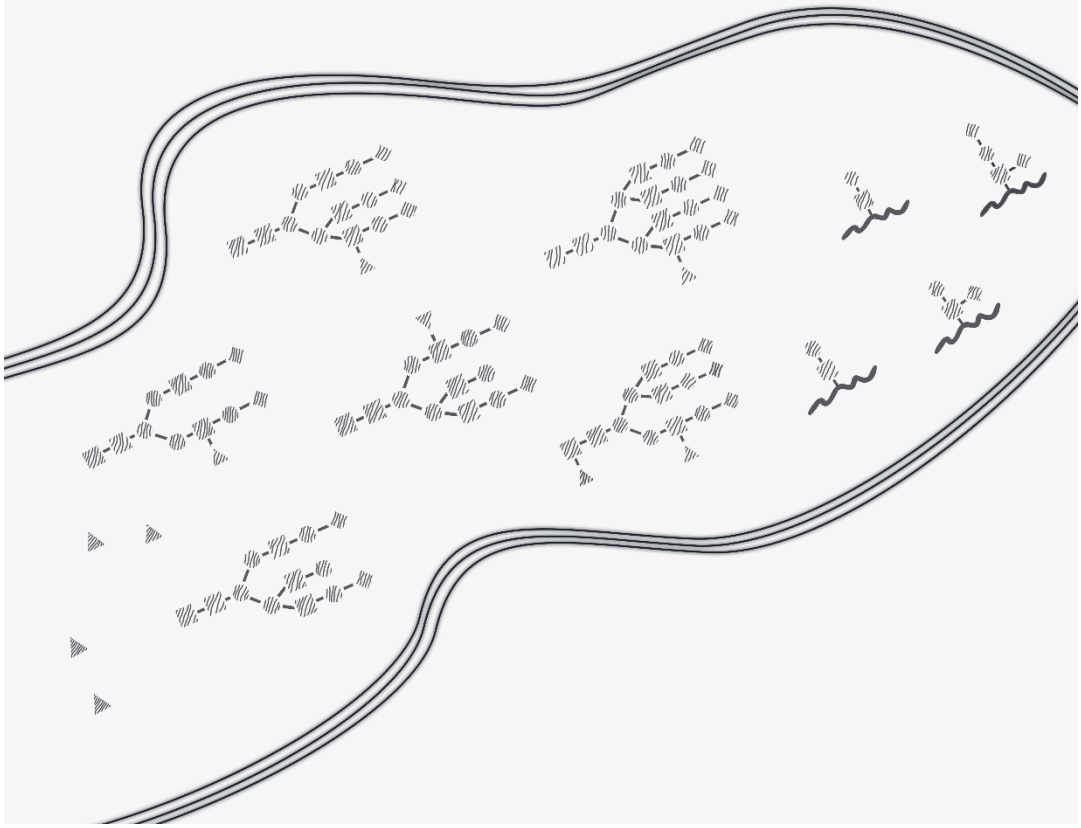
Biomarker discovery in diabetes mellitus and lipid metabolism: multi-platform glyco(proteo)mic approaches

Demus, D.A.

CHAPTER 7



7. Discussion and perspectives

Research presented in this thesis was performed as a part of the GlySign consortium, a highly collaborative programme between academia and industry initiated in 2016, which aimed at the advancement of medical glycomics and the establishment of novel biomarkers and diagnostic tools for precision medicine, within the therapeutic areas of allo- and autoimmune diseases, prostate cancer and diabetes. The initiation of this research was motivated by the already available research stressing the importance of glycosylation involvement in the pathophysiology of many diseases, the need of bringing robust assays for glycomics analysis in a clinical laboratory setting, and with facilitating the translation of current and new glycan-based biomarkers into the clinic.

In this thesis, development and optimization of analytical methods for the evaluation of diagnostic and prognostic glycan biomarkers were described. A range of challenges are faced while performing such evaluation, which derives from the technical limitations of analytical methods, the nature of research samples and biomolecules analysed, and with data processing. Some of these challenges were faced during the research presented in this thesis for which proposing and implementing solutions was required to secure the validity and quality of the results.

