Universiteit
Leiden
The Netherlands

Unraveling proteoform complexity by native liquid chromatography-mass spectrometry

Schaick, G. van

Citation
Schaick, G. van. (2023, January 24). *Unraveling proteoform complexity by native liquid chromatography-mass spectrometry*. Retrieved from https://hdl.handle.net/1887/3512617

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3512617

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

1

Introduction

Proteins are large and complex biomolecules that play key roles in most processes of life. Our bodies contain thousands of proteins, all with different functionalities varying from protection by naturally occurring antibodies to energy metabolism catalyzed by enzymes. Besides endogenous machineries, proteins are used all around us in everyday life. For instance, protein-based pharmaceutical products are used for treatment of various diseases and industrial enzymes can improve food quality.

1.1 The history of therapeutic proteins

Already in the late 19th century, the use of proteins had been proposed as a revolutionary approach to cure and prevent diseases and thereby, greatly improving public health and life expectancy. However, it was not until major advances were made in the field of biotechnology that (large-scale) production of protein-based medicines became possible [1]. Eventually, the first recombinant human product (insulin, Humulin®) was approved by the Food and Drug Administration (FDA) and entered the market in 1982, where it could be used to control glucose levels in patients with diabetes [2]. Another important milestone was the development and approval of recombinant human erythropoietin (EPO; Epogen®) to treat anemia in 1989 [3]. In the following decades, protein-based medicines grew into a successful drug class to fight various diseases, from cancers to autoimmune and inflammatory disorders [4]. Especially, the landmark discovery of monoclonal antibodies (mAbs) enabled the successful generation of precision tools due to their high antigen specificity and increased treatment efficacy [5-7]. For this, the development of hybridoma technology (i.e., fusion of specific B cell clones and immortalized myeloma tumor cells) by Köhler and Milstein in the early 1970s played a key role [8]. The first therapeutic mAb (rituximab, Rituxan®) targeting cancer was approved by the FDA in 1997 [9, 10], after which many more followed. In total over 100 therapeutic mAbs for a diverse range of clinical applications have already entered the market by 2022 [11]. Even though remarkable discoveries have been made over the past decades, the field of therapeutics is still evolving by searching for new targets, increasing effectiveness and reducing side effects [12].

1.2 The discovery of enzymes for industrial use

Enzymes have been a part of everyday processes since the dawn of history. One of the first examples was the application of naturally present enzymes in yeast for beer and wine production in ancient times [13]. From the 19th century onwards, underlying mechanisms of enzymes were unraveled and enzymes were successfully isolated from their host organisms all building

up to the release of the first commercially available enzyme-containing detergent in 1914 [14]. The use of enzymes was immediately embraced due to the reduced need for harsh chemicals, simple and efficient production and lower waste generation [15]. Nevertheless, it was only in the nineteen-sixties that (large-scale) production of industrial enzymes was boosted due to both new application fields and advances in biotechnological tools, particularly genetic engineering approaches that allowed production of enzymes with maximal productivity and optimized properties [16, 17]. While some enzymes are being still isolated from plant or animal tissue, most enzyme products are nowadays manufactured in large biological reactors using recombinant microorganisms. The majority of these proteins are produced by fungal species, such as *Aspergillus niger* and *Trichoderma reesei*, but there are also examples using bacterial or yeast strains [15, 18]. Enzymes produced in these (micro)organisms are known for their high specificity and stability which together with the ease of production makes them attractive for industrial purposes [19].

Today, industrially-produced enzymes have gained worldwide recognition and are used in over 700 commercial products in industries varying from household care and food to bioenergy. The largest industrial application area is the food and beverage industry [19, 20], where enzymes are used for baking, brewing, as well as beverage, dairy and oil production. During manufacturing, enzymes can be added at various stages, including preparation, packaging, transport or storage [16]. In the baking industry, dough flexibility, stability, volume and crumb structure can be improved using several different enzymes, such as amylases, cellulases, xylanases, lipases, glucose oxidases and lipoxygenases [19, 20]. For instance, amylases are known to retain moisture in bread flour leading to increased softness, freshness and shelf life. Moreover, lipases and xylanases enhance dough volume and stability [21, 22] and glucose oxidase and lipoxygenase improve dough strengthening and whiteness [23, 24]. Moreover, brewing of beverages has also highly benefited from the use of industrial enzymes to hydrolyze starch or proteins, solve filtration issues due to insoluble components of the malt, and prevent formation of chill-haze [25].

