differently and did not stabilize after day 14. Noticeable, in the spontaneously differentiated cells the expression of terminal blood group antigens carrying an α-1,2-fucoside stabilized at day 14, whereas most of these α-1,2-fucoside-carrying antigens in the butyrate-stimulated cells continued decreasing even after day 14. Fucoside moieties on glycans are increasingly recognized as critical attributes for cell-cell interactions and signaling processes[49]. The decrease of type 2 blood group H fucosylated glycans with a subsequent increase in type 1 blood group H and Lewis B (Le^B) glycans was previously correlated to the spontaneous differentiation of Caco-2 cells[46,50]. This effect was linked to the changes in the activity of specific enzymes responsible for type 1 blood group H biosynthesis such as β1-3-galactosyltransferase and α1-2-fucosyltransferase[51]. We did not find the blood group H type 1 antigens on *O*-glycans, which might indicate that other types of glycans are carriers of the antigen such as glycosphingolipids or *N*-glycans. For further studies it would be interesting to investigate the implications of this butyrate-induced fucose downregulation on cell properties. At the same time, upregulation of sialylated glycan species is induced upon differentiation, which was previously correlated to differentiation assessing the expression of corresponding glycosyltransferases such as ST3GAL6[52–54]. Previously, we compared the *O*-glycome of 26 CRC cell lines and revealed associations with differentiation where the well-differentiated cells expressed

Lewis-type fucosylation, whereas the undifferentiated cells, such as the Caco-2 cell line, showed expression of *O*-linked blood group H antigens[31]. However, it should be taken into account that, although developed from colon carcinoma, Caco-2 cells showed quite specific glycosylation, very rich in terminal blood group antigen expression compared to other CRC cell lines[31]. Moreover, they differentiate into small intestinal enterocyte-like cells, with a consequent loss of colonocyte properties[55]. This could explain why there was not an increase in Lewis type fucosylation observed in this study upon differentiation.

The *O*-glycan carrying a Cad antigen showed a downregulation over time for both groups, correlating to differentiation, but a significantly higher expression was observed for the butyrate-stimulated cells compared to the spontaneously differentiated cells. Structures with terminal Cad and Sda epitopes have previously been described as characteristic for normal colon mucin tissue[56]. However, in line with our findings, most studies report a downregulation of the Cad antigen with differentiation in CRC cell lines[32,56]. β1-4-GalNAc-transferase 2 is the *N*-acetylgalactosaminyltransferase (B4GALNT2) responsible for the addition of the terminal GalNAc residue in a β1-4 linkage to the sialylated galactose. It has previously been observed that the expression of the Cad antigen and the B4GALNT2 gene are regulated by promoter DNA methylation[57]. Recent literature reported the capability of butyrate to trigger DNA demethylation in CRC cell lines[58]. Promoter demethylation triggered by butyrate, followed by differential expression of the B4GALNT2 gene and corresponding Cad antigen, could explain the significantly higher expression of the Cad antigen in the butyrate stimulated group compared to the spontaneous differentiation group.

Another interesting *O*-glycan that shows an upregulation in the butyrate-stimulated samples compared to the spontaneously differentiated samples (after day 14), is the H1N1S3 *O*-glycan. Tandem MS revealed that this *O*-glycan carries a di-sialic acid (DiSia) unit (**Supplementary Information, Figure S7**), potentially with an α2-8 glycosidic linkage[59]. Previously an upregulation of α2-8-sialyltransferase *ST8SIA6* mRNA expression was reported in Caco-2 cells that underwent spontaneous differentiation[60]. Details regarding the biological function of this DiSia motif remains unknown, making it an interesting structure for future studies. Moreover, an increase in core 2 sialylated *O*-glycans was observed in differentiated cells, which is in line with

previously reported changes in the expression of α2-3-sialyltransferases (ST3GAL4 and ST3GAL6) as well as core 2 synthases (GCNT4 and GCNT3)[60]. Additionally, the previously reported upregulation of B3GNT3[60] could be responsible for the upregulation of the H3N3S2b glycan, by adding another LacNAc to the 6 arm of the *O*-glycan.

BSM and its *O*-glycan repertoire has been very well characterized by both tandem MS and NMR[38,39] and served in our study as a valuable reference standard allowing us to assign specific glycan linkages. We were able to confirm that the dominant isomers of *O*-glycans with composition H2N2F2 and H2N2F1 are carrying α1-2 linked terminal fucoses (**Supplementary Information, Figure S4** and **S5**, respectively) whereas a recent study reported the presence of an unconventional terminal α1-6-linked fucose[30]. However, it was not possible to fully determine the linkages of all *O*-glycan isomers, as several of them were not present in the BSM standard.

The stabilization of the *O*-glycome for spontaneously differentiated cells after day 14 is in line with the stabilization of the proteome as reported previously by Stierum *et al*., 14 days post-confluence[25]. In this study, we were able to study both the effect of butyrate stimulation and spontaneous differentiation on the same sample set used for profiling the *O*-glycome. In agreement with the dynamics of their *O*-glycomics data, the proteome of the butyrate-stimulated cells also showed further changes 14 days post-confluence, indicating an important role for butyrate-stimulated cellular regulation in the late phase of differentiation. In order to relate the previously mentioned *O*-glycomic changes to changes in the abundance of specific *O*-glycoproteins, we continued our analysis of the proteomic data by selecting the proteins involved in the GO ontologies – Glycosylation, GO Monosaccharide transport, as well as *O*-glycoproteins[42]. An upregulation of specific monosaccharide transporters SLC2A1 (GLUT1), SLC2A3 (GLUT3) and SLC2A5 (GLUT5) is seen in butyrate stimulated differentiated cells, important for uptake of glucose and other monosaccharides into the cell, which was previously associated with differentiated Caco-2 cells[61,62]. Interestingly, the only monosaccharide present in the cell culture media is glucose, therefore the availability of other monosaccharide precursors for glycan biosynthesis is dependent on their intracellular de novo biosynthesis from glucose. Changes in the expression of enzymes involved in metabolic conversion of monosaccharide precursors were observed with differentiation. Namely, hexokinase that converts

glucose into glucose-6-P (HK2), glutamine-fructose-6-phosphate aminotransferase 1 (GFPT1) and phosphoacetylglucosamine mutase (PGM3), both involved in the biosynthesis of UDP-GlcNAc, are upregulated with differentiation. These enzymes are important rate limiting enzymes of the hexosamine pathway, regulating the availability of precursors for *N*- and *O*-linked glycosylation of proteins. Studies on a molecular level for glycosyltransferases are known to be challenging, as many of them are low abundance proteins[63]. With our current method we were able to detect an increase in abundance of polypeptide GalNAc transferases 3 and 2 with differentiation (GALNT3, GALNT2 respectively), followed by a decline in the late stages of differentiation. Previously, knock out of GALNT2 and GALNT3 in human organotypic skin models caused impaired cell adhesion and decreased differentiation, respectively, implying the importance of these enzymes initiating *O*-GalNAc glycosylation for cell–extracellular matrix interactions and epithelial differentiation[64]. Additionally, proteins such as F11R, BCAM, ADAMTSL4, LGALS3BP, and laminin are found to be GALNT2 specific targets, whereas protocadherin (FAT2) and cathepsin (CTSD) are found to be a GALNT3 specific targets in skin cells[64]. Another study pinpointed apolipoprotein E (APOE), nucleobindin-2, lipolysis-stimulated lipoprotein receptor, laminin (LAMC2), coagulation factor X, and protein disulfide-isomerase A3 as GALNT3 and GALNT6 targets[65]. These observations could provide interesting perspectives to study the GALNT specific targets in the colon, and how the glycosylation of the specific proteins influences their function in cell adhesion during differentiation.

To identify associations between proteomics and glycomics changes with differentiation we continued using a data integration approach (MixOmics)[43]. The specific butyrate pattern shows higher abundances of cell adhesion proteins in later stages of differentiation. Those proteins show a similar change in abundance as sialylated glycan species indicating that specific glycosylation changes may be a consequence of changes in the abundance of specific proteins (**Figure 6**). Moreover, specific *O*-glycosylation of cell adhesion *O*-glycoproteins may play an important role in cell differentiation. The structure-function relationship of glycosylation in these proteins involved in adhesion is still to be deciphered.

Remarkably, we observed a general downregulation of abundances of transcription factors involved in epithelial differentiation such as HNF1A, GATA6, FOXA[19]. Whereas, abundance of HNF4A, the key regulator of intestinal genes in Caco-2

model[19], increased by day 14 in butyrate stimulated cells, in line with previous research where an upregulation on mRNA and protein levels was observed[66]. Moreover, it was previously shown that HNF4A also regulates the HNF1A and CDX2 promoter activity[66] and the functional interaction between HNF1A and CDX2 was demonstrated before in Caco-2 cells[67]. Although no unique peptides for CDX1 were identified in our study, the previously described target protein cytokeratin 20 showed upregulation with differentiation[68]. The activity of these transcription factors depends on a complex network of interactions, and not solely on the abundance of the protein product. The integrative analysis with transcription factors and *O*-glycans (**Supplementary Information 1, Figure S13** and **Figure S14)** reveals new potential regulators of glycosylation such as FOXA1, FOXP4 and STAT1, however, further studies are required to validate these hypotheses.

Interestingly, in the butyrate stimulated differentiation, the downregulation of HNF1A transcription factor correlated with fucosylated glycomics signatures such as H2N2F1S1c, H2N2F2b, H2N2F1d and H1N1F1 **(Figure 7).** HNF1A is a transcription factor expressed in several organs including the intestine and stomach. Genome-wide association studies by Lauc *et al.,* identified HNF1A as a key regulator of fucosylation[69]. It was demonstrated that HNF1A activates antennary FUTs (FUT3-11 in liver cell lines) and downregulated core FUTs (FUT8). Additionally, it showed that HNF1A upregulated GDP-mannose 4,6 dehydratase (GMDS), an important enzyme in de-novo d-fucose biosynthesis pathway which increases the availability of GDP-fucose precursor for antennary fucosylation of glycans[69]. While a trend in downregulation of GMDS was seen on the proteome level, it did not meet the significance threshold. Nevertheless, our data indicates that there might be an important role of HFN1A as a master transcriptional regulator of fucosylation in CacoCaco-2 cells, where downregulation of HNF1A correlates with the downregulation of fucosylation upon differentiation. HNF1A downregulation can lead to the limitation of d-fucose availability and consequently upregulation of terminal sialylation which is in direct competition with terminal blood group H antigen biosynthesis. Unfortunately, we did not detect the CDX1 transcription factor in our proteomics study, a previously reported transcriptional regulator of Lewis type fucosylation on *N*-glycans[32,70] and the detected CDX2 showed no statistically significant change. Moreover, the Caco-2 cell line, among other undifferentiated cell

lines, previously showed a predominant expression of *O*-linked blood group H fucosylation and weak associations could be seen with HNF4A transcription factor[31]. This further supports the hypothesis that HNF1A might be a key regulator of *O*-linked glycan fucosylation in Caco-2 cells during cell differentiation. Several other associations between specific glycan structures and protein expression are reported in this study and could potentially be the start of new endeavors which may have important implications for understanding normal biological functions such as colon differentiation.[69]

In this study an integrative approach was used to generate hypotheses about relationships between cell glycome and proteome that occurs with cell differentiation. However, to validate those hypotheses, *O*-glycoproteomic analysis is needed, which still remains a challenging task in complex biological samples. Previously, a genetic engineering approach that limits the glycan diversity to a single GalNAc or SiaGalNAc enabled enrichment of cell *O*-glycoproteome by lectin affinity chromatography and revealed the complexity of the cell *O*-glycoproteome[71]. Although this approach is invaluable for discovery of *O*-glycoprotein sites, it does not give information about the glycan microheterogeneity. In this study important hypotheses were made using an integrative approach which can be exploited for future research with targeted *O*-glycoproteomic approach of specific proteins. Moreover, functional studies can be employed to validate the suggested transcriptional regulators of glycosylation, and their role in cell differentiation. Which, eventually, will provide insights about the role of glycosylation cell adhesion and the role of fucosylation and sialylation in cell differentiation.

CONCLUSIONS

Differentiated CacoCaco-2 cells are an important and widely used model cell line. However, the glycome expression might change dependent on the growth- and bacterial metabolites available during the culturing process. In this study an untargeted in-depth screening of the CacoCaco-2 cell line was performed, identifying specific *O*-glycans that mark butyrate-induced epithelial differentiation and link them to potential protein carriers. While interesting structures were identified, further studies are required to identify their role in maintaining homeostasis in the epithelium. In the future this might allow us to gain a better understanding of the constantly changing *O*-

glycome at the gut-microbiota interface as correlated to bacterial hydrolase activities and its relevance to maintaining homeostasis or role in dysbiosis.

SUPPORTING INFORMATION

Supporting information is available upon request.-

DATA AVAILABILITY

The data in support of the findings of this study may be found within the manuscript and in the associated supplementary files. Mass spectrometry-based glycomics raw data were deposited in the Glycopost repository.

ACKNOWLEDGMENTS

This work was supported by the European Commission's Horizon 2020 program via the "GlyCoCan" project (Grant 676421), the ERC-2019-STG 852452 grant for K. Strijbis, and by the Netherlands Organization for Scientific Research (NWO) via a VIDI grant to T. Wennekes (723.014.005).

AUTHOR CONTRIBUTIONS

K.M. performed the glycomic analysis and proteomic data analysis. Y.M.C.A.L cultured the cells and performed alkaline phosphatase assay and microscopy. K.M. and O.A.M. performed the integrative data analysis and statistics. P.A.V. and G.M.C.J. performed the proteomics experiments. K.M., Y.M.C.A.L, G.S.M.L., K.S., T.W., M.W. conceptually designed the work. K.M., Y.M.C.A.L, K.S., T.W., G.S.M.L. and M.W. wrote the manuscript. All authors gave approval to the final version of the manuscript.

REFERENCES

1. Neish, A. S. Microbes in Gastrointestinal Health and Disease. (2009).
2. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. (2005).
3. Yang, B., Cao, L., Liu, B., McCaig, C. D. & Pu, J. The Transition from Proliferation to Differentiation in Colorectal Cancer Is Regulated by the Calcium Activated Chloride Channel A1. *PLoS ONE* **8**, e60861 (2013).
4. Martinez, K. B., Leone, V. & Chang, E. B. Microbial metabolites in health and disease: Navigating the unknown in search of function. (2017).

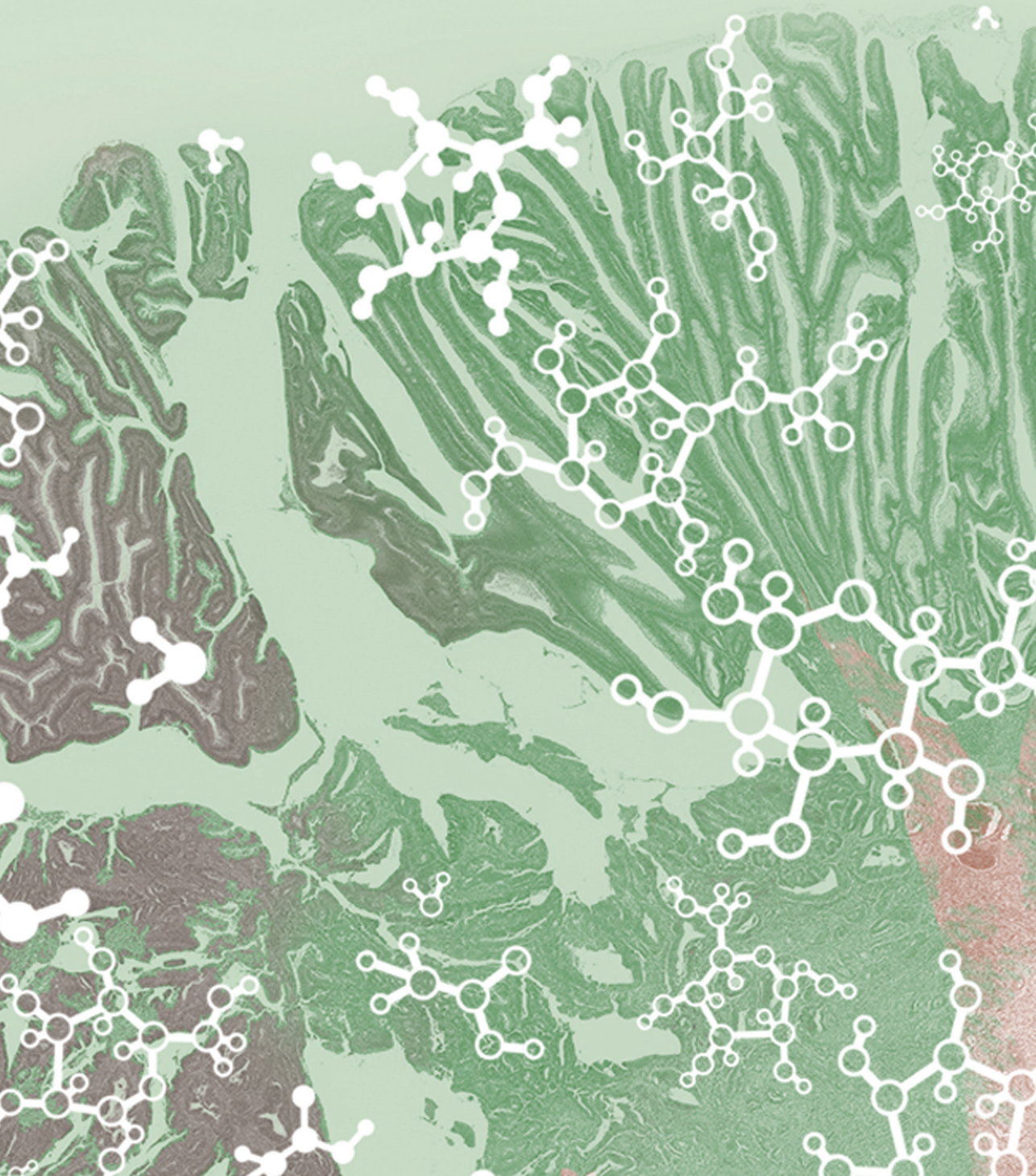
5. McIntyre, A., Gibson, P. R. & Young, G. P. Butyrate production from dietary fibre and protection against large bowel cancer in a rat model. *Gut* **34**, 386–391 (1993).
6. Lazarova, D. L., Chiaro, C. & Bordonaro, M. Butyrate induced changes in Wnt-signaling specific gene expression in colorectal cancer cells. *BMC Research Notes* **7**, 226 (2014).
7. Han, R., Sun, Q., Wu, J., Zheng, P. & Zhao, G. Sodium Butyrate Upregulates miR-203 Expression to Exert Anti-Proliferation Effect on Colorectal Cancer Cells. *Cellular Physiology and Biochemistry* **39**, 1919–1929 (2016).
8. Zuo, L., Lu, M., Zhou, Q., Wei, W. & Wang, Y. Butyrate suppresses proliferation and migration of RKO colon cancer cells though regulating endocan expression by MAPK signaling pathway. *Food and Chemical Toxicology* **62**, 892–900 (2013).
9. Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. (2016).
10. Willemsen, L. E. M., Koetsier, M. A., Van Deventer, S. J. H. & Van Tol, E. A. F. Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E1 and E2 production by intestinal myofibroblasts. *Gut* **52**, 1442–1447 (2003).
11. Flint, H. J., Scott, K. P., Louis, P. & Duncan, S. H. The role of the gut microbiota in nutrition and health. (2012).
12. Klepinina, L. *et al.* Colon cancer cell differentiation by sodium butyrate modulates metabolic plasticity of Caco-2 cells via alteration of phosphotransfer network. *PLoS ONE* **16**, e0245348 (2021).
13. Gaudier, E., Rival, M., Buisine, M. P., Robineau, I. & Hoebler, C. Butyrate enemas Upregulate Muc genes expression but decrease adherent mucus thickness in mice colon. *Physiological Research* **58**, 111–119 (2009).
14. Silva, Y. P., Bernardi, A. & Frozza, R. L. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. (2020).
15. Birchenough, G. M. H., Johansson, M. E. V., Gustafsson, J. K., Bergström, J. H. & Hansson, G. C. New developments in goblet cell mucus secretion and function. (2015).
16. Van Putten, J. P. M. & Strijbis, K. Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. (2017).
17. Paone, P. & Cani, P. D. Mucus barrier, mucins and gut microbiota: The expected slimy partners? (2020).
18. Bergstrom, K. S. B. & Xia, L. Mucin-type O-glycans and their roles in intestinal homeostasis. (2013).
19. Yin, S. *et al.* Functional genomics analysis of human colon organoids identifies key transcription factors. *Physiol. Genomics* **52**, 234–244 (2020).
20. Gaudier, E. *et al.* Butyrate regulation of glycosylation-related gene expression: Evidence for galectin-1 upregulation in human intestinal epithelial goblet cells. *Biochemical and Biophysical Research Communications* **325**, 1044–1051 (2004).
21. Shah, S. *et al.* n-Butyrate reduces the expression of β-galactoside α2,6- sialyltransferase in Hep G2 cells. *Journal of Biological Chemistry* **267**, 10652–10658 (1992).