Technical obstacles occur often during research performed in academia due to the novelty of research topics. Novel analytical approaches and research discoveries often emerge from academia and are eventually adopted by relevant industrial sectors. A demand for new therapies, including biomolecules with increasing functional and structural complexity, is high. This in turn brings about the need for innovative, higher throughput test methods, which are used to characterise these biomolecules. However, implementation of novel analytical approaches in biopharma, in particular their manufacturing sectors, is challenging. Normally, these sectors adhere to test methods proposed by regulatory bodies, which are outlined in regulatory guideline documents. In highly regulated environments, for instance GMP regulated laboratories of biologic manufacturing organisations, such test methods tend to employ well-

established and validated gold-standard analytical approaches, leaving little room for change or innovation[231-233]. Considering the increasing complexity of therapeutic molecules, this leads to a range of bottlenecks and hurdles in characterisation of biopharmaceuticals during manufacturing phases. Those aspects will be further discussed in the context of research and analytical approaches described in previous chapters of this thesis.

Glycosylation in healthcare and precision medicine

Precision medicine is an evolving field in healthcare, which is seeking to create and provide healthcare with improved diagnostic tools and therapies. New preventive approaches and highly efficient targeted treatment strategies, which are applicable to groups of individuals who meet certain characteristics, are expected to serve healthcare providers and support existing models of care. Improvement patient life quality and reducing the burden on public healthcare sectors are the major drivers of change, which is expected to occur with the introduction of precision medicine programs.

Research presented in this thesis was built on the basis of the latest developments in precision medicine, diagnostic biomarker discovery and bioanalytical innovations in the fields of glycomics and glycoproteomics [51, 82, 137]. Analytical technologies are evolving with multi-dimensional characterisation of large clinical sample cohorts becoming possible[234-236]. With developments of omics (e.g. proteomic, metabolomic, (epi-)genomic, glycomic) databases, powerful analytical methods for characterising patients, and a growing variety of software tools for processing large sets of data and data modelling, this field of science will be able to deliver new insights into novel personalised therapeutic approaches for disease prevention and treatment in healthcare. Currently, a growing number of such approaches, which employ novel types of biomarkers[237-239], are being proposed and investigated as potentially applicable in clinical practice. The majority of new diagnostic and therapeutic approaches is being proposed by research performed in academia, however, their translation into clinically relevant tools has been slow. There are challenges and bottlenecks that contribute to the slow pace and lack of success in bringing new precision medicine tools into clinical practice, which will be discussed further[240].

In **chapter two**, an inter-laboratory evaluation of an *N*-glycan biomarker for stratification of patients with HNF1A-MODY was described[52]. In this study an LC-based method was applied to measure antennary fucosylation levels in 320 cases with HNF1A-

MODY and type 2 diabetes. A major strength of this research was fit-for-purpose development of an LC-based method, which specifically targets antennary fucosylated *N*-glycans facilitated by semi-automated sample preparation and data processing. Enhanced analytical capabilities of the LC-based method allowed for an in-depth evaluation of the *N*-glycan biomarker for HNF1A-MODY (covering six single glycan traits and a derived antennary fucosylation trait), which complements discoveries from the previous studies researching this biomarker[51, 53]. A significant aspect of this research was its inter-laboratory character. A set of samples measured previously in another laboratory and tested as a part of this study (**chapter two**) was evaluated in the context of consistency of antennary fucosylation measurements and provided excellent results, as demonstrated by the correlation analysis for three best performing traits. The inter-laboratory character of this study brings an important value to the field of precision medicine and glycomics. Glycomics is a relatively young field and this study is the first of its kind in regards to inter-laboratory evaluation of a single *N*-glycan stratification biomarker. Lack of such inter-laboratory comparisons and the consequent lack of evidence for the in-depth performance of novel biomarkers results in little interest in commercialisation of such biomarker-based tests thus their poor translation into clinical practice.

