

2 Chapter Two: Literature Review

Glycosylation is the intracellular addition of sugar residues to the peptide-backbone of a protein. It is the most extensive post-translational modification in eukaryotic cells. There are multiple ways glycans are attached to the polypeptide, the three most common ways are:

- N-linked glycosylation,
- O-linked glycosylation or
- glycosylphosphatidylinositol (GPI) anchored glycosylation.

N-glycans are exclusively linked to the peptide at the tri-peptide sequence Asn-X-Ser/Thr (where X can be any amino acid). O-glycosylation-linkages are less well-defined but must occur at a serine or threonine hydroxyl group. Clusters of these amino acids can enhance O-glycosylation. Indications for GPI anchor sites are hydrophobic carboxyl-terminal peptides which are concurrently cleaved when the GPI is transferred from the rough endoplasmic reticulum (ER) to the nascent polypeptide [5].

Despite newer developments in the field of glycobiology, N-linked glycosylation represents by far the more important part of glycosylation occurring in mammalian-derived glycoproteins. But O-glycosylation is becoming more and more important [6]. O-glycans are normally much smaller in size than N-glycans.

Today's most recombinant produced glycoproteins in mammalian cells for pharmaceutical usage are N-glycosylated proteins. That is why biopharmaceutical companies spend a lot of attention in controlling specifically the N-glycosylation of their products. The following thesis is part of this industrial perspective and therefore will be focused on characterization and understanding of N-linked glycosylation of recombinant proteins from mammalian cell culture.

2.1 N-linked glycosylation pathway

N-linked oligosaccharides are produced in a series of enzyme-catalyzed reactions localized in several intracellular compartments. In short, the synthesis starts with a lipid-linked oligosaccharide moiety (Glc₂Man₉GlcNAc₂-P-P-Dol) [7]. This precursor bound on the lipid carrier dolichol phosphate is transferred into the ER. To build up complex oligosaccharides, the other necessary monosaccharides are added in a stepwise manner, first by addition of nucleotide sugars and next by conjugation with lipid intermediates (Figure 2) [8].