22. Jung, T. H., Park, J. H., Jeon, W. M. & Han, K. S. Butyrate modulates bacterial adherence on LS174T human colorectal cells by stimulating mucin secretion and MAPK signaling pathway. *Nutrition Research and Practice* **9**, 343–349 (2015).
23. Cornick, S., Tawiah, A. & Chadee, K. Roles and regulation of the mucus barrier in the gut. (2015).
24. Buhrke, T., Lengler, I. & Lampen, A. Analysis of proteomic changes induced upon cellular differentiation of the human intestinal cell line Caco-2. *Dev. Growth Differ.* **53**, 411–426 (2011).
25. Stierum, R. *et al.* Proteome analysis reveals novel proteins associated with proliferation and differentiation of the colorectal cancer cell line Caco-2. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1650**, 73–91 (2003).
26. Pshezhetsky, A. V. *et al.* Subcellular proteomics of cell differentiation: Quantitative analysis of the plasma membrane proteome of Caco-2 cells. *Proteomics* **7**, 2201–2215 (2007).
27. Tadjali, M., Seidelin, J. B., Olsen, J. & Troelsen, J. T. Transcriptome changes during intestinal cell differentiation. *Biochimica et Biophysica Acta - Molecular Cell Research* **1589**, 160–167 (2002).
28. Turck, N. *et al.* Proteomic analysis of nuclear proteins from proliferative and differentiated human colonic intestinal epithelial cells. *Proteomics* **4**, 93–105 (2004).
29. Tremblay, E. *et al.* Gene expression profiles of normal proliferating and differentiating human intestinal epithelial cells: A comparison with the Caco-2 cell model. *J. Cell. Biochem.* **99**, 1175–1186 (2006).
30. Xu, G., Goonatilleke, E., Wongkham, S. & Lebrilla, C. B. Deep structural analysis and quantitation of O-linked glycans on cell membrane reveal high abundances and distinct glycomic profiles associated with cell type and stages of differentiation. **14**, 2020.
31. Madunić, K. *et al.* Colorectal cancer cell lines show striking diversity of their O-glycome reflecting the cellular differentiation phenotype. *Cell. Mol. Life Sci.* **1**, 3.
32. Holst, S. *et al.* N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homebox 1 (CDX1)/Villin mRNA Expression. *Mol. Cell. Proteomics* **15**, 124–140 (2016).
33. 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* **16**, 355–363 (2020).
34. Madunić, K., Wagt, S., Zhang, T., Wuhrer, M. & Lageveen-Kammeijer, G. S. M. Dopant-Enriched Nitrogen Gas for Enhanced Electrospray Ionization of Released Glycans in Negative Ion Mode. *Anal. Chem.* acs.analchem.1c00023 (2021) doi:10.1021/acs.analchem.1c00023.
35. Paulo, J. A. & Gygi, S. P. Nicotine-induced protein expression profiling reveals mutually altered proteins across four human cell lines. *Proteomics* **17**, (2017).
36. McComb, R. B., Bowers, G. N., Jr & Posen, S. *Alkaline Phosphatase - Robert B. McComb, George N. Bowers, Jr., Solomon Posen - Google Books*. (1979).

37. Savage, A. V., Donohue, J. J., Koeleman, C. A. M. & van den Eijnden, D. H. Structural characterization of sialylated tetrasaccharides and pentasaccharides with blood group H and Lex activity isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **193**, 837–843 (1990).
38. Savage, A. V., D'Arcy, S. M. T. & Donoghue, C. M. Structural characterization of neutral oligosaccharides with blood group A and H activity isolated from bovine submaxillary mucin. *Biochem. J* **279**, 95–103 (1991).
39. Savage, A. V., Donoghue, C. M., D'Arcy, S. M., Koeleman, C. A. M. & van den Eijnden, D. H. Structure determination of five sialylated trisaccharides with core types 1, 3 or 5 isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **192**, 427–432 (1990).
40. Prieto, P. A. *et al.* Expression of human H-type α1,2-fucosyltransferase encoding for blood group H(O) antigen in chinese hamster ovary cells. Evidence for preferential fucosylation and truncation of polylactosamine sequences. *Journal of Biological Chemistry* **272**, 2089–2097 (1997).
41. Möginger, U. *et al.* Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and Squamous Cell Carcinoma. *Front. Oncol.* **8**, 70 (2018).
42. Steentoft, C. *et al.* Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488 (2013).
43. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
44. Carmona-Vicente, N., Allen, D. J., Rodríguez-Díaz, J., Iturriza-Gómara, M. & Buesa, J. Antibodies against Lewis antigens inhibit the binding of human norovirus GII.4 virus-like particles to saliva but not to intestinal Caco-2 cells. *Virology Journal* **13**, (2016).
45. Murakami, K. *et al.* Norovirus Binding to Intestinal Epithelial Cells Is Independent of Histo-Blood Group Antigens. *PLoS ONE* **8**, e66534 (2013).
46. Amano, J. & Oshima, M. Expression of the H type 1 blood group antigen during enterocytic differentiation of Caco-2 cells. *J. Biol. Chem.* **274**, 21209–21216 (1999).
47. Lea, T. Caco-2 Cell Line. in *The Impact of Food Bioactives on Health: in vitro and ex vivo models* (eds. Verhoeckx, K. et al.) 103–111 (Springer International Publishing, 2015). doi:10.1007/978-3-319-16104-4_10.
48. Walter, E. & Kissel, T. Heterogeneity in the human intestinal cell line Caco-2 leads to differences in transepithelial transport. *European Journal of Pharmaceutical Sciences* **3**, 215–230 (1995).
49. Li, J., Hsu, H. C., Mountz, J. D. & Allen, J. G. Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases. (2018).
50. Youakim, A. & Herscovics, A. Differentiation-associated decrease in the proportion of fucosylated polylactosaminoglycans of CaCo-2 human colonic adenocarcinoma cells. *Biochemical Journal* **247**, 299–306 (1987).
51. Amano, J., Kobayashi, K. & Oshima, M. Comparative study of glycosyltransferase activities in Caco-2 cells before and after enterocytic differentiation using lectin-affinity high-performance liquid chromatography. *Arch. Biochem. Biophys.* **395**, 191–198 (2001).

52. Link-Lenczowski, P. *et al.* A switch of N-glycosylation of proteome and secretome during differentiation of intestinal epithelial cells. *Biochimica et Biophysica Acta - Molecular Cell Research* **1866**, 118555 (2019).
53. Qi, F. *et al.* ST3GAL3, ST3GAL4, and ST3GAL6 differ in their regulation of biological functions via the specificities for the α2,3-sialylation of target proteins. *FASEB Journal* **34**, 881–897 (2020).
54. Pinto, R. *et al.* CDX2 homeoprotein is involved in the regulation of ST6GalNAc-I gene in intestinal metaplasia. *Lab. Invest.* **95**, 718–727 (2015).
55. Engle, M. J., Goetz, G. S. & Alpers, D. H. Caco-2 cells express a combination of colonocyte and enterocyte phenotypes. *J. Cell. Physiol.* **174**, 362–369 (1998).
56. Malagolini, N., Dall'Olio, F. & Serafini-Cessi, F. UDP-Ga1NAc:NeuAcα2,3Galβ-R (GalNAc to Gal) β1,4-N-acetyl-galactosaminyltransferase responsible for the Sda specificity in human colon carcinoma CaCo-2 cell line. *Biochem. Biophys. Res. Commun.* **180**, 681–686 (1991).
57. Wang, H. R., Hsieh, C. Y., Twu, Y. C. & Yu, L. C. Expression of the human Sda β-1,4- N-acetylgalactosaminyltransferase II gene is dependent on the promoter methylation status. *Glycobiology* **18**, 104–113 (2008).
58. Sun, X. & Zhu, M. J. Butyrate Inhibits Indices of Colorectal Carcinogenesis via Enhancing α-Ketoglutarate-Dependent DNA Demethylation of Mismatch Repair Genes. *Molecular Nutrition and Food Research* **62**, e1700932 (2018).
59. Sato, C. & Kitajima, K. Disialic, oligosialic and polysialic acids: Distribution, functions and related disease. (2013).
60. Park, D. *et al.* Characteristic Changes in Cell Surface Glycosylation Accompany Intestinal Epithelial Cell (IEC) Differentiation: High Mannose Structures Dominate the Cell Surface Glycome of Undifferentiated Enterocytes. *Mol. Cell. Proteomics* **14**, 2910–2921 (2015).
61. Mahraoui, L. *et al.* Presence and differential expression of SGLT1, GLUT1, GLUT2, GLUT3 and GLUT5 hexose-transporter mRNAs in Caco-2 cell clones in relation to cell growth and glucose consumption. *Biochem. J* **298 Pt 3**, 629–633 (1994).
62. Mahraoui, L. *et al.* Expression and localization of GLUT-5 in Caco-2 cells, human small intestine, and colon. *Am. J. Physiol.* **263**, G312-8 (1992).
63. R, S. *et al.* Targeted Proteomics Reveals Quantitative Differences in Low Abundance Glycosyltransferases of Patients with Congenital Disorders of Glycosylation. (2020) doi:10.1101/2020.09.15.291732.
64. Bagdonaite, I., Pallesen, E. M. H., Ye, Z. & Vakhrushev, S. Y. O-glycan initiation directs distinct biological pathways and controls epithelial differentiation. *EMBO* (2020).
65. Lavrsen, K. *et al.* De novo expression of human polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) in colon adenocarcinoma inhibits the differentiation of colonic epithelium. (2017) doi:10.1074/jbc.M117.812826.
66. Boyd, M., Bressendorff, S., Møller, J., Olsen, J. & Troelsen, J. T. Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC Gastroenterol.* **9**, 68 (2009).

67. Mitchelmore, C., Troelsen, J. T., Spodsberg, N., Sjöström, H. & Norén, O. Interaction between the homeodomain proteins Cdx2 and HNF1α mediates expression of the lactase-phlorizin hydrolase gene. *Biochem. J* **346**, 529–535 (2000).
68. Chan, C. W. M. *et al.* Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1936–1941 (2009).
69. Lauc, G. *et al.* Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1α as a master regulator of plasma protein fucosylation. *PLoS Genet.* **6**, e1001256 (2010).
70. Holst, S. *et al.* N-Glycomic and Transcriptomic Changes Associated with CDX1 mRNA Expression in Colorectal Cancer Cell Lines. *Cells* **8**, 273 (2019).
71. Schjoldager, K. T.-B. G. & Clausen, H. Site-specific protein O-glycosylation modulates proprotein processing — Deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 2079–2094 (2012).



COLORECTAL CANCER, BUT NOT HEALTHY COLON, EXPRESSES SPECIFIC CORE 2 SIALYLATED *O*-GLYCANS

Katarina Madunić[1], Oleg A. Mayboroda[1], Tao Zhang[1], Julia Weber[4], Geert-Jan Boons[4], Hans Morreau[3], Ronald van Vlieberghe[2], Tom van Wezel[3], Guinevere S.M. Lageveen-Kammeijer[1], Manfred Wuhrer[1]

[1] *Leiden University Medical Center, Center for Proteomics and Metabolomics, Postbus 9600, 2300 RC Leiden, The Netherlands*
[2] *Leiden University Medical Center, Department of Surgery, Postbus 9600, 2300 RC Leiden, The Netherlands*
[3] *Leiden University Medical Center, Department of Pathology, Postbus 9600, 2300 RC Leiden, The Netherlands*
[4] *Department of Chemical Biology and Drug Discovery, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands*

Manuscript submitted

Chapter 6

ABSTRACT

Current immunotherapy for colorectal cancer (CRC) shows limited patient benefit due to the relative paucity of the targeted molecules. Changes in protein glycosylation are a hallmark of CRC, and glycans are expressed widely on the cell surface, which makes them promising targets. Here we present an in-depth study of CRC mucin type *O*-glycosylation, derived from epithelial cancer regions compared to healthy colon mucosa from the same patients. We found a total of 100 tumor associated carbohydrate antigens (TACAs) which were solely observed in the cancer region. Core 2 O-glycans with Lewis, sialyl-Lewis and otherwise sialylated 6-arms were most discriminating, being found in most of the cancers whilst being absent in the neighboring healthy colon mucosa. These specific O-glycan structures present promising new targets for development of innovative cancer immunotherapeutics.

INTRODUCTION

Colorectal cancer (CRC) is one of the leading malignancies worldwide with over 900,000 deaths in 2020[1]. Conventional therapeutic strategies include chemotherapy, radiation therapy, and surgery. However, despite the recently introduced and highly successful pre-symptomatic population screening, many CRC cases are still detected at an advanced stage, leading to an unsuccessful treatment[2]. In the past decade, new CRC therapies emerged, such as specific inhibitors and antibodies targeting soluble proteins and cellular receptors[3]. These include the biologics anti-VEGF (bevacizumab), anti-EGFR (cetuximab and panitumumab), anti-PD-1 immune checkpoint inhibitors (pembrolizumab, nivolumab), BRAF V600E inhibitors (several combinations of vemurafenib, irinotecan, and cetuximab or panitumumab)[4], with many more in development. Unfortunately, these treatments are only efficient for a limited subset of patients as many patients develop therapy resistance[5]. Therefore, the identification of new specific molecular targets and the development of new targeted therapies for CRC are essential for providing better treatment for a larger group of patients.

Various studies have shown altered glycosylation patterns of proteins and lipids as a hallmark of cancer[6,7]. The changes in glycosylation are originating from a shift in the expression of glycosyltransferases, modifications of their enzymatic activity, mislocalization in the endoplasmic reticulum or Golgi apparatus, availability of substrates or nucleotide sugar donors as well as competition between the enzymes[6]. The characteristic glycan alterations in cancer include specific aberrant expression of incomplete carbohydrate structures or *de novo* expression of carbohydrate antigens also known as tumor-associated carbohydrate antigens (TACAs)[8]. Of note, aberrant cancer glycosylation has already found its way into the clinics as the CA19-9 serological marker for e.g. pancreatic cancer and CRC which is known to be the sialyl-Lewis A (sLe^A) carbohydrate antigen[9].

The expression of the heavily glycosylated intestinal mucins changes in CRC. It has been demonstrated that MUC2 and MUC4 expression is downregulated in CRC, whereas MUC1, MUC5AC and MUC17 are overexpressed[10], showing reduced expression of core 3 and core 4 structures with increased expression of *O*-linked truncated core 1 glycans, (sialyl-)T antigen, (sialyl-)Tn, and sialyl-Lewis X

(sLe^X)antigen [7], and several of these motifs are being evaluated as targets of immunotherapeutic monoclonal antibodies[11], despite their often limited specificity.

While the presence of these carbohydrate motifs in tumors is well-known, information on the underlying TACA structures is vastly lacking. The majority of the studies rely on detecting abnormal expression of glycosyltransferases in cancer and detection of terminal epitopes on all glycan types, and no differentiation is made between the *N*- and *O*-linked glycoproteins or glycolipids, nor are the specific TACA structures revealed[12–18].

In recent years, sensitive mass spectrometry (MS) based glycomics workflows have been developed enabling untargeted screening of potential glycan markers derived from complex biological samples. However, tumor tissue heterogeneity is a major obstacle for cancer glycomics studies. The tumor microenvironment is populated by non-cancer cells such as immune cells or fibroblasts which can hamper the detection of specific cancer associated glycan species. This issue can be largely overcome by sampling the areas of interest with laser capture microdissection (LCM) prior to glycan release as demonstrated in hepatocellular carcinoma by Hinneburg *et al.*[19,20].

In this study we focused on decoding the tumor specific *O*-glycan signatures of CRC. The *O*-glycome was determined from both epithelial regions of primary tumors and metastatic sites and was compared to healthy colon tissue from the same patients. Furthermore, we established a high throughput workflow that sequentially releases the *N*-and *O*-glycans from LCM formalin fixed paraffin embedded (FFPE) tissues, followed by analysis using porous graphitized carbon liquid chromatography (PGC-LC)-MS/MS in negative ion mode, enabling a powerful separation of isomeric species as well as in-depth structural characterization of TACAs. This approach allowed us to identify CRC specific TACAs as potential targets of innovative cancer immunotherapy such as anti-TACA antibodies, adoptive T-cell therapies and antibody-drug conjugates.

MATERIALS AND METHODS

MATERIALS AND REAGENTS

Hematoxylin (cat. nr. GHS232), sodium dodecyl sulfate solution 20% (cat. nr. 05030), trifluoroacetic acid (TFA; cat. nr. 1.38178.0050), tris(hydroxymethyl)amino-methane (cat. nr. 252859; lot#BCBM2559V), sodium borohydride (cat. nr. 452882; lot nr.

STBD8912V), hydrochloric acid (cat. nr. 258148; lot nr. SZBD3100V), DL-dithiothreitol (DTT; cat. nr. D0632; lot nr. SLBW0160), ammonium bicarbonate (cat. nr. 09830; lot nr. BCBQ6426V,) cation exchange resin Dowex 50W X8 (cat. nr. 217492; lot nr. MKCH2513), ammonium acetate (cat. nr. A1542) and polyvinylpirrolidone mw 40, 000 (cat. Nr PVP40, lot.nr. BCBM0949V) were purchased from Sigma Aldrich (St. Louis, MO). Ethanol (EtOH; cat. nr. 100983.1000), NaCl; cat. nr. 1.06404.1000) and methanol (MeOH; cat. nr. 1.06009.2500) were purchased from Merck (Darmstadt, Germany). Acetonitrile LC-MS grade (cat. nr. 01203502) was obtained from Biosolve (Valkenswaard, The Netherlands). Glacial acetic acid (cat. nr. A6283; lot nr. SZBG2660H) and potassium hydroxide (cat. nr. P1767) were purchased from Honeywell Fluka (Charlotte, NC). PNGase F (Flavobacterium meningosepticum recombinant in E. coli; Cat No. 11365193001) was obtained from Roche (Mannheim, Germany). SPE bulk sorbent Carbograph (cat. nr. 1769; lot nr. 5122145) was acquired from Grace Discovery sciences (Columbia, TN). MultiScreen® HTS 96-multiwell plates (pore size 0.45m) with high protein-binding membrane (hydrophobic Immobilon-P PVDF membrane) and 96-well PP Microplate (cat. nr. 651201; lot nr. E1708385) were purchased from Millipore (Amsterdam, The Netherlands), 96-well PP filter plate (cat. nr. OF1100) was obtained from Orochem technologies (Naperville, Il). MembraneSlide 1.0 PEN (cat.nr. 415190-9041-000) and Adhesive Cap 500 µL tubes (cat.nr. 415190-9211-000) were purchased from Carl Zeiss Microscopy (Göttingen, Germany). All buffers were prepared using Milli-Q water (mQ) generated from a ELGA system (Millipore, Amsterdam, Netherlands), maintained at ≥ 18 MΩ. Chemically synthesized glycopeptide standards were provided by Utrecht University (Weber et al.; manuscript in preparation).

FFPE TISSUE SECTIONING AND STAINING

Anonymized human CRC paraffin tissue blocks were from the Department of Pathology at Leiden University Medical Center (LUMC, Leiden, The Netherlands). All samples were handled in a coded fashion, according to the national ethical guidelines (“Code for Proper Secondary Use of Human Tissue”, Dutch Federation of Medical Scientific Societies). Paired primary tumor (T1-T12) and adjacent normal colon mucosa (C1-C12) from 12 patients were selected for the cohort. Additionally, 6 metastatic cancers obtained from liver metastases of different patients were included (M15- M21). The tissues were cut into 5 µm thick sections from paraffin blocks using

a microtome and mounted on glass slides for hematoxylin and eosin (HE) staining, or 10 µm thick sections on membrane PEN slides for LCM. All slides were dried overnight at 37 ˚C and stored at 4 ˚C.

Slides used for LCM were deparaffinized using xylene (three times for 5 min) and washed with absolute EtOH (two times for 2 min). The slides were then rehydrated in a series of EtOH by submerging the slides in 85% EtOH, followed by 70% EtOH and distilled water. Hematoxylin was applied for 20 sec. The slides were then washed with demineralized water for 2 min and dehydrated with increasing EtOH series by submerging in 70% EtOH, followed by 85% EtOH and finally absolute EtOH. The slides were briefly air dried and stored at 4 ˚C. The slides used for pathologist annotation were stained with HE following the routine protocols.

PATHOLOGIST ANNOTATION

Tumor and healthy epithelial tissue regions from HE slides were annotated by a pathologist by careful inspection under the microscope. Differentiation grade was determined by assessing glandular formation in the HE histological slides.

