Development and application of alternative methods for profiling proteins N-glycosylation

Alessandro Quaranta
Development and application of alternative methods for profiling proteins N-glycosylation

Alessandro Quaranta

Academic dissertation for the Degree of Doctor of Philosophy in Analytical Chemistry at Stockholm University to be publicly defended on Friday 4 October 2019 at 10.00 in Magnélisalen, Kemiska övningslaboratoriet, Svante Arrhenius väg 16B.

Abstract

Glycosylation is a post-translational modification (PTM) that exerts profound structural and functional effects on the modified protein. Glycan synthesis and conjugation to proteins are regulated by a myriad of factors, both genetic and environmental, and are also influenced by external stressors. Glycosylation patterns are known to vary in correlation to a large number of diseases; therefore, it is possible to study such alterations to identify reliable biomarkers and help elucidate mechanisms underlying the disease. For these reasons, the development of analytical methods able to investigate the glycosylation of proteins in complex samples and to measure and characterize disease-related alterations is of great importance.

In this thesis, the development and application of rapid and small-scale methods for the analysis of the glycosylation pattern on specific proteins in biological fluids, with a high degree of automation and potential for parallel sample treatment, is presented.

Paper I illustrates a profiling method based on a microfluidic compact disc (CD) and its application to humans serum samples. The workflow integrated all the sample preparation steps, allowing a high degree of automation and sample treatment parallelization, significantly reducing the required processing time. In Paper II, a bead-based procedure for the immunoaffinity extraction of selected proteins from complex biological matrices was developed. This procedure improved and extended the applicability of the microfluidic CD method, increasing the flexibility and maintaining a good potential for automation. Paper III included a derivatization procedure in the bead-based methodology, to stabilize sialic acids for matrix-assisted laser desorption/ionization (MALDI) and to discriminate between connectivity isomers. Additionally, the method was applied to different biological fluids in order to highlight interperson variations of glycosylation. To increase the sample throughput, the method was scaled to a multi-wells format in Paper IV and subsequently applied to the investigation of alterations in the glycosylation pattern correlated to Alzheimer’s disease.

Papers V and VI focus on applications based on electrospray ionization (ESI). In Paper V, a source for paper spray ionization (PSI) was modified to create a new set-up to extend the applicability of this mass spectrometry (MS) technique to large biomolecules. It was possible to measure intact proteins, identifying many glycoforms together with other PTMs, as well as to characterize released glycans, performing structural analysis by tandem mass spectrometry (MS/MS). In Paper VI ESI-MS and the bead-based sample preparation method developed in Papers II, III, and IV were used for quantification of various glycoforms of intact proteins. Additionally, a travelling wave ion mobility spectrometry (TWIMS) MS/MS method was developed to structurally characterize the related N-glycans after enzymatic release.

The methods proposed in this thesis show valid approaches, which could be applied to investigate alterations of glycosylation at different levels, with potential implementation for biomarker investigation and development.

Keywords: N-glycosylation, Glycomics, Glycosylation Biomarkers, Intact Glycoproteins, Glycoform Quantification, Mass Spectrometry, Ion Mobility Spectrometry, MALDI-MS, Paper Spray Ionization, Microfluidics, Magnetic Beads, Immunoaffinity Purification, Nanobodies.

Stockholm 2019

http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-171844

ISBN 978-91-7797-783-4
ISBN 978-91-7797-784-1

Department of Environmental Science and Analytical Chemistry

Stockholm University, 106 91 Stockholm
DEVELOPMENT AND APPLICATION OF ALTERNATIVE METHODS FOR PROFILING PROTEINS N-GLYCOSYLATION

Alessandro Quaranta
Development and application of alternative methods for profiling proteins N-glycosylation

Alessandro Quaranta
List of Papers


The author was responsible for planning and performing the experiments and contributed to data evaluation and to the writing of the manuscript.


The author contributed to planning and performing the experiments, evaluating data and writing of the manuscript.


The author contributed to planning and performing the experiments, evaluating data and writing of the manuscript.


The author was responsible for planning and performing the experiments, for data evaluation and writing of the manuscript. The author also contributed to data analysis.

The author made substantial contributions to the development of the analytical setup, experimental work, data analysis and evaluation, and writing of the manuscript.

VI. A, Quaranta, M. Spasova, E. Passarini, I. Karlsson, L. Ndreu, G. Thorsén, L. L. Ilag, *N-glycosylation profiling of selected intact proteins by high-resolution mass spectrometry (MS) and glycan analysis using ion mobility-tandem-MS*, submitted for publication.

The author was responsible for generating the idea, planning and performing the experiments, data analysis and evaluation, and writing of the manuscript.
Populärvetenskaplig sammanfattning


Denna avhandling fokuserar på utveckling och tillämpning av analysetekniker för glykaner från specifika proteiner extraherade från humana biologiska vätskor såsom serum och cerebrospinalvätska. Flera tekniker som vanligtvis används för glykananalys är snabba och enkla profileringsmetoder. För en mer fullständig information om glykosyleringen krävs oftast mer komplexa och mödosamma förfaranden. I avhandlingen beskrivs utveckling av metoder som ger både snabb och tillförlitlig information om glykosyleringen av specifika proteiner. De färdiga metoderna har applicerats framgångsrikt på humana biologiska prover och är därmed potentiellt värdefulla för biomarköridentifiering.
Contents

List of Papers ........................................................................................................... 1
Populärvetenskaplig sammanfattning ................................................................. 3
Abbreviations .......................................................................................................... 7
1. Background ........................................................................................................... 9
   1.1 Glycosylation ................................................................................................. 10
   1.2 N-Glycan synthesis ...................................................................................... 14
   1.3 Variations in glycosylation .......................................................................... 17
2. Glycoproteins .................................................................................................... 20
   2.1 Health and acute phase proteins ................................................................. 20
   2.2 Antibodies as biopharmaceuticals ................................................................. 21
3. Analysis of N-glycosylation ........................................................................... 22
   3.1 Strategies for N-glycan characterization ...................................................... 22
      3.1.1 Pre-concentration of glycoproteins ......................................................... 23
      3.1.2 Analysis of released N-glycans ............................................................... 24
      3.1.3 Analysis of intact glycoproteins ............................................................. 26
      3.1.4 Analysis of glycopeptides .................................................................... 27
   3.2 Automated methods for increased throughput ........................................... 28
      3.2.1 Microfluidics lab-on-a-chip devices ...................................................... 28
   3.3 Mass spectrometry analysis ......................................................................... 30
      3.3.1 Matrix-Assisted Laser Desorption Ionization ......................................... 30
      3.3.2 Electrospray ionization ......................................................................... 31
      3.3.3 ToF mass analyzer ................................................................................ 32
      3.3.4 Tandem MS for structural elucidation of N-glycans ......................... 33
   3.4 Ion Mobility Spectrometry ......................................................................... 35
      3.4.1 Application of IMS to oligosaccharide analysis .................................... 37
   3.5 Ambient MS and Paper Spray Ionization .................................................... 39
4. Results and Discussion ..................................................................................... 40

5
4.1 N-glycan profiling using microfluidic CDs ............................................. 41
  4.1.1 The Gyros CD .................................................................................. 41
  4.1.2 Method development and evaluation .............................................. 42
  4.1.3 Application for the identification of alcoholics .............................. 45
  4.1.4 Protein array CD .............................................................................. 47
4.2 Bead-based approaches ........................................................................ 48
  4.2.1 Method development and characterization ...................................... 48
  4.2.2 Glycoprofiling of biofluids from different individuals .................... 50
  4.2.3 Sialic acid derivatization ................................................................. 56
  4.2.4 Characterization of CSF glycosylation of AD patients ...................... 58
4.3 Intact proteins characterization ............................................................ 62
  4.3.1 SAPSI development and application to intact proteins .................... 63
  4.3.2 Application of SAPSI to intact TFN glycosylation ......................... 67
  4.3.3 Quantification of glycoforms of intact proteins .............................. 68
  4.3.4 Application to serum TFN and to TRA .......................................... 71
4.4 Structure elucidation of released N-glycans ....................................... 75
  4.4.1 SAPSI analysis of released N-glycans ............................................ 76
  4.4.2 ESI-IMS-MS/MS analysis of released N-glycans ............................ 78
  4.4.3 Application to neutral N-glycans released from TRA ...................... 79
  4.4.4 Application to acidic N-glycans released from TFN ........................ 81
5. Conclusion and perspectives ................................................................. 85
Acknowledgments .................................................................................... 87
References .................................................................................................. 90
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>α-1-Anti Trypsin</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>AGP</td>
<td>α-1-Acid Glycoprotein</td>
</tr>
<tr>
<td>AMS</td>
<td>Ambient Mass Spectrometry</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>ANTS</td>
<td>8-aminonaphthalene-1,3,6-trisulfonic acid</td>
</tr>
<tr>
<td>APP</td>
<td>Acute Phase Protein</td>
</tr>
<tr>
<td>APR</td>
<td>Acute Phase Response</td>
</tr>
<tr>
<td>APTS</td>
<td>9-aminopyrene-1,3,6-trisulfonic acid</td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>ATD</td>
<td>Arrival Time Distribution</td>
</tr>
<tr>
<td>CCS</td>
<td>Collisional Cross-Section</td>
</tr>
<tr>
<td>CD</td>
<td>Compact Disc</td>
</tr>
<tr>
<td>CDG</td>
<td>Congenital Disorder of Glycosylation</td>
</tr>
<tr>
<td>CDT</td>
<td>Carbohydrate-Deficient Transferrin</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese Hamster Ovary</td>
</tr>
<tr>
<td>CE</td>
<td>Capillary Electrophoresis</td>
</tr>
<tr>
<td>CFG</td>
<td>Consortium of Functional Glycomics</td>
</tr>
<tr>
<td>CID</td>
<td>Collision-Induced Dissociation</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal Fluid</td>
</tr>
<tr>
<td>DHB</td>
<td>2,5-dihydroxybenzoic acid</td>
</tr>
<tr>
<td>DMA</td>
<td>N,N-dimethylaniline</td>
</tr>
<tr>
<td>DTIMS</td>
<td>Drift-Tube Ion Mobility Spectrometry</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic Reticulum</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray Ionization</td>
</tr>
<tr>
<td>ETD</td>
<td>Electron Transfer Dissociation</td>
</tr>
<tr>
<td>Fuc</td>
<td>Fucose</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full-Width Half Maximum</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>Glc</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcNac</td>
<td>N-Acetyl Glucosamine</td>
</tr>
<tr>
<td>HAP</td>
<td>Haptoglobin</td>
</tr>
<tr>
<td>HILIC</td>
<td>Hydrophilic Interaction Liquid Chromatography</td>
</tr>
<tr>
<td>HSA</td>
<td>Human Serum Albumin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IMS</td>
<td>Ion Mobility Spectrometry</td>
</tr>
<tr>
<td>IPA</td>
<td>Isopropyl Alcohol</td>
</tr>
<tr>
<td>LacNac</td>
<td>N-acetyllactosamine</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid Chromatography</td>
</tr>
<tr>
<td>LIF</td>
<td>Laser-Induced Fluorescence</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix-Assisted Laser Desorption/Ionization</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>MSIA</td>
<td>Mass Spectrometric ImmunoAssay</td>
</tr>
<tr>
<td>NeuAc</td>
<td>N-Acetylneuraminic Acid</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PGC</td>
<td>Porous Graphitized Carbon</td>
</tr>
<tr>
<td>PLS-DA</td>
<td>Partial least Squares Discriminant Analysis</td>
</tr>
<tr>
<td>PNGase F</td>
<td>Peptide-N Glycosidase F</td>
</tr>
<tr>
<td>Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>PSI</td>
<td>Paper Spray Ionization</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-Translational Modification</td>
</tr>
<tr>
<td>QbD</td>
<td>Quality by Design</td>
</tr>
<tr>
<td>RP</td>
<td>Reverse Phase</td>
</tr>
<tr>
<td>RSD</td>
<td>Relative Standard Deviation</td>
</tr>
<tr>
<td>SAPSI</td>
<td>Solvent-Assisted Paper Spray Ionization</td>
</tr>
<tr>
<td>Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>SIMCA</td>
<td>Soft Independent Modelling of Class Analogy</td>
</tr>
<tr>
<td>TFN</td>
<td>Transferrin</td>
</tr>
<tr>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>ToF</td>
<td>Time of Flight</td>
</tr>
<tr>
<td>TRA</td>
<td>Trastuzumab</td>
</tr>
<tr>
<td>TWIMS</td>
<td>Traveling-Wave Ion Mobility Spectrometry</td>
</tr>
<tr>
<td>UOXF</td>
<td>Oxford Glycobiology Institute</td>
</tr>
</tbody>
</table>
1. **Background**

Proteomics addresses the large-scale and comprehensive identification and characterization of the entire set of proteins expressed in an organism, defined as the proteome, and its changes under the influence of external perturbations [1]. The study of the proteome involves handling a very high degree of complexity compared to the investigation of the genome. Whereas nucleic acids consist of pre-determined sequences of nucleotides encoding information, sequence alone cannot determine the function of a protein, since four levels of structure contribute to define it [2]. Moreover, proteins are subjected to a large variety of alterations that take place after their synthesis, further tuning and defining their functions [3]. Most proteins in fact are unable to perform their function as unmodified polypeptides, but require a permanent or transient chemical modification to acquire the desired activities [3].

Post-translational modifications are defined as covalent enzymatic or non-enzymatic modifications of proteins after their synthesis [4]. A common modification involves cleavage of a section of the synthesized protein, usually a signal peptide used to target the protein to its proper compartment in the cell, to generate the mature form of the protein [5].

Among the possible covalent modifications of amino acid side chains or to the termini of the polypeptide chain, methylation and acetylation are most typical, modulating gene expression and protein-DNA interactions [4,6]. Similarly, glycosylation is responsible for several structural and signaling processes, regulating the protein sorting in cellular compartments as well [3,4,7]. Phosphorylation, on the other hand, is a highly frequent reversible process that modulates cell signaling, activation and inactivation of enzymes, and molecular interactions [3,4,6].

Covalent additions often regulate structure and stability of proteins, together with signaling, recognition of receptors, cell localization, half-life, and enhancement of enzymatic activity, allowing the proper functioning of the cellular machineries [3,4,6,8]. As PTMs depend both on specific enzymes and on substrates, external and epigenetic factors can influence their consistency and outcome. At the same time, since PTMs influence important properties of proteins, an alteration could produce impairment in the protein function, potentially
causing diseases, as in the case of congenital disorders of glycosylation (CDG) [9]. For these reasons, the study of PTMs and their alteration can yield indications regarding mechanisms involved in a disease state, and open possibilities to develop complementary diagnostic approaches.

1.1 Glycosylation

The relevance of glycosylation has been evidenced in the recent years both due to its frequent occurrence and its implication in a large variety of biochemical processes [8]. Glycosylation consists in the covalent binding of a mono or oligosaccharide to the protein backbone, and is the most common permanent PTM: it is recognized that more than 50% of all the proteins are glycosylated [10]. Sugars can be attached to heteroatoms in the side chains of many amino acids, giving rise to different classes of glycans differing in type of bond, synthetic pathway, and functions [11]. With the exception of glycation, which consists in the non-enzymatic binding of glucose or fructose to the side chain of lysine residues [12], glycosylation occurs enzymatically. The two most common glycosylation types found on proteins are N- and O-glycosylation, but other types such as C-mannosylation and glyco-phosphatidylinositol (GPI) anchor attachments are also known [11]. N-glycans are bound to the amide nitrogen of asparagine (Asn) and have a shared pentasaccharide core (chitobiose core, figure 1), which is differentially substituted to give rise to the various final structures [13]. O-glycans, bound to the hydroxyl oxygen of serine (Ser) or threonine (Thr), can have eight core structures which can be further modified to produce their final structures (figure 1) [11].

![Figure 1](image-url)

**Figure 1.** Representation of the chitobiose core of N-glycans and of the eight possible core structures of O-glycans.
The present work focuses exclusively on N-glycans, as almost 90% of all the glycoproteins bind one or more N-linked glycan. The interest in the study of N-glycosylation has steadily increased during the last two decades, due to the involvement of N-glycans in a multitude of biological processes. Understanding the biological role of these sugars can lead to insight on mechanisms behind a large variety of diseases and to the development of precise N-glycans biomarkers [8].

The influence of N-glycans on chemical and biological properties of glycoproteins has been widely described in the literature [7,8]. Glycosylation plays an important role in the correct folding of glycoproteins [14], granting their correct function and issuing high stability and resistance to proteolytic and physicochemical degradation [15,16]. Serum half-life and mobility of glycoproteins are also glycosylation-mediated [17,18], as well as extracellular interactions, intracellular transport and communication, cell signaling, and tumor formations [19]. Moreover, glycosylation can modulate the affinity of binding partners, as has been illustrated for immunoglobulin G (IgG), for which modification to the glycosylation pattern have been shown to drastically reduce the activity and, in some cases, to cause severe immune reactions [20].

The linkage of N-glycans to proteins requires the presence of specific consensus sequences Asn-X-Ser/Thr (where X is any amino acid), defined as sequons [11]. Even though such sequons are quite frequent in protein sequences, not all the sequons are indeed glycosylated. The possibility that a sequon is glycosylated depends on many factors, especially on the identity of the amino acid “X” [21]. Additionally, the same sequon can bind a variety of N-glycan structures, giving rise to distinct protein glycoforms. A specific protein with various sites of glycosylation will thus have many glycoforms differing by site occupancy (macroheterogeneity) and by the different glycan structures present on each site (microheterogeneity) [22,23].