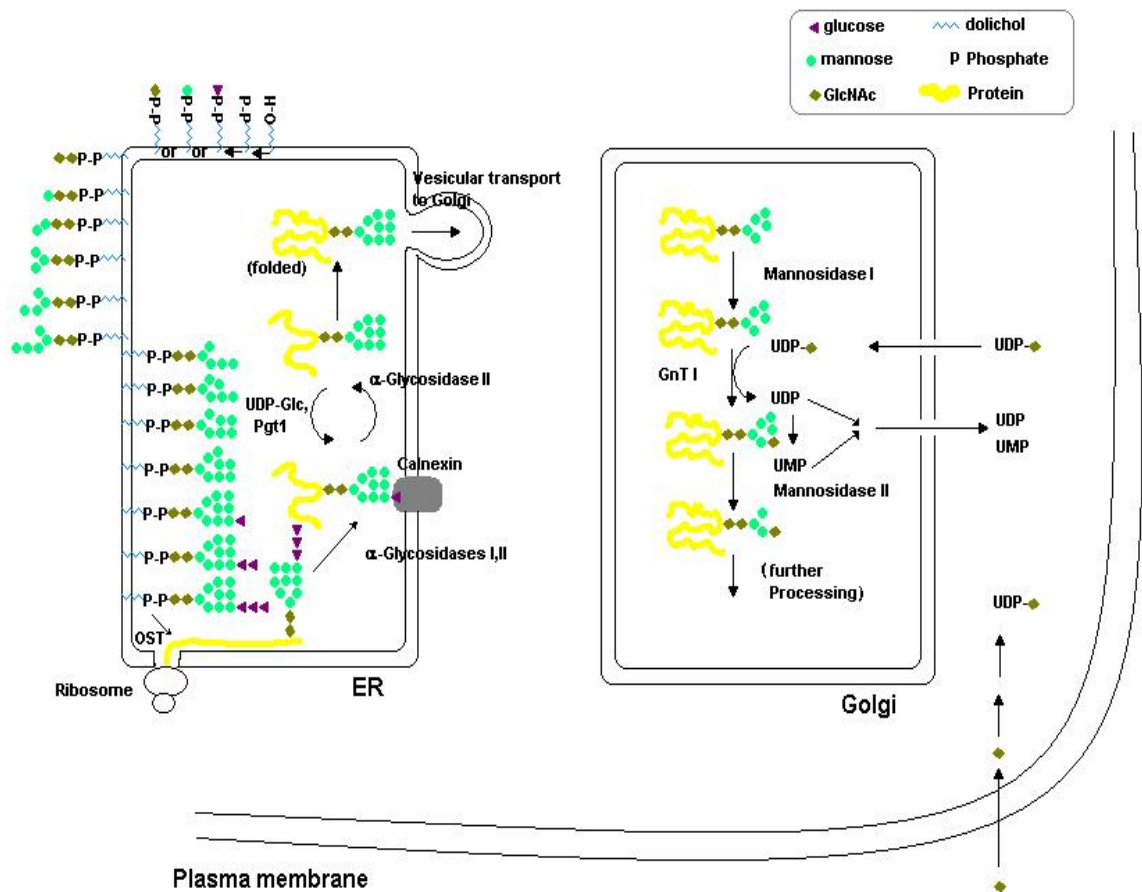


Figure 2: Biosynthetic pathway for N-linked glycosylation in CHO-cells¹

¹ Modified from Wong, C.-H., presented at Glycomics and Carbohydrates in Drug Development, March 2005, La Jolla - California

- and the complex structure type (Figure 5)

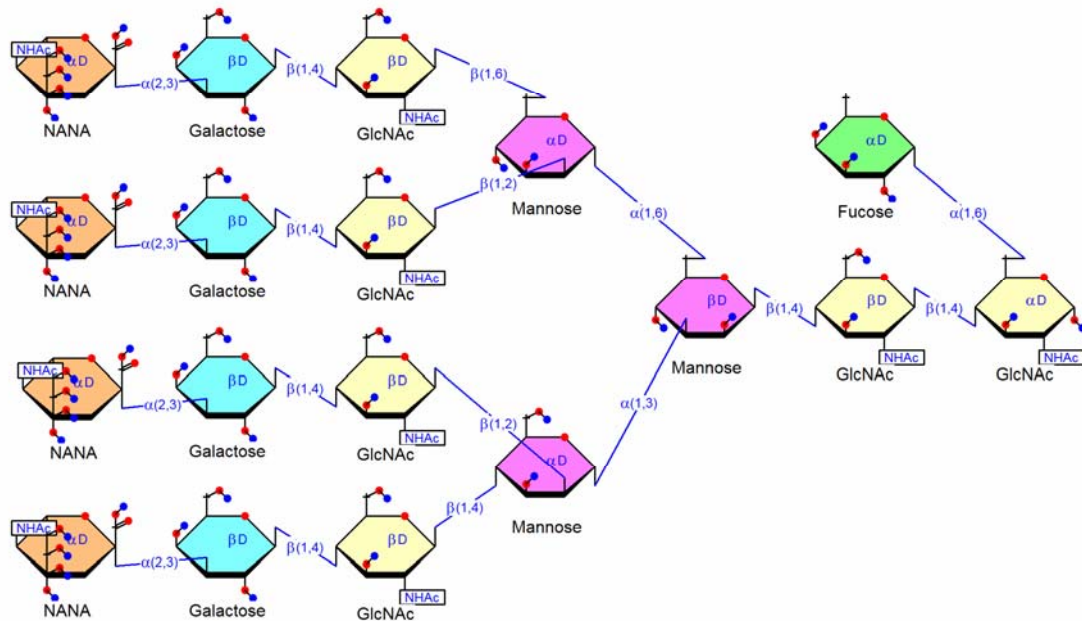


Figure 5: Complex oligosaccharide structure type²

These structures have a unique pentasaccharide core unit consisting of $(\text{GlcNAc})_2\text{Man}_3$ but they differ in the type and sequence of sugar moieties in the branches. Typical for complex type structures is the occurrence of the disaccharide $\text{Gal}\beta(1,4)\text{GlcNAc}$ in the antennae. The high-mannose type is characterised by having only mannose residues in its outer arms. Hybrid structures contain the oligosaccharides with only mannose residues in one arm and the complex structure (containing $\text{Gal}\beta(1,4)\text{GlcNAc}$) in the outer arm.