LASER CAPTURE MICRODISSECTION

Tissue sections mounted on PEN membrane slides for LCM were carefully inspected under the microscope. Target areas marked on HE slides by the pathologist were located, and the laser was positioned in this area (**Supplementary Fig. S2**). To ensure that comparable amounts of tissue were used for each sample, cells were counted in three different regions of the tissue in an area of approx. 2500 µm^2. An average of the count was used to extrapolate the area needed to be cut in order to obtain samples containing approximately 20,000 cells. The large difference in cell size between hepatic and colon cells was taken into account by performing the LCM on 10 µm thick tissue sections, enabling the extraction of sufficient material for the analysis. LCM was performed using PALM RoboSoftware and collected in adhesive cap 500 µL tubes. Considering the healthy colon mucosa is surrounded by infiltrated lymphocytes that could not be separated from the epithelial cells successfully, we also extracted lymphocyte follicles from different normal colon tissues and pooled them together to obtain a potentially confounding immune cell glycan profile (IM). Moreover, we extracted the stroma controls for the tumors with high stroma infiltration that might

have contamination coming from cancer associated fibroblasts and immune cells (ST4, ST6, ST7, ST11, ST12, SM19, and SM21).

GLYCAN RELEASE AND PURIFICATION

A lysis buffer (100 µL) containing 100 mM Tris HCl, 0.1 M DTT, 100 mM NaCl, and 1% SDS was added to microdissected tissues collected in adhesive caps. The pieces were carefully collected in lysis buffer from the lid of the adhesive caps and transferred to a 1.5 mL Eppendorf tube. Prior to sonication using a Branson sonicator rod (three times for 15 sec, with output power 2/10) the tubes were placed on ice. During sonication, the samples were kept on ice and left 20 sec to cool down between each cycle. Furthermore, the samples were incubated at 99 ˚C for 60 min with mild agitation (400 rpm). After incubation, the samples were slowly cooled down at room temperature (RT). In the meantime, PVDF membrane plate was preconditioned using 100 µL of 70% EtOH, until the membrane turned opaque, followed by an additional wash using 100 µL of MQ. Upon rewetting the membrane with 5 µL of 70% EtOH, 100 µL of carefully mixed sample lysates were loaded onto the membrane wells. Additionally, a tissue lysate (containing approximately 8 x 10^4 cells) was split into three separate samples (TECH1, TECH2, and TECH3) and processed independently (different wells of the PVDF membrane plate) to check for technical variability. The plate was incubated at RT on a shaker for 20 min to ensure binding. The unbound material was removed by centrifugation at 500 *x g* and washed with 100 µL of MQ. An additional, 40 µL of 0.5% PVP-40 in MQ was added to the PVDF membrane wells to block the membrane and incubated for 5 min on a shaker. Upon removal of the blocking agent by centrifugation, the membrane wells were extensively washed, two times with 100 µL of PBS, two times with 100 µL of 10 mM ammonium bicarbonate followed by two times 100 µL of MQ. Each time the washing agent was removed by centrifugation at 500 *x g* for 1 min. Since the addition of PVP-40 increases the water flux through the membrane, the hydrophilicity is also increased[21]. Therefore, to avoid enzyme mixture going through, 10 µL of MQ was added to each well of the PVDF membrane plate to soak the membrane and fill the space underneath it. The plate was incubated on a shaker with light agitation for 5 min. The enzyme mixture (15 µL) contained 2U of *N*-glycosidase F and was added to each well followed by an incubation of 15 min at 37 °C. Finally, 15 µL of water was added on top, to prevent drying out the membrane overnight, and placed in the 37 ˚C incubator. The next steps were

performed as previously described[22]. Briefly, the following day, upon recovery of *N*-glycans, 50 µL of 0.5 M sodium borohydride in 50 mM potassium hydroxide solution was added to each well and incubated for 16 h at 50 ˚C. The samples were cleaned using self-packed cation exchange columns in 96-well plates, followed by PGC-SPE. The samples were dried and stored at -20 ˚C until analysis.

PGC-LC-MS/MS MEASUREMENTS

Prior to analysis, the samples were resuspended in 15 µL of MQ and 1 µL of each sample was pooled together in a separate vial for quality check and method optimization (QC). The QC pool and a bovine submaxillary mucin standard were used to check instrument performance each day of measurement. A total of 6 µL of each tissue sample was injected for analysis (40%). A custom-made trap column (size 30 × 0.32 mm) packed with 5 µm particle size PGC stationary phase from Hypercarb PGC analytical column (size 100 × 4.6 mm, 5 µm particle size, Thermo Fisher Scientific, Waltham, MA) was used to load the samples using 100% buffer A (10 mM ammonium bicarbonate) at a loading flow of 6 µL/min. The glycans were separated on a custom-made PGC column (100 mm × 75 µm, 3 µm particle size obtained from Thermo Fisher Scientific) at a 0.6 µL/min flow rate by applying a linear gradient from 1% to 50% of buffer B (acetonitrile, 10 mM ammonium bicarbonate) over 73 min. A constant column temperature of 45 ˚C was maintained. A part of the samples were remeasured to resolve the isomers with composition H2N2F1S1 (isomer c and f). For this purpose a gradient was used ranging from 1% to 50% of buffer B over 110 minutes at 35 ˚C. The peak ratios of the resolved peaks was taken to extrapolate the peak areas from the original measurement (**Supplementary Table S11**). The LC system was coupled to an amaZon ETD speed ESI ion trap MS using the CaptiveSpray™ source (Bruker Daltonics) with an applied capillary voltage of 1000 V in negative-ionization mode. The drying gas (N_2) temperature was set at 280˚C and the flow to 3 L/min. The nebulizer gas pressure was kept at 3 psi enriched with isopropanol as described before[22]. MS spectra were acquired in enhanced mode within a mass to charge ratio (*m/z*) range of 380-1850, target mass of smart parameter setting was set to *m/z* 900, ion charge control (ICC) to 40,000 and maximum acquisition time to 200 ms. MS/MS spectra were generated by collision-induced dissociation on the three most abundant precursors, applying an isolation width of 3 Th. The fragmentation cut-off was set to 27% with

100% fragmentation amplitude using the Enhanced SmartFrag option (30–120% in 32 ms) and ICC was set to 150,000.

DATA PROCESSING

Extracted ion chromatograms including all observed charge states (1^- and 2^-) of the first three isotopes were used to integrate the area under the curve (AUC) for each individual glycan isomer using Compass DataAnalysis software (v.5.0). Peaks were manually picked and integrated. Relative quantitation was performed using the total area of all glycans within one sample as reference (100%). Glycan structures were identified by manual inspection of MS/MS spectra following known *O*-glycan biosynthetic pathways and available literature[23–27]. In the structural interpretations all hexoses (H) are assumed to be galactose, all deoxyhexoses a fucose (F), all internal *N*-acetylhexosamines (N) an *N*-acetylglucosamine except for the N2S1b, confirmed to be core 5 by comparison with bovine submaxillary mucin glycans standard. All terminal *N*-acetylhexosamines attached to a galactose also substituted with a *N*-acetylneuraminic acid (S) are assumed to be *N*-acetylgalactosamine as part of the Sda/Cad antigen. All terminal *N*-acetylhexosamines attached to a galactose also substituted with a fucose are assumed to be *N*-acetylgalactosamine as part of blood group antigen A. Sulphate modification position is not determined due to partial sulphate migration, and no presence of diagnostic fragment ions for the position on the Gal or GlcNAc residue. Therefore, its position in the glycan structures can be either on the Gal or the neighboring GlcNAc residue. Structures of the selected TACAs (**Fig. 1**) were confirmed by synthesized standards (J. Weber e*t al., manuscript in preparation,* 2022).

STATISTICS

Data analysis and visualization were performed in ‘’R’’ software using the following packages: tidyverse, readxl, caret, gridExtra, EnhancedVolcano, ggpubr, pcaMethods, Rcpm, ggrepel, tidyHeatmap, stringr and ComplexUpset. An imputation of the minimum positive value (0.0001) was performed to enable use principal component analysis (PCA). Variables with near zero variance were excluded before computing the PCA model using nearZeroVar function from caret package. Differences between groups were tested using Wilcoxon-Mann-Whitney non-parametric statistical test. P-values were adjusted for multiple testing using the “Benjamini-Hochberg” method.

TRANSCRIPTOMIC DATA

TCGA expression data was downloaded via the firebrowse.org website. The laser capture microdissected data GEO: GSE21815 was downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21815.

RESULTS AND DISCUSSION

A HIGH THROUGHPUT WORKFLOW FOR GLYCAN ANALYSIS FROM LASER CAPTURE MICRODISSECTED COLON TISSUES

First a high throughput workflow was established which enables the characterization of *O*-glycans from LCM FFPE tissue slides (**Supplementary Fig. S1**). This workflow is based on previously established protocols for glycan analysis from complex biological samples[22,28] as well as a formerly developed glycomic workflows using FFPE tissue sections[19]. Briefly, specific regions were extracted of the FFPE tissue sections by LCM, such as normal colon mucosa (**Supplementary Fig. S2a** and **b**), cancer epithelia (**Supplementary Fig. S2c** and **d)** and cancer stroma. Tissue lysates from approximately 2 x 10^4 up to 2.5 x 10^4 cells were used for protein immobilization on PVDF membrane plate wells. The *N*-glycans were released prior to *O*-glycan release to avoid interference. *O*-glycans were released from the immobilized proteins, purified and analyzed by PGC-LC-MS/MS. To ensure protein solubilization from formalin fixed material and the binding of the solubilized proteins to the PVDF membrane, the amount of detergent used for the lysis was decreased to 1%, compared to the 4% used in the original FFPE tissue glycomics approach[19]. This way we avoid the tedious chloroform/methanol protein precipitation step, as this gave unreproducible protein yields for small sample amounts. Regarding *O*-glycan detection, we used isopropanol enriched nitrogen gas to increase sensitivity and the quality of the MS/MS spectra, as described previously[22]. As depicted in **Supplementary Fig. S3** the method showed very good precision.

GLYCOSYLATION SIGNATURES IN COLORECTAL CANCER

The presence of TACAs was investigated using paired primary tumors from 12 CRC patients (T1–T12) and normal colon from the same patients (C1-C12). Additionally, six metastatic CRCs from liver metastases of six different patients (M15-M21) were analyzed. Clinical data related to the samples used in our cohort is available in **Supplementary Table S1** and **S2**. From the 12 cancer samples, eight were

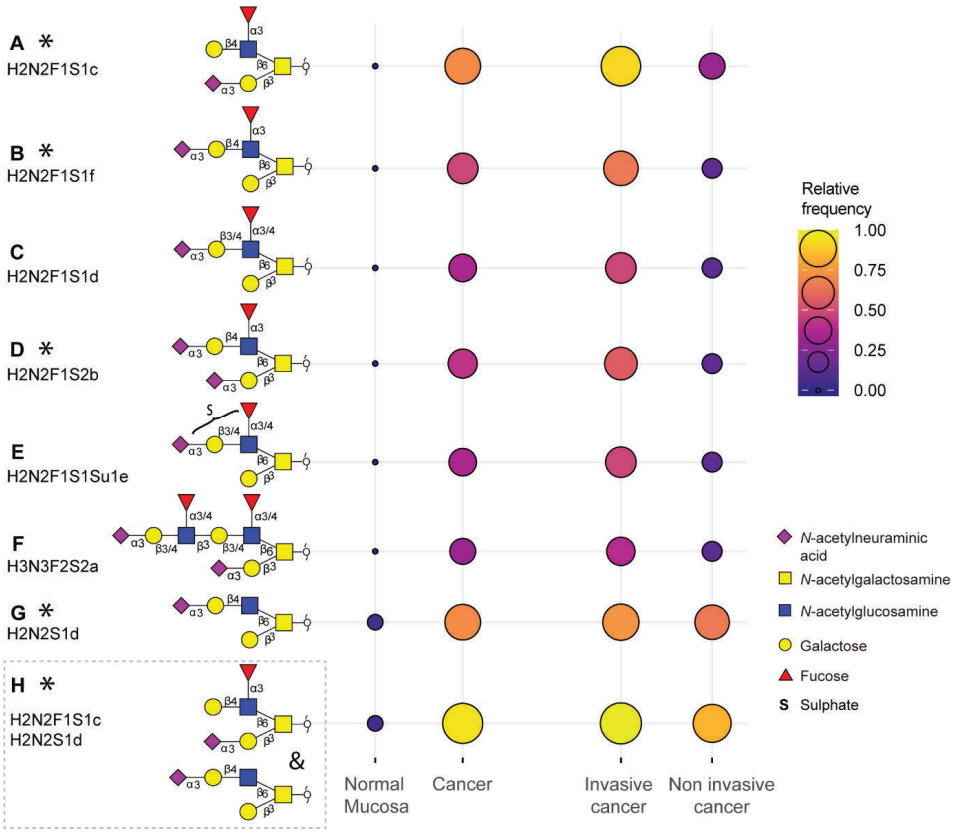


Fig. 1. Tumor associated carbohydrate antigens identified in colorectal cancer. *A subset of the O-glycans overexpressed in cancer which is detected in more than 33% of the cancers and undetectable in normal colon mucosa samples* ***(a, b, c, d, e, f)*** *or detected in maximum one normal mucosa* ***(g)****. Either one of the two TACAs in* ***(h)*** *is detected in 94% of the cancers. The selected TACAs show higher specificity for invasive cancer (Dukes stages C and D, n=12) than non-invasive cancer (Dukes stages A and B, n = 6) (right panel). Structures confirmed by synthesized standards (J. Weber et al., manuscript in preparation, 2022) are labelled with an asterisk. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

adenocarcinoma (AC), one was neuroendocrine carcinoma (NEC), and two were mucinous adenocarcinoma (MUC) of which one was partly signet-ring cell adenocarcinoma (T2). During the microdissection of the epithelial tumor regions and mucosa regions from normal colon it was ensured that the regions were well separated from the surrounding submucosal and muscle layers of the colon and tumor stroma. In addition, the stroma regions for a subset of stroma high cancers were also collected. *O*-glycosylation profiles were obtained for those regions separately.

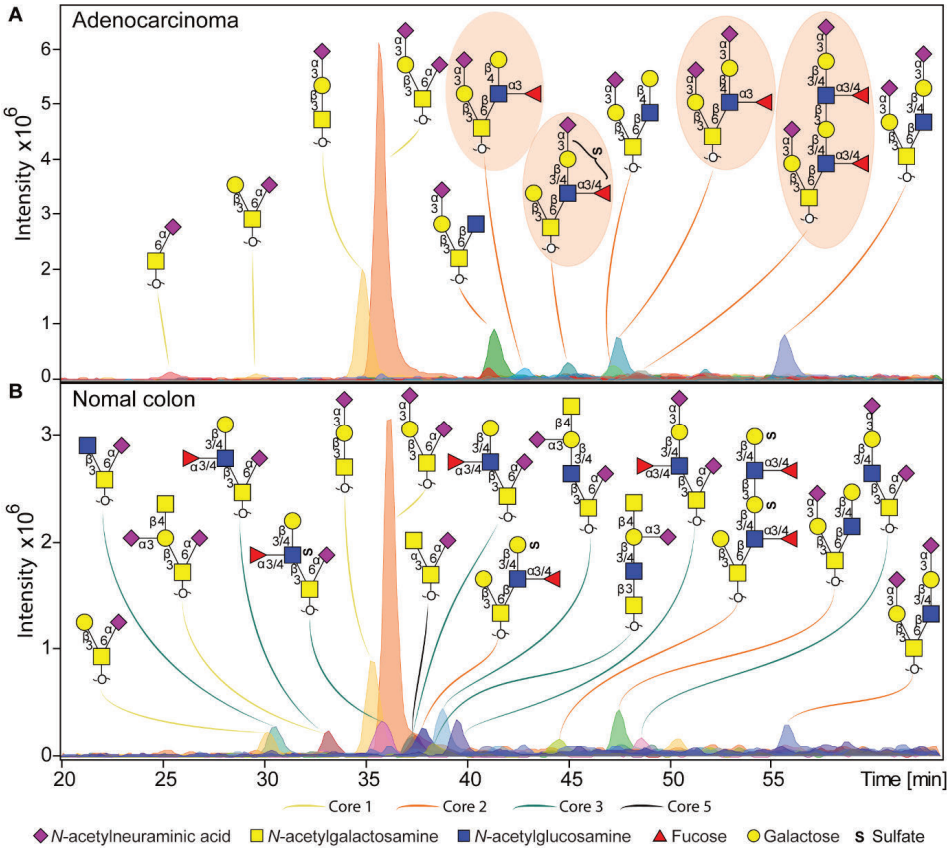


Fig. 2. Example of O-glycan chromatographic profile from (a) adenocarcinoma and (b) normal colon mucosa from the same patient (T8 vs C8). a) *The adenocarcinoma from patient 8 is characterized by specific expression of core 2 glycans, carrying terminal Le^{X} and sialyl-Lewis$^{X/A}$ antigens, or just terminal α2-3 sialylation linked to the galactose. Presence of TACAs (Fig. 1) are circled with orange background.* ***b)*** *Normal colon mucosa from the same patient shows expression of a diversity of core 3 structures, in most cases carrying an α2-6 linked sialic acid linked to the core GalNAc. Core 3 glycans are carrying terminal Sda antigens, as well as Le$^{X/A}$ and sulpho-Le$^{X/A}$ epitopes. The TACAs observed in adenocarinoma are not detected in the normal colon mucosa from the same patient. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

Overall, a total of 172 *O*-glycans were detected in the analyzed tissues (**Supplementary Table S3** and **S4**). From those, 108 were observed in the cancer and stroma, with no expression in normal colon mucosa (**Supplementary Table S5**). From those, 100 *O*-glycans were solely detected in cancer (TACAs), whereas only 20 *O*-glycans were shared between cancer and normal colon mucosa (**Supplementary**

Fig. S5). The majority of the TACAs have a core 2 structure carrying sialyl-Lewis$^{X/A}$ (sLe$^{X/A}$) antigens or an α2-3-linked sialic acid attached to the galactose (Gal) (**Supplementary Fig. S4, S5, S6b** and **S7**). Seven core 2 *O*-glycans that were not detected in normal mucosa, but exclusively found in more than six cancers (relative frequency > 33%) are depicted in **Fig. 1a-g** and listed in **Supplementary Table S6** and **S7**. MS/MS spectra of the selected TACAs are shown in **Supplementary Fig. S8.** Additional confirmation was found by performing standard addition experiments with available synthesized standards as exemplified in **Supplementary Fig. S9.** Sialylated core 2 *O*-glycan with terminal Le$^{X/A}$ antigen (**Fig. 1a**) was absent from normal mucosa and found in 72% of the cancers. One core 2 glycan with terminal α2-3 sialylation of Gal on the 6 arm is found in one normal mucosa sample, and in 72% of the cancers (**Fig. 1g**). In particular, at least one of the two TACAs depicted in **Fig. 1h** were detected in 94% of the cancers in our study. Interestingly, the only sample in which those glycans were not detected is a poorly differentiated adenocarcinoma, specifically classified as solid type, without glandular formation, usually associated with better patient prognosis[29]. An example of TACA expression is illustrated in **Fig. 2a**, which shows the expression of four specific TACAs (labeled with an orange background) in the adenocarcinoma tissue from patient 8, while those *O*-glycans were not detected in the normal colon mucosa (**Fig. 1b**). This pattern of TACA expression in cancer regions but not in presumably unaffected epithelium of the same patient was consistently observed across the entire dataset. Therefore, our data provides strong evidence for existence of highly specific *O*-glycans which are not present in the normal colon mucosa tissue. Moreover, significantly different *O*-glycomic traits (summarized in **Supplementary Fig. S4**) were found between CRC and normal mucosa (**Fig. 3**; **Supplementary Table S8**, **S9** and **S10**). While core 2 glycans are overexpressed, core 3 *O*-glycans are downregulated in cancer (**Fig. 3a** and **f**). A different expression is also observed in regard to sialylation, where terminal α2-3 sialylation is overexpressed, and α2-6 sialylation is downregulated in cancer (**Fig. 3b** and **g**). A core 2 *O*-glycan with terminal α2-3 sialylation on Gal is detected in the normal colon mucosa, (**Fig. 3e**) but it is strongly overexpressed in cancer. Differences were found also in regard to antigen expression, namely terminal sLe$^{X/A}$ and terminal α2-3 sialylation were upregulated in cancer in the context of core 2 *O*-glycans (**Fig. 3c** and **d; Supplementary Fig. S7** and **S10a**). Although the presence of sLe$^{X/A}$ was associated with several epithelial cancers[30–33], no specific targets were previously