Heterogeneity of N-glycans is a consequence of the chemical complexity of the monosaccharide building blocks. In mammals, only ten monosaccharide are commonly found (table 1) [24].
Table 1. Identity, structure, and molecular weight of the ten monosaccharides commonly found in mammalian N-glycans.

<table>
<thead>
<tr>
<th>Structure (linear)</th>
<th>Structure (cyclic)</th>
<th>Sugar name</th>
<th>Sugar class</th>
<th>Molecular weight (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Structure" /></td>
<td><img src="image2" alt="Structure" /></td>
<td>β-D-Xylose</td>
<td>Pentose</td>
<td>150.13</td>
</tr>
<tr>
<td><img src="image3" alt="Structure" /></td>
<td><img src="image4" alt="Structure" /></td>
<td>α-L-Fucose</td>
<td>Deoxy-Hexose</td>
<td>164.15</td>
</tr>
<tr>
<td><img src="image5" alt="Structure" /></td>
<td><img src="image6" alt="Structure" /></td>
<td>α-β-D-Glucose</td>
<td>Hexose</td>
<td>180.16</td>
</tr>
<tr>
<td><img src="image7" alt="Structure" /></td>
<td><img src="image8" alt="Structure" /></td>
<td>α-D-Galactose</td>
<td>Hexose</td>
<td>180.16</td>
</tr>
<tr>
<td><img src="image9" alt="Structure" /></td>
<td><img src="image10" alt="Structure" /></td>
<td>α-β-D-Mannose</td>
<td>Hexose</td>
<td>180.16</td>
</tr>
<tr>
<td><img src="image11" alt="Structure" /></td>
<td><img src="image12" alt="Structure" /></td>
<td>β-D-Glucuronic Acid</td>
<td>Uronic Acid</td>
<td>194.14</td>
</tr>
<tr>
<td><img src="image13" alt="Structure" /></td>
<td><img src="image14" alt="Structure" /></td>
<td>α-β-D-N-acetylglucosamine</td>
<td>N-acetyl hexose</td>
<td>221.21</td>
</tr>
<tr>
<td><img src="image15" alt="Structure" /></td>
<td><img src="image16" alt="Structure" /></td>
<td>α-β-D-N-acetylgalactosamine</td>
<td>N-acetyl hexose</td>
<td>221.21</td>
</tr>
<tr>
<td><img src="image17" alt="Structure" /></td>
<td><img src="image18" alt="Structure" /></td>
<td>α-N-acetyleneuraminic acid</td>
<td>Sialic acid</td>
<td>309.28</td>
</tr>
<tr>
<td><img src="image19" alt="Structure" /></td>
<td><img src="image20" alt="Structure" /></td>
<td>α-N-glycolyneuraminic acid</td>
<td>Sialic acid</td>
<td>325.27</td>
</tr>
</tbody>
</table>

However, this restricted number of monosaccharides is not a limiting factor to the existing possible structural combinations. Free sugars are linear polyhydroxyaldehydes or polyhydroxyketones that undergo cyclization reactions between the ketone or aldehyde carbonyl group and one of the hydroxyl groups in the chain, yielding hemiacetals. The specific –OH group reacting with the carbonyl defines the size of the ring, which for hexoses usually has six members (pyranosic ring) or, less frequently, five members (furanosic ring). By this cyclization, the former carbonyl carbon becomes chiral and generates two possible anomers (α, β), in equilibrium with the open form. When polymerizing, the anomic
carbon of one sugar can potentially bind any –OH group in another, giving rise to a large variety of glycosidic linkages, which are further distinguished as α- or β-bond, according to the state of the anomeric carbon. The availability of different branching points is a key factor influencing glycan flexibility, compactness, and also physical and biochemical properties [25].

To simplify this complexity and obtain immediate ways of visualizing such structures, nomenclature systems have been developed from the Consortium of Functional Glycomics (CFG) and the Oxford Glycobiology Institute (UOXF) [7,26]. Both systems make use of specific symbols to identify the various monosaccharides, with color systems and specific positioning to distinguish between isomers. UOXF nomenclature implies knowledge of the location of each linkage, which is then indicated by different positions of the line around the associated symbol, while the anomeric state of the carbon involved in each bond is specified by depicting the bond with a dashed (α) or a continuous (β) line (figure 2 a, b, c) [26]. Since specific information on the site of each glycosidic bond is not always available, the CFG nomenclature could represent a more general way of representing glycan structures. In this system, distinct classes of sugars are represented by different shapes, and isomers are distinguished by colors. If information on anomeric state or linkage positions are available, they can be added as a text note on the line representing the bond (figure 2 d, e) [7].
Figure 2. Nomenclature systems used to describe glycans structures. a) UOXF graphical description of a bi-antennary bisected glycan. b) Different sugars are defined by shapes, and c) linkage location is indicated by line position around the symbol, while anomeric state is indicated by dashed or continue lines. d) CFG graphical depiction of the same glycan with known linkage information noted as text on the respective glycosidic bond line and e) legend of the symbols and color used to describe monosaccharid. Panels a), b), and c) were modified from Royle et al., 2006 [26], panels d) and e) from Varki et al., 2009 [7].

In the present work, glycans will be further depicted according to the CFG nomenclature. Additionally, specific glycan species will be named in the text by referring to their monosaccharide composition with conventional abbreviations: H, hexose; N, N-acetyl hexose; F, fucose; S, sialic acid; P, pentose; U, uronic acid, followed by the number of respective units.

1.2 N-Glycan synthesis

The addition of N-glycans to proteins begins simultaneously to the protein synthesis and can be defined as a co- and post-translational modification [13]. Unlike DNA transcription or protein translation, glycan synthesis is not a template-driven process. The addition and
removal of sugar units, which is mediated by enzymes present in the cell where the protein is being translated, can be affected by several epigenetic and environmental factors [27]. The whole process is depicted in figure 3.

Figure 3. N-glycan synthesis taking place in the ER and Golgi apparatus. Enzymes responsible for the various steps, and various classes of N-glycans produced are indicated.

In the first step, a lipid-oligosaccharide precursor consisting of a dolichol pyrophosphate unit (acting as carrier molecule) and a GlcNac2-Man9-Glc3 sugar, is moved into the endoplasmic reticulum (ER) [28]. When the specific sequons are encountered on the nascent protein, the glycosidic portion of the precursor is moved from the lipid carrier to the protein, and covalently bound to the nitrogen in the Asn side chain by the oligosaccharyltransferase [21]. At this stage, the protein is still unfolded and the hydroxyl groups of Ser or Thr are in proximity of the rather inert
amide group of Asn, increasing its reactivity [29]. For this reason, if the X amino acid is a Pro the process is less favored and the sequon can be left non-glycosylated, as Pro would cause a kink in the protein secondary structure preventing the formation of a loop between Asn and Ser/Thr [21]. Following the binding, the saccharide precursor is trimmed by glucosidases to remove two of the three terminal glucose units. In the next stage, the precursor is recognized by the lectin chaperone proteins calnexin and calreticulin, which sequester the nascent glyco-polypeptide chains and protect its correct folding, preventing aggregation and excretion of the incompletely folded chains from the ER [14]. Once the protein is properly folded, the last glucose unit is cleaved, together with a further mannose sugar. The protein is consequently transferred to the Golgi apparatus, where further reactions creating glycosylation microheterogeneity take place. If glycans are not further modified, they are maintained as high-mannose. This kind of glycans is typical of unicellular yeast [30], but in most multicellular organisms a further trimming is performed to leave only a GlcNac2-Man5 core, which is then modified in several ways by several glycosyltransferases that use activated nucleotide-sugar donors to add different sugar units and produce a large diversity of final structures [31]. The most common modification to the core is the addition of “antennae” consisting in one or repeated units of the Galβ1-4GlcNAc block, referred to as a type-2 N-acetyllactosamine or “LacNac” sequence [13]. Antennae are frequently capped by the binding of α-linked sugar units, mostly sialic acids, which prevent further addition of monosaccharides to the glycan chain [13]. Fucosylation can occur both on the chitobiose core (on the GlcNac sugar which binds the protein Asn) or on the antenna, depending to the enzyme responsible for fucose binding [32]. The inclusion of a bisecting GlcNac in the β-1,4 position of the central mannose in the N-glycan core is performed by the N-acetylglucosaminyl transferase (Gn-T) III, which is in competition with other Gn-Ts and with (fucosyltransferase) Fu-T VIII, responsible for core fucosylation (figure 4) [13]. Fucosylation on the antenna is performed by Fu-T IX on the GlcNac residue of the LacNac unit. The resulting structure is defined as Lewis-X motif or Sialyl Lewis-X motif, if the antenna is terminally sialylated. N-glycans can thus be divided in four main groups, namely complex, high-mannose, hybrid,
and bisected structure classes [7]. Figure 4 illustrates the structures of the different N-glycans classes.

**Figure 4.** N-linked glycans structure classes: complex, high-mannose, hybrid, and bisected, with evidenced motifs and structures described in the text.

Complex glycans are characterized by a variable number of antennae and are named after their branching as bi-, tri-, and tetra-antennary [7], while bisected glycans are predominantly bi-antennary and show only antennary fucosylation. Features present on bisected glycans are often referred to as “brain-type” glycosylation, as these oligosaccharides are usually present on proteins synthesized inside the central nervous system [33].

### 1.3 Variations in glycosylation

Unlike proteins, where the sequence is template-generated by a definite series of nucleotides, glycans are the result of hundreds of genes, coding for numerous enzymes (figure 3), which activities are modulated by substrate amounts and environmental factors [34]. It is estimated that
around the 5% of the human genome codes for glycosylation-related processes [34,35]. Genetic alterations such as mutations or polymorphisms can affect the glycosyltransferases by altering their affinity for the activated sugar-nucleotide substrates or for specific positions in the oligosaccharide chain. Such alterations can also target the nucleotide-sugar transporters, responsible for the presence of those substrates in the Golgi [35], and the overall effects can produce meaningful variations in the glycome of different individuals. The expression and activity of the enzymes taking part in the biosynthesis of glycans can vary in response to several internal and external factors. Chemical conditions in the Golgi play a major role in the final glycosylation outcome: increases of 0.2 units of pH, for example, markedly inhibits α-(2,3)-sialylation by relocating the specific sialyltransferase [36]. Also, the levels of glycosylation in IgG were considerably decreased in connection to decreased oxygen levels: overall galactosylation was reduced as a result of lesser oxidative phosphorylation in the production of the galactose-related nucleotide-sugar donor uridine diphosphate-gal (UDP-Gal) and by its reduced transport from the cytosol to the Golgi [37]. Finally, both biological factors such as age, gender, ethnicity, and pregnancy, and environmental factors, such as diet, exposure to pollutants, smoking, and alcohol abuse seem to have a strong effect on glycosylation pathways [38-41].

The complex combination of all these factors leads to pronounced differences in the glycome among different individuals [39,40]. On the other hand, glycan profiles are very stable over time in single individuals in the absence of perturbing factors [39]. The connection between cancer and alteration of glycome has been extensively studied in the last years [42-45]. Proteins in cancer cells show modifications in their glycan structures depending on both cancer type and stage of disease. This is due to alterations in cellular metabolism, dysregulation at transcriptional levels, and changes in both expression and localization of glycosyltransferases in the Golgi apparatus [45,46]. Increases of sialylation and altered expression of α-2,6-sialyltransferase I, for example, were correlated with the differentiation state of colorectal tumors, and the amounts of α-2,6-sialylated LacNac produced were correlated with the survival rate of patients in stages I and II [47]. Other modifications of sialylation have been linked to stomach and ovarian
cancer [48], whereas changes in the fucosylation degree and position were observed in hepatocellular carcinoma, lungs, and breast cancer (increase in core fucosylation) [49-52] and pancreatic cancer (increase in the degree of fucosylation) [53]. Common glycan alterations also consist in increased branching and addition of poly-LacNac units to the antennae, further capped to form sialyl Lewis-X structures [44,54]. Similarly to cancer, many other conditions can alter the health state of the cells and affect the glycosylation machinery. Inflammatory processes take place in almost every disease, from bacterial infections and injuries, to autoimmune and neurodegenerative diseases, and can significantly modify the expression of glycans on several proteins [41,55-59]. Since most of the serum proteins are synthesized by the liver, diseases targeting this organ are widely reflected in their glycosylation profile. Alterations in branching were observed in viral liver diseases (hepatitis B and C), while fucosylation degree and position varied in liver cirrhosis and fatty liver disease [60]. Even though fucosylation changes in these cases followed a pattern similar to the one observed for hepatocellular carcinoma, their quantification allowed discrimination between cancerous and non-cancerous states [50].

The relationship between diseases and glycosylation alteration is not one-sided: increasing evidence shows that aberrant glycosylation is not only a phenotypical effect but can also be related to the progression of the disease. Increases of fucosylation and sialylation on N-glycans present on membrane proteins expressed by cancer cells have been shown to confer many functional features related to cell proliferation, tissue invasion, resistance to chemotherapy, and metastasis potential [61,62]. Additionally, impairment of protein functions caused by aberrant glycosylation could also act as main or indirect cause for the insurgence of a disorder. One example is fatty liver disease, where increased levels of bisecting GlcNac in apolipoprotein-B decreases its excretion, resulting in severe accumulation and in the formation of fat vesicles due to abnormal lipid accumulation in the hepatocytes [63]. A class of disorders caused by genetic defects in the glycosylation system is the congenital disorder of glycosylation (CDG). CDG is caused by abnormal N- and O-glycosylation of proteins and lipids and has a prominent nervous system involvement [55].
2. Glycoproteins

2.1 Health and acute phase proteins

Acute phase response (APR) is a prominent innate systemic reaction triggered by perturbation of the healthy state of organisms resulting from infections, inflammations, injuries, surgeries, neoplasia, or immunological disorders [64,65]. During APR, the vascular and inflammatory systems are activated by pro-inflammatory cytokines, which also activate receptors on several target cells, triggering a systemic reaction resulting in physiological responses such as fever, anorexia, muscle catabolism, and altered protein metabolism [65]. As a result of APR, the activity of hepatocytes is strongly influenced by the interaction with cytokines, resulting in drastic alterations of the synthesis of proteins taking place in these cells [64]. Proteins that have modified concentration level of at least 25% following APR are defined as acute phase proteins (APPs) [65], and are divided into positive and negative APPs if their concentration increases or decreases respectively. APPs are an evolutionary conserved group of proteins which functions include modulation of inflammation, nutrient transport, pathogen opsonization, anti-oxidation, and protection against pathogen proteases [66,67]. While positive APPs protect tissues, sustain the inflammatory response, and remove pathogens and nutrients in a very efficient way, negative APPs can have inferior performance or even opposite effects [65]. For these reasons their synthesis is minimized during APR and the spared amino acids are used for synthesizing positive APPs, with concentration levels that can increase from 25% to up to 1000-fold [68]. Since most APPs are glycosylated, this boost or drop in synthetic activity can have substantial effects on the glycosylation outcome, as the milieu for both enzymes and substrates necessary for the glycosylation is modified [69]. The study of glycosylation during APR on APPs can thus provide synergistic information highlighting disease-related alterations, especially for issues generating chronic APRs, such as cancer, autoimmune diseases, neurodegenerative diseases, and rheumatoid arthritis [55-58,70,71]. Despite the fact that alterations induced by APR affect different proteins in the same way, glycosylation is also protein-specific, for reasons mostly depending on structural factors [72]. Hence, observing disease-
specific effects on the glycosylation pattern of specific proteins, leaving others mostly unaltered, gives the possibility to identify valuable biomarkers [72]. An example of a disease-specific biomarker is the prostate-specific antigen glycoprofiling for the diagnosis of prostate cancer [73]. Disease-specific alterations were found in several types of cancer, with increases of sialyl Lewis-X motifs, fucosylation, and sialylation [70]. These potentially valuable disease markers can be lost if the total glycome of biological fluids is analyzed, as glycans coming from high abundance proteins may level out differences coming from species of low abundance. Immunoglobulins are among the most abundant glycoproteins in serum and, being a major component of the adaptive immune system [74], their concentration and glycosylation are very prone to change in the presence of generic inflammatory processes [58]. Immunoglobulin glycans could mask changes coming from proteins specifically targeted by disease mechanisms, leveling out differences and hiding useful information [72]. Such effects stress the importance of profiling individual proteins, where possible, in order to obtain specific information on type and stage of different diseases.