The glycosyltransferases responsible for N-glycosylation are very specific in

² Created with Prozyme® CarboDraw-Software

the type of sugars they connect. Each enzyme utilizes a specific nucleotide sugar as a cosubstrate to generate a defined carbohydrate linkage.

Because of multiple steps involved in the generation of N-glycans, glycoproteins generally exhibit both macro- and microheterogeneity in their glycosylation profile.

Macroheterogeneity, also termed as "variable site occupancy", refers to variability in the location and number of oligosaccharide attachments. On the other hand, microheterogeneity, also termed as "site microheterogeneity", refers to the variability in the oligosaccharide structures at specific glycosylation sites. Macro- and microheterogeneity in glycosylation result in a set of glycoforms for one specific glycoprotein, all of them have an identical amino acid backbone but are dissimilar in the structure and/or the disposition of their carbohydrate units. This heterogeneity leads to different physical, biochemical and biopharmaceutical properties and, therefore, also functional diversity.

2.2 Functions of carbohydrates in glycoproteins

More than 50% of all natural mammalian proteins are glycosylated proteins [9]. Many of them are potentially therapeutically important. The carbohydrate moieties of glycoproteins have been found to influence strongly many aspects of protein properties.

2.2.1 Influence on protein solubility and protein stability

Because of the hydrophilic character of oligosaccharides, they can dramatically increase the solubility of proteins and therefore prevent protein aggregation.

- the removal of oligosaccharides of erythropoietin (EPO) leads to a significant decrease in solubility [10]
- deglycosylated granulocyte colony-stimulating factor (G-CSF) tends to aggregation [11]. This can be explained by covering the hydrophobic

protein surfaces with the high hydrophilic carbohydrates and their strong affinity for water [12].

Concerning thermal stability, the presence of oligosaccharides can also dramatically increase glycoproteins' resistance to denaturation.

- deglycosylated EPO is much more susceptible to thermal denaturation [13] than the glycosylated form
- glycosylated bovine pancreatic ribonuclease is much more stable than the deglycosylated form because the oligosaccharides contribute 0.3 to 0.4 kcal/mol to its thermal stability [14].

2.2.2 Protection from protease attack

Carbohydrates play a major role in protecting glycoproteins from proteolytic attack.

Wang et al detected that porcine pancreatic ribonuclease is much more vulnerable to the attack by trypsin and subtilisin when its carbohydrate moiety is removed [15]

Olden et al found out that unglycosylated fibronectin from chicken embryo fibroblasts is more susceptible to pronase digestion [16].

The most impressive example is EPO. In 1974 Goldwasser et al reported that the removal of terminal sialic acid in EPO results in an increased susceptibility to proteolysis by trypsin [17]. A lot of work was done to explain and use this phenomenon, preferentially since the beginning of the biotechnology era in the 1980s. In 2002 Amgen released the first glycoengineered erythropoietin, called darbopoietin, with 3 additional N-linked glycosylation sites and with a 3-fold increased longevity in the human plasma [18]. This phenomenon was so remarkable that Amgen decided to develop a general glycoengineering approach for the improvement of mammalian-derived biologicals [19].

Yet et al proposed that oligosaccharides generally protect glycoproteins from proteolytic attack by masking potential cleavage sites on the protein surface [20].

2.2.3 Influence on biological activity and immunogenicity

In general, only completely glycosylated proteins have full biological activity. There are some examples that impressively demonstrate this:

- factor IX completely loses its clotting activity after removal of its sialic acids [21]
- deglycosylation of HIV-1 envelope glycoprotein gp120 reduces its binding affinity to CD4 receptors [22]
- interferon- β [23] needs a 10 times higher dosage for its deglycosylated form in MS therapy

However, there are quite a few exceptions which demonstrate no correlation between glycosylation and in vitro activity.