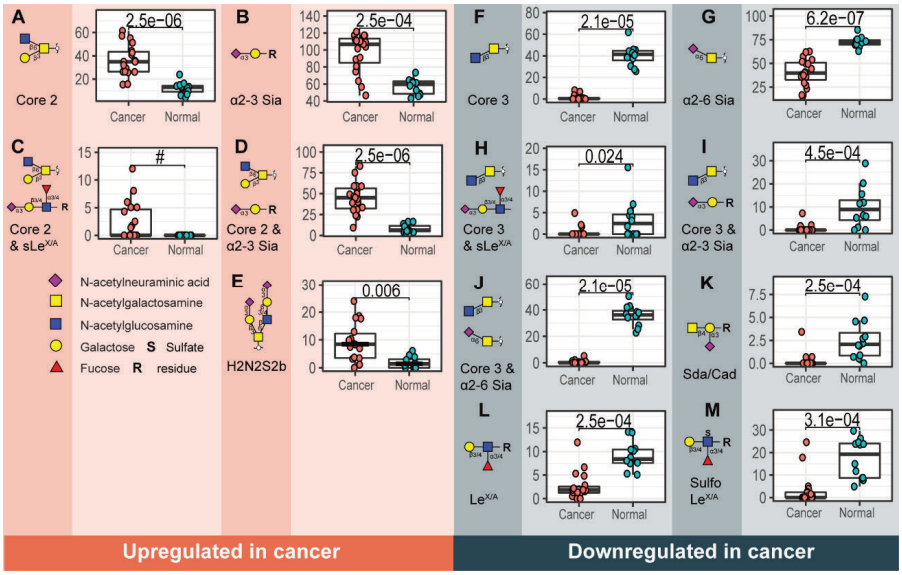


Fig. 3. Structural O-glycan features that differentiate between cancer and normal colon mucosa. *Statistically significant upregulation in cancer is found for various O-glycan features;* ***a)*** *Core 2 O-glycans,* ***b)*** *α2-3 sialylation,* ***c)*** *core 2 O-glycans with sLe$^{X/A}$ and* ***d)*** *core 2 O-glycans with terminal α2-3 sialylation linked to a galactose together with* ***e)*** *one individual O-glycan with composition H2N2S2. Whereas,* ***f)*** *core 3 O-glycans,* ***g)*** *α2-6 sialylation of the core GalNAc,* ***h)*** *core 3 O-glycans with sLe$^{X/A}$ antigen,* ***i)*** *core 3 O-glycans with terminal α2-3 sialylation,* ***j)*** *core 3 O-glycans with terminal α2-6 sialylation,* ***k)*** *Sda antigen,,* ***l)*** *Le$^{X/A}$ and* ***m)*** *sulfo-Le$^{X/A}$ antigen show statistically significant downregulation in cancer. Differences between groups were tested using Wilcoxon-Mann-Whitney non-parametric statistical test. Correction for multiple testing was made using the Benjamini-Hochberg method.* [#] *no p-value is reported as there is no variance in the control group. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

identified. Notably, the sLe$^{X/A}$ antigens and terminal α2-3 sialylation were also carried by core 3 structures in normal colon mucosa and mucinous adenocarcinomas but showed a downregulation in cancer (**Fig. 3h** and **i; Supplementary Fig. S10b** and **Table S10**). *O*-glycans carrying this antigen but with a core 4 structure did not show a difference between cancer and normal mucosa (**Supplementary Fig. S10c** and **Table S10**). Based on our results we suggest that the specificity of the immunotherapeutic antibodies should move towards TACAs with specific (s)Le$^{X/A}$ epitopes carried by core 2 *O*-glycans, and not those that are present on core 3 or core 4. Interestingly, a previous study evaluated the influence of glycan cores for the specificity of the sLe^{X} antibodies, and revealed that commonly used KM93 antibody binds to sLe^{X} on core 2

O-glycans, and not on core 3 *O*-glycans[34]. This reveals the potential for development of specific antibodies targeting a single core-type *O*-glycan in combination with cancer associated $sLe^{X/A}$ epitopes.

GLYCOSYLATION SIGNATURES IN NORMAL COLON MUCOSA

In the normal colon mucosa nearly 80 *O*-glycans were detected and 26 of them were specific for the mucosa, belonging mainly to *O*-glycans with a core 3, carrying an α2-6 sialylation on the core GalNAc, terminal Sda antigen ([GalNAcβ1-4(Neu5Acα2-3)Gal-R) and (sulfo)-$Le^{X/A}$ antigens, and showing statistically significant downregulation in cancer (**Fig. f-m)**. A total of 12 *O*-glycans were shared between cancer, mucosa and microenvironment controls, which were mainly sialylated core 1 *O*-glycans, such as ubiquitous sialyl-3T, and disialyl-T (**Supplementary Fig. S7** and **S12**) and core 2 types with terminal α2-3 sialylation on Gal (**Fig. 3e; Supplementary Fig. S5** and **S7**). Previous studies on the glycosylation of secreted mucus from normal colon mucosa reported expression of very similar glycosylation features, namely a high expression of core 3 structures carrying (sulfo)-$Le^{X/A}$ and Sda antigens. Yet the abundance of sialylated core 1 *O*-glycans was lower than observed here[35,36]. Potentially the glycosylation of the secreted mucins could differ from the cell glycosylation, however, those signatures could also partly originate from the immune cells infiltrated into the healthy colon mucosa, which could not be fully excised using LCM.

GLYCOSYLATION OF SPECIFIC CANCER TYPES

The observed glycomic signatures were further explored by taking into account various cancer types (**Supplementary Fig. S11** and **S12**). No clear clustering was found of cancers with the same differentiation grade, tumor stroma ratio and location origin in the colon (**Supplementary Fig. S13** and **S14**). The tumors with high immune infiltration did not show a specific glycan signature (**Supplementary Fig. S14b**). Moreover, MSI tumors, known for high immune infiltration, did not cluster together on the PCA plot (**Supplementary Fig. S13a**). Sialyl-Tn antigen was detected in both cancer and healthy mucosa (**Supplementary Fig. S12**), however, upregulated in cancer (no significance). It showed the highest expression in T11, a poorly differentiated stage 4 MSI tumor, the two mucinous MSI adenocarcinomas T2 and T3, and neuroendocrine carcinoma T5. Previous reports found that the expression is higher in poorly differentiated and mucinous colon adenocarcinomas, but no

associations with MSI were investigated[37]. It is yet to be determined whether MSI status is associated with expression of specific glycosyltransferases leading to higher expression of α2-6 sialylated Tn and T antigens.

Mucinous adenocarcinomas (T2 and T3) expressed 43 specific *O*-glycans (**Supplementary Fig. S11**), mostly carrying $sLe^{X/A}$ antigen, on core 2 but also core 4 *O*-glycans (**Supplementary Fig. S11** and **S12**). Both tumors were microsatellite instable (MSI) which are known to behave as low grade cancers[38]. Core 4 *O*-glycan expression can be explained by a strong downregulation of *ST6GALNAC1* and *3* in cancer, leading to the branching of the core 3 precursors in cancers with expression of core 4 synthase *GCNT3*[39]. Neuroendocrine carcinoma (T5) showed specifically high expression of Sd^a antigen, compared to other types with low or no expression (**Supplementary Fig. S12**).

Cancers with lymph node invasion (LNI) or invasion to distant organs, including Dukes stage C and D, showed a higher expression of the selected TACAs on core 2 *O*-glycans compared to cancers without invasion (Dukes stage B) (**Fig. 1; Supplementary Table S6**). Due to their absence or limited expression in the normal colon mucosa, and high specificity for invasive cancer, those TACAs can serve as promising targets for treatment of invasive CRC (**Fig. 2**).

PATHWAYS OF *O*-GLYCAN BIOSYNTHESIS IN COLORECTAL CANCER

We created a biosynthetic model (**Fig. 4**) explaining the glycosylation changes that occur in cancer compared to normal colon mucosa by integrating the glycomics results with the online available transcriptomics[40] (**Supplementary Fig. S15**). Downregulation of *B3GNT6*, core 3 synthase that adds β1-3-linked GlcNAc to the core GalNAc, leads to downregulation of core 3 structures, whereas upregulation of *C1GALT1* leads to upregulation of core 1 structures in cancer. Previously, expression of core 3 synthase was associated with better patient prognosis in pancreatic[41], prostate cancer[42] and colon cancer[43]. This competition between core 1 and core 3 synthase as well as core α2-6 sialylation in the context of sialyl-Tn has been described previously in colon cancer cells[44]. We did not observe statistically significant upregulation in the expression of sialyl-Tn antigen in cancer although the trend was present. The main enzyme responsible for the biosynthesis of sialyl-Tn antigen *ST6GALNAC1* is downregulated in cancer, however, a higher expression of sialyl-Tn antigens can be

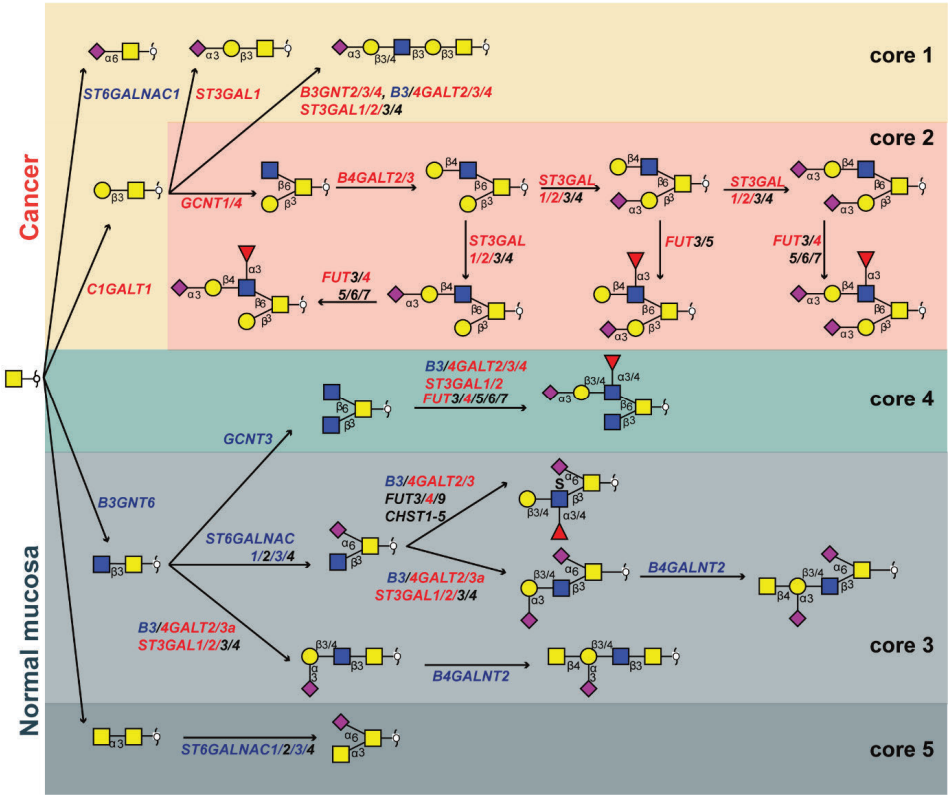


Fig. 4. Proposed biosynthetic model explaining the differences in glycosylation of colorectal cancer and normal colon mucosa. *The most abundant structures in cancer (upper red and yellow panel) and normal colon mucosa (lower green and gray panels) are depicted together with genes encoding for the GTs involved in their biosynthesis. The biosynthetic pathways of different core structures are labelled with different colors. Pathways upregulated in cancer are marked in orange and yellow, whereas pathways downregulated in cancer are marked in light and dark gray. Core 4 O-glycans show no statistically significant difference between normal mucosa and cancer, marked in green. Glycosyltransferase genes upregulated in cancer are labelled with red and those downregulated in cancer are displayed in blue. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

due to a mutation in *C1GALT1C1* gene encoding for a chaperone (Cosmc) required for the activity of core 1 synthase which could lead to a blockage of the alternative pathway[45]. On the other hand, upregulation of *GCNT1* and *GCNT4* (both core 2 synthases that add the GlcNAc to the core GalNAc in the β1-6 position, creating the 6-arm), leads to overexpression of core 2 structures. The upregulation of *GCNT1* and *4* is seen in the microdissected CRC tumors, but not in the TCGA dataset

(**Supplementary Fig. S15**). Addition of a galactose residue to the 6-arm by activity of β1-4-galactosyltransferases is upregulated in cancer [46,47], leading to the biosynthesis of the pentasaccharide with α2-3-linked sialic acid (**Fig. 1g; H2N2S1d**), a TACA with high specificity for cancer and a precursor for other highly specific TACAs (**Fig. 1b-e**). Additionally, the addition of another LacNAc creates a sialyl-dimeric $Le^{X/A}$ antigen glycan which is also specifically found in cancer (**Fig. 1f**). The fucosyltransferase *FUT4* involved in the biosynthesis of sLe^X antigen is overexpressed in colon cancer (**Supplementary Fig. S15**) and it was previously associated with lung adenocarcinoma metastasis and poor patient prognosis[48,49]. It has also been previously described that the upregulation of $sLe^{X/A}$ antigens in CRC is due to downregulation of *B4GALNT2* which adds β1-4-linked GalNAc creating the Sda antigen in the normal colon mucosa[50,51]. This enzyme shows a downregulation in The Cancer Genome Atlas (TCGA) dataset, and only a trend could be observed towards the same direction in LCM dataset (**Supplementary Fig. S15**). Additionally, the expression of 6-sulfo-Le^X and sialyl-6-sulfo-Le^X were previously associated with normal colon epithelia which is in contrast to the expression of sLe^X in CRC[13]. Additionally, a lower expression of $Lewis^A$ type antigens as well as sLe^A was found in CRC compared to normal colon mucosa due to a downregulation of β1-3-galactosyltransferase B3GALT5[52]. The elongation of core 1 structures is in competition with the core 2 biosynthetic pathway and the biosynthesis of sialyl-3T antigen (H1N1S1b). While this antigen is also detected in normal mucosa it shows an overexpression in cancer *(***Supplementary Fig. S12**) due to an upregulation of *ST3GAL1 (***Supplementary Fig. S15)** which is previously associated with lymph node invasion in CRC[53]. Finally, we detected core 5 structure H1N2S1b in the normal colon mucosa, and this has been described before[36]. However, with our PGC-LC-MS/MS approach we could confidently distinguish only one core 5 *O*-glycan (N2S1b) from isomeric core 3 glycans as this glycan was previously characterized by NMR from bovine submaxillary mucin[27]. It is unknown whether any of the remaining core 3 isomers originate from core 5.

In summary, our findings show that the downregulation of core 3 synthase (B3GNT6) in cancer leads to the overexpression of core 1 truncated *O*-glycans such as sialyl3-T by action of *ST3GAL1*, or branched *O*-glycans by the action of core 2 synthases *GCNT1* and *GCNT4.* The addition of a β1-4 galactose and α2-3 sialic acid on the 6

arm of the *O*-glycans forms a specific pathway in cancer, starting from a sialylated precursor (**Fig. 1g**) leading to biosynthesis of fucosylated TACAs carrying terminal sLe^X antigen by the addition of a α1-3-linked fucose by *FUT4* overexpressed in cancer.

FUTURE PERSPECTIVES

Analysis of specific molecular signatures from patient derived material is fundamental for understanding molecular mechanisms of disease onset and progression. FFPE tissue sectioning is part of standard care in pathology and many tissues are widely available and well preserved at room temperature for a long period of time. Yet, tissue heterogeneity poses a problem for glycomics and transcriptomics analysis masking cell specific signatures. Dissecting specific regions of the tissue by LCM, largely overcomes this issue and enables analysis of glycosylation signatures from specific cells of interest[54]. However, it is essential that further improvements are made in regard to sensitivity of our current MS methods, making single cell glycomics analysis possible to avoid any contamination from different cell populations.

It is of great importance to determine the specificity of different glycosyltransferases for core structures, different arms of branched glycans as well as expression of TACAs on different glycan classes. Targeting glycans may have several benefits compared to proteins. Namely, the TACAs are expressed on the cell surface, are directly accessible to therapeutics and can be carried by multiple proteins, reflecting the overall glycosylation phenotype of the cell, providing a broader tumor targeting strategy[55]. Ideally, the expression of TACAs is absent or limited in normal colon mucosa, however, some of the TACAs have been reported as important for the leukocyte extravasation through the endothelium, albeit only at very low relative abundance[56,57] serving as selectin ligands on P-selectin glycoprotein ligand-1 (PSGL-1). Cancer cells carrying the same ligands may employ this physiological mechanism for successful metastasis to distant organs which is why it is important to target these antigens, particularly in invasive carcinoma[30].

While there is no doubt that the differences in the TACAs patterns between the cancer and healthy mucosa are a consequence of changes in the glycosylation machinery of the cells, it remains unclear whether the TACAs are protein specific, and if the differences in their expression is related to mucin expression. With the aim to increase immunogenicity and specificity of glycan targets, they are often coupled to protein

carriers, either adjuvants or pathological carriers such as mucins. Aberrant expression of Tn antigen was found on MUC1 membrane glycoprotein, and anti-Tn and T antigen carrying MUC1 expressed on showed efficiency in phase I clinical trial on ovarian, breast and cervical cancer[58] but no effect on patient outcome in phase II clinical trial[59]. Another monoclonal antibody targeting Tn-antigen on MUC1 showed promising preclinical results for targeting breast cancer, however, the results need further validation by clinical trials[60]. Additionally, genetically modified T-cells expressing chimeric antigen receptors (CAR-s) targeting Tn glycoforms of MUC1 showed promising cytotoxicity in xenograft models of T-cell leukemia and pancreatic cancer[61]. Our study showed that core 2 sialylated or core 2 $sLe^{X/A}$ carrying *O*-glycans have the highest specificity for CRC, however, more research is needed to determine if these signatures are mucin specific, and whether bispecific antibodies or CAR-T cells targeting both the protein carrier and the cancer specific glycans will increase the specificity and immunogenicity of the developed therapeutics.

CONCLUSION

In this study we present a novel panel of highly specific TACAs, based upon changes in the *O*-glycomic profile between CRC and healthy colon mucosa. These TACAs are promising new targets for development of innovative cancer immunotherapeutics and lay the foundation for the treatment of invasive CRC.

SUPPORTING INFORMATION

Supporting information is available upon request.

DATA AVAILABILITY

The data in support of the findings of this study may be found within the manuscript and in the associated supplementary files. Mass spectrometry-based glycomics raw data were deposited in the Glycopost repository accessible via the following link: https://glycopost.glycosmos.org/preview/460809305620fc81b8a4e5 (pin code: 6390). Detailed MS/MS spectra peaklists with annotations are available in the Glycoworkbench files.

ACKNOWLEDGMENTS

The authors thank Meaghan Polack for the tumor stroma ratio assessment and Stijn A.S.L.P. Crobach for a discussion about cancer slides included in this study. Additionally, we thank Carolien A.M. Koeleman, Lisa A. de Neef and Agnes L. Hipgrave Ederveen for technical assistance. This work was supported by the European Commission's Horizon 2020 programme ''GlyCoCan'' project, grant number 676421. The results shown in Supplementary Fig. S15 are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

AUTHOR INFORMATION

CONTRIBUTIONS

K.M. performed the experiments. K.M., G.L., M.W. conceptually designed the work. K.M., G.L., M.W. wrote the manuscript. R.V. assisted with laser capture microdissections. T.Z assisted with glycan MS/MS spectra annotation. T.W. assisted with study design and access to patient material. O.M. assisted with statistical analysis. H.M. annotated HE slides. J.W. and G.B. prepared glycan standards. All authors read and commented on the manuscript.

CORRESPONDING AUTHOR

Manfred Wuhrer – Center for Proteomics and Metabolomics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; orcid.org/0000-0002-0814-4995

ETHICS DECLARATION

The authors declare no competing interests.