2.2 Antibodies as biopharmaceuticals

Biopharmaceutical protein drugs such as monoclonal antibodies, chimeric antibodies, and recombinant proteins are finding wide use in treatment of diseases such as cancer, asthma, rheumatoid arthritis, immune diseases, and as vaccines [75]. From the 1970s, the development and establishment of the recombinant DNA and hybridoma technologies allowed the production of large amounts of protein drugs from engineered cells. Producing the drug in living cells ensures high structural similarity with endogenous proteins, with less adverse reactions and higher specificity with respect to small molecules [76]. On the other hand, as a consequence of the complexity of protein structure and of the expression systems, structural heterogeneity can arise and affect the drug performance and safety [76]. Glycosylation was introduced as a critical quality by design (QbD) standard in cell culture for developing manufacturing processes [77], stressing the necessity of understanding how the glycosylation outcome is affected by production conditions. The glycosylation profile of a biopharmaceutical depends on many factors, related to the culture conditions, the host cell systems, the
clonal variation, and the glycoprotein itself [78]. The choice of host cells is extremely important, as the presence of non-human structures, such as α-1,3-linked galactoses, have been related with both reduced half-life and immune reactions [75,79]. Chinese hamster ovary (CHO) cell lines have become the predominant host cell system for the production, as their glycosylation systems produce human-like patterns [80]. Culture conditions such as pH, oxygen content, and nutrient supply strongly affect the glycosylation as well. Values of pH outside the range 6.9-8.2 result in minimum occupancy rate of the glycosylation sites [81], while low oxygen content causes reduced galactosylation and fucosylation [82]. Glucose starvation in the culture media produces a shortage of glucose-derived nucleotide sugar precursors, resulting in reduced glycosylation and increase of high-mannose glycans [83]. The level of glycosylation and sialylation, the presence of high-mannose glycans, and the distribution of the main glycoform are known to play a decisive role in determining the pharmacological properties of a protein drug [84,85]. High mannose glycans and reduced levels of sialylation have been linked to shorter serum half-life and quicker clearance rates of the drug from the bloodstream [85]. Close control over these conditions allows the production of drugs with highly consistent glycosylation profiles. Quality control of the glycoforms throughout the product development, regarding both micro- and macroheterogeneity, is necessary to ensure that the quality of the drug is maintained over time and that no unexpected change occurs [76].

3. Analysis of N-glycosylation

3.1 Strategies for N-glycan characterization

Given the high biological significance, the wide structural variety, and the inherent complexity of glycans in biological samples, the development of a broad assortment of analytical methodologies and strategies have been promoted for the analysis and characterization of glycosylation over time [86-88]. Several approaches can be considered depending on the type of information that is sought and the type of sample investigated. Pre-concentration of the target glycoconjugate, separation of all the glycoconjugates from other chemical species, or
removal of high abundance species which could mask target analytes at lower levels have to be considered at the beginning of the analysis. Depending on the required information, three major strategies can be adopted for characterizing glycosylation, which provide different perspectives, relying on different technologies, and requiring specific sample treatments [86]. Microheterogeneity can be described by analyzing glycans after chemical or enzymatic release from the protein backbone. This approach provides information on the sugar composition and structure, and can be quantitative. Macrobheterogeneity information can be obtained by analysis of glycosylation directly on intact glycoproteins, providing knowledge on occupancy rate and, by mass difference, on the sugar composition of the expressed glycans. Structural information is however limited unless instrumentation capable of fragmenting the intact protein is used (top-down approach). Finally, analysis of glycopeptides after proteolytic digestion can give insights on identity, sugar composition and structure, site occupancy, and macroheterogeneity.

3.1.1 Pre-concentration of glycoproteins

The broad concentration range in which glycoproteins occur is a hindrance to the high-sensitivity investigation of low-level disease biomarkers in complex biological fluids [89]. The study of minor components requires isolation, enrichment, and separation from major constituents, which can otherwise hide important information. Lectins are proteins with high affinity for specific sugar structures, and are well known for their specific interaction with many types of glycans. It is estimated that more than 160 lectins are easily obtainable, and more than 60 are commercially available [90]. By using these proteins, alone or in an array, it is possible to retain and enrich most kinds of oligosaccharide motifs [91]. This can be exploited for the selective investigation of specific structural motifs on multiple proteins [92-94], or coupled to immunoaffinity reagents, for the inspection of such motifs on targeted proteins [92]. Immunoaffinity purification can also be applied to study glycosylation micro and macroheterogeneity on specific proteins [95]. While extremely efficient, the applicability of this approach is dependent on availability, quality, and price of the required immunoaffinity binder, as well as on a preliminary knowledge about the investigated system.
The depletion of major protein components is a common procedure used for facilitating the study of minor components of biofluids without having to isolate a single glycoprotein. Several commercially available immunoaffinity columns, targeting different combinations of high-abundance proteins can be used as necessary. The depletion of the most abundant species, seldom quantitative, will anyway expose the next most abundant group of proteins, making this approach not valuable for the analysis of trace proteins [97].

3.1.2 Analysis of released N-glycans

Release of N-glycans from a glycoprotein can be performed either chemically or enzymatically. The most widely used methodology consists in the enzymatic cleavage using peptide N-glycosidase F (PNGase F). PNGase F hydrolyzes the β-aspartylglycosamine bond between the side chain of Asn and the core GlcNac of all N-glycans, unless an α-3 fucose is bound to that GlcNac residue, liberating complete oligosaccharides [98]. Such fucosylation is sometimes present in plants and insects, and can be hydrolyzed by employing PNGase A [99]. Other glycosidases are more selective: endoglycosidase D and H cleave between the two GlcNac in the chitobiose core of hybrid and high-mannose N-glycan respectively, leaving one sugar on the protein [86]. In earlier times, when the availability of these enzymes was low, chemical release of N-glycans by hydrazinolysis was the method of choice. Hydrazinolysis is the only chemical method for maintaining an intact reducing terminal, which is necessary for most derivatization procedures required for the subsequent analysis of glycans. On the other hand, the reaction relies on the toxic and explosive anhydrous hydrazine (H₂N₂), and results in the degradation of the protein backbone [100,101].

Once released, glycans are usually separated by liquid chromatography (LC) or capillary electrophoresis (CE) and detected by spectroscopy or MS. Alternatively they can be detected directly without separation by matrix-assisted laser desorption/ionization coupled to time-of-flight MS (MALDI-ToF MS). LC separation can be performed by hydrophilic interaction liquid chromatography (HILIC), classic reversed phase (RP) LC, or by using porous graphitized carbon (PGC) as LC stationary phase [78,86,87,102,103]. Due to the polar nature of glycans, HILIC and PGC are often the methods of choice, while separation by RP chromatography
requires derivatization with a hydrophobic label in order to increase retention. Nevertheless, the resolution between different glycan species separated by RP LC remains suboptimal [102]. HILIC retention on the other hand depends on the hydrophilic potential of each glycan, which in turn depends on size, sugar composition, branching, and charge [86]. As a consequence, HILIC methods often allow the separation of structural isomers, which is typically more difficult with other methodologies [104,105]. PGC has similar ability in separating hydrophilic species with the advantage of tolerating a broader pH range compared to silica-based stationary phases [106]. The resolution achieved with PGC is comparable to HILIC, allowing the separation of branching isomers and anomers [107-109].

CE separation of glycans is usually coupled to either laser-induced fluorescence detection (LIF) or MS [110,111]. CE-LIF has the advantage of rather high resolution and sensitivity, but requires derivatization to incorporate chromophores and ionisable charged functional groups. Commonly used reagents fulfilling these requirements are 9-aminopyrene-1,3,6-trisulfonic acid (APTS) and 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS) [112,113].

As mentioned before, UV and fluorescent detection of glycans require derivatization. The most frequently used reagent for the purpose is 2-aminobenzamide (2-AB), a fluorescent tag which can enable detection limits down to the femtomole level [26].

MALDI-ToF is characterized by high analysis speed and low sample consumption, and shows potential for automation [114,115]. These characteristics make it suitable for large scale studies and for high throughput methods. On the other hand, sialylated glycans are not stable under MALDI conditions, due to the labile nature of the α-glycosidic bond between Gal and sialic acid [86]. To overcome this issue, sialic acids can be enzymatically removed using sialidase enzymes or derivatized to stabilize the bond [116]. The most common derivatization procedure is permethylation, which generates methyl derivatives for each free hydroxyl and amino group, as well as a methyl ester from the carboxylic group of the sialic acid [117,118]. Linkage-specific derivatizations, able to discriminate between α2,3- and α2,6-linked sialic acids have also been developed [116,119,120].
The use of tandem MS allows for structure elucidation by collision induced dissociation (CID) fragmentation experiments. Depending on the ionization, different kinds of fragmentation can be achieved. In positive mode ionization, glycosidic cleavage is dominant, yielding limited information on the position of the bond, while in negative mode ionization there is a predominance of cross-ring fragmentation, which can provide linkage-specific and branching information [121,122]. Alternatively, information on the sugar sequence of glycans can be obtained by sequential exoglycosidase digestion. These enzymes have a specific affinity for single monosaccharides and for specific linkage positions. Arrays of exoglycosidases can thus be used to sequentially remove terminal sugar(s) from an oligosaccharide chain and, by monitoring the corresponding mass shifts information on sequence and structure can be inferred [112,123,124].

3.1.3 Analysis of intact glycoproteins

Analysis of proteins without any enzymatic digestion or chemical modification allows measuring the mass of different proteoforms at the same time. Multiple PTMs can then be described together with glycosylation without requiring extensive sample preparation [125,126]. Intact proteins are usually purified by immunoaffinity capture [127,128] or sorted from complex samples by LC or CE [128,129], and subsequently analyzed by high-resolution MS after ionization by ESI, nano-ESI, or MALDI [125,130,131]. MALDI ionization mostly produces mono-charged ions, making it challenging to discriminate between proteoforms of high molecular weight [131], whereas ES-based ionization results in multiply charged ions which are easily measurable by most mass analyzers [132]. The multiply-charged envelope generated in ESI can then be deconvoluted to zero-charge mass values. Identification of the different glycoforms is consequently obtained by mass difference, and relies on the instrumental resolution. These methods usually have the advantage of being quick, suitable for screening purposes, potentially quantitative, and independent from the variable yield of enzymatic or derivatization reactions [126,133]. However, the MS ionization efficiency of large glycoproteins is often a limiting factor [134]. Furthermore, if the glycoproteins are ionized in a denatured state, a high degree of glycosylation could hamper the ionization, due to the
stabilizing effect of the glycans on the structure, requiring higher energies for the denaturation [16].

### 3.1.4 Analysis of glycopeptides

Glycopeptides are typically obtained by protease digestion of isolated proteins [88,135,136]. Digestion with trypsin is widely used due to its ability to cleave the peptide bond at specific sites, allowing the prediction of the mass and composition of the produced peptides. Moreover, the sizes of the resulting peptides are typically manageable by most detection techniques. The potential generation of multiply glycosylated glycopeptides, the co-occurrence of non-glycosylated peptides, and the scarce reproducibility of the digestion makes this approach not always sustainable [137]. Other non-specific proteases, such as pronase and proteinase K, may be used to overcome some of these limitations, since their action produces shorter peptides, although that implies that the peptide fragments cannot be predicted beforehand [136,137].

After digestion, glycopeptides are separated using several LC or CE methods, as discussed for released glycans in section 3.1.2 [135,136]. In addition, lectin arrays can be used to enrich glycosylated peptides separating them from non-glycosylated ones, avoiding ion suppression due to the higher ionization efficiency of the latter [138]. Structural characterization is then performed by fragmentation experiments in tandem MS, mainly by CID or electron-transfer dissociation (ETD). CID fragmentation takes place mostly on the more labile bonds, producing preferential fragmentation of the oligosaccharide chain leaving the peptide unaltered. If more than one glycan structure is present on the same glycosylation site data interpretation can be complicated [139,140]. ETD fragmentation is a “reaction-type” fragmentation and it involves interaction with a gas-phase electron transfer agent, typically fluoranthene. This kind of fragmentation is effective in breaking the N–C$_\alpha$ peptide bond and, since the vibrational energy of the parent ion is not increased in the process, PTMs generally remain on the peptide [141]. In this way it is possible to determine the glycosylation sites by mass difference, providing information on macroheterogeneity. Combination of CID and ETD experiments can thus fully describe glycosylation heterogeneity at both micro and macro levels [139,140].
MS of glycopeptides from digested glycoproteins provides complementary information to the other approaches and is the more straight-forward way to obtain site-specific information on glycosylation [135]. On the other hand, time-consuming sample preparation and separation steps are required, making it unsuitable for high-throughput set-ups. Moreover, extremely reproducible protein digestion is difficult to achieve for highly glycosylated proteins due to the stabilizing effect of glycans on protein structure [142].

3.2 Automated methods for increased throughput

The complex sample preparation workflows and the long incubation and analysis times for glycosylation profiling have hampered the possibility of performing large scale glycomics studies for a long time. In the last two decades, the development of fast and reliable methods for parallel treatment of large number of samples was possible thanks to the advancements in miniaturization, automation, and integration of multiple processes. Multi-well formats employing derivatized beads or functionalized pipette tips (as in MS immunoassay, MSIA [127]) were implemented to increase sample throughput and decrease the reaction times of derivatization procedures and enzymatic digestions [143,144]. Additionally, the use of multi-channel automated pipettes and liquid handling robots allowed the parallel manipulation of a large number of samples and accelerated the sample preparation [145]. Microfluidics solutions were also proposed to integrate and parallelize entire workflows, as described in more detail in the following section.

3.2.1 Microfluidics lab-on-a-chip devices

Microfluidic devices consist of integrated and miniaturized processing units for handling samples and reagents in the nano- to micro-liter scale. Miniaturization introduces several important advantages, such as extremely limited sample and reagent consumption, high degree of automation, possibility of treating large number of samples in parallel, and reduced reaction times. The behavior of fluids in microchannels shows specific features, such as the pronounced importance of hydrophilic/hydrophobic interactions, surface tension, and capillarity.
Additionally, flows become laminar affecting the mixing mechanism of fluids, which then takes place mainly by diffusion [146].

Several microfluidic designs have been introduced in the last two decades to tackle many issues, and can be divided as passive and active, depending on the mechanisms used to move liquids. Passive devices use capillary forces and hydrophilic/hydrophobic interactions, whereas active devices make use of external equipment, such as motors and pumps. Among active microfluidics, centrifugal systems constitute a separate class, as they require a simple rotating motor to operate all the necessary liquid transfers [147].

The first applications of microfluidic systems for characterization of glycosylation consisted in the analysis of native and derivatized N-glycans by means of chip-based LC and CE [148,149]. In 2009 Bynum et al. developed a fully integrated lab-on-a-chip for enzymatic release, clean-up, separation of the released N-glycans on PGC LC, and nano-electrospray into a ToF MS. In contrast to typical overnight procedures, the chip made use of immobilized PNGase F reducing the enzymatic deglycosylation time to only 6 seconds and confining the whole procedure to 10 minutes, also allowing for quantitative analysis [150]. An application that includes a derivatization step and a PGC LC separation for N-glycans released from immobilized glycoproteins was also developed for MALDI-MS detection. This chip provided the separation of N-glycan isomers and the detection of unique glycan species from serum and tissue samples [151].

Centrifugal microfluidic discs have found broad application in biomedical analyses, but application to the study of glycosylation has been limited. In previous studies from our group, a microfluidic CD designed for MALDI-MS application and produced by Gyros AB (Sweden), has been successfully used to characterize the glycosylation of therapeutic antibodies [152,153].

Disadvantages of the microfluidics technology mostly affect the versatility of the different platforms. Valves designed to control flows and liquid routers for example, are affected by the surface tension so that their performance with liquids other than the one they were designed for can be worse. Moreover, the strong influence of the surface tension can limit the variety of solvents and mixtures that can be used in microfluidic channels. Additionally, channels for protein-based applications are often
chemically modified to become more polar, to avoid hydrophobic adsorption of proteins on the wall, making subsequent use of non-polar solvent or reagents difficult.

**3.3 Mass spectrometry analysis**

MS is probably the most widespread and universally applied detection technique in analytical chemistry, both for qualitative and quantitative purposes. MS measures the mass-to-charge (m/z) ratio of ionized analytes and can be used to obtain structural information of unknown species from fragmentation experiments in tandem MS. Ionization can be achieved in several ways, depending on the physicochemical properties of the analyte, on the complexity of the sample, and on sample pre-treatment, which uses specific solvents. Major mechanisms for ionization are electron, chemical, electrospray, and photoionization. Several ion sources able to perform ionization under vacuum or in ambient conditions and in gas, liquid, and solid phase have been developed. The two most common sources employed for the ionization of glycans are MALDI and ESI [87], which are discussed in the following subsections. After ionization, analytes may be separated according to their m/z ratio by several types of mass analyzers, which can be discriminated according to their resolving power and mass accuracy. The resolving power is defined for a single peak as \( \frac{M}{\Delta M} \), where \( M \) is the measured m/z and \( \Delta M \) is the measured width in m/z at “full-width half maximum” (FWHM). With sufficiently high-resolving, well-calibrated (high-mass accuracy) MS it is possible to obtain accurate m/z and elemental composition of the ion. For peptides, this usually requires at least 10000 FWHM, which can easily be obtained by modern ToFs, orbitrap, or Fourier transform ion cyclotron resonance (FT-IC) [87].

**3.3.1 Matrix-Assisted Laser Desorption Ionization**

Since its introduction in the late 1980s by Karas [154], MALDI has been extensively used for the ionization of biomolecules such as proteins, peptides, glycans, glycoconjugates, and oligonucleotides. The possibility of ionizing such large molecules depends on the mild ionization mechanism, which relies on a matrix compound for embedding the analytes and transferring the energy in a homogeneous and soft way.
MALDI is a solid state photoionization technique, performed under high vacuum. Samples are mixed with a large excess of a suitable matrix compound, allowed to dry and co-crystallize, and inserted in a high vacuum chamber, where a pulsed laser hits the crystals. The energy provided by the laser is mainly absorbed by matrix molecules, which expand into the gas phase causing intact analyte molecules to follow into gas phase in the expanding matrix plume. Ionization is thought to take place in gas phase through proton transfer mechanisms between photoionized matrix and analyte molecules, producing mostly singly charged ions. However, the exact MALDI ionization mechanism is still debated and not completely understood [155].