- unglycosylated EPO showed very similar specific activity compared to its native counterpart in-vitro [24, 25]

E. Coli-derived human IFN- γ has full antiviral and antiproliferative activity in vitro [26].

Carbohydrates can mask existing antigenic sites on the peptide backbone and thus affect the immunogenicity of a protein. Examples for that:

- a mutant form of the H3-influenza virus can escape monoclonal antibody recognition by forming a new glycosylation site [27]

On the other hand, oligosaccharides can also act as an immunogen itself and many mammalian circulating antibodies can target specific oligosaccharide determinants. For example:

- about 1% of human IgG is specific for the terminal Gal α (1,3)Gal β (1,4)GlcNAc epitope [28]

Meanwhile, the presence of oligosaccharides may indirectly impact the glycoproteins' antigenicity. Consequently, the immunogenicity of a glycoprotein could be altered as a result of oligosaccharide-protein interactions. For example:

- there are differences in antigenicity between glycosylated and unglycosylated Semliki forestvirus protein and bovine luteinizing hormone [29]

2.2.4 Influence on in vivo circulatory half-life

Oligosaccharides on proteins play also a dominant role in defining the in vivo clearance rate of glycoproteins. Since high in vitro specific activity will be of no significance if the injected protein is eliminated from the circulatory system too fast, the circulatory half-life is one of the most important aspects of intravenous injected therapeutic glycoproteins. Glycoproteins are generally cleared out of the human blood stream by three well-known mechanisms. Those glycoproteins with exposed galactose and N-acetylglucosamine, including desialylated complex N-glycans, bind to the asialoglycoprotein receptor which can be found on hepatocytes and consequently are eliminated from the circulatory system [30]. The second clearance pathway are mannose receptors on the surface of liver endothelial cells and resident macrophage cells which bind glycoproteins with high-mannose structures [31]. The last clearance mechanism is the continuous removal of human proteins with molecular weights less than 70 kD in the glomeruli of the kidney. Therefore protein tertiary structure and molecular weight as well as the presence of surface charge can all have a major impact on the protein filtration rate.

Examples are:

- glycosylated tissue plasminogen activator (t-PA) has a much higher in vivo activity than its unglycosylated forms [32], although its in vitro activity is much lower [33-37]
- the in vivo activity of the glycosylated EPO-forms is much higher than that of unglycosylated forms due to a prolonged circulatory half-life [24, 25]
- human chorionic gonadotropin (hCG) and lutropin have a much prolonged half-life and therefore also in vivo activity than their deglycosylated forms [38, 39]

2.3 Factors influencing protein glycosylation

Protein glycosylation can be influenced by many different factors. In this chapter the three most important determinants will be discussed: the polypeptide structure, the host-cell phenotype and the environment in which the cells are cultured.

2.3.1 Influence of protein structure

The protein structure can influence the carbohydrate content in many ways, particularly the primary, tertiary and quaternary structures of the polypeptide play significant roles in the occurrence and conformation of glycosylation due to protein-carbohydrate interactions. As a consequence, glycans from different proteins derived from the same cell line can be dramatically different. Two well-known examples are the N-glycan structures of interferon- β from CHO cell culture compared with the glycan-structures of erythropoietin. IFN- β N-glycans consist mainly of complex biantennary structures, whereas EPO N-glycans possess only 6% complex biantennary and mainly complex tetraantennary structures [40, 41].

Even the oligosaccharides from different glycosylation sites of the same protein can be dramatically different. The study of t-PA glycosylation is a good example of site-specific glycosylation heterogeneity [12]. t-PA has four potential glycosylation sites: Asn-117, Asn-184, Asn-218 and Asn-448. Asn-218 with the amino acid sequence of Asn-Pro-Ser is never glycosylated. Asn-184 is subjected to variable site occupancy depending on the mammalian cell type it is derived from. Regarding the microheterogeneity of the glycan structures, Asn-117 oligosaccharides are mostly high-mannose type, and the structures at both Asn-184 and Asn-448 are complex type. These site-associated differences in glycosylation of t-PA clearly demonstrate the impact of local protein environment on the product glycosylation pattern.