Despite the major advances in enzyme technology that have led to the wide spectrum of available enzymes, continuous development is needed to create more diverse, versatile and robust products. Currently, society is moving into an era in which elegant and innovative ways must be found to meet the rising demand for food supply and deal with the exhaustion of natural resources. Industrial enzymes are powerful tools to tackle these challenges by for example the discovery of enzymes to facilitate biofuel

production from (non-edible) biomass and thereby supplementing or even replacing fossil fuels [26]. All in all, industrial enzymes will be crucial to guide us into a cleaner and more sustainable future.

1.3 The path from gene to fully functional protein

The production process of proteins starts from specific genes, which can be described as the instruction manual to build a protein. While around 25.000 of these genes are identified in human [27], millions of proteins and protein variants exist [28]. This fascinating number can be reached due to the fact that the construction of proteins is more complex than simply assembling amino acids to form polypeptide chains. After the synthesis of amino acid chains, these polypeptides are folded and further finetuned by post-translational modifications (PTMs) to eventually obtain functional proteins. Especially due to the presence of different PTMs, proteins arising from a single gene often do not occur as single molecular entities but rather as a heterogeneous pool of so-called proteoforms. The term proteoform includes all different molecular forms of a protein product from a single gene, including genetic variation, alternative splicing, and variants generated by PTMs [29, 30]. Examples of commonly observed PTMs are glycosylation, phosphorylation, acetylation, proteolytic truncation, deamidation, oxidation, and glycation [31]. The type, nature and complexity of PTMs depend on several factors, such as the organisms that produced the protein, storage conditions, and function that it should fulfill.

One of the most widespread and diverse PTMs is glycosylation, which is essential for proper protein folding [32] as well as plays a key role in many biological processes, including molecular recognition, signaling, and inflammation [33, 34]. Glycosylation can be described as the enzymatic attachment of covalent assemblies of monosaccharide building blocks to the protein backbone. The most regularly observed monosaccharides are hexose, N-acetylhexosamine, deoxyhexose and sialic acid (Sia) (**Fig. 1a**). These assemblies, so-called glycans, can either be attached to the nitrogen of an asparagine residue (N-glycosylation) or the oxygen of a serine or threonine residue (O-glycosylation). N-glycosylation occurs at distinct sequons (NXT/S, where X cannot be proline) and N-glycans have a conserved core structure consisting of two N-acetylglucosamine (GlcNAc) and three Mannose (Man) moieties. This core structure can be extended with either only Man residues (high mannose glycans) or with a combination of building blocks (complex or hybrid glycans) (**Fig. 1b**). Moreover, the structure of N-glycans can be further complicated by decoration with additional modifications,

such as acetylation or phosphorylation. On the contrary, O-glycosylation lacks a distinct amino acid consensus sequence and no universal O-glycan core structure is known. The most common type of O-glycosylation is the N-acetylgalactosamine (GalNAc) type (or mucin-type) of which the four most occurring core structures are depicted in **Figure 1c**. All these structures start with one GalNAc unit, whereafter the glycan can be further elongated with galactose (Gal), GlcNAc, fucose (Fuc) or Sia residues [35].

The type of glycosylation attached to proteins is dependent on the host organism. For biopharmaceuticals where glycosylation is relevant for the mechanism of action, often mammalian cell lines are used for production resulting in mainly complex type glycans. An additional advantage of creating human-like glycosylation for biopharmaceutical products is the reduced risk of immunogenicity [36]. Moreover, glycosylation can significantly impact the stability, folding and secretion of these proteins as well as alter their biological activity. For instance, the number of Sia and Gal-GlcNAc (N-acetyllactosamine; LacNAc) units in the glycans influence the *in vivo* bioactivity and half-life of erythropoietin (EPO) [37, 38]. For therapeutic mAbs, N-glycan core fucosylation has a great impact on activation of antibody-dependent cellular cytotoxicity [39]. Moreover, the presence of high mannose glycans on therapeutic mAbs is often unfavorable due to higher clearance [40, 41]. Where high mannose glycans levels on mAbs should be minimized, this type is highly abundant on industrial enzymes secreted by fungi [42]. For these proteins, high mannose glycosylation is often even required for proper enzyme secretion or performance [33]. For example, deglycosylation of a glucoamylase expressed in *Aspergillus niger* caused reduced thermostability and increased enzyme flexibility and aggregation [43]. Similar effects were observed for a xylanase expressed in *Pichia pastoris*, where lower specific activity was detected after deglycosylation or modification of a glycosylation site by site-directed mutagenesis [44].