The choice of the matrix is critical for effective MALDI. A suitable matrix must be able to absorb the energy of the laser at the specific wavelength, be stable under vacuum, co-crystallize or effectively dissolve the analytes, be chemically inert, and promote ionization of the analytes. Released glycans are commonly analyzed in positive ionization mode as [M+Na]$^+$ ions, after removal or stabilization of the labile sialic acids, by using 2,5-dihydroxybenzoic acid (DHB) or derivatives as matrix compounds. Negative ionization of neutral glycans is not so common due to the extreme difficulties in forming negative carbohydrate ions under normal MALDI conditions. The use of 2′,4′,6′-trihydroxyacetophenone (THAP) as matrix, with the addition of dopants consisting in nitrate, chloride, or phosphate ions has resulted in satisfactory formation of negative ions, thus expanding the applicability of the technique [156].

In the early 2000s, Armstrong et al. introduced a new class of ionic liquids (salts with melting point below room temperature) as suitable matrix compounds [157]. Such compounds form more homogeneous surfaces than conventional matrices, avoiding the formation of “sweet spots”, and improving resolution, beneficial for both MALDI imaging and quantitative applications.

### 3.3.2 Electrospray ionization

ESI occurs at atmospheric pressure in liquid phase or during transfer to gas phase. Analytes are dissolved in a proper solvent nebulized by a needle on which a high voltage is applied (± 2-5 kV). Nebulization results in the formation of small charged droplets, further desolvated by
drying gases and source heating. As the droplet shrinks, its surface charge density increases reaching its limit value (Rayleigh limit), thus overcoming the surface tension. Consequently, the charge repulsion causes the droplet to “explode” into smaller charged droplets (Coulomb fission). The formation of ions from these droplets is explained by two possible mechanisms: the charged residue model and the ion evaporation model. According to the former, fissions are repeated until the generated droplets contain only one single ion, while the latter suggests that the field strength at the surface is strong enough to “expel” a bare ion [158]. ESI is a soft ionization method which produces singly and multiply charged ions, including adducts with small cations or anions. Typically, large molecules form multiply-charged ions in ESI as a consequence of the ion production mechanisms. The charge present on the surface of the final droplet may end up on any of the ionisable sites on the molecule. Multiple charges allow for the measurement of intact macromolecules, as the resulting m/z falls within the operative mass range of most mass analyzers [158]. Compared with MALDI, ESI is a cooler process so that stabilization of sialic acids is not an issue. Glycans are ionized both in positive and negative modes as doubly and triply charged protonated/deprotonated ions. Due to their preferential coordination of sodium, Na adducts are also present both in positive and in negative mode. Single net charge results from the coordination of a sodium ion by a doubly-charged negative glycan ion. In general, such ions are not efficiently produced by ESI of glycans [122].

3.3.3 ToF mass analyzer

ToF separates ions according to the time they require to travel through a high-vacuum field-free flight tube. After the ionization, the ions are accelerated into the analyzer by the application of a potential pulse, with a velocity depending on their m/z ratio. The time between the application of the pulse and the registration of the signal by the detector is thus proportional to the m/z of each species. ToF resolution depends on many factors, such as the length of the flight tube, the focus of the ion beam, and the spread of the initial kinetic energy of ions with the same m/z. Pulsed delayed extraction, reflectrons, and orthogonal design are commonly used to correct for these factors and improve the instrumental
resolution. Multiple reflectrons can be used to further increase the resolution, but in each reflection ions are lost, and thus sensitivity. Today, the resolving power of modern ToFs can often be increased to over 40000 FWHM, at least in particular operation modes. Moreover, these analyzers can cover a very wide mass range and can be operated at very high frequencies, thus giving the possibility of obtaining mass spectra over a broad ion mass range extremely rapidly [159].

3.3.4 Tandem MS for structural elucidation of N-glycans

As mentioned in section 3.1.2, the structural elucidation of glycans can be performed by sequential exoglycosidase digestion or by tandem MS (MS/MS) fragmentation experiments. The drawbacks of the former approach are several (elevated cost, long time required, no information on the linkages between sugar units). MS/MS experiments can overcome these limitations and provide full structural information, but the data provided are typically more difficult to interpret. In 1988, Domon and Costello introduced a systematic nomenclature system for designating glycan fragment ions, in order to help the interpretation of fragmentation spectra and to facilitate the discussion and the evaluation of the results [160]. This nomenclature, illustrated in figure 5, has analogies to the one used for peptides, but introduces symbols for the description of branching and cross-ring fragments.

Figure 5. Domon and Costello nomenclature applied to a linear trisaccharide. Obtained with permission from [88], copyright 2013 American Chemical Society.

Fragments containing the reducing end are named X_i, Y_i, and Z_i, where i is the number of the glycosidic bond cleaved starting from the terminal
GlcNac. The respective fragments obtained from the loss of the portion with the reducing end are named $A_j$, $B_j$, and $C_j$, where $j$ is the number of the glycosidic bond starting from the non-reducing end. In case of branching, Greek letters are assigned to each antenna in order of decreasing molecular weight ($\alpha>\beta>\gamma$…). If the glycosidic bond is cleaved, fragments are named according to the retention of the glycosidic oxygen: Y and B if the oxygen stays on the reducing end and Z and C otherwise. Cross-ring fragments A and Z are preceded by two superscript figures indicating the sugar ring bonds that have been broken.

Fragmentation of glycans is frequently performed by CID, but recently electron-based methods such as ETD are finding broader application, due to the complementary information they provide [122]. In CID, ions are accelerated by a potential that increases their kinetic energy, and collide with molecules of an inert gas. After redistribution of the increased internal energy, a covalent bond at one of the most energetically favorable sites in the molecule breaks [161]. CID glycan fragmentation patterns differ considerably according to the type of parent molecular ion and to the ionization mode. Fragmentation of protonated ions in positive mode produces mostly glycosidic bond cleavage, with a predominant formation of B and Y ions. These ions provide information on composition and sequence of a glycan, but cannot be used to identify the linkage position. Additionally, loss or rearrangements of monosaccharides units have been reported for protonated species [162,163]. Fragmentation of sodiated species in positive mode on the other hand, is not affected by such phenomena, and results in increased rates of A and Z cross-ring fragments, which can be used to obtain structural information on linkage position [163,164]. Fragmentation of negative ions produces more cross-ring fragments for both deprotonated ions and anionic adducts. This effect was observed for both sialylated and anion-adducted neutral glycans [164,165].

ETD, the mechanism of which has been briefly outlined in section 3.1.4, produces fragmentation of the parent ion by the transfer of an electron from a radical anion to the analyte. Due to its non-ergodic mechanism, this kind of fragmentation is much faster and affects different bonds in the analyte compared to CID. ETD fragmentation of oligosaccharides has been reported to yield extensive cross-ring fragmentations, enabling the determination of linkage types and branching positions [166].
3.4 Ion Mobility Spectrometry

Ion mobility spectrometry allows separation of ions in the gas-phase according to their mobility in an inert buffer gas, under influence of a low energy electric field. Under these conditions, different ions interact with the field according to their charge, mass, and shape. The velocity of an ion in the tube is proportional to the field intensity according to the mobility constant $K$:

$$v_d = K \cdot E$$

where $v_d$ is the velocity of the ion in the drift tube and $E$ is the intensity of the applied electric field. The mobility constant $K$ is related to the ion and is defined as:

$$K = \frac{3q}{16N} \sqrt{\frac{2\pi}{kT}} \sqrt{\left(\frac{m + M}{mM}\right) \frac{1}{\omega}}$$

where $q$ is the charge of the ion, $N$ is the number density of the buffer gas, $k$ is the Boltzmann’s constant, $T$ is the absolute temperature, $m$ is the mass of the buffer gas, $M$ is the mass of the ion, and $\omega$ is the collisional cross-section (CCS) of the ion [167]. According to the equation, the mobility constant of an ion ($K$) depends on its mass and charge, but also on its shape, described by the $\omega$ parameter. The CCS is a molecular property representing the rotationally averaged surface area of a molecule in its ionic form. The dependence of $K$ on the CCS allows the IMS to separate isomeric species of the same charge state according to the ion shape. IMS provides a further dimension of discrimination to that of chromatography and MS [168].

The most widespread IMS technique is the drift-tube ion mobility spectrometry (DTIMS), which is the only technique providing a direct relationship between mobility and CCS, permitting its direct calculation [169]. The setup of DTIMS consists of a drift tube filled with a gas (typically helium, argon, or nitrogen) where ring electrodes generate a weak electric field (2-20V) [170]. The drift gas is commonly operated at low pressure (<15 mbar), but higher pressures can be used (up to 1 atm) if the electric field is increased in order to improve ion transmission. Ions are introduced in packets through an ion gate or a trap ion funnel in
discrete 100-200 µs time intervals [171] and separated according to their velocity under the drift tube conditions (figure 6a).

**Figure 6.** Schematic representation and separation mechanisms of different IMS technologies. a) drift-tube IMS; b) traveling-wave IMS or TWIMS; c) trapped IMS or TIMS; d) cyclic TWIMS. Modified from Jiang et al., 2019 [169].

In 2004, Waters Corporation developed a modification of DTIMS consisting in the application of a radiofrequency (RF) potential to move the ions through the drift tube [172]. In this setup, named travelling-wave IMS (TWIMS or T-wave), the drift tube consists of a series of stacked ring electrodes in which a RF voltage is applied to confine ions between two consecutive rings. A rapidly switched direct current (DC) voltage is superimposed to the RF fields to generate a wave, which is used to move ions in the tube. Ions are thus confined between two spikes and separation occurs since higher mobility species roll over the spikes and have a faster transit through the cell, while lower mobility species cannot keep up with the wave and roll back to a subsequent cycle (figure 6b). Compared to DTIMS, in TWIMS ion transmission and resolution are improved and can be further tuned by optimizing wave-related parameters. Moreover, ions will enter and exit the drift cell at the same potential (the electric field being transient), simplifying coupling to MS analyzers. On the other hand, CCS values cannot be directly calculated, but require careful calibration of the instrument with well-characterized systems [169].
In recent years, IMS resolution has drastically increased (20-50 times) by the introduction of new generation instrumentation, using different technologies such as trapped ion mobility spectrometry (TIMS), developed by Bruker (figure 6c) [173], and cyclic TWIMS, developed by Waters (figure 6d) [174].

3.4.1 Application of IMS to oligosaccharide analysis

The ability of IMS for the separation of isomeric species has found a broad application in the analysis of oligosaccharides in recent years. As described in section 1.2, glycans show high isomerism at multiple levels (figure 7): compositional isomers containing monosaccharides differing in the stereochemistry along a single bond (e.g. glucose and galactose), connectivity isomers differing in the binding site of two or more residues (e.g. α-1,3 and α-1,4 linkage), and configurational isomers differing in the state of the anomeric carbon (e.g. α-1,3 and β-1,3 linkage).

Figure 7. Classes of isomers in oligosaccharides: a) compositional isomers, containing monosaccharide building blocks differing in the stereochemistry by a single group; b) connectivity isomers, differing in the position of one or more glycosidic bonds; c) configurational isomers, differing in the anomeric state of the carbon involved in the glycosidic bond. Obtained from Hofmann et al., 2015 [175], with permission from Springer Nature.

Differences in connectivity and configuration have an impact on the three dimensional shape of an oligosaccharide, thus affecting its CCS.
2015, Hoffman et al. were able to separate connectivity and configurational trisaccharides [175]. IMS has since been successfully employed for the resolution of isomeric forms of small saccharides, but current instrumental resolution has not shown satisfactory separation for larger glycans [176-178]. Two main strategies have been employed to circumvent this issue: prior fragmentation of glycans and subsequent IMS separation of the generated fragments [179,180], and adduct formation with metal cations [181,182]. Prior fragmentation can be performed in instruments equipped with fragmentation cells placed between the ion source and the IMS cell. One example is the Waters Synapt G2-S, which is equipped with a tri-wave cell, allowing fragmentation both before and after IMS (figure 8).

![Figure 8. Tri-wave cell allowing for ion fragmentation before and after the IMS separation. If the parent ions are fragmented before IMS separation, the arrival time distribution of daughter ions is obtained. If the fragmentation is performed after the IMS cell, the arrival time distribution of daughter ions will reflect the one of the respective parent. Obtained from Hoffman et al., 2017 [176], copyright 2017 American Chemical Society.](image)

Fragmentation of a large glycan before IMS separation will produce smaller saccharides, where the difference in a single linkage or anomeric state will have a higher impact on the overall 3-D structure. Separation of these fragments by IMS is then easier and allows the resolution of the isomeric features and the structure assignment [180]. Coordination with group I and II cations has been used to improve the IMS resolution between oligosaccharides. Glycans have a preference for coordination with alkali cations, especially sodium, due to the ideal
distance between adjacent hydroxyl groups. This has been utilized to distinctly alter the three-dimensional shape of different species, thus improving the IMS resolution. For instance, penta and hexasaccharide isomers were resolved by coordination with Ca\(^{2+}\) ions [182]. Other than isomers resolution, T-wave IMS has also been used to separate N-glycans from other compounds in complex samples by selecting and allowing only species with glycan-related mobilities inside the mass spectrometer. In this way, contaminant compounds were not acquired and the signal-to-noise ratio highly improved [183].

3.5 Ambient MS and Paper Spray Ionization

Ambient mass spectrometry (AMS) includes a group of ionization techniques carried out in ambient and open air conditions, often coupled to high-resolution MS. Due to the minor or lack of sample preparation required, the time for such techniques is often determined by the MS analysis and is thus extremely reduced, enabling the development of fast and reliable screening methods with possibility of quantitation [184]. Among AMS techniques, PSI has become popular due to its simplicity, versatility, time- and cost-effectiveness, and still high sensitivity. The mechanism of ionization is the same as the one illustrated for ESI in section 3.3.2, with the basic difference that the emitter is a triangular tip of paper connected to a potential source by a metallic clip. In PSI the sample is spotted on the paper tip and allowed to dry. Subsequently an appropriate solvent is deposited and the potential applied. This leads to the formation of a Taylor cone generating the spray towards the MS inlet and consequently to ion formation [185]. Among the factors influencing the efficiency of the ionization, paper thickness, mechanical resistance, surface chemistry, and fiber morphology are the most important. Additionally, the shape and the angles of the paper tip and the sample-to-paper ratio are known to influence the emission of electrospray. Since its development in 2009, PSI has been mainly applied for the analysis of drugs and metabolites in biofluids, forensics, and food safety, and only a few studies have dealt with intact standard proteins and protein complexes [186].
4. Results and Discussion

The interest in glycosylation generated by the increased understanding of its biological role and its implications in biomarkers discovery has led to the development of a plethora of diverse approaches for the characterization of this PTM, as discussed in section 3.1. The possible techniques span from quick and simple profiling methods to complex procedures yielding full structural elucidation of a sample. Currently, there is still a lack of efficient methods, amenable for the analysis of many samples in parallel, for determining the glycosylation of individual proteins in complex matrices. Screening methods are typically applied for the characterization of glycosylation in total biofluids. In this way, information based on the altered glycosylation of single proteins, which could provide evidence of organ-specific issues for proteins predominately produced in specific organs (e.g. bone marrow, liver, or heart), is often neglected. The study of glycosylation on selected proteins typically entails long and laborious procedures to extract the target protein and perform deep structural characterization of the expressed glycans. While these approaches provide extremely important information, their implementation to assess changes induced by diseases in a clinical lab set-up is not optimal. Development of methods enabling the monitoring of such changes on selected proteins extracted from relevant biological fluids could help discover and validate glycan-based biomarkers.

This work has focused on the development and application of rapid and small-scale methods for N-glycans analysis of specific proteins, with a high degree of automation and potential for parallel sample treatment. All the developed methods employed microfluidics or bead-based multiwells immunoaffinity purification procedures to extract the target proteins from complex biological matrices. Glycosylation was subsequently investigated both from the analysis of released glycans and at the intact protein level, by MALDI-MS and ESI-MS detection avoiding time-consuming separation and derivatization steps. Additionally, methods allowing for both structural elucidation of released glycans and quantification of intact proteins glycoforms were developed.
4.1 N-glycan profiling using microfluidic CDs

Centrifugal devices are a class of active microfluidics systems that makes use of centrifugal forces to move liquids, without the need for external pumps or voltage supply, as all the operations can be performed by means of a simple rotating motor. These devices are commonly circular, shaped as compact discs (CDs), and suitable for parallelization [147].

In Paper I, a microfluidic CD produced by Gyros AB and designed for MALDI-MS application was the basis for developing a method for the characterization of desialylated N-glycans from transferrin extracted from human serum. In previous studies conducted by our research group, this device has been used to study and characterize glycosylation and other PTMs in biopharmaceutical antibodies [152,153]. The aim of Paper I was thus to adapt the method to a complex matrix such as human serum, and to demonstrate the applicability to the study of glycan biomarkers in serum samples. The target protein, transferrin (TFN), is a 77 kDa iron transporter protein synthesized in the liver with two glycosylation sites. Alteration in its glycan pattern in serum has been implicated in hepatocellular and stomach cancer [42,187], Alzheimer’s disease (AD) [55], and chronic alcoholism [188,189]. In particular, carbohydrate-deficient transferrin (CDT) is used as biomarker for chronic alcoholism. In Paper I, the glycosylation pattern of TFN in serum of healthy control and chronic alcoholic individuals was analyzed and the possibility of discriminating between the two sample sets was tested. The results were then compared to established TFN-based diagnosis methods for chronic alcoholism.