Normally N-glycosylation can occur exclusively at the tripeptide sequence Asn-X-Ser/Thr. But the presence of this consensus sequence does not necessarily result in the glycosylation of a possible glycosylation site. For

example, Thr at position 3 leads to an increased chance of glycosylation compared to Ser at this position, and a proline residue within or near this sequence reduces the likelihood of glycosylation. Furthermore, perspective glycosylation sites near the N-terminus are more likely to be glycosylated, which suggests temporal competition between protein folding and the initiation of glycosylation [42].

2.3.2 Influence of expression system

The expression system has a significant influence on the glycosylation pattern of a protein. That is due to the fact that different expression systems have different glycosylation capabilities. Glycosylation capability depends on the post-translational enzymatic machinery of glycosyltransferases and glycosidases. The machineries of different expression systems differ in the concentration, kinetic characteristics and compartmentalization of the individual glycosyltransferases and glycosidases.

Common bacteria systems are generally incapable to glycosylate proteins, whereas yeast, insect, plant and mammalian cells share the features of N-linked oligosaccharide processing in the ER, including the attachment of the precursor $\text{Glc}_2\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$ and subsequent truncation to a $\text{Man}_8\text{GlcNAc}_2$ -structure. However, oligosaccharide processing inside the Golgi apparatus varies in different cell types because of their differences in the glycosylation machinery. For example, plant-derived glycoproteins are generally not sialylated and frequently contain xylose, a monosaccharide normally not found in mammalian N-linked oligosaccharides. Oligosaccharide chains in yeast cells are elongated in the Golgi through stepwise addition of mannose, resulting in super-high-mannose structures that often contain more than 100 mannose monomers. Another expression system with mammalian-unlike glycosylation characteristics is the baculovirus-infected insect cell system that has become popular for recombinant protein production due to both, its short process development time and its potential high yields. However, the glycosylation of insect cells is significantly different from mammalian cells because they are believed to be unable to process complex-type oligosaccharides [11] and are limited to producing only simple

oligomannose-type oligosaccharides of $\text{Man}_{3-9}\text{GlcNAc}_2$ [43].

Mammalian-derived glycan structures are closer to humans because they have more common glycosylation machinery. However, distinctive glycosylation disparities exist for proteins derived from mouse cells, human cells and transgenic animals. N-glycolylneuraminic acid (NeuGc), a derivative of the human sialic acid N-acetylneuraminic acid (NeuAc), has been found to be more prevalent than NeuAc in antibodies derived from mouse or human-mouse hybridoma [44]. Since human glycoproteins generally do not contain NeuGc, a high level of NeuGc in the glycoprotein does not only lead to a quick removal of the molecule from the circulatory system but also can induce an immune response which is characterized by high titres of anti-NeuGc antibodies [45]. Despite this major drawback of mouse-derived producer cell lines, there are also some advantages regarding glycosylation of IgG1-type-antibodies that concern their therapeutic efficacy. It was found that the ADCC (antibody dependent cellular cytotoxicity) depends on particular N-glycan-structures that are located at the conserved Fc-glycosylation site at Asn-297. The two favourable glycan structures are the biantennary N-glycan-structure with a bisecting GlcNAc and a non-core-fucosylated biantennary structure which are predominantly produced by mouse-derived cell lines [46-48]. NeuAc is the dominant sialic acid present in CHO cell-derived glycoproteins although small amounts of NeuGc have also been detected [49]. However, CHO cells are one of the few expression systems which are almost human-like. Small differences appear due to lack of a functional $\alpha 2,6$ -sialyltransferase (ST) enzyme in CHO cells which is typical for humans. Thus, they are only able to synthesize $\alpha 2,3$ -linked terminal sialic acids via $\alpha 2,3$ -ST [50]. In contrast to CHO cells, human and mouse cell lines have both enzymes and, therefore, express both sialic acid linkages. However, CHO cells have been shown to be capable of producing glycoproteins with both $\alpha 2,3$ -linked and $\alpha 2,6$ -linked sialic acid after genetic glycoengineering, particularly after they have been transfected by a cloned rat $\alpha 2,6$ -ST gene [50, 51]. Mouse-human hybridomas have been found to perform the glycosylation characteristics of the mouse parental line with all its drawbacks e.g. NeuGc-production [52], but also its advantages concerning

ADCC. For the glycoproteins derived from the milk of transgenic animals, low percentages of complex-type glycans have been observed. For example, a greater proportion of truncated and high-mannose structures has been observed for interferon- γ expressed in transgenic mice [53].