Unfortunately, the use of LC-based methods for clinical testing is poorly recognised in clinical practice. Therefore, the chances of applying such methods for clinical testing are low. In response to this situation, an enzymatic plate-based assay was developed, which is described in **chapter three**[55]. This assay was developed on the basis of research results on *N*-glycan biomarker for HNF1A-MODY using LC methods[51-53]. A significant consideration in designing this plate-based assay was to remove major bottlenecks, such as expensive instrumentation, low throughput and data processing, which limit translation of LC methods into clinically-relevant tests. The diagnostic performance of the assay, which has been optimised, automated and validated using a thousand clinical samples, was close to the performance of LC-based assay for this biomarker with an AUC of 0.87 and an AUC of 0.90, respectively. The study described in **chapter three** is

the continuation of the evaluation of this glycan-based stratification biomarker for HNF1A-MODY and, together with the study results from **chapter two**[52], provides valuable insights into a process of glycan biomarker evaluation and validation, and the inter-laboratory and translational aspects of these studies. There is a strong evidence on applicability of the glycan biomarker-based diagnostic assay for selected patients suspected of having MODY, which was presented in **chapter two and three**. However, a low frequency of HNF1A-MODY in the general population (approximately 1–4% of diabetes cases)[241], rather low risk of misdiagnosis-related consequences (commonly misdiagnosed as type 2 diabetes)[81] and the on-going crisis in public healthcare sectors diminish the chances of bringing this assay into clinics. Nevertheless, this study can serve as a proposition of a workflow for evaluation of novel biomarkers and design of an assay, which is a prototype of a diagnostic/screening test with a potential to be further developed and applied in clinical practice. Such studies are expected to facilitate glycan biomarker research by opening the door for similar studies testing technical aspects of novel biomarkers, such as their performance across laboratories.

A significant aspect, which has to be considered while commercialisation of diagnostic tests and is typically not taken into consideration in academic research on clinical biomarker testing technologies, is adherence to *in vitro* diagnostic (IVD) regulations. IVD regulations are guidelines that establish the safety and effectiveness standards for diagnostic medical devices, enforced by regulatory bodies such as the FDA and EMA. While both ISO 9001:2015 and ISO 13485:2016 are international standards that provide guidelines for quality management systems (QMS), ISO 13485:2016 is more specific to the medical device industry. ISO 13485:2016 includes additional requirements for manufacturers of IVDs to ensure that medical device products meet appropriate standards of safety and performance, outlining a commitment to continuous improvement, risk management, and the documentation of processes and procedures. The ISO 13485 framework includes a set of standards for the development, production, and distribution of these products. Compliance with the IVD regulations is critical to obtain regulatory approvals. Adhering to these regulations ensures the quality and

reliability of biomarker-based tests, strengthening trust among healthcare providers and patients, and ultimately promoting adoption and usage of these tests in clinical practice. Of note, machine learning and AI-based modelling approaches that are developed with an intention to be applied for clinical decision-making and diagnosis also fall under medical devices regulations thus are controlled by the ISO 13485 framework. One of the requirements stated in the ISO 13485 framework is the implementation of Quality Assurance systems during R&D stages of such biomarker- or software-based tests. Fulfilling this requirement is critical, however, brings an overall escalation of development costs, and subsequently the costs of the final product. This aspect is considered as a significant impediment in the translation of novel biomarkers and diagnostic approaches into commercially-available clinical tests[242, 243].

Cohort case-control studies are a type of observational study, in which individuals are observed or certain outcomes are measured. In experimental studies researchers would introduce some form of intervention meaning that one or more factors are altered, and their effects examined. In observational studies, scientific questions are addressed by finding and evaluating associations, such as between either disease and exposure or analytically measurable changes that occur in relation to characterised clinical factors: a diagnosed disease or disease-related complication. Such analytically measurable changes that are found through association analysis could be further evaluated as potential new biomarkers[244].

Three well described sample cohorts were used in the studies described in **chapters two - six** of this thesis. A cohort of cases with HNF1A-MODY and type two diabetes collected in the UK and Croatia was studied in **chapter two and three[52, 55]**. In the following chapters, the DiaGene cohort was analysed to obtain apolipoprotein CIII (apoC-III) *O*-glycosylation profiles[141]. These profiles were then subjected to the association analysis, the outcomes of which were meta-analysed using GWAS data generated from the Hoorn DCS cohort (**chapter five and six)**[180, 245]. These three sample cohorts had been previously characterised in multiple other studies using various