REFERENCES

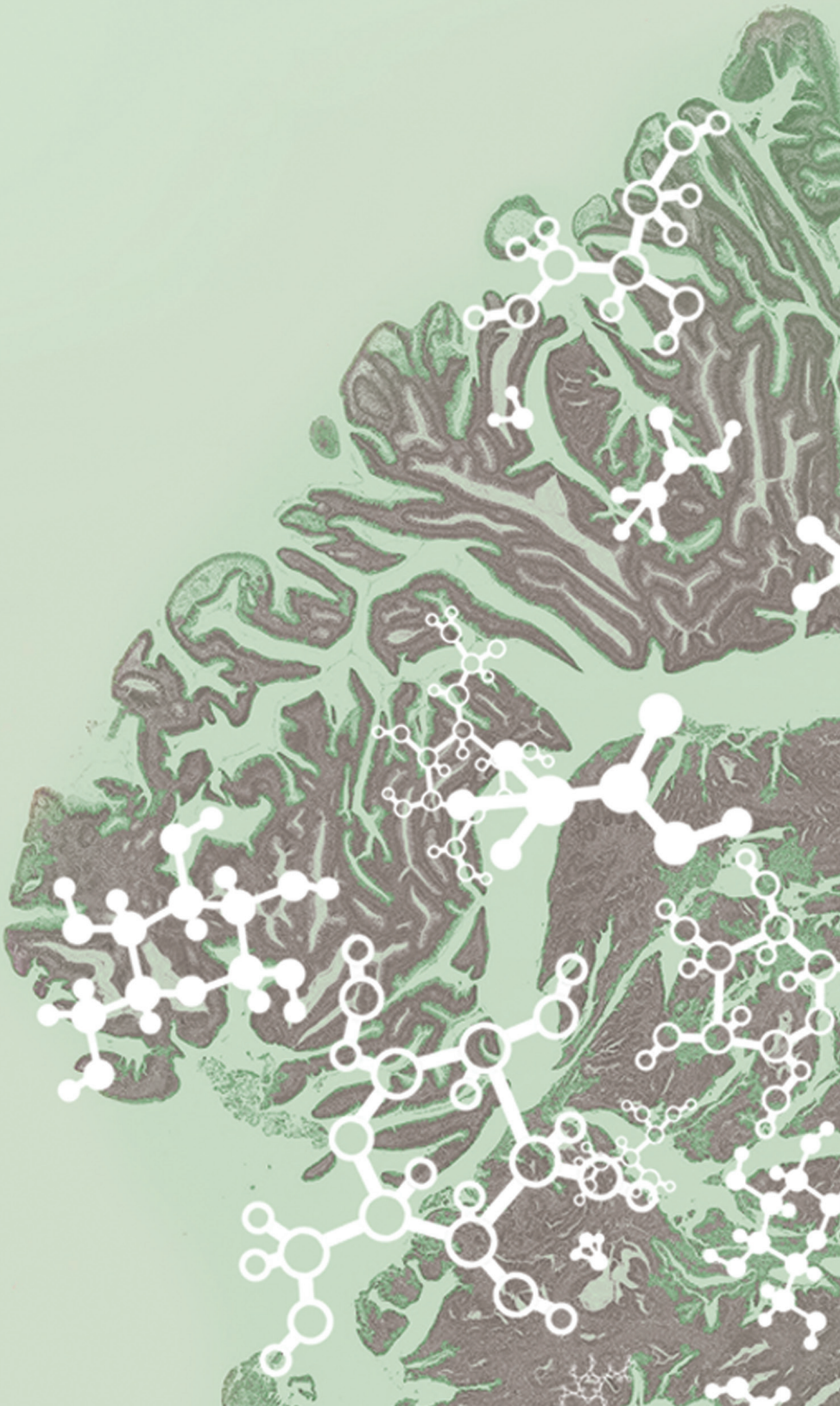
1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**, E359–E386 (2015).
2. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
3. Rosenblum, D., Joshi, N., Tao, W., Karp, J. M. & Peer, D. Progress and challenges towards targeted delivery of cancer therapeutics. *Nature Communications* vol. 9 1–12 (2018).
4. Piawah, S. & Venook, A. P. Targeted therapy for colorectal cancer metastases: A review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer. *Cancer* **125**, 4139–4147 (2019).

5. Ayyar, B. V., Arora, S. & O'Kennedy, R. Coming-of-Age of Antibodies in Cancer Therapeutics. *Trends Pharmacol. Sci.* **37**, 1009–1028 (2016).
6. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
7. Holst, S., Wuhrer, M. & Rombouts, Y. *Glycosylation characteristics of colorectal cancer*. vol. 126 203–256 (Elsevier Inc., 2015).
8. Kannagi, R., Yin, J., Miyazaki, K. & Izawa, M. Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alteration of carbohydrate determinants-Hakomori's concepts revisited. *Biochimica et Biophysica Acta - General Subjects* **1780**, 525–531 (2008).
9. Chen, W.-S., Chang, H.-Y., Li, C.-P., Liu, J. M. & Huang, T.-S. Tumor beta-1,4-galactosyltransferase IV overexpression is closely associated with colorectal cancer metastasis and poor prognosis. *Clin. Cancer Res.* **11**, 8615–8622 (2005).
10. Krishn, S. R. *et al.* Mucins and associated glycan signatures in colon adenoma-carcinoma sequence: Prospective pathological implication(s) for early diagnosis of colon cancer. **374**, 304–314 (2017).
11. Costa, A. F., Campos, D., Reis, C. A. & Gomes, C. Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends in Cancer* vol. 6 757–766 (2020).
12. Jass, J. R., Allison, L. J. & Edgar, S. G. Distribution of sialosyl Tn and Tn antigens within normal and malignant colorectal epithelium. *J. Pathol.* **176**, 143–149 (1995).
13. Izawa, M. *et al.* Expression of sialyl 6-sulfo Lewis X is inversely correlated with conventional sialyl Lewis X expression in human colorectal cancer. *Cancer Res.* **60**, 1410–1416 (2000).
14. Itzkowitz, S. H. *et al.* Expression of Tn, Sialosyl-Tn, and T Antigens in Human Colon Cancer. *Cancer Res.* **49**, (1989).
15. Fukasawa, T. *et al.* Associated expression of α2,3sialylated type 2 chain structures with lymph node metastasis in distal colorectal cancer. *Surg. Today* **43**, 155–162 (2013).
16. Ito, H. *et al.* Altered mrna expression of specific molecular species of fucosyl-and sialyl-transferases in human colorectal cancer tissues. vol. 71 556–564 1097-0215(19970516)71:4 (1997).
17. Kudo, T. *et al.* Up-regulation of a set of glycosyltransferase genes in human colorectal cancer. *Lab. Invest.* **78**, 797–811 (1998).
18. Shimodaira, K. *et al.* Carcinoma-associated Expression of Core 2 β-1,6-N-Acetylglucosaminyltransferase Gene in Human Colorectal Cancer: Role of O-Glycans in Tumor Progression. *Cancer Res.* **57**, (1997).
19. Hinneburg, H. *et al.* Unlocking Cancer Glycomes from Histopathological Formalin-fixed and Paraffin-embedded (FFPE) Tissue Microdissections. *Mol. Cell. Proteomics* **16**, 524–536 (2017).
20. Hinneburg, H., Schirmeister, F., Korać, P. & Kolarich, D. N- and O-Glycomics from Minor Amounts of Formalin-Fixed, Paraffin-Embedded Tissue Samples. in 131–145 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-6493-2_11.

21. Guo, Y. W. *et al.* Effect of PVP Hydrophilic Additive on the Morphology and Properties of PVDF Porous Membranes. *Adv. Mat. Res.* **981**, 891–894 (2014).
22. Madunić, K., Wagt, S., Zhang, T., Wuhrer, M. & Lageveen-Kammeijer, G. S. M. Dopant-Enriched Nitrogen Gas for Enhanced Electrospray Ionization of Released Glycans in Negative Ion Mode. *Anal. Chem.* acs.analchem.1c00023 (2021) doi:10.1021/acs.analchem.1c00023.
23. Robbe, C. *et al.* Evidence of regio-specific glycosylation in human intestinal mucins: Presence of an acidic gradient along the intestinal tract. *J. Biol. Chem.* **278**, 46337–46348 (2003).
24. Holmén Larsson, J. M., Karlsson, H., Sjövall, H. & Hansson, G. C. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756–766 (2009).
25. Savage, A. V., Donohue, J. J., Koeleman, C. A. M. & van den Eijnden, D. H. Structural characterization of sialylated tetrasaccharides and pentasaccharides with blood group H and Lex activity isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **193**, 837–843 (1990).
26. Savage, A. V., D'Arcy, S. M. T. & Donoghue, C. M. Structural characterization of neutral oligosaccharides with blood group A and H activity isolated from bovine submaxillary mucin. *Biochem. J* **279**, 95–103 (1991).
27. Savage, A. V., Donoghue, C. M., D'Arcy, S. M., Koeleman, C. A. M. & van den Eijnden, D. H. Structure determination of five sialylated trisaccharides with core types 1, 3 or 5 isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **192**, 427–432 (1990).
28. 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* **16**, 355–363 (2020).
29. Kazama, Y. *et al.* Microsatellite instability in poorly differentiated adenocarcinomas of the colon and rectum: relationship to clinicopathological features. *J. Clin. Pathol.* **60**, 701–704 (2007).
30. Trinchera, M. *et al.* Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. *Biology* **6**, 16 (2017).
31. Burdick, M. D., Harris, A., Reid, C. J., Iwamura, T. & Hollingsworth, M. A. Oligosaccharides expressed on MUC1 produced by pancreatic and colon tumor cell lines. *J. Biol. Chem.* **272**, 24198–24202 (1997).
32. Gomes, C. *et al.* Carcinoembryonic antigen carrying SLe X as a new biomarker of more aggressive gastric carcinomas. *Theranostics* **9**, 24 (2019).
33. Trinchera, M. *et al.* The biosynthesis of the selectin-ligand sialyl Lewis x in colorectal cancer tissues is regulated by fucosyltransferase VI and can be inhibited by an RNA interference-based approach. *Int. J. Biochem. Cell Biol.* **43**, 130–139 (2011).
34. Löfling, J. & Holgersson, J. Core saccharide dependence of sialyl Lewis X biosynthesis. *Glycoconj. J.* **26**, 33–40 (2009).

35. Mihalache, A. *et al.* Structural Characterization of Mucin O-Glycosylation May Provide Important Information to Help Prevent Colorectal Tumor Recurrence. *Front. Oncol.* **5**, 217 (2015).
36. Robbe C. Capon C. Coddeville B. & Michalski, J. C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J* **384**, 307–316 (2004).
37. Grosso, M., Vitarelli, E., Giuffrè, G., Tuccari, G. & Barresi, G. Expression of Tn, sialosyl-Tn and T antigens in human foetal large intestine. *Eur. J. Histochem.* **44**, 359–363 (2000).
38. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *Journal of Gastrointestinal Oncology* vol. 3 153–173 (2012).
39. Yeh, J. C., Ong, E. & Fukuda, M. Molecular cloning and expression of a novel β-1,6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I branches. *J. Biol. Chem.* **274**, 3215–3221 (1999).
40. Kogo, R. *et al.* Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* **71**, 6320–6326 (2011).
41. Doi, N. *et al.* Clinicopathological significance of core 3 O-glycan synthetic enzyme, β1,3-N-acetylglucosaminyltransferase 6 in pancreatic ductal adenocarcinoma. *PLoS One* **15**, e0242851 (2020).
42. Lee, S. H. *et al.* Core3 O-glycan synthase suppresses tumor formation and metastasis of prostate carcinoma PC3 and LNCaP cells through down-regulation of α2β1 integrin complex. *J. Biol. Chem.* **284**, 17157–17169 (2009).
43. Iwai, T. *et al.* Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4572–4577 (2005).
44. Barrow, H., Tam, B., Duckworth, C. A., Rhodes, J. M. & Yu, L. G. Suppression of Core 1 Gal-Transferase Is Associated with Reduction of TF and Reciprocal Increase of Tn, sialyl-Tn and Core 3 Glycans in Human Colon Cancer Cells. *PLoS One* **8**, e59792 (2013).
45. Ju, T. *et al.* Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res.* **68**, 1636–1646 (2008).
46. Isshiki, S. *et al.* Cloning, Expression, and Characterization of a Novel UDP-galactose:β-N-Acetylglucosamine β1,3-Galactosyltransferase (β3Gal-T5) Responsible for Synthesis of Type 1 Chain in Colorectal and Pancreatic Epithelia and Tumor Cells Derived Therefrom *. *J. Biol. Chem.* **274**, 12499–12507 (1999).
47. Salvini, R., Bardoni, A., Valli, M. & Trinchera, M. β1,3-Galactosyltransferase β3Gal-T5 Acts on the GlcNAcβ1→3Galβ1→4GlcNAcβ1→R Sugar Chains of Carcinoembryonic Antigen and Other N-Linked Glycoproteins and Is Down-regulated in Colon Adenocarcinomas *. *J. Biol. Chem.* **276**, 3564–3573 (2001).
48. Lu, H. H. *et al.* Fucosyltransferase 4 shapes oncogenic glycoproteome to drive metastasis of lung adenocarcinoma. *EBioMedicine* **57**, 102846 (2020).

49. Park, S. *et al.* Altered expression of fucosylation pathway genes is associated with poor prognosis and tumor metastasis in non-small cell lung cancer. *Int. J. Oncol.* **56**, 559–567 (2020).
50. Groux-Degroote, S. *et al.* B4GALNT2 gene expression controls the biosynthesis of Sda and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. *Int. J. Biochem. Cell Biol.* **53**, 442–449 (2014).
51. Pucci, M., Ferreira, I. G., Orlandani, M. & Malagolini, N. High Expression of the Sd a Synthase B4GALNT2 Associates with Good Prognosis and Attenuates Stemness in Colon Cancer. **3**, (2020).
52. Miyazaki, K. *et al.* Loss of disialyl Lewisa, the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (Siglec-7) associated with increased sialyl Lewisa expression on human colon cancers. *Cancer Res.* **64**, 4498–4505 (2004).
53. Schneider, F. *et al.* Overexpression of sialyltransferase CMP-sialic acid:Galbeta1,3GalNAc-R alpha6-Sialyltransferase is related to poor patient survival in human colorectal carcinomas. *Cancer Res.* **61**, 4605–4611 (2001).
54. Lageveen-Kammeijer, G. S. M., Kuster, B., Reusch, D. & Wuhrer, M. High sensitivity glycomics in biomedicine. *Mass Spectrom. Rev.* e21730 (2021) doi:10.1002/mas.21730.
55. Houvast, R. D. *et al.* Targeting glycans and heavily glycosylated proteins for tumor imaging. *Cancers* vol. 12 1–26 (2020).
56. Lo, C. Y. *et al.* Competition between core-2 GlcNAc-transferase and ST6GalNAc-transferase regulates the synthesis of the leukocyte selectin ligand on human P-selectin glycoprotein ligand-1. *J. Biol. Chem.* **288**, 13974–13987 (2013).
57. Wilkins, P. P., McEver, R. P. & Cummings, R. D. Structures of the O-glycans on P-selectin glycoprotein ligand-1 from HL- 60 cells. *J. Biol. Chem.* **271**, 18732–18742 (1996).
58. Fiedler, W. *et al.* A phase i study of PankoMab-GEX, a humanised glyco-optimised monoclonal antibody to a novel tumour-specific MUC1 glycopeptide epitope in patients with advanced carcinomas. *Eur. J. Cancer* **63**, 55–63 (2016).
59. Ledermann, J. *et al.* A double-blind, placebo-controlled, randomized, phase 2 study to evaluate the efficacy and safety of switch maintenance therapy with the anti-TA-MUC1 antibody PankoMab-GEX after chemotherapy in patients with recurrent epithelial ovarian carcinoma. *Ann. Oncol.* **28**, v626 (2017).
60. Lavrsen, K. *et al.* Aberrantly glycosylated MUC1 is expressed on the surface of breast cancer cells and a target for antibody-dependent cell-mediated cytotoxicity. *Glycoconj. J.* **30**, 227–236 (2013).
61. Posey, A. D. *et al.* Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. *Immunity* 44, 1444–1454 (2016).



DISCUSSION AND PERSPECTIVES

Chapter 7

Despite the accomplishments presented in this thesis, several challenges and limitations were encountered in the research which will be further addressed in this chapter. In addition, it will be discussed which models are (most) suitable for studying glycosylation changes in cancer as well as the integration of multiple omics approaches for a better understanding of the mechanisms underlying cancer pathogenesis. Finally, current therapeutic strategies that utilize glycosylation as well as future perspectives will be discussed.

TECHNICAL CHALLENGES

O-GLYCAN RELEASE

A major bottleneck in *O*-glycan analysis is the lack of a broadly acting enzyme that would release all types of *O*-GalNAc cores from the proteins. Therefore, the release is largely performed chemically under reducing conditions. The latter results in glycan alditols which lack the reducing end for labelling with fluorescent tags that allows fluorescent detection in combination with separation mechanisms such as liquid chromatography (LC) or capillary electrophoresis. Recently, a microwave assisted non-reductive release with 2-AB labelling for *O*-glycans was developed, however, further improvements are needed to minimize peeling and enhance separation[2]. Other studies have attempted to make use of mucin specific endoglycosidases to facilitate mucin specific epitope analysis[3]. However, the limitation of this approach is the partial loss of valuable information such as core structures that are not amenable to enzymatic release. Despite the fact that discovery of a mucin-specific protease for glycoprotein analysis is a great improvement[4], it still creates glycopeptides that contain multiple glycans per peptide backbone. This makes the analysis of specific glycan signatures and obtaining information about site occupancy very challenging. Therefore, the quest for a novel broad acting *O*-glycosidase is still ongoing, and its identification will for sure change the field for studies in which a near comprehensive release is warranted. Additionally, this would also allow to detect acetylation of sialic acids, as this modification is currently lost due to the high pH β-elimination step that is required for the chemical release method. As *O*-acetylation is assumed to have important implications in the immunomodulatory interaction with Siglecs[5] the exploration of mucin glycan acetylation is still largely pending but is expected to be

scientifically highly rewarding. Moreover the availability of such an enzyme will facilitate region-specific *O*-glycan profiling directly on tissue using MALDI-MS imaging (MSI), as it was applied previously for the analysis of *N*-glycans in CRC[6].

GLYCAN ANALYSIS

Despite the many advantages of porous graphitized carbon nano-liquid chromatography coupled to mass spectrometry for the analysis of mucin type glycans, the stationary phase does not retain monosaccharides, and has limited retention of neutral disaccharides. This makes the analysis of well-established cancer associated Tn- and T-antigens very challenging if not impossible[7]. We attempted to increase the retention of the single HexNAc, by changing parameters in the sample trapping phase and sample cleanup with PGC-SPE, yet our attempts were unsuccessful. Thus, different methods for detecting the Tn antigen were used. First, we enabled the unambiguous detection of the permethylated HexNAc in the matrix region of MALDI-FT-ICR-MS spectra, by using adsorption mode data post-processing[8]. This was confirmed by internal standards. Next, we contributed to establishing a semi-automated, high throughput approach for permethylated *O*-glycosylation profiling of cell lines by MALDI-FT-ICR-MS[9]. We applied this approach on a set of 21 CRC cell lines, aiming to compare the expression of Tn antigen. However, no conclusive restuls were obtained about the origin of the detected HexNAc, as no specific internal fragment ions were found that would allow confident differentiation between the GalNAc or GlcNAc (unpublished data). Therefore, we established a RP-C18 separation platform to analyze permethylated *O*-glycans and achieved separation of *O*-GalNAc from *O*-GlcNAc, confirmed by commercially available standards. This approach provided a complimentary strategy to PGC-LC-MS/MS enabling detection of small glycan structures, such as the important cancer markers Tn and T antigen, and allowed for their separation from the *O*-GlcNAcylation that are known to be mainly present in cytosolic proteins[10]. Unfortunately, neither of these methods allowed the distinction between the *O*-GalNAc modification of intracellular proteins, intracellular mucin type glycan precursors and the plasma membrane expressed cancer associated Tn antigen (*O*-GalNAc). Since this type of analysis is not possible once the glycans are released from proteins, an enrichment strategy would be highly beneficial to exclude contributions from the intracellular glycan precursors in the profile. While previous studies used ultracentrifugation for enrichment of membrane fractions[11,12], it

remained unclear how much contamination is present from the other cell compartments. Therefore, different approaches were attempted in our group to enrich for plasma membrane proteins, however, the procedures were time-consuming, laborious and with low recovery making them unsuitable for subsequent glycomic analysis (unpublished data). Further studies are required to establish an efficient plasma membrane protein glycosylation profiling from complex biological samples. Additionally, to tackle this issue a glycoproteomic approach was developed to distinguish *O*-GalNAc and *O*-GlcNAc bearing glycopeptides based on the specific ratio of oxonium ions in MS/MS[13] relying on the literature-based preselection of plasma membrane associated and secreted proteins. This, however, has limitations as many proteins can be located in different cell compartments. Significant insights into the cell *O*-glycoproteome were achieved using the “Simple Cell” methodology developed by the Copenhagen Center for Glycomics. This technology allows for the discovery of *O*-linked glycosylation sites upon simplification of the cell *O*-glycome at the level of Tn and Sialyl-Tn antigen, enabling enrichment and bottom-up analysis of cell *O*-glycoproteome[14]. It significantly broadened the map of human *O*-glycoproteome identifying more than 600 glycoproteins and almost 3000 distinct sites from protein extracts of various cancer cell lines[14]. Moreover, an additional 649 glycoproteins were successfully identified from human blood including plasma proteins, endothelial cells and platelets[15]. These valuable studies gave us insights into over 10.000 *O*-GalNAc glycosylation sites, and the knowledge that up to 80 % of all proteins going through the secretory pathway are *O*-glycosylated. However, despite the advantages, this approach might not fully represent the *O*-glycosylation sites in nongenetically engineered cells, as polypeptide GalNAc transferases have lectin binding domains and the simplification of the cell glycome might cause a difference in their glycosylation capacity of specificity[16]. Additionally, further insights into the occupancy of the individual *O*-glycan sites are needed, as the current lectin enrichment approaches create a bias towards the occupied peptides, and cannot give a full picture about their relative occupancy[16].