As mentioned above, scientific work is directed to overcome natural hurdles in different expression systems incapable of human-like glycosylation by genetic glycoengineering approaches. Mammalian cell culture processes are very cost intensive due to low productivity and high safety demands. Therefore some companies use transgenic plants as alternative and cost-saving production systems [54-56]. Yeast cells are also in the focus of intensive research in pathway manipulation to create robust cell lines with high productivity and human-like glycosylation [57-59]. But until now, a lot of problems have to be solved and it is questionable if the resulting manipulated plant or yeast cell line really enables the biopharmaceutical industry to establish cheaper and safer production processes with a comparable product efficacy.

2.3.3 Culture environment

There have been done intensive studies examining glycosylation profiles of mammalian cell culture-derived glycoproteins in dependence on culture conditions. There are a lot of contradictory results, that prevent from describing general rules how the glycosylation pattern can be influenced by environmental parameters of the production process. The specific glycosylation of a product is more dependent on the expression system and the genetic variability of the producer clones rather than on the relative stable and standardized culture conditions of biotechnological production processes which often influence the glycosylation pattern only in a minor way. However, many scientific articles in this special field are worth to be mentioned.

For example, pioneering work has demonstrated that macroheterogeneity of IFN- γ glycosylation changes dramatically during batch culture of recombinant CHO cells [60]. Early work on t-PA showed that culture environment can influence both the macroheterogeneity and microheterogeneity of

oligosaccharide structures of glycoproteins [61]. Lots of other results support these observations, but for getting practical implications for improvements of production processes it is necessary to fully understand the culture's environmental impacts on product glycosylation. One optimization task is the minimization of glycoprotein heterogeneity to facilitate the product characterization within the regular drug approval process, to improve product consistency between different batches (quality control of biological drugs), and to accelerate the development process of biosimilars (biogenerics) in the pharmaceutical research and development pipelines.

2.3.3.1 Cell culture medium composition

The content of many nutrition media used in cell culture may have an impact on product glycosylation. There are some studies that have suggested that product glycosylation is affected by the glucose concentration in the culture medium. In one study, the degree of glycosylation of monoclonal antibodies produced by human hybridomas in batch culture has been reported to be influenced by the availability of monosaccharides [62]. It was observed that in glucose-limited chemostat cultures, cells at low growth rates produce product with a lower degree of glycosylation compared to faster-growing cells [63]. One hypothesis is that a change in the glucose metabolism (due to limiting the culture on glucose) or a lack of key nucleotide sugars (e.g. UDP-GlcNAc) is essential to the assembly of oligosaccharides on glycoproteins. Hayter examined further that pulsed additions of glucose caused a rapid improvement in the proportion of fully glycosylated IFN- γ . Secretion and cell growth also increased, but the glycosylation deteriorated significantly once the glucose had been depleted [64]. Lipid supplements, also key ingredients in nutrition media, are particularly important for the biosynthesis of dolichol, a key carrier for the oligosaccharide moiety (Glc₂Man₉GlcNAc₂-P-P-Dol) before it is transferred to the nascent protein. The addition of lipid supplements alone or in combination with lipoprotein carriers has been shown to improve the N-glycosylation site occupancy of IFN- γ [65, 66]. Protein glycosylation is also altered by the supplementation of precursors for cytidine and uridine using the availability of certain nucleotide sugars, e.g. shown for rat