Another frequently observed PTM is glycation, which can be described as the non-enzymatic reaction between reducing sugars and amine groups of proteins (i.e., lysine (lys) and arginine (arg) residues or the N-terminus) (**Fig. 2**) [45]. The first step of this reaction is the condensation of the carbonyl group of the sugar and the primary amine of a protein resulting in the formation of the unstable Schiff base, which then rearranges to the open chain Amadori product and eventually to the stable Amadori Product [46]. No specific sequence motif has been reported for glycation, but so-called glycation hotspots have been identified [47]. These hotspots have often hydrophobic and acidic amino acids in close proximity to the glycated lys or arg residues

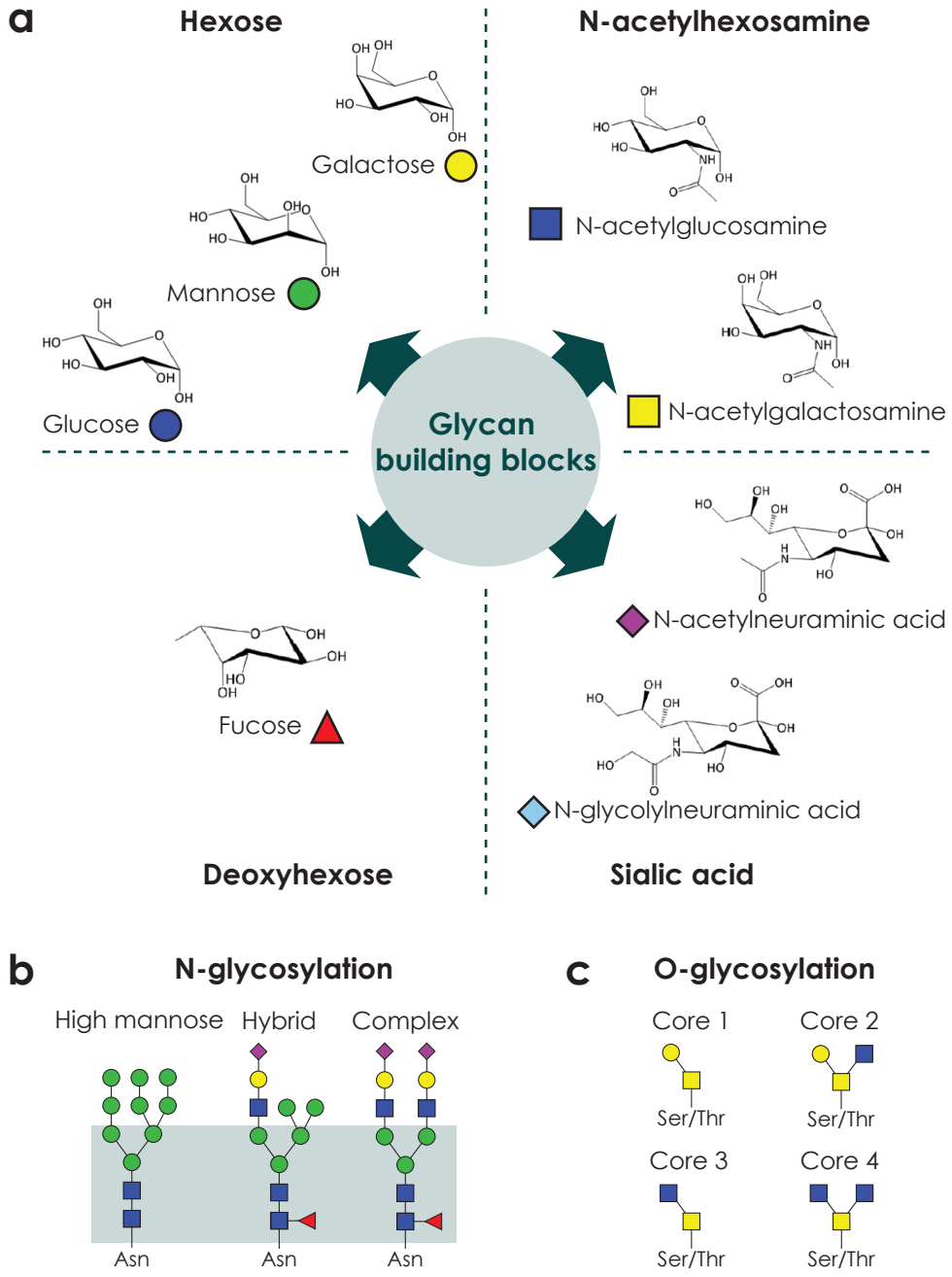

Figure 1 | Overview of glycan building blocks and examples of often observed types of N-glycans and O-glycans. (**a**) Most common monosaccharide building blocks for glycans, where the monosaccharides are depicted with both their structure and symbol (including color code). (**b**) Different types of N-glycosylation, including high mannose, hybrid and complex glycans. The common N-glycan core ($Man_3HexNAc_2$) is indicated. (**c**) Representation of frequently observed core structures of mucin-type O-glycosylation.

[48]. The total glycation levels of protein products can be greatly impacted by processing and storage conditions, such as presence of sugars, elevated temperatures, basic pH and storage time [49, 50]. Nevertheless, if glycation is present, it can have a tremendous effect on protein stability and activity [51]. For therapeutic mAbs, glycation in the complementary-determining regions potentially affects antigen binding or can cause further degradation leading (possibly) to severe aggregation [52]. Where for biopharmaceuticals glycation most often negatively impacts the activity and stability, the effect glycation on industrial enzyme performance can also be positive. For instance, glycated glucoamylases allowed more efficient conversion of maltose and maltodextrin into glucose compared to the non-glycated variant [53]. Conclusively, glycation can greatly alter protein functionality making reliable monitoring of this low-abundant modification crucial. Under well-controlled manufacturing conditions, the degree of glycation in biopharmaceutical and biotechnological products is commonly low. Therefore, to study this modification proteins are often stressed to induce glycation by storage at elevated temperatures or by incubation with high concentrations of reducing

