Unraveling mucin type o-glycosylation signatures of colorectal cancer

Madunić. K.

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).