hepatocytes [67]. A major obstacle of cell culture and upstream processing today is the adaptation of serum-containing (BSA or HSA) culture to serum-free culture conditions because of possible prion contaminations in the product. But proteins derived from serum-containing culture and serum-free culture have shown differences in glycosylation patterns. Patel et al examined that monoclonal IgG produced by mouse hybridomas in a serum-free medium had higher levels of terminal sialic acid and galactose residues compared to that produced using serum [4]. However, in another study better galactosylation was observed for antibodies produced from serum-containing culture of CHO cells [68]. When recombinant IL-2 producing BHK-21 cells were adapted from serum-containing to serum-free medium, this resulted in substantial changes to its glycosylation. The glycan chains increased their complexity, showing a higher number of arms and higher levels of terminal sialylation and proximal α 1-6 fucosylation. The overall level of glycosylation also increased [69]. Another interesting study showed a more consistent glycosylation for a protein produced in cell culture in comparison to the production in ascites fluid [70]. Growth factors are another important group of medium constituents which are often added to promote cell proliferation and increase the cell productivity. But it is not unlikely that the expression levels of glycosyltransferase may also be influenced by the cells' growth rate. For example, the activity of GlcNAc-transferase V has been correlated with the growth rate of HepG2 cells [71]. In 1990, Nakao et al reported that the addition of growth factor IL-6 to a myeloma cell line reduced the activity of N-acetylglucosaminyltransferase III (GnT-III), but increased the activity of GnT-IV and GnT-V, leading to altered oligosaccharide structures [72].

2.3.3.2 Other intracellular events

Since nearly three decades, it has been known that glycosylation of a protein usually occurs during translation [73]. The spatial and temporal competition among translation and glycosylation has been proposed, and investigated by several groups. It was suggested that there is only a brief moment in time when potential glycosylation sites on a nascent polypeptide are near the specific region of space where oligosaccharyltransferase active sites are

located [74]. Shelikoff et al could show that by lowering the protein synthesis rate with cycloheximide, the glycosylation site occupancy of recombinant prolactin produced by C121 cells is improved [75]. However, taking tPA as model protein produced in CHO cells does not give such convincing results. Bulleid et al showed nearly no influence between the rate of protein synthesis and protein glycosylation [76].

Another type of competition takes place between glycosylation and protein folding, for example the normally unglycosylated potential glycosylation site of ovalbumin becomes glycosylated after denaturation [77]. Protein folding is often arranged through disulfide bond formation. And the time window which is open for the formation of disulfide bonds influences the efficiency of N-linked glycosylation in certain proteins. When a potential glycosylation site is masked by the folding process then afterwards glycosylation is prevented. This hypothesis is also confirmed by the examination that low concentrations of the reducing agent dithiothreitol prevent cotranslational disulfide bond formation in the endoplasmic reticulum and lead to complete glycosylation of a t-PA-sequon that normally undergoes variable glycosylation [78]. Miletich et al examined the partial glycosylation site occupancy of a non-standard Asn-X-Cys-sequon in protein C which was supposed to be the result from site obstruction by disulfide bond formation [79].

Another important intracellular event that may affect the extent of protein glycosylation is the expression level of chaperone proteins inside the ER. Chaperone proteins (chaperones) represent one major class of ER-proteins that facilitate protein folding. One important group of chaperone proteins that have an influence on product glycosylation are the glucose-regulated binding proteins (BiPs). BiPs may affect the oligosaccharide profile by selective retaining of non-glycosylated proteins. Dorner et al extensively studied the influence of BiPs on the glycosylation ability of different cells. In 1987 they correlated the N-linked glycosylation site occupancy of three different products (tPA, Factor VIII and von Willebrand Factor) and associated it with the level of BiPs in the producer cell line [80]. Further studies indicated that levels of BiP were decreased by co-expressing antisense BiP genes, secretion of nonglycosylated tPA increased proportionally [81]. Association

with BiP may reflect aggregation or inefficient folding of nonglycosylated proteins. Therefore, overexpression of BiP would be expected to increase the selective retention of unglycosylated protein.

2.3.3.3 Physical parameters of culture

The effects of changing the physical parameters of cell culture have also been extensively studied. But as mentioned above the results are partially controversial. The concentration of oxygen (pO_2) is a critical parameter for biotechnological production processes regarding productivity and viability of the cells, but also regarding glycosylation. There are several studies which describe a decrease in the level of sialylation of a recombinant protein when the producer cells are in an hypoxic state. For example, CHO-derived follicle stimulating hormone (FSH) was influenced by hypoxia in this way [82]. Though Lin et al described nearly no effects on the glycosylation of tPA under hypoxic conditions in CHO cells [83], the majority of published studies described a significant influence in the way that the glycosylation process was disturbed and that more not fully-processed oligosaccharide structures were released [84-86].