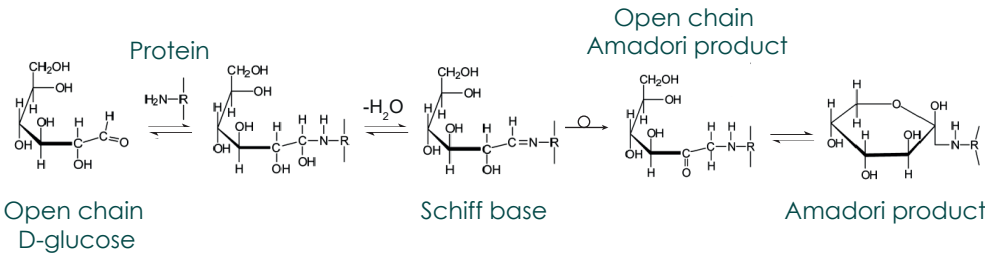

Figure 2 | Schematic overview of the Maillard reaction leading to protein glycation. An open chain glucose molecule reacts with an amine groups of a protein, specifically with lysine and arginine residues or the N-terminus leading to the formation of the unstable Schiff base. Subsequently, this Schiff base will rearrange into the more stable Amadori product.

sugars [47].

Besides glycosylation and glycation, also other PTMs can impact protein functionality. One clear example illustrating the potential influence of minor modifications on protein activity is the deamidation of asparagine residues of therapeutic proteins, which results only in a mass shift of 1 Da (conversion from asparagine to aspartic acid) but may greatly affect receptor binding and thereby, bioactivity and clearance [54, 55]. Another modification that does not alter the protein mass that can substantially change protein function (causing potentially immunogenic responses), is isomerization of an asparagine residue

into isoaspartate [54]. Additionally, oxidation of mAbs can often be linked to altered FcRn receptor binding and decreased half-life [56]. Finally, aggregate formation or protein precipitation is a known mechanism of inactivation of protein-based products [50, 57]. For therapeutic proteins, this could lead to loss of efficacy and altered bioavailability, and even immunogenic responses [57].

Altogether, a heterogeneous and dynamic landscape of proteoforms is generated along the path from gene to protein and even further during production processes, formulation or storage for therapeutic proteins and industrial enzymes. Creating a better understanding of the protein structures and their effect on function is essential, not only to unravel physiological processes in human but also to ensure the product quality of biotherapeutics in health care or industrial enzymes in the food industry.