4.1.1 The Gyros CD

The microfluidic device used in Paper I is a 1.2 mm thick polyolephin CD, treated with oxygen plasma to become hydrophilic, coated with a polyelectrolyte solution to avoid protein adsorption to the surface, and covered with a lid by heat lamination. It consists of 6 sections, each containing 9 microstructures, for a total of 54 sample treatment positions (figure 9).
Each microstructure is composed of two microcolumns in series, connected by a liquid router, which is used to divert effluents from the first column either to the second column or to waste. Fluid motion is driven by centrifugal force and controlled by hydrophobic barriers that restrain the access to certain sections at low rotational speeds. Two inlets allow loading reagents, samples, and solvents onto the two columns selectively. A MALDI-MS dedicated sample spot is placed directly downstream of the second column. In this way, analytes can be eluted with a solvent containing the MALDI matrix and co-crystallized at the same time. The section containing the MALDI spots, sputtered with a conductive gold layer to avoid charge build-up during the ionization, is then cut and loaded in specific adaptors to fit commonly used mass spectrometers.

4.1.2 Method development and evaluation

The open architecture of the microfluidic CD allows a high versatility in the choice of the packing materials for the two microcolumns. To extract TFN from serum, streptavidin-coupled particles were packed in column 1 and were subsequently derivatized with a biotinylated anti-TFN affibody. Affibodies are 6.5 kDa single domain proteins which are engineered to have high affinity and specificity to the protein target [190]. Unlike many recombinant antibodies, affibodies are not glycosylated. The use of non-glycosylated binders permits direct deglycosylation of the immobilized target protein, without the necessity of performing a preliminary deglycosylation of the immunoaffinity binder or of eluting and recapturing the target protein. The biotinylated affibodies were captured.
on the stationary phase of column 1 by exploiting the strong streptavidin-biotin interaction. Excess of reagent was washed away by directing the effluent to the waste channel after the liquid router by spinning the CD at speeds lower than 1800 rpm. In this way the hydrophobic barrier on top of the second column would prevent hydrophilic solution to pass through, leaving the waste channel as the only available path. After sample incubation and TFN capture on the affibody, the N-glycans were directly cleaved on the immobilized protein by incubation with a mixture of PNGase F and neuraminidase, the latter used to cleave the sialic acids. The reaction was performed in 4 steps for a total of 60 minutes at room temperature, and the released desialylated glycans were directed to the second column by spinning the CD at speeds higher than 3500 rpm. At such speeds the centrifugal force is able to overcome the hydrophobic barrier and the flow is directed according to the rotational direction of the CD. Column 2, packed with PGC, was used to recapture the glycans, which were subsequently eluted with a solution of acetonitrile (ACN)/water (1:1) containing the MALDI matrix. A solution of 2.1M 2,5-dihydroxybenzoic acid (DHB) and 0.2M N,N-dimethylaniline (DMA) in ACN/H₂O (1:1) was used both for eluting the glycans and as MALDI matrix. Since the common MALDI matrix DHB tends to crystallize as large spiked crystals that can obstruct the second column outlet, DMA was added to the solution to produce small granular and homogeneous crystals, amenable for the microfluidics setup. Analysis was performed in positive MALDI-ToF MS, and the glycans were detected as sodiated adducts. A representative spectrum for glycans released from standard TFN is shown in figure 10.

Figure 10. MALDI-ToF MS spectrum of desialylated N-glycans obtained from standard transferrin. Glycans are acquired as [M+Na]⁺ ions and depicted according to the CFG symbols.
The observed glycosylation pattern of TFN consisted mostly of di- and tri-antennary complex glycans and their analogue fucosylated species, and was consistent with previous literature studies [191].

Sialic acids were cleaved in order to remove heterogeneity from the obtained spectra, and to increase the sensitivity for poly-antennary species. Heterogeneity can arise due to the fact that poly-antennary glycans with the same core structure, but with a different number of sialic acids, will give rise to several peaks. The removal of sialic acids causes merging of these peaks and an increase in the signal of larger glycans. The absence of derivatization procedures avoids increasing time, costs, and complexity of the methodology. In this way information concerning the sialylation level of transferrin is lost, but this approach is still useful if the information sought pertains to the core structure or branches of the glycans [192].

The method was evaluated regarding the selectivity of the immunoaffinity capture and the inter- and intra-CD precision. Mouse serum was analyzed as blank matrix to test for generic contamination showing the total absence of glycan-related signals. TFN was also spiked in mouse serum and the results showed no deviation from the relative intensity of the standard protein. In order to test cross-reaction with human proteins, TFN was spiked in a synthetic protein mixture, containing human serum albumin (non-glycosylated), IgG, and haptoglobin (HAP) at relevant serum concentration levels. In this case, minor traces of glycan signals arising from non-specifically bound proteins were found. These proteins were identified as IgG from the presence of fucosylated non-galactosylated glycans, such as H3N4F and H4N4F, typical of immunoglobulins. However, the intensities related to such contaminations never reached more than 10% of typical transferrin positive controls, and the relative area patterns were comparable to the ones obtained from the standard.

Short term and intermediate precision were evaluated, and a nested analysis of variance (ANOVA) with random effects was used to identify the most significant source of imprecision. Overall, the relative standard deviation (RSD) of the method was calculated to be 7.8%, 12.8%, 11.7% and 12.4% for H5N4, H5N4F, H6N5, and H6N5F glycanes respectively. The largest source of variance was identified in the variation of intra-CD
replicates, probably related to differences in the manually-packed PGC columns, which may influence the glycan elution.

4.1.3 Application for the identification of alcoholics

The TFN glycosylation profile was investigated in serum samples from healthy control (n=12) and chronic alcoholic (n=19) individuals. Although no statistical difference between the average values of the different glycan signals could be observed in the two sample cohorts, the variation appeared to be larger for the chronic alcoholics, especially considering the H5N4 and H6N5 glycans (figure 11).

Figure 11. Average values of the relative areas of N-glycans obtained from TFN in the two sample cohorts. Error bars are expressed as standard deviation of the average value (N=36 for control samples; N=48 for chronic alcoholics).

The observed alterations indicated a higher variability in the glycan branching in alcoholics compared to healthy controls, with extremely variable levels of the tri-antennary glycan H6N5. CDT, typically used as alcoholism biomarker, is mainly characterized by the absence of a complete N-glycan chain, but the presence of tri-antennary structures with a low degree of sialylation has been reported as well [189].
The dataset was analyzed by means of principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA). Due to the larger data variation in the alcoholics group, SIMCA was deemed a more suitable classification method than PCA. The SIMCA model was built on the control group, and the samples were classified as belonging or not belonging (figure 12).

Figure 12. a) PCA and b) SIMCA plots of the analyzed samples. The control samples (red x) have been used to create a SIMCA model. Samples that had sufficiently low Mahalanobis distance to the center of the model, and sufficiently low residual, were classified as non-alcoholics. Here 14 of the 19 samples from the alcoholics (blue circles) fell outside the decision limit, and were thus correctly classified.

A repeated fourfold segmented cross-validation of the SIMCA model gave 4% false positives, and 40% false negatives, resulting in a clinical sensitivity (ratio between number of true positives and the sum of false negatives and true positives) of 74%. This performance is similar to state of the art biomarkers, which have clinical sensitivity of 80% [188].

This microfluidics approach method is characterized by very short experimental time, allowing the analysis of up to 54 samples in a total time of 3.5 hours. Moreover, this approach implies a high level of automation, which could be further optimized by using pre-packed CDs, minimizing solvent transfer times and possibility of human error. A pronounced cost reduction is also achieved, due to the use of extremely small amounts of enzymes and immunoaffinity reagents, which are commonly very expensive. The sample processing time achieved with the very fast lab-on-a-chip device developed by Bynum et al. is 10 minutes per sample [150]. The CD enables the treatment of 54 samples in parallel reducing the preparation time to 3.8 minutes per sample, plus
the MALDI-ToF MS analysis time. Additionally, the immunoaffinity purification step integrated in our device is not provided in lab-on-a-chip systems, where total samples were directly deglycosylated on immobilized PNGase F.

4.1.4 Protein array CD

The use of a streptavidin-coupled substrate allows the coupling of any biotinylated immunoaffinity binder without major modifications to the general method. This versatility can be exploited to simultaneously characterize the glycosylation profile of several proteins belonging to a sample in the same CD, thus creating protein arrays targeting specific issues.

An array CD targeting six proteins, namely TFN, HAP, c1-inhibitor, α-1-acid glycoprotein (AGP), IgG, and α-1-antitrypsin (AAT) was proposed. Alterations in the glycosylation of these proteins have been linked to hepatocellular and ovarian cancer (TFN, HAP, and AAT) [50,123,193], acute inflammation, rheumatoid arthritis, and acute pancreatitis (TFN, AAT, AGP, IgG, and c1-inhibitor) [57,58,71].

Figure 13 shows spectra obtained from the analysis of the glycosylation of the six target proteins in standard solution [194].

![Figure 13. MALDI-MS spectra obtained from the analysis of standard proteins using the microfluidic CD protein array.](image)

The application did not require any modification of the general single-protein method, except for the loading of different biotinylated immunoaffinity binders in the different sections, performed
automatically by the liquid handling robot. Application to serum samples would enable the parallel glycoprofiling of the N-glycans core structures of up to six proteins in nine samples within 3.5 hours, simultaneously screening for several potential biomarkers, requiring only small amounts of samples and reagents.

4.2 Bead-based approaches

Bead-based methods have been extensively used to study protein complexes or interactions between proteins, but have found limited application in the study of glycosylation \[88,195\]. These approaches can be easily automated, multiplexed, and parallelized, with noticeable time and reagents savings. These advantages are evident compared to more traditional techniques, but are not as pronounced as in microfluidics and lab-on-a-chip devices. On the other hand, bead-based techniques gains in versatility, as they pose no limitations regarding the polarity of solvents or the type of reagents, and may also lead to lower limits of detection since the volume of the biological samples can be increased.

In Paper II, a magnetic bead-based immunoaffinity purification procedure was developed to purify target proteins from complex biological fluids. N-glycans were enzymatically cleaved and desialylated directly on-bead and then analyzed by MALDI-ToF MS. The procedure was improved in Paper III, with the introduction of a compatible process to derivatize and stabilize the sialic acids. The method was used to profile the glycosylation of six target proteins in serum and cerebrospinal fluid (CSF) from healthy donors (Paper II and Paper III), highlighting the presence of naturally occurring interpersonal variations and analyzing for the first time the individual glycosylation profiles of several proteins in CSF. In Paper IV the method was scaled to a multi-well format and applied on the investigation of alterations in the glycosylation caused by Alzheimer’s disease, both to look for potential biomarkers and to investigate mechanisms underlying the disease.

4.2.1 Method development and characterization

In order to keep the versatility in the choice of the immunoaffinity binders, streptavidin-coated magnetic beads were chosen as a substrate for the development of the immunoaffinity purification method.
Magnetic beads offer several advantages in handling procedures over other kinds of beads, as they can be retained with a magnet during all the washing operations thus providing a quicker and more efficient way of removing supernatants than long centrifugation steps. Moreover, as they possess a solid magnetic core, the possibility of non-specific adsorption of proteins is minor compared to beads displaying a “net” structure, which can act as a molecular sieve trapping proteins of compatible size. Commercially available biotinylated camelid homodimeric heavy-chain antibodies (VHH nanobodies) were chosen as immunoaffinity reagents. These engineered reagents are the equivalent to the Fab domain of single-chain antibodies from Camelidae and have extremely high affinity for their target antigen. Additionally, they are not glycosylated and show several properties that make them valuable for immunoprecipitation procedures, such as small size (15 kDa), long shelf life, and structural stability under harsh conditions [196].

The method developed in Paper II consisted in coupling of the desired VHH on the beads, extraction of the target protein from the matrix of interest, extensive washing to remove undesired components, and on-beads deglycosylation and desialylation. The released N-glycans were subsequently concentrated on PGC solid-phase extraction (SPE) columns and analyzed by MALDI-ToF MS.

The method was evaluated to calculate the incubation times required for binding the target protein and performing the deglycosylation. Binding efficiency was assessed by quantifying the protein in the beads supernatant at specific time intervals using a CE-LIF method developed for the purpose, showing saturation of the VHH after 30 minutes.

The deglycosylation efficiency was assessed by analyzing the supernatant after 2, 4, and 18 hours of enzymatic reaction; the three time points had no significant difference, both in terms of absolute intensity and comparing the normalized relative intensities between different glycans. Specificity of the immunoaffinity capture was evaluated by incubating the bead-bound VHHs with protein mixtures lacking the related target protein. While cross-reaction of the VHH was negligible (accounting for less than 5% of the total), the largest source of contamination was identified as the plastic of the vessels used for the procedure. Glycoproteins adsorbed on the plastic surface of the tube, in fact, would not be removed during washing steps and their glycans...
would be cleaved together with the target protein glycans in the presence of PNGase F. To minimize this effect, the beads bearing the target proteins were transferred to clean tubes just before the addition of the enzymes mixture. Inter- and intra-day variations were calculated for glycans released from AAT to estimate the method precision. Inter-day variation, expressed as RSD, was found to be smaller than intra-day variation, which was less than 10% for all the examined species.

The developed procedure had a total duration of 6 hours, including the SPE clean-up, allowing complete execution in a single working day. In Paper IV the method was scaled to a multi-well format, introducing automation and parallelization. An electronic 12-channel pipette and a DynaMag96 magnet, specifically designed for 96-wells plates, reduced the time required for solvent transfer steps. This time scale is comparable to other bead-based procedures for released glycan characterization, but also includes an additional protein purification step [144].

4.2.2 Glycoprofiling of biofluids from different individuals

In Paper II, the developed method was applied to the characterization of the glycosylation profile of TFN and AAT in five serum (individuals S1-S5) and eight CSF samples (individuals C1-C7), obtained from healthy donors. Paper III extended the application to four other proteins, namely HAP, c1-inhibitor, AGP, and IgG. The glycosylation profile in CSF of these proteins (except for IgG) was, to the best of my knowledge, observed here for the first time. Additionally, total serum and CSF were also analyzed by performing an in-solution deglycosylation on the crude samples and the results were compared to single protein variations.

N-glycosylation of the target proteins in serum consisted mostly in di- and tri-antennary mono- or non-fucosylated structures. Tetra-antennary species were observed in AGP, HAP, and c1-inhibitor, but not for every individual. IgG displayed a more varied pattern, with a large number of species having partial galactosylation. The varied pattern obtained for IgG in Paper III greatly differs from the one presented in section 4.1.3, obtained with the protein-array CD. The reason for this difference can be identified in the fact that the protein-array CD analyzed a single standard IgG, whereas in serum many different IgG proteins are present. Figure 14 displays the relative abundances of the identified glycan structures in the five analyzed serum samples. The measured areas are expressed as
the average of two replicate samples analyzed in duplicate, and are normalized to the sum of all the areas of each spectrum. The identity of each glycan can be retrieved in table 2, together with the measured m/z value and possible structure.

**Figure 14.** Relative abundances of various glycan species for the target proteins in the different serum samples and N-glycosylation of total serum. The x axis represents the different glycan species, detected as sodium adducts [M+Na]⁺, while in the y axis the normalized area values are reported. The identity of each glycan signal is displayed in table 2. Mean values were generated from two replicates analyzed in duplicate. Error bars indicate standard deviations.
Table 2. Identification of the desialylated glycan species discussed in Paper II, III, and IV, including m/z values of the sodiated adduct ions, possible structure, and sugar composition. Sugar composition described with conventional abbreviations: H, Hexose; N, N-acetyl hexose; F, Fucose. Structures depicted according to the CFG: Mannose, green circle; Galactose, yellow circle; N-acetyl glucosamine, blue square; Fucose, red triangle. Possible structures were obtained by searching the CFG, Carbbank, GlycomeDB, and Glycosciences databases, and drawn by Glycoworkbench software [197].

<table>
<thead>
<tr>
<th>[M+Na]+</th>
<th>Sugar composition</th>
<th>Possible structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1257.42</td>
<td>H5N2</td>
<td></td>
</tr>
<tr>
<td>1282.45</td>
<td>H3N3F</td>
<td></td>
</tr>
<tr>
<td>1339.48</td>
<td>H3N4</td>
<td></td>
</tr>
<tr>
<td>1444.51</td>
<td>H4N3F</td>
<td></td>
</tr>
<tr>
<td>1485.53</td>
<td>H3N4F</td>
<td></td>
</tr>
<tr>
<td>1501.53</td>
<td>H4N4</td>
<td></td>
</tr>
<tr>
<td>1647.59</td>
<td>H4N4F</td>
<td></td>
</tr>
<tr>
<td>1663.58</td>
<td>H5N4</td>
<td></td>
</tr>
<tr>
<td>1688.61</td>
<td>H3N5F</td>
<td></td>
</tr>
<tr>
<td>1809.64</td>
<td>H5N4F</td>
<td></td>
</tr>
<tr>
<td>1850.67</td>
<td>H4N5F</td>
<td></td>
</tr>
<tr>
<td>1866.66</td>
<td>H5N5</td>
<td></td>
</tr>
<tr>
<td>1891.69</td>
<td>H3N6F</td>
<td></td>
</tr>
<tr>
<td>1955.70</td>
<td>H5N4F2</td>
<td></td>
</tr>
<tr>
<td>1996.72</td>
<td>H4N5F2</td>
<td></td>
</tr>
<tr>
<td>2012.72</td>
<td>H5N5F</td>
<td></td>
</tr>
<tr>
<td>2028.71</td>
<td>H6N5</td>
<td></td>
</tr>
</tbody>
</table>
CSF N-glycosylation exhibited a completely different pattern, as illustrated in *figure 15*.
Figure 15. Relative abundances of various glycan species for the target proteins in different CSF samples, and N-glycosylation of total CSF. Three out of the eight samples were compared for each protein to illustrate the interpersonal differences. The x axis represents the different glycan species, detected as sodium adducts [M+Na]+, while in the y axis the normalized area values are reported. The identity of each glycan signal is displayed in table 2. Mean values were generated from two replicates analyzed in duplicate. Error bars indicate standard deviations.
Approximately fifteen glycan species could be identified for each target protein, with a high incidence of potentially bisected structures, typical of “brain-type” glycosylation. These structures accounted for about 6-15% of all the target proteins, except for AGP, where the amount was lower, and for TFN, where they reached 25%. Additionally, doubly fucosylated species such as H4N5F2 were detected at low relative abundance (<10%) in all proteins except IgG. The glycosylation profiles observed in CSF for IgG and TFN were in agreement with previous literature studies [198,199], while no literature data could be found for AAT, AGP, HAP, and c1-inhibitor glycosylation in CSF.