analytical approaches to obtain GWAS and *N*-glycome profile datasets for thousands of individuals that provided blood plasma samples for these studies[28, 33, 246]. In **chapter four**, apoC-III glycosylation profiles were measured by a highly-automated method employing a liquid-handling platform and MALDI-FTICR MS[179]. A similar workflow had been previously applied to obtain *N*-glycosylation profiles from big cohorts of blood plasma samples[28, 247]. Unlike *N*-glycosylation profiles, which are stable and not affected by freeze-thaw cycles[248, 249], protein analyses reveal other challenges. Fresh human samples would be the most desirable in certain analyses, however, this is a rather unusual case scenario. Sample cohort collection is a lengthy process initiated by multi-centre collaborative efforts. Moreover, most of large sample cohorts are often used multiple times in various studies, which requires their long-term storage and transfer. It has been shown in the past that the freeze-thaw cycles cause greater protein degradation than due to long-term storage at -70^{o}C[250] and sample stability assessment has always been a part of bioanalytical method validation. Jian at al. tested the influence of long-term storage, freeze-thaw cycles and room temperature exposure on plasma samples in a study on apoC-III and it was reported that ratios of apoC-III proteoforms remain constant regardless of the exposure to various sample-handling conditions[251]. Despite the good stability of apoC-III proteoforms in stored plasma samples, these analytes might undergo certain modifications. For instance, repeated freeze-thaw cycles have an effect on protein oxidation in plasma samples. Due to the nature of the study analysing a glycoprotein and considering a large number of cases (n = 771, n = 2318 and n = 1571 in **chapter four, five and six**, respectively), new challenges arose, which put a new perspective on sample preparation optimisation and dataset processing in this study. Those challenges also drew attention to data quality of this and similar studies on apoC-III using MALDI MS.

Clinical biomarker discovery and validation relies on studying large number of samples, which require repeatable high-throughput analytical methods. The study described in **chapter four** addresses challenges posed by large-scale clinical research, overcoming a problem of plasma samples heterogeneity due to varying levels of protein oxidation

arising from storage conditions and sample processing[179]. A major strength of the method presented in this chapter is reduced methionine oxidation heterogeneity, which allowed for high-throughput analysis of apo-CIII glycosylation. The limitations of this method arise from the use of MALDI-FTICR MS which is high-end instrumentation and is not commonly found in research laboratories. Sophisticated analytical instrumentation is commonly found in centralised research facilities, which are accessible to many scientists (through collaborations) working in different research fields. These technologies, which are not meant for routine clinical applications, are used to overcome the limitations provided by commonly used, less performing instrumentation. 15T MALDI-FT-ICR MS provides mass spectra of superior quality compared to MALDI-TOF MS although the latter technology is more diffuse. In MALDI-TOF MS measurements, apoC-III glycoforms are detected as broad peaks due to low resolution and apoC-III oxidoforms are not resolved. This leads to the broadening and distortion of the apoC-III glycoform signals which can eventually overlap with other apoC-III proteoforms or other proteins affecting the quantification. On the basis of previous studies on apoC-III *O*-glycosylation by MALDI-TOF MS and the study described in **chapter four**, the severity of this problem depends on the method used for apoC-III enrichment from plasma/serum samples. The chance of overlaps between signals is higher when apoC-III is enriched using solid-phase extraction (SPE), such as C18- or C4-based SPE. In this case, other small plasma/serum proteins are co-enriched and measured in the same m/z-range of apoC-III glycoforms. The problem of overlaps is reduced, but not solved when a more specific enrichment method is used, for example using the immunocapture of apoC-III. Trenchevska and co-workers reported on an MS-immunoassay for the analysis of apoC-III proteoforms including glycoforms and truncated forms[252]. In this study, mono-oxidation of apoC-III glycoforms and overlapping of apo-$CIII_1$ with its truncated form apoC-$III_{1\text{-}A}$ are clearly visible. Since both oxidation and truncation of apoC-III glycoforms can occur to a different extent even in fresh samples, overlapping species might affect the accurate quantification of apoC-III glycoforms. Unlike the MS data curation step applied in the study described in **chapter**

four, the evaluation of peak shape of apoC-III glycoforms detected in hundreds of MALDI-TOF mass spectra has never been reported. Such evaluation would highlight the limitation of MALDI-TOF MS compared to MALDI-FT-ICR MS regarding quantification accuracy of species with similar m/z-values. The quality provided by MALDI-TOF MS seems to be sufficient to report differences between the expression of the apoC-III glycoforms in current studies (e.g. researching congenital disorders affecting *O*-glycosylation)[253, 254], however, for other applications that quality may not be sufficient as it may fail to detect more subtle differences.