In the past decades, there has been a rapid growth of new technologies measuring molecular signatures of single cells. While RNA sequencing already offers high throughput measurements of single cell transcriptomes, the proteomic characterization of single cells is a rapidly emerging field[17]. Moreover, the

development of mass cytometry (CyTOF, cytometry by time-of-flight) allowed simultaneous detection of around 50 proteins revealing cell surface molecular signatures on single cells[17,18]. Although glycomic analysis of single cells is still not possible, recent advances in *N*-glycan profiling by MSI offer reasonable resolution that could soon enable dissection of glycan signatures of only a few cells within a pixel[19]. Similarly, capillary electrophoresis electrospray ionization-mass spectrometry (CE-ESI-MS) approach offers excellent sensitivity for the analysis of labeled *N*-glycans from complex biological samples at the attomole level[20]. However, developments to enable highly sensitive detection of *O*-glycans are still lagging behind. In **Chapter 2** we describe the downscaling the cell amount needed for a robust analysis of both *N*- and *O*-glycans from cell lines following the previous published protocols for MALDI-TOF *N*-glycan analysis[21]. In **Chapter 3** we describe our research to increase sensitivity of our analysis by using different dopant solvents. We showed that alcohol-based dopant-enriched nitrogen (DEN) gas outperforms acetonitrile and approaches without dopants. In addition, isopropanol was found to increase the fragment spectra (MS/MS) intensity, while methanol provided the best results in terms of signal to noise ratios on the MS level. Additionally, the bias in ionization that was created by the low organic solvent content in the elution gradient was corrected for the early eluting species, significantly increasing the abundance of small *O*-glycans, important tumor associated antigens such as sialyl Tn antigen. Finally, in **Chapter 6** we successfully analyzed the glycome of approximately 2 x 10^4 tissue-derived cells, from different regions of the tumors and healthy mucosa. However, further developments in regard to sensitivity are needed to facilitate the analysis of single cells, thereby fully excluding the possible contamination coming from stroma and immune infiltrate, as well as analysis of specific cells in the invasive front of the tumor.

STRUCTURAL IDENTIFICATION OF GLYCANS

Due to a complex, non-template driven biosynthesis glycans can be presented in various isomers. MS alone is unable to provide complete information for *de-novo* full structural elucidation of the glycans. The same theoretically possible fragment ions can be present in different glycan molecules; therefore, current prediction of glycan MS/MS fragmentation is not reliable to allow for confident differentiation and identification of glycan species. Substantial improvements are made with ion mobility MS, which enables the separation of isomeric glycan species in the gas phase,

providing a good perspective for their identification by determining gas phase conformer distributions (CDs) upon building a library of collision cross section distributions (CCSDs)[22,23]. Despite the fact that negative ion mode collision induced dissociation (CID) generates insightful cross-ring fragmentation, full structural elucidation only by MS/MS is not yet possible. In **Chapters 4, 5** and **6** we elucidated the *O*- and *N*-glycan structures based upon a combination of previously established fragmentation patterns as well as known biosynthetic pathways and elution orders in the PGC column. Additionally, we also relied on the information obtained from well-characterized mucin glycoproteins such as bovine submaxillary mucin and porcine stomach mucin[24,25]. Additional exoglycosidase digestion was performed in order to support structural elucidation of separated isomers in **Chapter 4** where we describe the detection of more than 150 different *O*-GalNAc glycans. Nevertheless, not all glycans were able to be fully unambiguously assigned, as the tandem MS spectra did not contain enough diagnostic ions necessary for the full structural identification. Negative ion CID of multiply sialylated glycans results in a spectrum dominated by B- and Y-ions which originate from the loss of sialic acids. Unfortunately, these ions are not informative for the elucidation of the glycan structures. Previously, we attempted to use sialic acid derivatization to obtain more informative fragments from multi-sialylated *N*-glycan species (data not shown). Since these particular glycans did not show sufficient derivatization efficacy and specificity, additional peaks of the same glycan isomers were observed due to various underivatized sialic acids and the investigation was discontinued. Nevertheless, the spectra showed more informative fragment ions which resulted from the fragmentation of the glycan branches and cores which could potentially be used for less ambiguous glycan annotation. The formation of multiply charged glycan precursors is also an important aspect facilitating informative MS/MS. In **Chapter 3** we attempted to use different dopant solvents to increase charges for *O*-glycan species in the mass region from 900 - 1400 Da. While, the intensities of the lower charged species were boosted intensively using different polar protic dopants, no shift was observed towards the higher charged species in this mass region. This effect is likely due to the protic nature of the dopants. Further research is needed to investigate if other dopants could shift the precursors towards higher charged species.

In **Chapter 4** we characterized 26 different CRC cell lines and detected more than 150 different *O*-GalNAc glycans. In the process all MS/MS spectra were manually assigned, as fully automated data analysis software for negative mode MS/MS of glycans is still not available. To overcome this bottleneck, we have made our MS/MS spectra available in the repository Unicarb DR and contributed to a global project to facilitate future robust automated glycan assignments[26]. This library will be further expanded by the glycans identified in the study presented in **Chapter 6.** Glycomic databases which contain fully annotated glycan spectra are an important tool, which will facilitate automated identification of tandem MS in the future. However, unambiguous *de novo* glycan sequencing is not possible without the presence of a larger library of chemically synthesized standards which are well characterized by NMR spectroscopy. Moreover, isotopically labelled internal standards will allow for absolute quantification, which will overcome the current limitations in normalization methods for biomarker discovery[27]. This will provide further insights into the MS/MS fingerprints of specific glycans and provide deeper understanding of the fragmentation patterns of mucin type *O*-glycans in negative ion mode. Fortunately, a set of standards was recently obtained which enabled confident identification of tumor associated glycans in CRC that is presented in **Chapter 6**.

MODELS FOR STUDYING GLYCOSYLATION CHANGES IN CRC

Analysis of specific glycosylation signatures directly from tissue specimens is fundamental to obtain a better understanding of the molecular mechanisms responsible for carcinogenesis. However, the limited availability of human tissues and ethical concerns are a limiting factor. Therefore, suitable cell models are crucial to ensure unlimited availability and functional studies, although studies have questioned comparability of results from different laboratories of the same cell lines due to differences in culturing conditions, as well as their genomic instability. Nevertheless, it has been demonstrated that cell lines do recapitulate the recently established consensus molecular subtype (CMS) classification of CRC tumors based on their gene expression[28,29]. Especially, CRC cell lines are widely used models in functional studies investigating cancer pathogenesis. However, their *N*- glycosylation was characterized just recently[21]. In **Chapter 4** we have completed the characterization of the same set of cell lines focusing on mucin type *O*-glycosylation. We demonstrated that the *O*-glycosylation, of a panel of 26 different CRC cell lines, is very diverse and mainly

associated to the differentiation status of the cells. The diversity revealed on the glycome level reflects the overall heterogeneity of the tumors from where those cell lines were developed. The specific glycosylation phenotypes of the various CRC cell lines should be taken into account when selecting *in vitro* model systems. Additionally, differences were observed in glycosylation for the same cell lines cultured in different laboratories and media, mainly regarding the relative abundance of different glycans, whereas the differences were minor regarding the qualitative expression of different epitopes (**Chapter 4**). 2D cell line models are often criticized because they cannot represent the overall microenvironment of the tumors. However, simplification is necessary and appropriate when studying cell biological processes, avoiding complexity arising from different types of cells in the microenvironment. In **Chapter 5** we studied the glycosylation changes in the CaCo-2 cell line of a 2D model upon stimulation with butyrate and compared it with spontaneous differentiation. Our previous study (**Chapter 4**) showed that well differentiated CRC cell lines had higher expression of Lewis type antigens and I-branching, however, differentiation of the CaCo-2 cell line led to a decrease in terminal blood group H-antigen expression and an increase in terminal α2-3-sialylation. No increase in the expression of terminal Lewis type antigens was observed upon differentiation. It must be noted that CaCo-2 cell line showed unique glycomic signatures compared to the rest of the cell lines in **Chapter 4**, visible on the principal component analysis (PCA) scores plot. However, this cell line was chosen as a model because of its unique capability to differentiate in culture.

Interestingly, in the culture media the only monosaccharide that was supplied to the cells was glucose. This results in a process that is often overlooked, namely, the ability to convert glucose into other types of monosaccharide precursors providing the availability of different monosaccharides for the glycan biosynthesis. This is a particularly important mechanism in the colon, since the cells can take up different monosaccharides from their environment using monosaccharide transporters. Therefore, it would be interesting to study changes in the glycosylation of CRC cell lines when cells are supplied with a mix of different monosaccharide precursors in their medium in order to mimic the microenvironment of colon cells in the gut. This simple adaptation would likely increase the glycobiological and functional relevance of these *in vitro* models.

Despite the advances in 3D cell culture, patient derived organoids have proven to be much more robust preclinical models[30]. Compared to 2D cell models they have the multicellular complexity and tissue structure, facilitating functional studies. Phenotypic and genotypic profiling showed that they resemble the tumors they originated from, and initial studies have demonstrated they can be used as preclinical models for patient stratification and therapy choices in personalized medicine[31]. Recently, organotypic 3D skin cell models have been developed allowing for high throughput genetic engineering unravelling specific functions of different types of glycans on the differentiation of human skin[32,33]. Moreover, specific targets for different polypeptide GalNacTs have been revealed using this approach, unraveling the substrates for *O*-GalNAc glycosylation and related to specific phenotypes of GALNT knockouts[32]. These models provide a great potential to obtain further understanding into the physiological and pathological roles of glycosylation. Due to great improvements in the field of patient derived colon organoids[30,31,34], a better understanding of the role of glycosylation in colon crypt differentiation and malignancies can be obtained from these models. Despite their advantages, current organoids frequently do not contain cells from the tumor microenvironment such as immune cells and fibroblasts. Therefore, these complex two-way interactions are often excluded from consideration. Consequently, patient derived tissue material is still advantageous to capture cancer associated changes in glycosylation. However**,** tissue heterogeneity often poses a problem, as glycomic signatures cannot be traced back to different cell types. To tackle this issue, an enrichment of different tissue regions containing cells of interest such as epithelial mucosa or cancer cells excluding the muscle layers, stromal cells or immune infiltrate is an essential step prior to glycomics analysis. In **Chapter 6** we have developed such an approach for the analysis of *N*- and *O*-glycans from patient derived formalin-fixed, paraffin-embedded (FFPE) tissues, using laser capture microdissection to enrich for tissue regions of interest. We first identified the *O*-GalNAc linked cancer associated glycans from paired CRC tissues, which showed no expression in the normal colon mucosa. Moreover, we are currently expanding the study by the analysis of *N*-glycan signatures from the same set of samples and the primary cell lines derived from the same patient material. These cell lines are an important source of tumor associated antigens present in the tumors, that could be used for validation of anti-tumor antibody binding. In fact, we observed that the tumor associated carbohydrate antigens (TACAs) expressed by the tumors are present in most of the cell lines derived

from the analyzed tumors as well (data not shown), which will be an important tool for further therapeutic development. Interestingly, the cells show similar glycosylation to the tumors from which they were derived from. However, also some substantial differences were observed, in particular for mucin secreting adenocarcinoma. The possible explanation could be that, when analyzing tissue glycosylation, the proteins secreted in the microenvironment of the cells could also be captured, whereas only cell pellets were analyzed from the cultured cells. Analysis of the cell line secretome is possible, however, this remains challenging as the protein yields are very low and can be contaminated by the presence of bovine serum proteins added to the culture medium. Recently, differential glycosylation of the intracellular and secreted proteins was demonstrated, emphasizing the importance of studying the glycoproteome and understanding the functions of differential glycosylation[13].

INTEGRATION OF MULTIPLE OMICS APPROACHES

The advances in high throughput omics approaches have led to an increasing amount of publicly available data from different sources such as genomics, transcriptomics, proteomics and metabolomics. Combining the insights of such datasets lead to a better understanding of mechanisms behind the development of different diseases. Since it has been demonstrated before that cell lines can be good representatives of tumors based on gene mutation, gene expression and protein expression[29], we took advantage of the published transcriptomics datasets of the gene expression in overlapping cell lines in **Chapter 4**. We attempted to correlate the glycosyltransferases known to be differentially expressed in colon like *versus* undifferentiated cell lines to validate our glycomic findings and to gain further understanding of the impact of and link between transcriptomic expression and glycomic phenotypes. Despite the fact that the data was not obtained from the same cell pellets, we observed some expected associations with the glycosylation traits such as the correlation of FUT3 and FUT6 with Lewis type antigen expression, and GCNT3 with I-branching. However, it must be taken into account that protein glycosylation is not only influenced by the mRNA transcripts of the enzymes, but also the protein abundance, protein activity, different enzyme competition and substrate availability[16]. This can explain why some of the expected associations were not found in our data, such as the expression of FUT2, the enzyme responsible for blood group H expression in the gut, and blood group H antigen expression on *O*-glycans.

Moreover, the regulatory mechanisms of protein O-glycosylation are still poorly understood. We, therefore also analyzed well known transcription factors involved in colon differentiation. Some new associations were discovered that still needs further validation, for example the association of transcription factors MYB and ETS2 in regulation of glycosylation. Despite our expectations to find associations with CDX1 transcription factor, previously shown to be associated with regulation of *N*-glycan fucosylation[21,35], no strong associations were found with the *O*-glycome fucosylation. Thus, in **Chapter 5** we chose to investigate this further and analyze both the *O*-glycome and the proteome of the cells undergoing differentiation in culture in an integrative manner. With this approach we tried to understand if changes on the glycome level are mainly a consequence of changes in the cell glycosylation machinery or changes of the abundance of specific proteins. Despite the fact that our analysis covered over 5,000 different proteins, the majority of them were not related to mucin type *O*-glycan biosynthesis. Nevertheless, we did observe a decrease in the level of transcription factor HNF1A with differentiation, a key regulator of fucosylation, which was in line with a decrease in fucosylation of Caco2 cells with differentiation. Additionally, many of the identified proteins that showed statistically significant change with differentiation are *O*-glycoproteins. This indicates that most probably the changes in the cell glycome originate both from changes in the glycosylation machinery, and changes in the abundances of specific proteins. Interestingly, in the proteome we observed changes in the metabolic conversion and transport of monosaccharide precursors, indicating that studying changes in the cell metabolome could be the missing link for explaining changes in the cell glycome. Additionally, in **Chapter 6** we took advantage of a large dataset revealing transcriptomic signatures of CRC from over 100 patients (The Cancer Genome Atlas (TCGA) dataset). The changes in the cancer glycome reflected the changes in the cancer induced expression of mainly pathway specific core initiating enzymes such as core 1 synthase (C1GALT1), core 2 synthases (GCNT1 and GCNT3), and core 3 synthase (C3GNT6), whereas the expression of terminal epitopes showed less clear associations. Despite the success in supervised associations of glycan data with different enzymes, further development in systemic biology approaches is necessary to allow for a comprehensive analysis and integration with biosynthetic pathways. Recently an approach was developed that decomposes glycan structures into substructures taking into account the shared biosynthetic pathways between different individual glycans, as well as their

interdependence[36]. This enables integration with enzyme expression data, taking into account the enzyme activities of each step in the substructure biosynthesis and the known biosynthetic pathways. However, fully identified glycan structures are needed for the analysis in order to understand changes in linkage specific manner. We envision that the PGC-LC-MS/MS methodology in combination with the availability of synthetically produced standards will become the ideal approach to reveal isomer specific signatures to enable integration of multi-omics data with the glycan biosynthetic pathways. In summary, integration of different omics approaches provides a better understanding of the mechanisms behind differential glycosylation in diseases, validating the glycomic signatures revealed by mass spectrometry.

THERAPEUTIC STRATEGIES EMPLOYING GLYCANS/FUTURE PERSPECTIVES

Profiling overall cell glycomic signatures is an important approach for understanding disease related changes in the cell glycome, which can then be linked to changes in the cell glycosylation machinery such as expression or activity of glycosyltransferases, glycosidases and activated sugar donors. Previous studies showed that cancer cells express increased sialylation on their cell surface, resulting in the first targeted therapeutic strategy that used neuraminidases to cleave off the sialic acids of glycans [37,38]. However, as shown in our study in **Chapter 6**, healthy colon cells also express many mucin type sialylated glycans. Interestingly, MS revealed that this type of sialylation is different, core 3 related α2-6 sialylation, that should not be targeted by anti-cancer drugs. This illustrates that targeting sialylation as a whole, can give misleading results[39]. This highlights the importance of targeting specific glycans, expressed solely or predominantly by the tumors, with no or limited expression in the surrounding normal tissue. In **Chapter 6** we selected seven of the most promising targets and initiated the validation of their potential as therapeutic targets. Of note, we are currently working on expanding this study, which focused on mucin type *O*-glycans, by exploring the *N*-glycan signatures of the same set of samples which will give a more comprehensive overview of the cancer associated changes in CRC. TACAs can be targeted with antibody-dependent cellular cytotoxicity (ADCC)-inducing or complement-dependent cytotoxicity (CDC)-inducing antibodies or with antibody-drug conjugates[40]. Similarly, bispecific antibodies could be used to direct immune cells

to TACA-expressing tumors especially in the cases where TACAs are expressed in normal tissues of another organ system. This may be a beneficial approach, for increasing cancer specificity and avoiding off-target effects. In the case of the TACAs identified in **Chapter 6**, the blocking of tumor associated core 2 SLe^X glycans on leukocytes might be limiting physiological leukocyte homing. Alternatively, chimeric antigen receptor (CAR) expressing immune cells could be redirected towards TACA-presenting tumors. However, there are still many obstacles to overcome before effective CAR-T cell therapies for solid tumors such as CAR-T cell trafficking and infiltration, immunosuppressive tumor microenvironment and CAR-T cell associated toxicities[41]. Additionally, TACAs can engage immune receptors including inhibitory Siglec receptors on T-cells and myeloid cells and different antibody and sialidase based strategies can be employed for improving anti-cancer immunity by preventing these interactions[42]. A number of truncated *O*-glycan (Tn, sialyl-Tn and T antigen) targeting antibodies were developed, however, they displayed poor affinity to the tumors[40,41]. Similarly, a vaccine targeting sialyl-Tn antigens was developed for the treatment of breast cancer and showed high titers of antibodies, but did not provide survival benefit for the patients[43]. Although targeting individual glycans can be beneficial, because of their high relative expression on the cell surface, they are poorly immunogenic. Therefore, targeting glycans together with specific protein/peptide carriers could increase the immunogenicity. To pursue this, further studies are required which evaluate the protein carriers of specific TACAs. *O*-glycopeptide targeting antibodies carrying Tn , T and sialyl-T antigens on MUC1, MUC4 and MUC2 have been developed to overcome this issue[41,44]. While they demonstrate a higher affinity they may exhibit reduced efficiency due to heterogenous expression and relative paucity of the protein in different tumors[41,44]. Other than the expression of specific glycans on specific proteins, it is also important to understand where the specific glycans are present on the protein carrier and how a change in the site occupancy or a change in the glycan structure affects the function of the proteins. Only from that point onwards we can begin to understand which glycans are essential for solely protecting proteolytic cleavage and which ones are important for interaction with other cells in the environment[16]. A good example is PSGL-1 where it was discovered that core 2 sLe^X glycans bind selectins only on the *N*-terminal tip of the protein in close proximity of sulfated tyrosines[45]. This will allow development of inhibitors that specifically target this interaction, which will increase the specificity and affinity to a

large extent. However, despite the developments of “Simple Cell” technology to facilitate a faster comprehensive analysis of *O*-glycoproteome site occupancy[32,33], further research is needed for matching specific glycans to specific proteins together with their specific sites in complex biological samples to reveal their functions in physiological and pathological conditions.

In summary, this thesis largely contributed to the development of higher throughput methodologies to analyze both *N*-and *O*-glycans from complex biological samples such as cell lines and tissues. Moreover, it revealed the mucin type *O*-glycan repertoires from commonly used CRC cell lines as well as patient matched CRC tissues and healthy mucosa controls leading to the identification of CRC specific glycan signatures that will be further validated as potential therapeutic targets for antibody-based immunotherapy. Importantly, the research opened up new perspectives for the investigation of TACAs, and their functional role in the development and progression of CRC using suitable colon organotypic models as well as patient derived organoids.