Due to the large complexity of the glycosylation patterns of proteins purified from CSF, differences between individuals were more noticeable if the focus was placed on selected glycan species. Moreover, for many species the differences observed in single proteins were not reflected when the total CSF was analyzed. An example of this is reported in figure 16.

![Figure 16](image)

**Figure 16.** Comparison of the glycosylation profile of the total CSF between two individuals. Differences across single proteins for the H6N5F species are highlighted (inset; N=3).

The total CSF N-glycan pattern of two individuals was compared, and single protein differences were highlighted for the tri-antennary H6N5F glycan to show potentially obscured differences. Even though no
significant difference could be observed in the levels of H6N5F in the total CSF, individual C7 had much higher rates of this species in HAP and AGP compared to individual C6. On the other hand, only individual C6 presented this glycan on IgG. Since IgG is the most abundant glycoprotein in serum and CSF, total fluid glycosylation patterns are strongly influenced by the ones of IgG. This highlights the risk associated with total biological fluid glycan profiling, as highly abundant proteins may obscure the changes in glycosylation of less abundant ones. Additionally, if IgG is considered, its glycosylation pattern is prone to be modified by more general inflammatory processes, potentially masking disease-specific effects taking place at the same time on other proteins.

4.2.3 Sialic acid derivatization

Alterations of sialylation have been reported in many types of cancer, both as consequences of the disease and as mechanistic mediators of, for example, cell proliferation, tissue invasion, and metastatic potential [48,62]. Unfortunately, as mentioned in paragraph 1.6.2, analysis of native sialylated glycans by MALDI-MS is extremely difficult. The presence of a carboxylic acid moiety on the anomeric carbon of sialic acids makes the glycosidic bond between them and galactose labile and prone to cleavage under MALDI conditions. As a result, the charged sialic acid unit is lost and the remaining uncharged glycan cannot be detected by MS. Among the many derivatization procedures implemented to stabilize this bond, Reiding et al. proposed in 2014 a selective esterification which allowed to discriminate connective isomers having α-2,6 and α-2,3 bound sialic acids [116]. By this procedure, 1-ethyl-3-(3-(dimethyl amino) propyl)-carbodiimide (EDC) and 1-hydroxybenzotriazole (HOBt) were used as activators in ethanolic solution (ethanol 95% v/v) to form ethyl esters from α-2,6 and lactones from α-2,3 bound sialic acids, introducing a mass difference of +28 Da (ethyl esterification) for the former and -18 Da (loss of water) for the latter (figure 17).
Figure 17. Differential derivatization of sialic acids with EDC and HOBt in EtOH 95% v/v. a) ethyl esterification of α-2,6 linked N-acetyleneuraminic acid, resulting in a mass increase of 28 Da and b) lactone formation from α-2,3 linked N-acetyleneuraminic acid, resulting in a mass decrease of 18 Da. Obtained from Reiding et al., 2014 [116], with permission from ACS publications.

In Paper III this procedure was adapted to the bead-based method developed in Paper II and included as optional step for the analysis of glycosylation-related issues where sialylation is a key factor. Briefly, after capture of the target protein, N-glycans were esterified by incubation with 0.25M EDC and 0.25M HOBt in ethanol 95% v/v for 1 hour at room temperature. After washing, the enzymatic release was performed by incubation with PNGase F alone. The procedure was applied to TFN, HAP and IgG extracted from serum samples (Figure 18).

Figure 18. Sialic acid esterification of N-glycans from purified serum proteins, with differentiation between α-2,3 and α-2,6 linked residues. a) HAP; b) TFN; c) IgG.
Sialylated TFN N-glycans consisted in only di-antennary species, differing by the linkage and number of sialic acids and by the possibility of fucosylation. H5N4S2 was the main species, having predominantly both sialic acids linked via an α-2,6 bond. No tri-antennary species could be observed for TFN, but they were detected in HAP with varying degrees of sialylation and at very low levels. As in TFN, the main species in HAP was H5N4S2 with both α-2,6 sialic acids. IgG were chosen as positive controls, since they express mostly non-sialylated glycans. However, in addition to the glycan structures already detected without derivatization, four additional sialylated glycans were identified, all with only one 2,6-linked sialic acid.

By comparing the results of the derivatization of TFN and HAP N-glycans to the analysis of desialylated N-glycans, a pronounced decrease in the number of core structures was observed. Only two core structures were identified for esterified N-glycans coming from TFN compared to four in desialylated samples. For HAP this effect was even more evident, with only three core structures identified versus eight in desialylated samples. Larger core structure can be lost due to variable degrees of sialylation and different types of linkage causing the signal to split into several peaks. For low abundance species this could cause the signals to fall below the detection limit. Therefore, when studying alterations related to fucosylation, bisection, and branching, desialylation remains the most valid approach. However, derivatization alternatives remain valid for issues related to sialylation or to where the linkage position carries discriminating power, as in colorectal cancer, where alteration of α-2,6-sialylation was correlated with the differentiation state [47].

Attempts to introduce derivatization procedures for the analysis of sialic acids to the microfluidic CD workflow were also performed, without positive outcome. This was either due to too fast evaporation of the reaction solvents or to the incompatibility of the required solvents with the polar channels of the device.

4.2.4 Characterization of CSF glycosylation of AD patients

The successful application of the bead-based method to biological fluids with the identification of naturally occurring interpersonal variations of the glycosylation in healthy individuals, demonstrated the possibility of employing the method to search for glycan biomarkers. In Paper IV,
alterations in the glycosylation profile of five selected proteins correlated to Alzheimer's disease were investigated by comparing CSF samples obtained from 17 AD patients and 20 healthy controls. The selected proteins, namely TFN, AAT, AGP, IgG, and c1-inhibitor, were chosen both for the availability of VHH antibody fragments and for their potential involvement in AD. Indications of AD-induced changes in the glycosylation have been reported for TFN in serum, with an increase of brain-type structures [41,200]; for AAT in CSF a decrease of the overall glycosylation levels was observed [201]. Changes in the glycosylation of AAT, AGP, and IgG in serum have been related to acute and chronic inflammation [58,202]. It has been demonstrated that AD is correlated with a chronic state of neuroinflammation caused by the triggering of the brain innate immune system by the presence of misfolded and aggregated proteins on glial cells [203]. Furthermore c1-inhibitor, which belongs to the same family of serpine protease inhibitors as AAT, was linked to ovarian cancer based on alteration of its glycosylation in serum [193].

In recent years, the importance of glycosylation and its role in the correct functioning of the central nervous system has been recognized. Studies on the influence of glycosylation in AD have also been conducted, both directly on amyloid precursor and tau proteins and generally on the enzymes responsible for glycosylation. Three main alterations to the glycosylation machinery were observed: underexpression of OST [204], overexpression of Gn-T III [205], and alterations in the activities of Fu-T VIII [206] and Fu-T IX [204]. As a consequence, one can expect to observe a decrease in the overall glycosylation, an increase in the amount of bisected species and an enhancement of antennary fucosylation. Consequently, poly-antennary N-glycans and core fucosylated species are expected to decrease. Since these modifications affect the core structure of N-glycans, neuraminidase was included in the enzymatic mixture to cleave the sialic acids.

No significant differences could be observed in the glycosylation profiles of TFN, AAT, and IgG between AD and control samples. AGP and c1-inhibitor, on the other hand, showed indications of AD-induced changes in glycosylation. The most abundant glycan from c1-inhibitor was different on average in the two sample cohorts, with H5N4 most present in AD samples and its equivalent mono-fucosylated H5N4F in control samples. This result was not anyway straightforward, since one third of
the control samples did not conform to the trend, giving rise to high standard deviation around the average value (figure 19a). N-glycans coming from AGP were di-, tri-, and tetra-antennary, mono- or non-fucosylated. AD samples showed decreased amount of tri- and tetra-antennary species, which did not result in a statistically significant difference with the control group, but was observable as a trend (figure 19b).

It can be observed that the number of glycans considered for Paper IV is limited if compared to what is shown in figure 15, related to Paper III. This is due to the selection made in Paper IV to include only relevant species in terms of intensity and presence in an appropriate number of samples. This was made after having verified the lack of specific presence/absence trends in the two sample sets.

Figure 19. Relative abundances of the different desialylated glycan species for the AD (red) and control (blue) sample sets obtained from a) c1-inhibitor and b) AGP. On the x axis, the identity of the identified glycans is reported (m/z calculated as sodium adducts, refer to table 2 for detailed information); on the y axis the area, normalized to the most abundant glycan for each protein, is reported. Error bars are reported as standard deviation of the normalized area values obtained for each glycan in each sample set. Possible structures of the identified glycans are included over each bar. Scores plots for the principal component analysis (PCA) of data obtained are reported in c) for c1-inhibitor and d) for AGP.
Analysis of data from c1-inhibitor and AGP by PCA revealed a partial separation (in the case of AGP) and a partial grouping for AD samples (c1-inhibitor). PCA results alone were not conclusive for the perspective of the observation of biomarker, due to the lack of clear separation between the groups. As these proteins seemed to carry discriminant power, partial least squares discriminant analysis (PLS-DA) models were built first on the single proteins matrices, and successively on the data set resulting from the concatenation of the scores from the two individual models. This final PLS-DA model was able to correctly classify 74±7% of the AD patients and 71±4% of the healthy subjects, leading to a total non-error rate of 72±3% (figure 20).

![Figure 20](image)

**Figure 20.** Scores plot for the mid-level data fusion PLS-DA model in cross validation. The model is built on the data set resulting from the concatenation of the scores from the two individual PLS-DA models related to AGP and c1-inhibitor. Control (16) samples are represented as blue boxes and AD (12) samples are represented as red dots.

The data generated for c1-inhibitor was limited by the fact that glycans were below the detection limits in one third of the samples; this was the case for 8 AD samples and 3 controls. This can be explained by the lower concentration of c1-inhibitor in CSF compared to the other target proteins, and by impaired glycosylation caused by genetic alterations in AD [201,204]. Alterations to c1-inhibitor fucosylation are less
straightforward, as the position of the fucose could not be assigned. Similarly, many potentially bisected species were observed, but the structures could not be confirmed by tandem MS due to the low intensities of the corresponding signals in the individual samples. The decrease in branching of N-glycans obtained from AD AGP could be related to the overexpression of Gn-T III reported by Akasaka-Manya [205].

Even if the method was not performing like the used AD biomarkers, the observed alterations in glycosylation were in agreement with previous results about enzymatic alterations in AD. This approach can thus be used to add insight to potential mechanisms mediating the disease. In this case, alterations were observed only in the two least abundant proteins. If glycosylation of total CSF had been studied, such changes would probably have been masked by IgG predominant pattern, stressing again the value of extracting information from single proteins.

4.3 Intact proteins characterization

The analysis of intact proteins by high-resolution MS provides information on the exact mass of different proteoforms at the same time. This approach enables the simultaneous characterization of the different glycoforms without the necessity of enzymatic reactions, derivatization procedures, or denaturation to expose buried glycosylation sites to enzymes or chemicals. Glycan masses are obtained by taking the difference between the measured mass and the mass of the deglycosylated protein, and can be used to calculate the sugar composition. However, since structural characterization is not immediate at this level, advanced instrumentation capable of fragmenting the intact protein is required.

In Papers V and VI, two methods for the characterization of glycosylation at the intact protein level were proposed. Paper V described the modification of a common AMS technique and its application to the analysis of large biomolecules. In Paper VI the bead-based method developed in Papers II-IV was modified and validated for the analysis and absolute quantification of the glycoforms of TFN extracted from the serum of healthy individuals and of the biopharmaceutical trastuzumab (TRA). In both methods the target proteins were ionized by ESI or ESI-like mechanisms. The charge
envelopes obtained were deconvoluted by the Waters MaxEnt1 deconvolution software, which retains quantitative information, allowing absolute quantification of the identified glycoforms under certain conditions.

4.3.1 SAPSI development and application to intact proteins

PSI has found wide application to the analysis of small molecules in complex matrices such as biofluids, but it has seldom been employed for the characterization of larger biomolecules [186]. Traditional PSI uses triangular paper tips (wicks) connected to a power supply as ESI emitters; the sample is deposited on the wick, allowed to dry and, after the addition of a proper solvent, ionized by application of a potential to the wick. The generated spray is directed towards the MS inlet for a time spanning between some seconds and a few minutes, depending on the solvent used for ionization. Solvent evaporation and spray consumption take place during the ionization, altering the conditions over time and causing signal instabilities [207]. This is especially the case for large and multiply charged molecules such as intact proteins, ultimately leading to a lower MS peak resolution.

In Paper V a modification of PSI that overcomes such limitations making this technique suitable for the analysis of large biomolecules was realized. This set-up, named SAPSI (solvent-assisted paper spray ionization), integrated the PSI support with the power supply and the fluidics system of a Waters Synapt G2S nano-ESI source. In this way it was possible to obtain precise and direct control over the applied potential, the solvent composition and the flow rate. A schematic overview of the SAPSI source is illustrated in figure 21.
SAPSI of intact proteins was first tested on human serum albumin (HSA) and TFN. Source parameters, solvent flow rate, and solvent composition were tuned concerning signal-to-noise ratio, resolution of single charge states, and resolution and reproducibility of the deconvolution. The signal stabilization effect generated by the continuous solvent flow is illustrated in figure 22.
As a result of solvent evaporation the resolution decreased and the charge increased, shifting the charge envelopes towards lower m/z (figure 22a, 22b). The application of a constant flow restored the signal and considerably extended the time of analysis, allowing acquisition of a larger number of scans, which improved both resolution and signal-to-noise ratio. Several solvent mixtures were tested, considering different parameters such as surface tension, dielectric constant (ε), viscosity, and boiling point. The best solvent for protein analysis was isopropanol (IPA)/water (1:1), with the addition of formic acid (1% v/v) to improve the ionization in positive mode. Such solvent is characterized by low surface tension (24 mN m$^{-1}$, at 25 °C), facilitating the desolvation, and medium-high values of ε (43, at 25 °C), improving the conductivity of the wetted wick. Changes in solvent composition had drastic effects on the ionization efficiency, as shown in figure 23 for TFN. By replacing IPA with ACN the surface tension of the solvent mixture increased (from 24 to 33 mM m$^{-1}$, at 25 °C), leading to less efficient desolvation and to a significant decrease in resolution, resulting in unresolved and inaccurate deconvolution (figure 23a, b).

Figure 23. MS spectrum of TFN with a) water/ACN (1:1), formic acid 1% v/v and b) MaxEnt1 deconvolution in the range 75-85 kDa. c) MS spectrum of TFN with water/IPA (1:1), formic acid 1% v/v d) MaxEnt1 deconvolution in the range of 75-85 kDa.
The developed conditions were applied to standard proteins and crude biofluids. The resolution of the obtained deconvolution spectra allowed the full separation of proteoforms with a mass difference of at least 15 Da. Figure 24 shows the deconvolution spectrum of standard HSA.

**Figure 24.** MaxEnt1 deconvolution of standard HSA performed with a resolution of 0.1 Da.

Several proteoforms produced by oxidation and cysteinylation of HSA were identified, as well as species resulting from coordination of alkali cations and from glycation. The presence of these species have already been reported in blood [208] and in commercially available HSA standards [209]. The intense signals related to cation adducts were due to the presence Na$^+$ and K$^+$, retained on the cellulose of untreated paper, which can cluster with the analytes.
4.3.2 Application of SAPSI to intact TFN glycosylation

Several glycoforms were identified by deconvolution of the spectra obtained by SAPSI analysis of standard TFN (figure 25).

![Deconvolution of standard TFN](image)

**Figure 25.** MaxEnt1 deconvolution of standard TFN performed with a resolution of 0.1 Da.

The main glycoform consisted of two di-antennary di-sialylated N-glycans H5N4S2, as expected from literature reports [191]. The other glycoforms always maintained one H5N4S2, while the second glycan differed for various sialylated and fucosylated bi- and tri-antennary species. Most glycoforms had intense signals due to coordination of one or two sodium ions, giving rise to M+23 and M+46 peaks. Sodium coordination was enhanced by the presence of large amounts of cations on untreated paper, as mentioned for HSA in 2.3.2.