Genome-wide association studies (GWAS) help to identify variations, called single nucleotide polymorphisms (SNPs), in genes statistically associated with a particular trait or disease. By studying genetic variants across many genomes, these significant associations emerge, and it becomes possible to gain insights into molecular mechanisms that are likely involved in disease development and identify novel potential drug targets[255]. In **chapter five**, apoC-III *O*-glycosylation including sialylation, and a genomic variants datasets from 2318 cases with type 2 diabetes being part of the DiaGene cohort were subjected to a genome-wide association analysis[221]. Additionally, associations between genetic variants and lipid biomarkers were meta-analysed against the Hoorn DCS cohort. In **chapter six**, in addition to the associations between apolipoprotein-CIII glycosylation levels and micro- and macrovascular complications of type 2 diabetes within DiaGene cohort, the associations of apoC-III glycosylation-linked genetic variants with the prevalence and incidence of type 2 diabetes micro- and macro-vascular complications were meta-analysed against the Hoorn DCS cohort[245]. Despite the potential of GWAS findings for uncovering novel associations between genetic variants and disease or other measurable traits, there are outstanding questions in these types of studies. In the instance when identified genetic variants are located outside coding regions, which most likely will not lead to direct changes in the structure of a protein, interpreting GWAS associations requires additional experimental validation. Commonly proposed approaches for the validation of GWAS associations

are combining GWAS with gene editing and cellular phenotyping or integrating GWAS with various omics data to investigate the functional manifestations[256].

Due to the immense complexity and inter-connectivity of molecular mechanisms in human physiology and, therefore, inability to fully mimic *in vivo* pathophysiological conditions under *in vitro* conditions, no functional experimental validation was performed in the studies presented in **chapter five and six[221, 245].** Nevertheless, various reports on molecular and genetic aspects of apo-CIII glycosylation, diabetes and lipid metabolism-related mechanisms aid in understanding the causality of research findings presented in these chapters. Research described in **chapter five** revealed genetic variants in *GALNT2* and *IFT172* genes, which had previously been directly or indirectly linked to *O*-glycosylation mechanisms, to be associated with apoC-III glycosylation and lipid levels[221]. It was concluded that the occurrence of these genetic variants could modulate lipid levels by altering glycosylation of apo-CIII or other proteins involved in lipid metabolism. Up to date, the presence of specific apolipoproteins on triglyceride-rich lipoprotein (TRL) particles, such as hepatic VLDL, has been recognised to regulate TRL metabolism by modulating their lipidation and secretion, lipoprotein lipase (LPL) mediated TRL hydrolysis, and TRL remnant uptake by liver[123]. There is a growing evidence of the involvement of glycosylation in critical mechanisms of lipid metabolism. For instance, the glycosylation status of apoC-III, specifically sialylation, affects the binding of TRL-rich particles to biglycan, a chondroitin or dermatan sulfate proteoglycan present in the arterial wall [136]. Further, it was demonstrated that the sialylation status of apo-CIII impacts its inhibitory capacity against LPL[135] and its affinity to hepatic receptors that play a role in TRL remnant uptake[59]. LPL, an enzyme degrading circulating triglycerides, contains two N-glycosylation sites, with glycosylation at N70 being crucial for LPL activity and intracellular trafficking[257]. Altogether, the presence or absence of varying glycoforms on both apo-CIII and LPL might significantly impact lipid metabolism through the interactions between LPL, apo-CIII containing TRL particles and hepatic TRL uptake receptors. However, the involvement of other unstudied protein intermediates, their glycosylation status or epigenetic factors cannot be dismissed. The