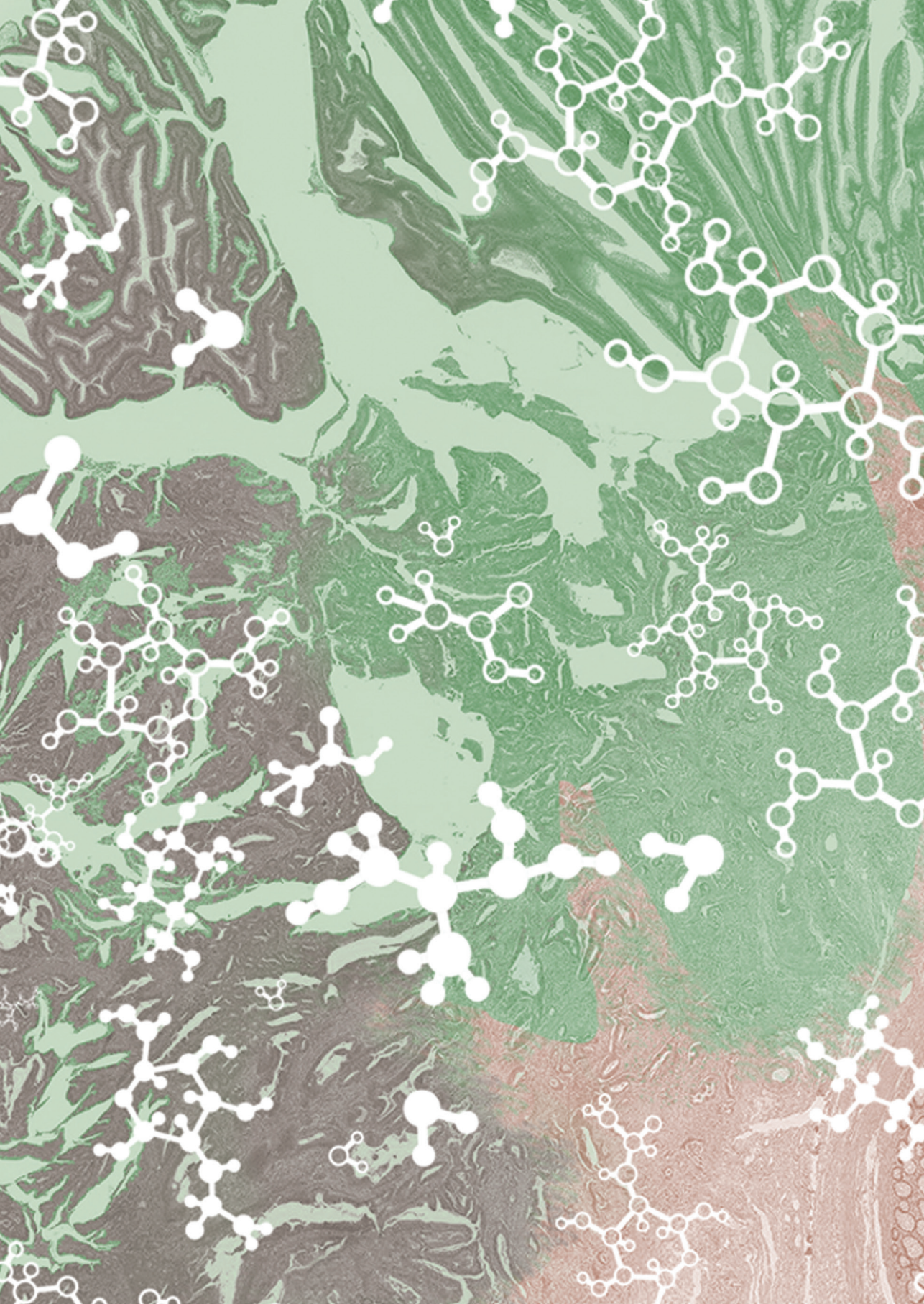
REFERENCES

1. Wilkinson, H. et al. The O-Glycome of Human Nigrostriatal Tissue and Its Alteration in Parkinson’s Disease. J. Proteome Res. 20, 3924 (2021).
2. Crouch, L. I. et al. Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. Nat. Commun. 11, (2020).
3. Malaker, S. A. et al. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. Proc. Natl. Acad. Sci. U. S. A. 116, 7278–7287 (2019).
4. Kelm, M. et al. Targeting epithelium-expressed sialyl Lewis glycans improves colonic mucosal wound healing and protects against colitis. JCI Insight 5, (2020).
5. Boyaval, F. et al. N-glycomic signature of stage II colorectal cancer and its association with the tumor microenvironment. Mol. Cell. Proteomics 25, mcp.RA120.002215 (2020).
6. Jensen, P. H., Karlsson, N. G., Kolarich, D. & Packer, N. H. Structural analysis of N- and O-glycans released from glycoproteins. Nat. Protoc. 7, 1299–1310 (2012).
7. Vreeker, G. C. M. et al. O- and N-glycosylation analysis of cell lines by ultrahigh resolution MALDI-FTICR-MS. Int. J. Mass Spectrom. 448, 116267 (2020).
8. Kotsias, M. et al. A semi-automated, high throughput approach for O-glycosylation profiling of in vitro established cancer cell lines by MALDI-FT-ICR MS. Glycoconj. J. 1–10 (2021).

9. Comer, F. I. & Hart, G. W. O-Glycosylation of Nuclear and Cytosolic Proteins: DYNAMIC INTERPLAY BETWEEN O-GlcNAc ANDO-PHOSPHATE*. J. Biol. Chem. 275, 29179–29182 (2000).
10. An, H. J. et al. Extensive determination of glycan heterogeneity reveals an unusual abundance of high mannose glycans in enriched plasma membranes of human embryonic stem cells. Mol. Cell. Proteomics 11, M111.010660 (2012).
11. Xu, G., Goonatilleke, E., Wongkham, S. & Lebrilla, C. B. Deep structural analysis and quantitation of O-linked glycans on cell membrane reveal high abundances and distinct glycomic profiles associated with cell type and stages of differentiation. 14, 2020.
12. Pirro, M. et al. Oxonium Ion Guided Analysis of Quantitative Proteomics Data Reveals Site-Specific O-Glycosylation of Anterior Gradient Protein 2 (AGR2). Int. J. Mol. Sci. 22, (2021).
13. Steentoft, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 32, 1478–1488 (2013).
14. King, S. L. et al. Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. Blood Adv 1, 429–442 (2017).
15. Wandall, H. H., Nielsen, M. A. I., King-Smith, S., Haan, N. & Bagdonaite, I. Global functions of O-glycosylation - Promises and Challenges in O-glycobiology. FEBS J. 1–30 (2021).
16. Vistain, L. F. & Tay, S. Single-Cell Proteomics. Trends Biochem. Sci. 46, 661–672 (2021).
17. de Vries, N. L., Mahfouz, A., Koning, F. & de Miranda, N. F. C. C. Unraveling the Complexity of the Cancer Microenvironment With Multidimensional Genomic and Cytometric Technologies. Front. Oncol. 10, 1254 (2020).
18. Ijsselsteijn, M. E., van der Breggen, R., Farina Sarasqueta, A., Koning, F. & de Miranda, N. F. C. C. A 40-Marker Panel for High Dimensional Characterization of Cancer Immune Microenvironments by Imaging Mass Cytometry. Front. Immunol. 10, 2534 (2019).
19. Heijs, B., Potthoff, A., Soltwisch, J. & Dreisewerd, K. MALDI-2 for the Enhanced Analysis of N-Linked Glycans by Mass Spectrometry Imaging. Anal. Chem. 92, 13904–13911 (2020).
20. Lageveen-Kammeijer, G. S. M. et al. Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. Nat. Commun. 10, 2137 (2019).
21. Holst, S. et al. N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homebox 1 (CDX1)/Villin mRNA Expression. Mol. Cell. Proteomics 15, 124–140 (2016).
22. Sastre Toraño, J. et al. Identification of Isomeric N-Glycans by Conformer Distribution Fingerprinting using Ion Mobility Mass Spectrometry. Chemistry 27, 2149–2154 (2021).
23. Pallister, E. G. et al. Utility of Ion-Mobility Spectrometry for Deducing Branching of Multiply Charged Glycans and Glycopeptides in a High-Throughput Positive ion LC-FLR-IMS-MS Workflow. Anal. Chem. 92, 15323–15335 (2020).
24. Savage, A. V., D'Arcy, S. M. T. & Donoghue, C. M. Structural characterization of neutral oligosaccharides with blood group A and H activity isolated from bovine submaxillary mucin. Biochem. J 279, 95–103 (1991).

25. Savage, A. V., Donoghue, C. M., D'Arcy, S. M., Koeleman, C. A. M. & van den Eijnden, D. H. Structure determination of five sialylated trisaccharides with core types 1, 3 or 5 isolated from bovine submaxillary mucin. Eur. J. Biochem. 192, 427–432 (1990).
26. Rojas-Macias, M. A. et al. Towards a standardized bioinformatics infrastructure for N- and O-glycomics. Nat. Commun. 10, 3275 (2019).
27. Uh, H.-W. et al. Choosing proper normalization is essential for discovery of sparse glycan biomarkers. Molecular Omics (2020) doi:10.1039/c9mo00174c.
28. Sveen, A. et al. Colorectal cancer consensus molecular subtypes translated to preclinical models uncover potentially targetable cancer cell dependencies. Clin. Cancer Res. 24, 794–806 (2018).
29. Berg, K. C. G. et al. Multi-omics of 34 colorectal cancer cell lines - a resource for biomedical studies. Mol. Cancer 16, 116 (2017).
30. Matano, M. et al. Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. Nat. Med. 21, 256–262 (2015).
31. Vlachogiannis, G. et al. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. Science 359, 920–926 (2018).
32. Dabelsteen, S. et al. Essential functions of glycans in human epithelia dissected by a CRISPR-Cas9-engineered human organotypic skin model. Dev. Cell 54, 669-684.e7 (2020).
33. Bagdonaite, I., Pallesen, E. M. H., Ye, Z. & Vakhrushev, S. Y. O-glycan initiation directs distinct biological pathways and controls epithelial differentiation. EMBO (2020).
34. Fujii, M. et al. A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. Cell Stem Cell 18, 827–838 (2016).
35. Holst, S. et al. N-Glycomic and Transcriptomic Changes Associated with CDX1 mRNA Expression in Colorectal Cancer Cell Lines. Cells 8, 273 (2019).
36. Bao, B. et al. Correcting for sparsity and interdependence in glycomics by accounting for glycan biosynthesis. Nature Communications 2021 12:1 12, 1–14 (2021).
37. Bagshawe, K. D. & Currie, G. A. Immunogenicity of L 1210 murine leukaemia cells after treatment with neuraminidase. Nature 218, 1254–1255 (1968).
38. Büll, C. et al. Targeted delivery of a sialic acid-blocking glycomimetic to cancer cells inhibits metastatic spread. ACS Nano 9, 733–745 (2015).
39. Cornelissen, L. A. M. et al. Disruption of sialic acid metabolism drives tumor growth by augmenting CD8 + T cell apoptosis. International Journal of Cancer 144, 2290–2302 (2019).
40. Mantuano, R. Tumor-associated carbohydrates and immunomodulatory lectins as targets for cancer immunotherapy. Journal for ImmunoTherapy of Cancer 8, 1222 (2020).
41. Steentoft, C. et al. Glycan-directed CAR-T cells. Glycobiology 28, 656–669 (2018).
42. Bärenwaldt, A. & Läubli, H. The sialoglycan-Siglec glyco-immune checkpoint–a target for improving innate and adaptive anti-cancer immunity. Expert Opinion on Therapeutic Targets vol. 23 (2019).
43. Miles, D. et al. Phase III multicenter clinical trial of the sialyl-TN (STn)-keyhole limpet hemocyanin (KLH) vaccine for metastatic breast cancer. Oncologist 16, 1092–1100 (2011).

44. Rodrigues Mantuano, N., Natoli, M., Zippelius, A. & Läubli, H. Tumor-associated carbohydrates and immunomodulatory lectins as targets for cancer immunotherapy. J Immunother Cancer 8, (2020).
45. Leppänen, A., Yago, T., Otto, V. I., McEver, R. P. & Cummings, R. D. Model glycosulfopeptides from P-selectin glycoprotein ligand-1 require tyrosine sulfation and a core 2-branched O-glycan to bind to L-selectin. J. Biol. Chem. 278, 26391–26400 (2003).



Appendices

LIST OF ABBREVIATIONS

ABC	Ammonium bicarbonate
AC	adenocarcinoma
ADCC	antibody- dependent cellular cytotoxicity
ANOVA	Analysis of Variance
AP	alkaline phosphatase
Asn	asparagine
AUC	area under the curve
B3GALT	β1-3 galactosyltransferase
B3GNT	β1-3-*N*-acetylglucosaminyltransferase
B4GALNT	β1-4 acetylgalactosaminyltransferase
B4GALT	β1-4 galactosyltransferase
BCA	Bicinchoninic acid
BSM	bovine submaxillary mucin
CaCo-2	Cancer *coli-2*
CAR	chimeric antigen receptor
CCSD	collision cross section distribution
CDC	complement-dependent cytotoxicity
CDX1	Caudal Type Homeobox 1
CE-ESI-MS	capillary electrophoresis electrospray ionization-mass spectrometry
CID	collision induced dissociation
CIM	clustered image map
CIMP	CpG island methylator phenotype
CIN	chromosomal instability
CMP	cytidine monophosphate
CMS	Consensus molecular subtype
CRC	Colorectal cancer
CRCSC	Colorectal cancer subtyping consortium
CV	coefficient of variation
CyTOF	cytometry by time-of-flight
CZE	capillary zone electrophoresis
DEN	dopant enriched nitrogen

DiSia	di-sialic acid
DMEM	Dulbecco's modified Eagle's medium
DTT	dithiothreitol
EMT	epithelial to mesenchymal transition
ESI	electrospray ionization
ETD	Electron-transfer dissociation
EtOH	ethanol
FA	formic acid
FBS	fetal bovine serum
FFPE	formalin-fixed, paraffin-embedded
Fuc	fucose
FUT	α1-4 fucosyltransferases
GAG	glycosaminoglycan
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GalNAcol	reduced reducing end of GalNAc
GALNT	polypeptide-*N*-acetylgalactosaminyltransferase
GCNT	β1-6-*N*-acetylglucosaminyltransferase
GDP	guanidine diphosphate
GI	gastrointestinal
Glc	glucose
GlcNAc	N-acetylglucosamine
GPI	glycosylphosphatidylinositol
GSL	glycosphingolipid
GST	glycosyltransferase
GuHCl	guanidine hydrochloride
HCl	hydrochloric acid
HE	hematoxylin and eosin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HILIC	hydrophilic interaction liquid chromatography
ICC	ion charge control
IGS	image guided surgery
IPA	isopropanol

IT	ion trap
KOH	potassium hydroxide
KOH	potassium hydroxide
LacNAc	*N*-Acetyllactosamine
LCM	laser capture microdissection
Le	Lewis
LNI	lymph node invasion
MALDI	matrix assisted laser desorption/ionization
Man	mannose
MeCN	acetonitrile
MeCN	acetonitrile
MeOH	methanol
MGC	mesoporous graphitic carbon
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSAC	mass spectrometry average composition
MSI	microsatellite instability
N_2	nitrogen
NaBH4	Sodium borohydride
NaCl	sodium chloride
NEC	neuroendocrine carcinoma
NeuAc	*N*-acetylneuraminic acid
NeuGc	*N*-glycolylneuraminic acid
NK	Natural killer
NMuMG	mouse mammary gland
PBS	phosphate buffered saline
PCA	principal component analysis
PCMF	post column make-up flow
PCR	polymerase chain reaction
PEN	polyethylene naphthalate
PGC	porous graphitized carbon
PGC-nano-LC-ESI-MS/MS	porous graphitized carbon nano-liquid chromatography electrospray mass spectrometry

PNGase F	Peptide -*N*-Glycosidase F
PP	polypropylene
Pro	proline
psi	pound per square inch
PVDF	Polyvinylidene difluoride
PVP	polyvinylpirrolidone
qPAC	micro pillar array
RP	reverse phase
rpm	rounds per minute
RT	room temperature
RT	retention time
S/N	signal to noise
SDS	Sodium dodecyl sulfate
Ser	serine
SPE	solid-phase extraction
SPS	smart parameter setting
ST3GAL1	α2-3 sialyltransferase
ST6GALNAC	α2-6 sialyltransferase
ST8SIA	α2-8 sialyltransferases
TACA	tumor-associated carbohydrate antigen
TCGA	The Cancer Genome Atlas
TFA	Trifluoroacetic acid
TGF	Transforming growth factor
Thr	threonine
TMT	tandem mass tag
Tris base	tris(hydroxymethyl)amino-methane
Tyr	tyrosine
UDP	uridine diphosphate
UHPLC	ultra-high performance liquid chromatography
Xyl	xylose

ENGLISH SUMMARY

The surface of eukaryotic cells contains a very dense layer of oligosaccharides called glycans that are linked to protein and lipid carriers and play an important role in cell-cell and cell-extracellular matrix interactions. Aberrant glycosylation is a hallmark of cancer and plays a role in various cellular processes that are involved in cancer pathogenesis, such as cell proliferation, adhesion, migration and immune surveillance. Cancer-induced changes in glycosylation have an impact on the function of major glycoproteins in the human colon, therefore studies focused on colorectal cancer (CRC)-specific glycosylation signatures can provide novel insights into onset and progression of this disease. The major focus of this thesis was to investigate mucin type *O*-glycosylation signatures of CRC. For this purpose, a protocol for in-depth analysis of *N*- and *O*-glycans obtained from cell lines was developed (**Chapter 2**) using nanoscale porous graphitized carbon liquid chromatography coupled to mass spectrometry (PGC-nano-LC-MS). The sample preparation protocol was adapted to a robust 96-well plate format to increase throughput of release and purification compared to previous cell line approaches. This strategy, furthermore, allowed glycan isomer separation and detailed structural identifications. In **Chapter 3** additional conditions were optimized in the MS methodology by using polar protic dopant (methanol and isopropanol) enriched nitrogen gas to increase sensitivity on the MS and tandem MS level. This allowed for a more confident structural identification as well as correction for the ionization bias for early eluting species in the LC-gradient. In **Chapter 4** we applied the methodology developed in Chapter 2 to the analysis of *O*-glycosylation signatures of 26 different CRC cell lines. This analysis resulted in the characterization of more than 150 O-glycan structures and increased our understanding of glycan expression in the analyzed cell lines. The glycophenotypes showed correlation mainly to cell differentiation. Namely, colon-like well-differentiated cell lines expressed Lewis type antigens and I-branched glycans, whereas undifferentiated cell lines expressed sialylated core 1 glycans and partially blood group H-antigens. From pathway analysis it was determined that differences in expression of the glycosyltransferases and transcription factors in those groups associated with the glycan signatures. This revealed that these changes reflect the differences in the glycosylation machinery, and not per se the abundances of specific proteins. To gain further understanding in the mechanisms underlying glycomic changes with colon cell

differentiation, we explored changes in the cell line glycome and proteome upon spontaneous and butyrate-stimulated differentiation in in vitro cell culture (**Chapter 5**). Glycomic analysis revealed a decrease of blood group antigen H fucosylation with cellular differentiation, together with increased sialylation levels. Increased expression of specific cell adhesion proteins was found on the proteome level. By performing an integrative approach, we generated hypotheses about glycosylation signatures of specific cell adhesion proteins, which may play an important role in cancer progression. The localization of glycans on the cell surface and their role in biological processes are important in cancer pathogenesis, making them potential candidates for glycan targeting immunotherapy. Therefore, we further optimized the methodology to enable comprehensive analysis of *N*- and *O*-glycans from specific regions of formalin-fixed, paraffin-embedded tissues using laser capture microdissections and applied it for the analysis of selected regions of CRC tissues and their patient-matched colon mucosa controls (**Chapter 6**). We identified specific tumor-associated carbohydrate antigens (TACAs) that show expression only in the tumor samples, with no or limited expression in the normal colon mucosa. These cancer signatures were found to be associated with dysregulation of biosynthetic machinery in the cancerous regions, and largely correlated with transcriptomics of the relevant glycosyltransferases revealing a CRC overexpressed biosynthetic pathway. Since TACAs are present in high abundance on the surface of cancer cells which are linked to many different proteins, these are very promising targets for the development of tumor-specific immunotherapy.

NEDERLANDSE SAMENVATTING

Het oppervlak van eukaryote cellen bevat een compacte laag van oligosacchariden, ook wel glycanen genoemd. Deze glycanen zitten covalent gebonden aan zowel eiwitten als lipiden en spelen een belangrijke rol bij interactie tussen cellen onderling, maar ook tussen cellen en de extracellulaire matrix. Omdat glycosylering een belangrijke rol speelt bij verschillende cellulaire processen zoals cel proliferatie, -adhesie, -migratie en controle van de immuunrespons, worden veranderingen in glycosylering vaak aangetroffen tijdens het ontstaan en progressie van ziekten. Zo worden afwijkingen in de glycosylering bijvoorbeeld beschouwd als één van de pijlers van verschillende typen kanker, waaronder dikke darmkanker. Studies gericht op de analyse van de veranderingen in glycosylering in dikke darmkanker kunnen daarom nieuwe inzichten geven in het ontstaan en het verloop van deze ziekte.