The application proposed in Paper V showed for the first time the possibility of analyzing PTMs and protein modifications at the intact protein level by PSI. Several of the proteoforms observed in HSA could be used as biomarkers for different pathologies; for example, increases in cysteinylated and glycated HSA have been related to chronic kidney and liver diseases and diabetes mellitus respectively [210,211]. Characterization of these proteoforms by AMS could result in fast, cost-effective, and reliable screening methods to identify altered physiological states.
Attempts at investigating the glycosylation profile of TFN directly in biofluids by SAPSI were hampered by the presence of dominant protein species. Even though TFN is considered to be a high abundance protein, its concentration levels in blood are substantially lower than the concentration of HSA (1-3 mg/mL for TFN, 35-50 mg/mL for HSA). Analysis of diluted biofluids by SAPSI highlighted the presence of two protein species, namely HSA and apolipoprotein A1. The observation of less abundant proteins would require extensive sample pre-treatment such as immunoprecipitation or depletion of the most abundant species from the samples. Alternatively, the paper could be derivatized with a suitable immunoaffinity binder and thoroughly washed in order to remove unwanted components. The target protein could then be recovered by elution from the binder with an appropriate buffer and directly ionized and sprayed towards the MS inlet. Such solution could be easily implemented in a SAPSI setup but is not presented in this work.

4.3.3 Quantification of glycoforms of intact proteins

The absolute quantification of glycoforms of intact proteins extracted from biofluids by ESI-MS was explored in Paper VI. For the sample preparation, the bead-based immunopurification procedure developed in Papers II-IV was modified. The on-bead enzymatic deglycosylation was replaced with the elution of the protein from the bead-bound VHH. After elution, solvent exchange was performed by size exclusion spin columns to ammonium acetate (150 mM), an ESI-compatible solvent, and proteins were directly injected into the high-resolution mass spectrometer. No chromatographic separation was employed, except for a C18 guard column used to focus the injection band and to avoid ion suppression, resulting in a total analysis time of three minutes.

TFN was chosen as target protein due to its biological relevance and to its low degree of glycosylation, favoring the ionization under denaturing conditions. Additionally, the method was applied to the biopharmaceutical TRA in cell supernatant of production systems, as a possible application for quality control in drug manufacturing. In this case, given that TRA is a major component of cell supernatant, no immunoaffinity purification was required and the protein was analyzed just following dilution of the cell supernatant and solvent exchange to 150 mM ammonium acetate. TRA, sold under the commercial name of
Herceptin, is a recombinant humanized IgG1 with a molecular weight of 147 kDa, produced in Chinese hamster ovary (CHO) cells. It is used for the targeted therapy of HER2-positive metastatic breast cancer. TRA has two glycosylation sites, one in each heavy chain of the fragment crystallizable region [212].

Analysis of the two proteins in standard solution resulted in the identification of several glycoforms, as shown in figure 26.

**Figure 26.** Representative MS spectra of intact TFN and TRA, and subsequent deconvolution across the relevant mass range. **a)** MS full scan of TFN standard solution (0.1 mg/mL) and **b)** related MaxEnt1 deconvolution and identification of the different glycoforms. **c)** MS full scan of TRA standard solution (0.1 mg/mL) and **d)** related MaxEnt1 deconvolution and identification of the different glycoforms.

Five main TFN glycoforms could be identified: the most abundant expressing two H5N4S2 glycans (named TFN 1) and the others having one H5N4S2 together with different di- and tri-antennary mono- and non-fucosylated species (TFN 0 to TFN 4). The results are similar to the ones obtained by SAPSI in **Paper V** except for the peaks relative to sodium coordination. TRA showed mainly di-antennary N-glycans with different levels of galactosylation and possibility of fucosylation, giving rise to three major (named TRA 1-3) and five minor glycoforms (named TRA a-e). Details are reported in **table 3**.
Table 3. Sugar composition of the glycans associated with the target intact glycoproteins, theoretical mass, mass measured by deconvolution, and accuracy expressed as relative error in ppm. Theoretical mass values were obtained by the addition of glycan masses obtained with the Glycoworkbench software to the mass of the deglycosylated protein. Measured masses are average values of three standard replicates analyzed in triplicate. Sugar composition abbreviations: H, Hexose; N, N-acetyl hexose; S, Sialic acid; F, Fucose.

<table>
<thead>
<tr>
<th>Glycoform</th>
<th>Identified glycans</th>
<th>Theoretical mass (Da)</th>
<th>Measured mass (Da)</th>
<th>Accuracy (RE, ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFN 0</td>
<td>H5N4S2</td>
<td>77349.2</td>
<td>77348.3 ± 2.1</td>
<td>11.2</td>
</tr>
<tr>
<td>TFN 1</td>
<td>H5N4S2; H5N4S2</td>
<td>79555.0</td>
<td>79555.1 ± 1.3</td>
<td>-2.0</td>
</tr>
<tr>
<td>TFN 2</td>
<td>H5N4S2; H5N4S2F</td>
<td>79701.0</td>
<td>79701.6 ± 3.1</td>
<td>-7.5</td>
</tr>
<tr>
<td>TFN 3</td>
<td>H5N4S2; H6N5S3</td>
<td>80211.2</td>
<td>80211.3 ± 0.7</td>
<td>-0.9</td>
</tr>
<tr>
<td>TFN 4</td>
<td>H5N4S2; H6N5S3F</td>
<td>80357.2</td>
<td>80356.6 ± 1.0</td>
<td>7.6</td>
</tr>
<tr>
<td>TRA a</td>
<td>H3N3F; H3N4F</td>
<td>147852.2</td>
<td>147851.6 ± 0.9</td>
<td>4.1</td>
</tr>
<tr>
<td>TRA b</td>
<td>H3N4; H3N4F</td>
<td>147909.2</td>
<td>147907.4 ± 1.3</td>
<td>12.2</td>
</tr>
<tr>
<td>TRA c</td>
<td>H4N3F; H3N4F</td>
<td>148014.2</td>
<td>148011.1 ± 0.7</td>
<td>20.9</td>
</tr>
<tr>
<td>TRA 1</td>
<td>H3N4F; H3N4F</td>
<td>148055.3</td>
<td>148055.7 ± 0.4</td>
<td>-2.8</td>
</tr>
<tr>
<td>TRA 2</td>
<td>H4N4F; H3N4F</td>
<td>148217.1</td>
<td>148216.9 ± 0.5</td>
<td>1.3</td>
</tr>
<tr>
<td>TRA 3</td>
<td>H4N4F; H4N4F</td>
<td>148379.4</td>
<td>148378.5 ± 0.5</td>
<td>6.0</td>
</tr>
<tr>
<td>TRA d</td>
<td>H5N4F; H4N4F</td>
<td>148541.4</td>
<td>148539.6 ± 0.4</td>
<td>12.1</td>
</tr>
<tr>
<td>TRA e</td>
<td>H5N4F; H5N4F</td>
<td>148703.5</td>
<td>148700.8 ± 0.7</td>
<td>18.1</td>
</tr>
</tbody>
</table>

For TFN, quantification was performed on TFN1, and for TRA on the sum of its three major glycoforms. The quantitative method was validated in terms of linearity, sensitivity, accuracy, precision, selectivity of the immunopurification method, and matrix effect. Calibration curves built for the two proteins are displayed in figure 27.
Figure 27. Calibration curves in the 0.01 – 0.1 mg/mL concentration range for a) TFN and b) TRA. Each curve point results from the average area value obtained from a triplicate analysis. Error bars are expressed as standard deviation.

4.3.4 Application to serum TFN and to TRA

The analysis of TFN in five healthy donors serum samples showed the presence of the same glycoforms observed in the standard proteins. One additional glycoform present in all samples was produced by the loss of a sialic acid from TFN 1, a phenomenon naturally taking place with aging of the protein, which ultimately leads to galactose exposure and recognition by liver Gal-binding lectins and removal of the protein from circulation [213]. Representative deconvoluted spectra related to TFN in the five samples analyzed are reported in figure 28.
Figure 28. Representative MaxEnt1 deconvoluted spectra for TFN purified from the 5 target serum samples and identification of other forms not found in standard TFN. a) individual S1; b) individual S2; c) individual S3, with stress on the +32 protein peaks; d) individual S4; e) individual S5.

The results obtained from absolute quantification of the different glycoforms are reported in table 4.

Table 4. Concentration of the different TFN glycoforms in the analyzed serum samples (N=3).

<table>
<thead>
<tr>
<th></th>
<th>TFN 0 (mg/mL)</th>
<th>TFN 1 (mg/mL)</th>
<th>TFN 2 (mg/mL)</th>
<th>TFN 3 (mg/mL)</th>
<th>TFN 4 (mg/mL)</th>
<th>TFN TOT (mg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.11 ± 0.01</td>
<td>2.38 ± 0.19</td>
<td>0.31 ± 0.02</td>
<td>0.19 ± 0.01</td>
<td>0.12 ± 0.01</td>
<td>3.11</td>
</tr>
<tr>
<td>S2</td>
<td>0.12 ± 0.01</td>
<td>2.16 ± 0.16</td>
<td>0.23 ± 0.01</td>
<td>0.30 ± 0.02</td>
<td>0.07 ± 0.01</td>
<td>2.88</td>
</tr>
<tr>
<td>S3</td>
<td>0.13 ± 0.01</td>
<td>1.71 ± 0.16</td>
<td>0.26 ± 0.02</td>
<td>0.16 ± 0.01</td>
<td>0.10 ± 0.01</td>
<td>2.37</td>
</tr>
<tr>
<td>S4</td>
<td>0.08 ± 0.01</td>
<td>1.44 ± 0.13</td>
<td>0.22 ± 0.02</td>
<td>0.14 ± 0.01</td>
<td>0.07 ± 0.01</td>
<td>1.95</td>
</tr>
<tr>
<td>S5</td>
<td>0.11 ± 0.01</td>
<td>2.33 ± 0.14</td>
<td>0.27 ± 0.02</td>
<td>0.22 ± 0.01</td>
<td>0.09 ± 0.01</td>
<td>3.03</td>
</tr>
</tbody>
</table>
The absolute concentration values obtained for total TFN were in agreement with expected serum reference values [191]. Individuals S3 and S4 had lower total TFN levels. For individual S3, this could be due to the presence of intense signals with a mass shift of +32 Da for every glycoform, accounting for almost 50% of the total intensity (figure 28c). These peaks could be related to oxidation of the side chain of a methionine (to methionine sulfone) or of a proline (to glutamic acid) and their areas were not included in the quantification. The possibility of observing other protein modifications together with the target PTM represents a great advantage of using intact protein approaches. It is important to note that the presence of high amounts of oxidized amino acid side chains have been related to oxidative stress [214].

A histogram for visually comparing the obtained concentrations is shown in figure 29a, while in the inset (figure 29b) these values are normalized against TFN1 to highlight interpersonal variations.

![Figure 29](image)

**Figure 29.** Quantification of the different identified TFN glycoforms in the 5 analyzed serum samples. a) Concentration values for all the glycoforms, expressed in mg/mL as average of triplicate analysis of duplicate samples. Error bars expressed as standard deviation of the obtained average values. b) Normalization to TFN 1 highlighting relative variations in the other TFN glycoforms.

Naturally occurring interpersonal variations could be observed both in terms of branching (higher levels of tri-antennary species in individual
S2) and of fucosylation (decreased fucosylation in individuals S2 and S5). The observed differences were similar to what was already seen in Paper III (the study was conducted on the same samples), but a direct comparison is not so straightforward due to the difference in comparing glycoforms to released glycans.

Application to TRA was performed after spiking a standard of this protein in non-transfected cell supernatant at relevant concentrations, compatible with biopharmaceutical production systems. Table 3 reports the obtained mass values and details on the identification of the glycoforms. Concentration values are illustrated in figure 30.

Figure 30. Identified glycoforms of TRA and related concentration values obtained after quantification of spiked standard protein in non-transfected matrix. Each value is the average of triplicate samples analyzed in triplicates. Error bars expressed as standard deviation of the average concentration values.

Quantification of biopharmaceuticals in their production medium is not common, since their determination in biological fluids is considered more interesting for the study of their post-administration behavior. Nevertheless, the possibility of absolute quantification of glycoforms with minimal sample preparation could be beneficial for quality control procedures to highlight alterations that could decrease its efficiency or harm the health of patients.

At present, quantitative methods for glycan analysis are not extensively established, since most applications measure relative variations in the
glycan profiles in a sample [72]. Absolute quantification of protein glycoforms is commonly achieved by multiple-reaction monitoring (MRM) on glycopeptides [215]. The method developed in Paper VI performed similarly to MRM methods in terms of detection limits, linearity, precision, and accuracy, but did not require enzymatic digestion or chromatographic separation. Quantification methods at the intact protein level typically use LC or CE separation prior to MS detection, thus increasing the time of analysis [72,133]. Qualitative description of TFN glycoforms by high resolution MS at the intact protein level was performed by Van Scherpenzeel in 2016 [216] and was applied to the diagnosis of CDG. The method, however, was more time-consuming and did not allow for absolute quantification.

Overall, the method developed in Paper VI combined the multi-well parallelization and automation of the bead-based approach developed in Papers II-IV exploiting the advantages of high-resolution MS for the identification of different glycoforms, and the possibility of fast and reliable absolute quantification directly at the intact protein level. Such features make this approach suitable for both implementation in clinical laboratories and for the quality control of large batches of biopharmaceuticals. Additionally, as will be further described in sections 4.4.2 and 4.4.3, structural elucidation by IMS-MS/MS experiments could be included in the procedure, potentially allowing for qualitative description of all the quantified glycoforms.

4.4 Structure elucidation of released N-glycans

Structural characterization of glycans at the intact protein level poses many challenges in terms of required instrumentation and data interpretation. A simpler approach is to perform fragmentation experiments by tandem MS on released glycans, after ESI or MALDI ionization. As discussed in section 3.3, MALDI entails loss of sialic acids unless they are derivatized, but the obtained spectra are easily interpreted for the presence of almost only singly charged ions. ESI spectra are more complex due to the production of doubly or triply-charged ions, but labile residues can be maintained by tuning the source parameters.

In Paper V, the ability of SAPSI to ionize biomolecules was applied to the structural characterization of N-glycans, both in standard solutions
and after enzymatic release from proteins. Sialylated (acidic) and non-sialylated or desialylated (neutral) N-glycans were successfully analyzed by SAPSI. To my knowledge this was the first time that glycans were analyzed and their structure characterized by a PSI-based technique.

**Paper VI** proposed an IMS-MS/MS method to run in parallel to the glycoform quantification on intact protein, in order to identify the structures of the calculated sugar compositions. The method employed the bead-based approach developed in **Paper II-IV**, adapted to the ESI ionization of the released N-glycans.

### 4.4.1 SAPSI analysis of released N-glycans

Application of SAPSI to N-glycan analysis was performed in **Paper V**. The extended analysis time (described in section 4.3.1) proved to be extremely beneficial, allowing the tuning of the instrumental parameters for both MS full scan and tandem MS fragmentation in a single run, providing an advantage that, given the high costs of small amounts of N-glycan standards, is not negligible.

Initially, the system was tuned using a standard solution of the sialylated H5N4S2 glycan in positive ionization mode. The best performances were obtained using ACN/water (1:1), formic acid 0.1% v/v as ionization solvent, resulting in several doubly-charged ions produced by different combinations of sodium and potassium coordination.

Acidic N-glycans were obtained after PNGase F enzymatic release from TFN, while neutral N-glycans were cleaved from AAT by adding neuraminidase to the enzymatic solution. Both classes were successfully analyzed and fragmentation experiments were carried out on the most intense species. A full scan and a fragmentation spectrum obtained for acidic N-glycans released from TFN are shown in **figure 31**. It is possible to observe that the TFN charge envelope is still present. The resolution, however, is poor due to the use of ACN as ionization solvent. The resulting envelope is thus comparable to the one shown in **figure 23a**, with a lower overall charge due to the lower percentage of formic acid used in the glycan application. The deconvolution of the obtained envelope resulted in mass values compatible with the deglycosylated TFN, but the quality of the obtained spectrum was extremely poor and resembled the one shown in **figure 23b**.
Figure 31. MS full scan of acidic N-glycans released from TFN and fragmentation spectrum of the species of interest. a) MS full scan showing the charge envelope of TFN ionizing together with the released N-glycans. In the inset a zoom over the N-glycan region, highlighting all the different H5N4S2 species produced by different coordinated alkali cations, is provided. b) MS/MS spectrum of the [H5N4S2]^{2+} ion and identification of the generated fragments. Fragments are depicted using the Domon and Costello nomenclature.

In Paper VI four different glycan species in intact TFN were observed, but no glycan other than H5N4S2 could be identified after enzymatic release and SAPSI analysis. It is possible that the large number of peaks generated by alkali cation coordination would split the signal of less abundant species to below the detection limits, thus hampering their detection. From this perspective, paper pre-treatment to reduce the presence of alkali cations would be beneficial to extend the applicability of the technique.
4.4.2 ESI-IMS-MS/MS analysis of released N-glycans

Structural characterization of the glycoforms identified in intact proteins in Paper VI was achieved by a developed TWIMS-MS/MS method. The target protein was purified using the bead-based procedure developed in Paper II-IV and, after enzymatic deglycosylation and SPE clean-up, the released N-glycans were directly injected in the mass spectrometer. IMS was then used to separate the different glycan species, avoiding a time-consuming chromatography step. Similarly to intact protein analysis, a guard column was employed to focus the injection band and to separate the analytes from the front of elution. In this case the stationary phase was PGC and the analysis time was reduced to one minute. PGC is commonly used in LC separation of glycans, due to its ability in separating isomers, its tolerance to a wide range of solvents, as well as the retention of relatively hydrophilic compounds. The choice of PGC permitted the use of ammonium hydroxide as additive in the mobile phase to decrease sodium levels and limit the peak splitting effect observed in Paper V.