impact of glycosylation on lipid metabolism-related pathologic conditions might be larger than currently proposed in literature. As described in **chapter five[221],** all discovered genetic variants were directly or indirectly linked to glycosylation mechanisms, such as the initiation of mucine-type *O*-linked glycosylation for the genetic variants in *GALNT2* and subcellular trafficking between the endoplasmic reticulum and Golgi apparatus for the genetic variants in *IFT172* via a crosslink to the *NRBP1* gene. **In chapter six[245]**, interconnections between the previously studied genetic variants in *GALNT2* and *IFT172*, their effect on apo-CIII glycosylation status and development of type 2 diabetes micro- and macro-vascular complications were investigated. The most pronounced interconnection was found for retinopathy and genetic variants in *GALNT2*, suggesting that modulating apo-CIII glycosylation via *GALNT2* might be an important target for further research and the development of new preventive approaches for retinopathy. Of note, epigenetic factors play a critical role in gene activity without altering the DNA sequence[258]. Behavioural and environmental changes are known to introduce genetic modifications through epigenetic changes, with DNA methylation and non-coding RNA sequences playing key roles in the regulation of gene expression[259], which also have been specifically investigated in the pathogenesis of type 2 diabetes[260]. Moreover, certain medications used in type 2 diabetes treatment have been shown to induce epigenetic changes, potentially contributing to their therapeutic effects[261, 262]. Although sensitivity analysis and confounding approaches were applied in the studies presented in **chapters five and six**, which may partially diminish the potential effect of epigenetic changes related to such factors as medications, obesity or smoking, epigenetic aspects should be taken into account in future studies and during development of new therapeutic strategies.

Taking into consideration fit-for-purpose optimisation of the analytical method used to obtain apo-CIII *O*-glycosylation profiles and rigorous, systematic data curation applied in the studies on apo-CIII glycosylation described in **chapter four, five and six**[179, 221, 245], a new approach for the analysis and validation of novel biomarkers in large sample cohorts was proposed. By applying this approach, it was possible to assess the quality

of data. Although reporting detailed data quality metrics adds a significant value to research discoveries, it is not a standard practice and depends on awareness of research centres and scientific journals including editors and publishers. As stated earlier, MALDI-FTICR MS use can be a bottleneck in translating this type of biomarker into clinics. Nevertheless, results obtained in such large-scale studies with a focus on data quality and validity can serve as a base for further research, allowing for performing inter-study or inter-laboratory evaluations and developing simpler assays, such as the enzymatic plate-based assay proposed and described in **chapter three**[55].

In a future perspective, the collection of multi-dimensional datasets backed by strong data quality metrics will enable in-depth investigation of novel associations, for instance between glycosylation and GWAS or *N*-glycosylation and *O*-glycosylation, and facilitate studies, in which datasets (e.g. *N*-glycosylation profiles) obtained by different analytical methods or/and in various research laboratories are a subject of comparison and technical evaluation. Up to date, the performance of methods for high-throughput analysis of serum and IgG *N*-glycome have been reported in several studies, in which methods accuracy, repeatability and throughput were assessed[263, 264]. Such comparative studies are of high significance, especially in relatively young scientific fields such as glycomics. **Chapter two** presents development of a novel analytical method, which was applied to obtain antennary fucosylation profiles in blood plasma from a set of patients with HNF1A-MODY and type 2 diabetes[52]. These profiles were subjected to the comparative analysis and correlated with antennary fucosylation levels measured by a different method in another independent laboratory. By performing this research, it was possible to present the validity of both analytical approaches and confirm the validity of a glycan-based stratification biomarker, which both contribute to strengthen the value of glycosylation analysis and clinical applications of glycans.

In human (cohort) research studies, *in vivo* functional studies are usually unavailable and *in vitro* functional studies cannot fully represent the true *in vivo* scenario. In

chapter five[221], by combining GWAS and glycoproteomics data, it was possible to find a link between genetic variants and their functional manifestation that confirms that glycosyltransferases might be promising diagnostic and therapeutic targets, which is often indicated in cancer research.[265, 266] Further, in **chapter six**[245], the meta-analysis of genetic variants and type 2 diabetes complications allowed to strengthen the link between the previously investigated genetic variant in the *GALNT2* gene, its functional manifestation of apo-CIII glycosylation and the incidence of retinopathy. Altogether, the large cohort and comparative studies carry not only the advantage of statistical power and the potential to provide unique research discoveries in the field, but also valuable insights into the technical performance of analytical methods for glycosylation analysis.