1.4 Scope

To allow in-depth structural and functional characterization of proteoforms, novel analytical approaches are needed that resolve proteoform heterogeneity while preserving structural and functional integrity. This thesis focuses on the development of these approaches by combining native liquid chromatography (LC) separation with mass spectrometric (MS) detection and activity assays. Proteoforms can display very diverse physicochemical properties often requiring complementary analytical techniques for their in-depth characterization. For this purpose, various native LC modes exhibiting different selectivities are exploited in this thesis depending on the sample characteristics. The separation of proteoforms permits their subsequent structural (and if needed also functional) characterization as highlighted in this thesis for different biotherapeutic and biotechnological applications.

In **Chapter 2**, existing native separation techniques are introduced and benefits of coupling to MS are emphasized. Practical guidelines for performing native separation-MS analyses are provided with special attention to mobile phase composition, flow rate and available hyphenation strategies. Moreover, applications are described using native separation-MS methods to address biomedical and biopharmaceutical questions. **Chapter 3** focuses in more detail on improvement of hyphenation strategies, particularly in the use of dopant-enriched nitrogen gas to improve protein ionization after native LC separation. The influence of organic modifiers to enrich the dopant gas is systematically evaluated for proteins with varying physicochemical properties showing that this approach is most beneficial for highly glycosylated proteins.

Charge-based separations are considered the gold standard for monitoring

charge variants in biotherapeutic products. Online coupling of these methods to MS comes with the great benefits, such as less laborious workflows and lower risk of introduction of unintended modifications. However, additional serious hurdles should be taken to enable direct hyphenation. Most importantly, the non-volatile salts in the mobile phase should be replaced with MS-compatible alternatives, since these cause severe signal suppression and adduct formation in the MS instrument. In **Chapter 4**, we focus on the development of an MS-compatible anion exchange chromatography (AEX) method allowing direct and in-depth characterization of erythropoietin biopharmaceuticals. A single AEX-MS measurement allows analysis of various critical quality attributes, including charge heterogeneity, sialic acid content and number of LacNAc units, as well as additional modifications, such as O-acetylation and deamidation. To address the influence of PTMs on protein conformation, the native analytical toolbox can be further expanded with additional biophysical techniques, including ion mobility (IM) and collision induced unfolding (CIU). **Chapter 5** describes the online coupling of cation exchange chromatography (CEX) with CIU to study gas-phase conformations and unfolding of mAbs proteoforms. This novel approach enables mAb subclasses classification as well as unveils glycoform-specific information on the gas-phase stability of mAbs.

While functional characterization of proteins and proteoforms in biopharmaceutical products is often extensively reported, this information is investigated to a lesser extent for industrial enzymes. Therefore, innovative analytical strategies are developed in **Chapters 6**, **7**, and **8** using native separation techniques in parallel with MS detection and activity assays. This permits to characterize structural features and simultaneously assesses their impact on enzyme performance. The studied enzymes vary in size, complexity, and functionality. The prolyl-alanyl specific endoprotease of **Chapter 6** is a heavily glycosylated protein decorated with neutral and phosphorylated high mannose glycans. For the separation and identification of proteoforms with this particular modification, an AEX-MS method is developed. After separation, digestion specificity and efficiency of specific proteoforms were evaluated showing minor impact of phosphoglycans on protease activity. For the characterization of the endo-xylanase in **Chapter 7**, a combination of native LC modes enabled complete proteoform separation based on size with size exclusion chromatography (SEC), charge with AEX and glycation with boronate affinity chromatography (BAC). Activity measurements of the enriched proteoforms revealed that glycation negatively impacts the enzyme functionality. Finally, the influence of glycation on activity is further explored for an industrial lipase using BAC with parallel MS detection and

functional assays (**Chapter 8**). In contrast to the endo-xylanase, in this case, glycation stabilizes lipase activity instead of reducing functionality.

The final part of this thesis (**Chapter 9**) provides a general discussion of approaches to characterize proteoforms without compromising their functional state. Central themes in this section will be the concept of protein nativity and the balance that must be found between improving analytical performance and maintaining protein functionality. Moreover, specific attention will be paid to the separation principle of BAC for enrichment of glycated proteins.