IMS separation was carried out in the tri-wave cell of a Waters Synapt G2S instrument, which allows MS/MS fragmentation by CID before and/or after the TWIMS (figure 8). In Paper VI, N-glycans were first separated by TWIMS and subsequently fragmented in the transfer cell of the tri-wave. In this way, parent ions and fragments are time aligned and share the same arrival time distribution (ATD). Baseline separation by TWIMS of different species or positional isomers would yield pure fragmentation spectra, enabling structural assignment of each species unequivocally.

The method was applied to N-glycans released from the two target proteins of Paper VI: TFN and TRA. In the case of TRA, no immunoaffinity purification was performed, and the protein was deglycosylated directly in the production medium.
4.4.3 Application to neutral N-glycans released from TRA

Analysis of neutral N-glycans released from TRA resulted in the identification of the six species producing the eight glycoforms described in section 4.3.3. Glycans were ionized in positive mode resulting in singly charged ions, with high amounts of mono-sodiated ions. TWIMS separation and tandem MS fragmentation were subsequently performed on the sodiated ions, due to the more stable fragmentation pattern, except that of sodiated H5N4F, which too low intensity. In contrast to protonated species, sodiated ions are not affected by artifact sugar rearrangement during the CID fragmentation, and can produce a higher rate of A and Z cross ring fragments, which are usually not common when glycans are ionized in positive mode [163,164]. Details on the glycans identified after release from TRA are reported in Table 5.

Table 5. Identification of the glycans released from TRA. Structures were based on the composition and from MS/MS experiments; theoretical m/z value were calculated by using the Glycoworkbench software; mass accuracy is expressed as relative error in ppm between measured and theoretical mass value.

<table>
<thead>
<tr>
<th>TRA</th>
<th>Sugar composition</th>
<th>Ionization form</th>
<th>Measured mass (m/z)</th>
<th>Theoretical mass (m/z)</th>
<th>Mass accuracy (RE, ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3N3F</td>
<td>[M+Na]^+</td>
<td>1282.454</td>
<td>1282.454</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>H3N4</td>
<td>[M+Na]^+</td>
<td>1339.476</td>
<td>1339.476</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>H4N3F</td>
<td>[M+Na]^+</td>
<td>1444.407</td>
<td>1444.507</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>H3N4F</td>
<td>[M+Na]^+</td>
<td>1485.534</td>
<td>1485.534</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>H4N4F</td>
<td>[M+Na]^+</td>
<td>1647.685</td>
<td>1647.587</td>
<td>-60</td>
<td></td>
</tr>
<tr>
<td>H5N4F</td>
<td>[M+H]^+</td>
<td>1787.198</td>
<td>1786.755</td>
<td>-248</td>
<td></td>
</tr>
</tbody>
</table>

When considering protonated ions, separation could not be achieved for different glycans with the same number of sugar units. The heptasaccharides H3N3F and H3N4 could however be separated from the octasaccharides H4N4 and H3N4F, while partial separation was achieved for the larger H4N4F (nonasaccharide) and H5N4F.
(decasaccharide). Sodiated species on the other hand, showed peak splitting for the glycans with asymmetric antennae, namely H3N3F, H4N3F, and H4N4F. This partial separation is related to the different position of the antenna on the core, either on the same side or on the opposite side of the fucose residue (position 6 and 3 respectively), generating two structures with distinct CCS values. The impossibility of a complete resolution of all the glycan species prevented obtaining pure fragmentation spectra without setting the quadrupole as mass filter, to select and separately transmit each precursor glycan ion to the mobility cell. In this way, IMS was used to collect only the fragments reflecting the ATD of the target parent, avoiding interferences from compounds sharing the m/z with target glycans. Mobilograms of neutral glycans released from TRA and subsequent tandem MS spectra acquired in positive ionization mode are reported in figure 32.

**Figure 32.** Structural elucidation of neutral glycans after enzymatic release from TRA and subsequent TWIMS separation. a) TWIMS mobilograms illustrating the resolved mobilities of the six target glycans and b) related MS/MS spectra obtained by using the quadrupole as mass filter and fragment identification. An intensity threshold limit of 3% of the base peak intensity was set to exclude low-intensity fragments.
Positive mode fragmentation took place mostly at the glycosidic bond, even if sodiated precursors were selected, providing information only on the sequence. Fucosylation was assigned to the core for all the fucosylated species, except H5N4F. In this case, antennary fucosylation was considered for the fragment at m/z 618.218, consisting of the $^{3,5}X_{6,α^-}C_{3,α^-}$ fragment of the fucosylated antenna. Since this species was fragmented in the protonated form, however, assignment of fucosylation is made with a lower degree of confidence, as this diagnostic fragment could be the result of rearrangement of fucose under CID conditions. This possibility has to be taken into consideration also in the light of the prevalent core fucosylation in the other identified species.

4.4.4 Application to acidic N-glycans released from TFN

Analysis of acidic N-glycans released from TFN resulted in the identification of the four species producing the five glycoforms described in section 4.3.3. Acidic glycans were ionized in negative mode, and produced mostly doubly charged deprotonated ions. Details on the glycans identified after release from TFN are reported in table 6.

Table 6. Identification of the glycans released from TFN. Structures were obtained from the composition and from MS/MS experiments; theoretical m/z value were calculated by using the Glycoworkbench software; mass accuracy is expressed as relative error in ppm between measured and theoretical mass value.
Similarly to what was discussed in the case of TRA, baseline separation between all the species by TWIMS was not possible, due to the partial overlap between species differing by only a fucoose unit. Baseline separation was anyway achieved for di- and tri-antennary species. Thus, the quadrupole was used as mass filter to obtain pure fragmentation spectra, as described in the previous section for TRA. Resulting mobigrams and related tandem MS spectra are reported in **figure 33**.

---

**Figure 33.** Structural elucidation of acidic glycans after enzymatic release from TFN and subsequent TWIMS separation. **a)** TWIMS mobilograms illustrating the resolved mobilities of the four target glycans and **b)** related MS/MS spectra obtained by using the quadrupole as mass filter and fragment identification. An intensity threshold limit of 3% of the base peak intensity was set to exclude low-intensity fragments. In the case of H6N5S3 and H6N5S3F, due to the high intensity of the base peak, the threshold was set on the intensity of the parent ion.

Fragmentation resulted in mostly cross-ring fragments, as expected in negative mode, yielding information on both sequence and linkage, but complicating the interpretation. Fucosylation was assigned to the antenna for both the fucosylated species, due to the presence of specific fragments: H5N4S2 showed the $^{0,4}X_{6,6}B_{4,6}$ ion at m/z 843.688,
consisting of a fragment of the fucosylated antenna, while H6N5S3 showed the C₆Y₆,α fragment at m/z 1265.012, caused by the loss of a sialic acids and of the non-fucosylated GlcNac at the reducing end. It has to be considered that such fragmentation could be affected by a certain degree of rearrangement, as discussed in the previous section. Serum-like glycosylation usually calls for a prevalence of core fucosylation, and an increase of antennary fucosylation to generate sialyl-Lewis X motifs, is usually associated with pathologies such as cancer [70]. In the case of TFN antennary fucosylation has already been observed, although in that study the protein was purified from amniotic fluid [191,217].

The developed method was subsequently applied to glycans released from TFN purified from serum samples. Figure 34 shows representative spectra obtained from individual S3. The same species as in the standard were identified, and the glycoforms observed at the intact protein level were confirmed. Individual S3 was characterized by high levels of oxidation in all the glycoforms (figure 31). The absence of glycan-related M+32 peaks in full scan confirmed that such modification probably takes place on amino acid side chains. Structural elucidation of tri-antennary species was not possible in any sample due to the low concentration levels. This problem was not encountered on standard TFN due to the higher amount of protein that was deglycosylated. When applying the bead-based procedure to serum samples, the amount of immunoaffinity binder used and the TFN concentration levels became limiting factors. Thus, the obtained spectra had on average 10-50 fold lower intensities, and fragmentation of minor species was not possible.
The developed ion mobility method was able to separate glycans differing for more than one saccharide unit in size, but failed to discriminate between different sugars with the same number of monosaccharides, as well as for positional isomers. Only in the case of the neutral glycans released from TRA, positional isomers of sodiated glycans were partially separated. IMS separation, while quite common for small oligosaccharides, is not developed in the application to complex N-glycans. Differences in linkage, position of a unit, or anomeric state have higher impact on the three-dimensional shape of smaller molecules, but can be negligible in larger species with large structure similarity.

It was concluded that, even though IMS could be used to obtain more pure fragmentation spectra and provide indications on the glycans released from the target proteins, the resolution provided by our setup was not enough to fully implement the methodology. New generation IMS instrument, such as cyclic IMS or TIMS, may abate these issues, as the resolution is about 20 and 50 times higher when compared to the instrument used in this work.
5. Conclusion and perspectives

The intricacy of structural characterization of N-glycans and the inherent complexity of biological samples require laborious analytical methods and sophisticated technology to provide full and accurate description of glycosylation. Alternatively, fast and cheap profiling methods can be used for specific tasks, but with a limited scope.

The methods proposed in this thesis combine screening approaches with the possibility of increasing the amount of information in terms of single protein profiling, structure assignment, and quantification. All the methods, characterized as having the potential of parallel sample treatment, fast analysis time, and a high degree of automation, were successfully applied to the analysis of glycosylation in complex biofluids, such as serum and CSF. The interrogation of such complex matrices was conducted by targeting specific proteins, wherein glycosylation was deemed to be relevant for the targeted health issue.

Two main approaches were presented in this work for describing the N-glycosylation of selected proteins. The microfluidic method proposed in Paper I established a fully automated and integrated sample preparation procedure, which enabled consistent time reduction and minimized manual processing workload and the probability of errors associated with human operators, providing a fast, cheap, and reproducible screening method. The bead-based approach described in Paper II and further developed and applied in Papers III and IV provided increased flexibility while maintaining amenability for automation. Analysis of selected proteins extracted from healthy human serum and CSF highlighted the different glycosylation types, characteristic to the different biofluids and the naturally occurring variations in glycosylation between different individuals. Moreover, the glycosylation profiles of many of the target proteins were analyzed in CSF for the first time.

The importance of protein-targeted analysis was demonstrated in Papers III and IV: in Paper III alterations of the glycosylation of single proteins were found not to be reflected in the total biofluid, due to the masking effect of the glycosylation of high abundance glycoproteins, while in Paper IV the information obtained from two selected proteins was combined to better discriminate between two sample cohorts, and to highlight effects that would have been missed otherwise.
Application of the methods to investigate disease-induced alterations in glycosylation showed promising results in terms of discrimination between the sample cohorts and in describing alterations reflecting disease mechanisms. These results were however limited by the small number of samples available for the analysis and by the choice of the target proteins, restricted by the availability of immunoaffinity binders. The creation of panels targeting proteins relevant to specific health issues and their application to proper sample cohorts could lead to the development of precise diagnostic tools.

Currently, reliable quantitative methods for glycan analysis are not so extensively developed, thus limiting the validation of glycan biomarkers. In Paper VI the bead-based method developed in Papers II, III, and IV was adapted to the analysis of intact protein glycoforms by ESI-MS, and a method for their absolute quantification directly from MS spectra was developed and validated. The performance of the method was comparable to more established quantitation approaches, but did not have the necessity of enzymatic digestions and long chromatographic separations. In parallel, structural elucidation of the observed glycans was provided by an IMS-MS/MS method, which provided pure fragmentation spectra in a direct infusion set-up. Extension of the study to a larger number of target glycoproteins could be easily implemented by exploiting the versatility of the streptavidin-coated magnetic beads. Additionally, the use of biotinylated immunoaffinity reagents other than VHHs could improve the yield of the protein elution step, thus increasing the method sensitivity without significantly affecting the selectivity. Improvement of the IMS separation would also be required in order to fully implement the methodology.

Finally, the possibility of using PSI for the ionization of released glycans and for the characterization of glycosylation and other PTMs on intact proteins was demonstrated for the first time in Paper V. The versatility of the developed SAPSI setup could be further explored: integrated sample preparation systems could be developed by coupling the paper to specific immunoaffinity reagents for the capture and purification of low abundance glycoproteins directly from crude biofluids, or lectins for the isolation of glycoconjugates. Alternatively, in-situ reactions like labeling and/or chemical or enzymatical removal of glycans from the immobilized protein could be implemented.
Acknowledgments

First, I would like to thank the Analytical Chemistry unit at ACES for giving me the possibility of being a PhD student and for the financial support. I also want to thank Gyros AB, the Berzelii Technology Centre for Neuroscience, and Medimmune for the financial support.

This thesis would not have been possible without the guidance of my supervisors (in order of appearance) Gunnar Thorsén and Leopold L. Ilag. Thank you for accepting me as a PhD student, for all the advice, guidance, and support, as well as for all the interesting conversations and the scientific and philosophical discussions. I would also like to express all my gratitude to my co-supervisor Isabella Karlsson: your help and advice and constant support have been invaluable during all this time.

The papers on which this thesis is based could not have been performed without the collaboration of Anna Sroka-Bartnicka and Erik Tengstrand (Paper I), Federico Marini and Martin Ingelsson (Paper IV), Nicoló Riboni, Nicklas Österlund, and Hitesh Motwani (Paper V). Thanks also to all the diploma students that have helped me and have contributed to the published works: Angela de Waard, Lorena Ndreu, Matthijs Pijnappel, Neslihan Akkus, Maya Spasova, and Elena Passarini.

Special thanks to Åsa Emmer at the department of Chemistry at KTH for granting me access to the MALDI instrument used in papers II-III-IV. I would also like to thank the PhD students at KTH, Leila Josefsson and Linus Svenberg, for the help with the instrument and the nice time during courses and conferences.

Huge thanks to the present and past staff at the unit: Ulrika Nilsson (for your kindness in helping and advising as well as for your great feedback on my thesis), Roger Westerholm, Conny Östman, Jan Holmbäck (thank you for the coffee and to be always nice and helpful), Magnus Åberg, Ingrid Granelli (I really enjoyed assisting in the Bioanalysis course), Anders Colmsjö, Mohammed Abdel Rehim, Cindy de Wit, and Anneli Kruve-Viil. My gratitude goes also to Carlo Crescenzi for all the insightful collaborations and discussions, as well as for all the suggestions that made my research progress.
I would like to express my deep gratitude to the unit staff: Lena Elfver, for your nice attitude and the irreplaceable help in everything, Jonas Rutberg, for having taught me many strange and typical Swedish things and for the help with the language, and Zesi Wang, for your enthusiasm and kindness.

I would like to extend my gratitude to Stefan Karlsson from Waters, for the precious help with the instruments and the great technical advice, and to the people from the other side of the corridor: Margareta Törnqvist, Lillemor Asplund, Hitesh Motwani (for the very good collaboration and for the very useful feedback to my thesis), Johan Eriksson (for all your help with instrumentations and technical advice), Jonathan Martin, Ioannis Athanassiadis (for the nice food conversations and the help with technical problems), Jenny Aasa, Jana Weiss, Hongyu Xie, Giulia Martella and Stathis Vryonidis.

I would like to thank the wonderful people I shared the room with: Liying Jiang, Javier Zurita, Ahmed Ramzy, and Hatem Elmongy. I really enjoyed sharing the room with you and I want to thank you for being very nice people as well as good friends. To the past and present PhD students, and to the other staff, I want to thank you for creating a nice and cheerful environment at the unit: Aljona Saleh, Christoffer Bergvall, Emilia Eklund, Erik Tengstrand, Farshid Mashayekhy Rad, Hwanmi Lim, Ioannis Sadiktsis, Jonas Fyrestam, Josefine Carlsson, Karin Englund, Nadia Zguna, Nikola Radoman, Pedro Sousa, Petter Englund, Rozanna Avagyan, Rudolf Andrýs, Silvia Masala, Trifa Mohammed Ahmed, and Vibhu Rinwa.

A special mention to the huge Italian colony that has over time populated the department: Ciccio Iadaresta, Giovanna Luongo, and Nicolò Riboni, together with all the erasmus, guests, and exchange students: Niccoló Parenti, Donato Varalli, Emanuele Moccia, Silvia Spindici, Paola Savino, Alessia Viola, Davide Soricelli, Serena De Chiara, Roberto Di Meo, Debora Tommaseo, and Elena Passarini, as well as to Mila Amendola, Pau Benetó Vallés, Lorena Ndreu e Nesrine Mansouri, per il tempo passato insieme, l’amicizia, le serate e per l’avermi molte volte fatto sentire a casa.
To Prof. Francesco Buiarelli, to Patrizia Di Filippo, Carmela Riccardi, and Donatella Pomata, for all the fundamental teachings and the wonderful time during my experience in your lab. You made me understand that I really wanted to do this.

To all my friends in Sweden: Cristiano, Federico, Piero, Claudia, Chiara, Enrico, and in Italy: Jacopo, Alessandro, Marta, Giada, Andrea, Valentina, Andrea, and all the others. Grazie di tutto, vi voglio bene!

A final dedication to my family and, last but not least, to Silvana, per avermi sempre supportato (e sopportato), per l’esserci sempre stata e per quello che fai ogni giorno per me. Senza di te nulla sarebbe stato possibile.
References