Glycosylation in industry and biopharma research

Despite developments and increasing availability of diagnostic tools, which are accurate and allow early diagnosis, there is an urgent need for safer and more efficient therapeutic approaches. Taking into consideration the burden on public healthcare budget brought by the COVID-19 pandemic, increasing numbers of other high-cost chronic diseases and the economic crisis, a high demand for safer and broadly accessible therapies (including biopharmaceutical products) is warranted[267, 268].

Glycosylation is considered as one of the critical quality attributes in biologic manufacturing. Developments in IgG-type monoclonal antibodies, following bispecific antibodies and bispecific derivatives have shown that by designing and adjusting glycosylation patterns, it is possible to influence the half-life, immunogenicity and pharmacokinetics of therapeutic molecules. Glycosylation patterns have to be closely monitored throughout biologic development and manufacturing. Cell line engineering and adjusting fermentation process conditions by media supplementation approaches are common practices to control molecule glycosylation patters, alongside more sophisticated and expensive upstream-downstream processing strategies that employ downstream enzymatic remodelling and affinity-based separation of protein glycovariants[269].

Alternatively to the genetic disruption of glycan biosynthesis, the use of small molecules as glycosylation enzyme inhibitors has been proposed and evaluated by Shasha Li and co-workers[270]. This proposition falls under the media supplementation approaches and is considered operationally simple as well as cost effective, however, its application requires extensive case-to-case evaluation. The small molecules, which are applied in this approach, have to possess a range of critical characteristics - their efficiency must be maintained while the stability of expression systems and expressed proteins remain not affected. Furthermore, computationally-informed strategies can potentially be employed to increase the relative abundance of targeted glycan structures thus aid the production of more homogenous biologics with respect to

glycosylation[271]. Fisher et al. have reported a computational model, which has the potential to be utilised for glyco-engineering of biologics as well as further elucidation and advancing treatments for glycosylation-related diseases[272]. Nevertheless, achieving complete homogeneity of glycoforms of biologics remains a significant challenge.

A demand for characterisation and matching of glycosylation patterns, which is key to demonstrate batch-to-batch consistency and obtain regulatory approvals, is also increasing in a field of biosimilar manufacturing and has been proven challenging. Those challenges arise due to the increasing complexity of therapeutic biomolecules and insufficient power of current analytical test methods[273].

In **chapter two**, development of an LC-MS/MS method for analysing antennary fucosylation of blood plasma protein has been described. LC-MS is a standard test method to characterise glycosylation profiles of biologics in industry, often supported by exoglycosidases. **Chapter two** presents stages of method development, highlighting technical limitations, such as differentiating core and antennary fucosylation of *N*-glycans using LC-MS/MS without exoglycosidases, and low throughput of such methods when considering sample and data processing without the use of automated liquid-handling platforms and software. The limitations of LC-MS methods for glycan profiling, which are advised by regulatory bodies and routinely used in biologic development and manufacturing to characterise products and collect data for regulatory purposes, are significant. Those methods are not universal and often cannot fully characterise more complex therapeutic molecules, without being further optimised. Exoglycosidases, such as the one employed in the development of the LC-MS/MS method in **chapter two**, have served as a useful tool for the optimization and development of LC-MS methods for glycan profiling in biopharma, ensuring that these methods are fit for purpose. In many cases this optimisation approach may still be insufficient taking into consideration the increasing complexity of therapeutic proteins and their glycan patterns, as well as limitations of chromatographic separation (e.g. coeluting glycan

structures) applied in UHPLC systems. Thus, differentiation and quantification of critical glycosylation features, e.g. bisection or fucosylation, when *N*-glycan antennary fucosylation is also considered, may pose a challenge in the biologics manufacturing. In a study by Moran and co-workers, a more complex quantification approach of combining fluorescence and ESI mass spectrometry signals has been applied to procainamide-labelled blood plasma *N*-glycans with sialic acid linkage-specific derivatization via ethyl esterification separated using C18 reversed-phase UHPLC column[20]. This study addresses the problem of co-eluting glycan structures and benefits from differentiation of α2,3- and α2,6-sialylated *N*-glycans. Although the complexity of data analysis and the subsequent increase in time required for the analysis might limit the use of this novel quantification approach in industry, this approach can serve as an additional or orthogonal method to current analytical approaches.