Het belangrijkste doel van het onderzoek zoals beschreven in dit proefschrift was het bepalen van kenmerkende mucine type *O*-glycosylering in relatie tot dikke darmkanker. Ten eerste werd hiervoor een protocol ontwikkeld waarbij, in tegenstelling tot eerder ontwikkelde methoden, 96 monsters van tumorcellijnen tegelijkertijd kunnen worden opgewerkt (**Hoofdstuk 2**). Het protocol werkt zowel voor de analyse van *N*- als *O*-glycanen. Door het te combineren met een zeer gevoelig analytische platform (poreuze grafietkoolstof vloeistofchromatografie gekoppeld met massaspectrometrie (PGC-nano-LC- MS) kan isomeerscheiding van glycanen worden bereikt met behoud van volledige structuuridentificatie. In **Hoofdstuk 3** werden verscheidene condities van de methode verder geoptimaliseerd teneinde de gevoeligheid op MS en MS/MS niveau te verbeteren. Door bijvoorbeeld stikstofgas te verrijken met een polair en protisch oplosmiddel (methanol en isopropanol) konden er meer unieke fragmenten worden waargenomen waardoor de glycaan structuren met grotere zekerheid konden worden vastgesteld (isopropanol). Daarnaast zorgde het gebruik van methanol voor een correctie van de ionisatieverschillen van glycanen die met de gekozen gradiënt snel van de chromatografische kolom elueerden.

In **Hoofdstuk 4** werd de geoptimaliseerde methode zoals beschreven in *hoofdstuk 2* toegepast voor de analyse van 26 dikke darmkanker-cellijnen. Dit resulteerde in de karakterisering van meer dan 150 unieke *O*-glycaanstructuren en gaf inzicht in de (relatieve) hoeveelheid van de verschillende glycanen in deze cellijnen. De verschillen

in glycosyleringsprofielen (glycofenotypes) correleerden voornamelijk met celdifferentiatie. Met name de goed gedifferentieerde cellijnen brachten Lewis-type antigenen tot expressie en daarnaast glycanen met LacNAc-vertakkingen. De ongedifferentieerde cellijnen daarentegen vertoonden voornamelijk bloedgroep H-antigenen en type 1 *O*-glycanen met sialylering. Aan de hand van informatie uit bestaande databases bleken de glycofenotypes te kunnen worden gecorreleerd met de expressie van bepaalde glycosyltransferases en transcriptiefactoren. Dit impliceert dat de veranderingen in glycosylering de verschillen in de biosynthese van de glycanen weerspiegelen en niet slechts de hoeveelheid van specifieke eiwitten. Teneinde meer inzicht te krijgen in het onderliggende mechanisme dat deze veranderingen veroorzaakt in relatie tot celdifferentiatie, onderzochten we de veranderingen in het glycoom en proteoom van een cellijn na spontane, en door butyraat gestimuleerde differentiatie *in-vitro* (**Hoofdstuk 5**). Onze analyse liet op glycoom niveau een afname zien van gefucosyleerd bloedgroep H-antigeen met cellulaire differentiatie, in combinatie met een toename van sialylering. Op proteoom niveau werd een verhoging van specifieke adhesiemoleculen waargenomen. Door deze bevindingen te combineren hebben we nieuwe hypotheses opgesteld m.b.t. glycosylering van adhesiemoleculen in relatie tot progressie van kanker.

De lokalisatie van mucinen op het celoppervlak en hun rol in biologische processen is belangrijk bij de pathogenese van kanker. Hierdoor zijn het potentiële kandidaten voor glycaan-gerichte immunotherapie. Om dit verder te onderzoeken hebben we de methodologie geoptimaliseerd zodat het mogelijk werd de *N*- en *O*-glycanen uit specifieke gebieden van (tumor)weefsel van een dikke darm te bepalen. Hiervoor zijn formaline-gefixeerde en paraffine-ingebedde weefsels gebruikt waarbij met behulp van *laser-capture* microdissectie de gebieden werden geselecteerd voor onze analyses (**Hoofdstuk 6**). Hiermee zijn er specifieke tumor-geassocieerde glycaan-antigenen geïdentificeerd die (bijna) uitsluitend tot expressie komen in de tumor weefsels en niet of nauwelijks in het gezonde weefsel. Deze kenmerken waren geassocieerd met een ontregeling van de biosynthese van glycanen in de tumorweefsels, en konden grotendeels gecorreleerd worden met *transcriptomics* data van relevante glycosyltransferasen. Deze resultaten bieden een aanzet voor de ontwikkeling van tumor specifieke immunotherapie.

CURRICULUM VITAE

Katarina Madunić was born on March 29th, 1989 in Split, Croatia. Her interest in medicine and biochemistry began early in her education and she wrote her elementary school thesis about tumors in 2008. She followed her passion and joined an integrated 5-year master program in Pharmacy from the Faculty of Pharmacy and Medicinal Biochemistry, University of Zagreb, Croatia in 2012. During her studies she was particularly interested in courses such as molecular biology and biochemistry and here it was the first time she heard about glycans from prof. Gordan Lauc. She was immediately intrigued by the importance of these sugars in human diseases and was determined to do her thesis project about glycosylation. In 2011 she joined Genos Ltd as a master thesis intern and worked on the optimization of a high throughput HILIC analysis of fluorescently labelled *N*-glycans isolated from human leukocytes under supervision of dr. Irma Mahmuljin (Redžić) and prof.dr.sc. Gordan Lauc. After graduation in 2012, she worked in a pharmacy and later she joined the clinical research organization Optimapharm to work as a clinical research associate in phase 3 clinical trials (2015). Here she realized that she wanted to pursue a scientific career, and in 2016, she got the opportunity to start her PhD project at the Center for Proteomics and Metabolomics, Leiden University Medical Center in the Netherlands. In this project she explored glycosylation signatures of colorectal cancer using PGC-nano-LC-MS/MS under the supervision of prof. Manfred Wuhrer and dr. Guinevere Lageveen-Kammeijer. As part of the Marie Curie European Training Network, "GlycoCan", she visited the group of prof. Daniel Kolarich at Max Planck Institute for colloids and interfaces (Berlin, Gemany), where she received PGC-nanoLC-MS/MS method transfer training and applied it for glycomic analysis of proteins isolated from laser captured tissue microdissections. Additionally, she visited the industrial collaborator Ludger Ltd (Oxfordshire, UK) where she obtained experience with automated high throughput glycan analysis using Hamilton liquid handling robot and MALDI mass spectrometry of permethylated *O*-glycans. During her PhD she received extensive interdisciplinary scientific and soft skills training and was involved in various projects of her fellow PhD-students. She presented her research at eight Dutch and international conferences, for which she received three young investigator travel grants. The results of her work and many collaborations were published in peer-reviewed journals resulting in three first author and ten co-author publications.

Currently, she is employed within the same research group in Leiden, where she continued her research in the field of glycoproteomics. After her PhD defense she will spend 6 weeks in the lab of prof. Mohamed Abdel-Mohsen at the Wistar institute in Philadelphia, USA. There she will study the role of the gut glycome in regulating the homeostatic relationship between the host and its gut microbiota, during HIV infection. Finally, she is going to continue her academic career as a postdoctoral researcher at the Copenhagen Center for Glycomics in Denmark where she will focus on understanding cellular processes related to the function of cadherin-specific *O*-mannosylation in health and disease.

PHD PORTFOLIO

MANDATORY COURSES

- PhD Introductory Meeting 2019
- BROK Course- regulations for conducting clinical research in the Netherlands 2017
- Basic Methods and Reasoning in Biostatistics 2017

GENERIC/DISCIPLINARY COURSES

- Glycobiology and Glycochemistry International ELearning Course (NOVA Medical School, Lisbon, Portugal) 2017
- Glycan analysis workshop (Glycocan ETN workshop, Ludger, Oxfordshire, UK) 2016
- High-throughput data processing of MALDI-TOF-MS data (Glycocan ETN workshop, Leiden, The Netherlands) 2018
- Storytelling and Stagecraft for Scientists (Glycocan ETN workshop, Leiden, The Netherlands) 2018
- Business Process and Quality Management Workshop (Glycocan ETN workshop, Leiden, The Netherlands) 2018
- Breaking up with Excel: An Introduction to the R Statistical Programming Language (MSACL 2018 short course) 2018
- Introduction to Cytoscape (LUMC) 2018
- Leadership framework, Business Model Canvas, Quality management, SIPOC workshop (Glycocan ETN workshop, Ludger, Oxfordshire, UK) 2019
- Using R for data analysis (LUMC) 2019
- Academic Writing for Phds (Leiden University) 2019
- Advanced Data Science course in R Statistical Programming language (MSACL 2019 short course) 2019
- Clinical Internship at Pathology Department (LUMC) 2019
- MultiOmics Data integration in R (LUMC) 2019
- Writing an excellent research grant proposal (Leiden University) 2020
- Job Orientation for PhDs (Leiden University) 2020
- Job interview skills Training (Leiden University) 2020
- Develop Successful Methods for Identification and Quantitation in Complex Matrices (MSACL 2020 short course) 2020
- Basic management and leadership skills (Leiden University) 2020

CONGRESS ATTENDANCE AND POSTER OR ORAL PRESENTATIONS

- 27th Joint Glycobiology meeting (Nijmegen, The Netherlands) 2016
- 28th Joint Glycobiology meeting (Aachen, Germany) 2017
- NVMS Spring meeting- New Talents in Mass Spectrometry (Amsterdam, The Netherlands) 2017
- Clinical Mass Spectrometry Imaging symposium (Leiden, The Netherlands) 2017
- Mutanome analysis finding neoepitopes in cancer-mini-symposium (Leiden, The Netherlands) 2018
- CESI-MS Symposium (Leiden, The Netherlands) 2018
- Data science seminar LUMC (Leiden, The Netherlands) 2018
- Mass Spectrometry Applications to the Clinical Lab (MSACL) 2018 EU (Salzburg, Austria)- oral presentation 2018
- Mini-symposium: Glycobiology at the LUMC and beyond (Leiden, The Netherlands) 2019
- Gordon Research Seminar on Glycobiology (GRS) (Lucca, Italy)- oral presentation 2019
- Gordon Research Conference on Glycobiology (GRC) (Lucca, Italy)- poster presentation 2019
- Mass Spectrometry Applications to the Clinical Lab (MSACL) EU 2019 (Salzburg, Austria)- oral presentation 2019
- Annual Oncode-CGC conference (Amsterdam, The Netherlands)- poster presentation 2019
- Spinoza Symposium Glyco-Science and its Medical Implications (Amsterdam, The Netherlands) 2019
- NVMS Fall meeting- New Talents in Mass Spectrometry (Leiden, The Netherlands)- poster presentation 2019
- 2020 Society for Glycobiology Virtual Meeting (online)- poster presentation 2020
- Joint Warren and Beilstein Symposium on Glycosciences 2021 (online) 2021
- 37th International Symposium on Microscale Separations and Bioanalysis (online)- oral presentation 2021
- MSACL 2021 EU 7th European Congress (online)- oral presentation 2021

SECONDMENTS

- Max Planck institute of Colloids and Interfaces, Berlin, Germany- prof.dr. Daniel Kolarich' lab 2016
- Ludger Ltd, Oxfordshire, UK- industrial collaboration 2018
- Sahlgrenska Academy, Gothenburg, Sweden- prof.dr. Niclas Karlsson's lab 2018

AWARDS AND GRANTS

- MSACL young investigator travel grant 2018
- MSACL young investigator travel grant 2019

LIST OF PUBLICATIONS

1) Zhang, T.‡, **Madunić, K**.‡, Holst, S., Zhang, J., Jin, C., Ten Dijke, P., Karlsson, N.G., Stavenhagen, K. and Wuhrer, M., 2020. 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*, *16*(4), pp.355-363.
‡ These authors contributed equally to this paper.

2) **Madunić, K**., Zhang, T., Mayboroda, O.A., Holst, S., Stavenhagen, K., Jin, C., Karlsson, N.G., Lageveen-Kammeijer, G.S.M. and Wuhrer, M., 2021. Colorectal cancer cell lines show striking diversity of their *O*-glycome reflecting the cellular differentiation phenotype. *Cellular and Molecular Life Sciences*, *78*(1), pp.337-350.

3) **Madunić, K**., Wagt, S., Zhang, T., Wuhrer, M. and Lageveen-Kammeijer, G.S.M., 2021. Dopant-Enriched Nitrogen Gas for Enhanced Electrospray Ionization of Released Glycans in Negative Ion Mode. *Analytical chemistry*, *93*(18), pp.6919-6923.

4) Biwi, J., Clarisse, C., Biot, C., Kozak, R.P., **Madunić, K**., Mortuaire, M., Wuhrer, M., Spencer, D.I.R., Schulz, C., Guerardel, Y. and Lefebvre, T., 2019. OGT Controls the Expression and the Glycosylation of E-cadherin and Affects Glycosphingolipid Structures in Human Colon Cell Lines. *Proteomics*, *19*(21-22), p.1800452.

5) Rojas-Macias, M.A., Mariethoz, J., Andersson, P., Jin, C., Venkatakrishnan, V., Aoki, N.P., Shinmachi, D., Ashwood, C., **Madunić, K**., Zhang, T. and Miller, R.L., 2019. Towards a standardized bioinformatics infrastructure for *N*- and *O*-glycomics. *Nature communications*, *10*(1), pp.1-10.

6) Vreeker, G.C., Nicolardi, S., **Madunić, K**., Kotsias, M., van der Burgt, Y.E. and Wuhrer, M., 2020. *O*- and *N*-glycosylation analysis of cell lines by ultrahigh resolution MALDI-FTICR-MS. *International Journal of Mass Spectrometry*, *448*, p.116267.

7) Crouch, L.I., Liberato, M.V., Urbanowicz, P.A., Baslé, A., Lamb, C.A., Stewart, C.J., Cooke, K., Doona, M., Needham, S., Brady, R.R. and Berrington, J.E., **Madunić**, **K.**, Wuhrer, M., Chater, P., Pearson, J.P., Glowacki, R., Martens, E.C., Zhang, F.,Linhardt, R.J.,Spencer, D.I.R. andBolam, D.N. 2020. Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. *Nature communications*, *11*(1), pp.1-13.

8) Pirro, M., Mohammed, Y., de Ru, A.H., Janssen, G., Tjokrodirijo, R.T., **Madunić, K**., Wuhrer, M., van Veelen, P.A. and Hensbergen, P.J., 2021. Oxonium Ion Guided Analysis of Quantitative Proteomics Data Reveals Site-Specific *O*-Glycosylation of Anterior Gradient Protein 2 (AGR2). *International journal of molecular sciences*, *22*(10), p.5369.

9) Kotsias, M., **Madunić, K**., Nicolardi, S., Kozak, R.P., Gardner, R.A., Jansen, B.C., Spencer, D.I. and Wuhrer, M., 2021. A semi-automated, high throughput approach for *O*-glycosylation profiling of in vitro established cancer cell lines by MALDI-FT-ICR MS. *Glycoconjugate Journal*, pp.1-10.

10) Blöchl, C., Wang, D., **Madunić, K**., Lageveen-Kammeijer, G.S.M., Huber, C.G., Wuhrer, M., Zhang, T., 2021. Integrated *N*- and *O*-Glycomics of Acute Myeloid Leukemia (AML) Cell Lines. *Cells* 10, 3058.

11) Rodriguez, E., Boelaars, K., Brown, K., **Madunić, K**., van Ee, T., Dijk, F., Verheij, J., Li, R. J. E., Schetters, S. T. T., Meijer, L. L., Le Large, T. Y. S., Driehuis, E., Clevers, H., Bruijns, S. C. M., O'Toole, T., van Vliet, S. J., Bijlsma, M. F., Wuhrer, M., Kazemier, G., van Kooyk, Y. 2022. Analysis of the glyco-code in pancreatic ductal adenocarcinoma identifies glycan-mediated immune regulatory circuits. *Communications Biology*, *5*(1), 41.

12) Zhang, J., Zhang, Z., Holst, S., Blöchl, C., **Madunić, K**., Wuhrer, M., Ten Dijke, P. and Zhang, T., 2022. "Transforming Growth Factor-β Challenge Alters the *N*-, *O*-, and Glycosphingolipid Glycomes in PaTu-S Pancreatic Adenocarcinoma Cells." *The Journal of Biological Chemistry* 0 (101717): 101717.

13) Wang, D., Zhang, T., **Madunić, K**., de Waard, A.A., Blöchl, C., Mayboroda, O.A., Griffioen, M., et al. 2022. "Glycosphingolipid-Glycan Signatures of Acute Myeloid Leukemia Cell Lines Reflect Hematopoietic Differentiation." *Journal of Proteome Research*, 2022.

ACKNOWLEDGMENTS

PhD trajectory is a great collection of ups and downs, and applying for one was the best decision I ever made. I am so grateful for the things I have learned, the people I have met and the person I became because of it. It would have been impossible without the people I was surrounded by, who gave me love, patience, support, and knowledge.

First, I would like to thank prof. Gordan Lauc whose lecture sparked my interest in glycosylation. Thank you for giving me the opportunity to work in your lab and make my first steps in glycobiology. Special thanks goes to Irma who sent me an encouraging email to apply for the PhD position. Thank you so much for believing in me and supporting me in my master thesis project.

Next, I would like to thank prof. Daniel Kolarich for selecting me for the position, as well as his team members Falco and Andreia for helping me in my first steps of becoming a carbon girl. My sincere gratitude goes to my promotor prof. Manfred Wuhrer for giving me the opportunity and freedom to perform the research in his group. He was more than a supervisor, as he always tried to show me the positive side of everything, especially my own results. Thank you for always being so accessible and giving valuable feedback in record times. Many thanks to prof. Peter ten Dijke for being so inspirational and sharing his enthusiasm for my research.

Although it was difficult having many different supervisors, from each one of them I learned something different. Kathrin, thank you for guiding me through my first mass spectrometry steps. Stephi, it was incredible to work with you as we share the same passion for translational science. Finally, the third lucky charm, Guinevere, who successfully supervised so many students and with her super G magic helped me make my manuscripts much clearer and beautiful. Sorry for all the things I sent you at the last minute before the deadline and thank you for checking them regardless!

To all my dear friends I met in Leiden, who I consider my second family; Marti, Fanny, Alan, Alessio, Elena, Ana, Maja, Ieva, Leria, Roc, Thomas, Tamas and Sander. You have made rainy days sunny. I will always remember our Christmas dinners, conferences, wedding trips and boat rides. Most of all I thank my paranymphs, who have loved me and supported me no matter what. I could not imagine a better trio.

Marti, my loving, energetic, enthusiastic supporter who made me try so many crazy things. Fanny my zen-master, whose calmness and kindness always filled my heart with comfort.

My crazy PGC team, Tao and Di, we went through a lot together to make this method work. Tao thank you for all your help setting things up. Di, I enjoyed being part of your scientific journey and personal development. Big thanks goes to Carolien, my mentor in my first CPM days, when I still thought that the instruments would explode whenever I touched them. Agnes, Jan, Hans, Lisa and Wenjun I am so grateful for your help with troubleshooting. Oleg, thank you for guiding me through the fun world of statistics, R-scripts, and German sayings. Big thanks to Karli, Noortje, Vika, Manu, Noortje, Albert, Bas, Rosina, Florent, Bram, Paul, Peter and George for always being available to discuss science, and much more. Simone, you are a great teacher, I enjoyed so much to learn from you. Thank you Guusje, Wei, Steffen, Suzanne, Riemke and especially Christoph for your enthusiasm in preparing Christmas boxes and organizing parties.

To my greatest GlycoCan team- Martina, Ines, James, Max, Giulia, Thanos, Fanny, Roberta, Jelena, Ana, Leria, Hyun Il. We learned so much about each other and how to appreciate our diversity, which was an important life lesson. Thank you so much Helen and Daniel for guiding us in this process. Richard, Milli and Rad, thank you for your guidance during my secondment at Ludger, and special thanks to Max, who made it fun and relaxing. Sorry I made you work so hard. I would also like to thank my collaborators, who inspired me and trusted my enthusiastic research plans.

To all of my friends in Croatia, who always welcome me with their arms open, especially Teica, Bilja, Sandra, Mateja, Aco, Mandi, Ines, Zeljka, Nela, Ivana and Cuska. Darko, thank you for your selfless support during my PhD. To my greatest supporters, my sisters Gabi, Meri, Zrinka and Viki, I am so lucky to have you! Wouter, my love, thanks for being my designer, this thesis is so beautiful because of you. Thank you for being my steady rock, always bringing peace to my drama. Mama i tata, hvala što ste uvijek podrzavali moj izbor, iako se niste s njime slagali.