As previously mentioned, low throughput of sample processing and data analysis in glycoprofiling are major bottlenecks of current standard test methods in biologic manufacturing, but also methods used during biologic discovery phases. Moreover, processing and interpreting MS and MS/MS data from glycan profiling analyses require highly specialised and trained operators, and this also contributes to the final product's cost. In **chapter three**, an exoglycosidase plate-based assay for detecting α1-3,4 fucosylation in blood plasma proteins was proposed as a diagnostic tool for patients with HNF1A-MODY[55]. Development of this type of assay, which enables the analysis of glycosylation features of proteins in an automated high-throughput manner and benefits from simplified data processing, is of great industrial importance. The employment of liquid handling platforms in industry is more and more common. Moreover, enzyme discovery together with protein engineering approaches will facilitate developments of such assays[85, 274]. In **chapter three**, α1-3,4 fucosylation levels in complex mixtures of proteins were measured using the exoglycosidase plate-based assay. Regarding the characteristics of the exoglycosidase used in this study, the enzyme allowed for measuring of α1-3,4 fucosylation levels from glycans attached to proteins in blood plasma matrix. Previously, a similar assay has been developed and

applied to IgG samples to measure *N*-glycan galactosylation levels, which also allowed for quantification of sialylation levels[31]. Considering the complexity of blood plasma samples and relatively low complexity of biologic samples, which are highly purified biomolecules and their matrix can be easily adjusted, this type of a plate-based assay might eventually find a broader application in biopharma industry. Not only can the time and cost of development and manufacturing phases be reduced, aiding delivery of new therapeutics and biosimilars thus increasing their accessibility to patients. Employment of such assays measuring absolute levels of glycosylation features will add another analytical dimension to biomolecule characterisation. This may be essential for a more comprehensive evaluation of the quality of biopharmaceutical products and to ensure their safety for use in patients..

MALDI MS, which was applied in studies described in **chapter four, five and six[179, 221, 245]**, has still rather limited application in biologics manufacturing. Peptide mapping and intact mass analysis for regulatory approval purposes are performed routinely using the gold-standard LC-MS/MS method. However, those methods still suffer from major drawbacks, for instance, introducing artificial protein modifications in peptide mapping[275]. Despite the growing throughput in areas of sample preparation and processing, which is largely due to the introduction of liquid handling platforms, data processing and data analysis remain major bottlenecks in large-scale delivery of biologics. This aspect draws attention to the availability of commercially available data processing software, which aids high-throughput processing of LC (MS) and MALDI MS data for a large number of samples and provides a range of data quality metrics, similar to the software applied in studies described in **chapters two[52], four, five and six**. Developments in commercially available semi-automated data processing software are being noticeable (e.g. BioPharma Finder by Thermo Scientific or Protein Metrics by Dotmatics), however, manual intervention is still required in the majority of glycan and peptide mapping analysis. Currently, the generation and collection of large omics datasets (genomics, transcriptomics, proteomics and metabolomics) is becoming a more common practice in industry. The growing interest of industry in multi-omics

technologies and benefits associated with their use, in regards to robust drug discovery as well as process monitoring, optimization and control, makes MALDI MS a method of interest. With the accessibility of state-of-art sequencers and mass spectrometry, and developments in the fields of bioinformatics and artificial intelligence (AI), it is now becoming possible to approach diseases by probing more complex and transient disease- and treatment-related molecular changes to select new drug targets and validate them. As described in **chapter five**[221], combining large data sets, partially generated by MALDI-FTICR MS, can lead to important discoveries linking glyco(proteo)mics and genomics. In the field of glycomics, where large data sets are being delivered but still little is known about molecular mechanisms of aberrant protein glycosylation, new discoveries on interconnections between glycomics and other types of omics are considered as drivers for establishment of new research and innovations. Since the evaluation of glycan composition and the optimization of glycosylation patterns for specific therapeutic applications are crucial during biologics delivery stages, innovations brought to the field of glycomics will contribute to the progress of both academic and industrial sectors, and thus aid the delivery of new precision therapies and biotherapeutics.