Another physical parameter with a significant influence on glycosylation is the pH value. Changes within a range of 6.9 - 8.2 in the cell culture medium do not have a dramatic effect on the glycosylation profile of recombinant placental lactogen expressed in CHO cells. But leaving this range can be critical for glycosylation [87]. Similarly, increases in the concentration of ammonium ions in the culture medium resulted in reduced sialylation. This may be explained by the fact that ammonium ion concentrations above 2mM compromise sialyltransferases present in the Golgi, shown on G-CSF with a reduced sialylation degree produced by recombinant CHO cells [88].

2.3.3.4 Degradation of product glycosylation

There are a lot of reports which describe the effect of oligosaccharide degradation within a biotechnological production process. Culture time can play a critical role in defining product quality. For example, Sliwkowski et al have reported the loss of sialic acid from CHO-derived human deoxyribonuclease I over the course of a batch culture [89]. Later studies on interferon- γ confirmed this observation [90]. This phenomenon could be explained by studies of Goochee et al who reported a high sialidase activity in the medium due to cell lysis. Several types of glycosidase activity have been measured in CHO cell lysate and culture supernatant, and sialidase has been reported to be of the greatest activity [91]. Further studies showed that extracellular sialidase activity arising from cell lysis is capable of desialylating exogenously-added glycoproteins in CHO cell culture [92, 93]. Although sialidase, β -galactosidase, β -hexosaminidase, and fucosidase can be detected at low levels in supernatants from mouse 293, NSO and hybridoma cells, the sialidase activity in these cell lines is much lower than that found in CHO cells [94]. Later studies by Gu et al showed that the degradation process can be stopped by adding the sialidase inhibitor 2,3-dehydro-2-deoxy-N-acetylneuraminic acid to the CHO cell culture so that nearly no loss in sialylation over the culture time occurred [95].

2.4 Bio-analytical techniques for glycosylation analysis -

Overview

With the increasing number of new approved biological drugs in the market, there was a parallel development in analytical techniques for their characterization. Companies all over the world are driven by the great interest of producing recombinant therapeutic glycoproteins with a consistent homogeneous glycosylation to meet the increasing regulatory demands during the drug approval process and the regular quality control. As a result since the early 1980s much advancement has been made in the development of analytical methodologies to characterize both the macro- and

microheterogeneity in protein glycosylation, and not only regarding the establishment of new methods, but also regarding their qualification and validation in a pharmaceutical GMP-/GLP-environment.

2.4.1 Analytical techniques for glycosylation macroheterogeneity

To get information about the general presence of glycans on a protein, several analytical tools from classical protein analysis are useful. SDS-PAGE and Western blots combined with endo- and exoglycosidase treatment can give indications if the protein is glycosylated or not. For example, in a non-reduced SDS-PAGE, glycosylated molecules normally give a group of bands at the expected molecular weight, whereas nonglycosylated proteins give only one sharp band. However, there are some drawbacks associated with these techniques. Preparation of gels can be time-consuming, and at times relatively unreliable. In addition, no site-specific information about product glycosylation can be obtained by these techniques. Isoelectric focusing gel electrophoresis has been extensively used for assessing gross heterogeneity in product glycosylation, primarily the degree of sialylation but also with the drawbacks mentioned above.

When it was realized that gel electrophoresis is useful for glycosylation analysis in principle, a lot of research went into the development of techniques with higher resolution. One emerging newer technique is the capillary electrophoresis (CE), a separation procedure facilitated by high voltage-induced migration inside narrow-bore capillaries. CE has become a favorable technique for profiling glycoprotein macroheterogeneity due to its capability of separating even large molecules like monoclonal antibodies with low percentages of glycosylation into their different glycoforms (Figure 6).

Confirmation of NGHC by CGE

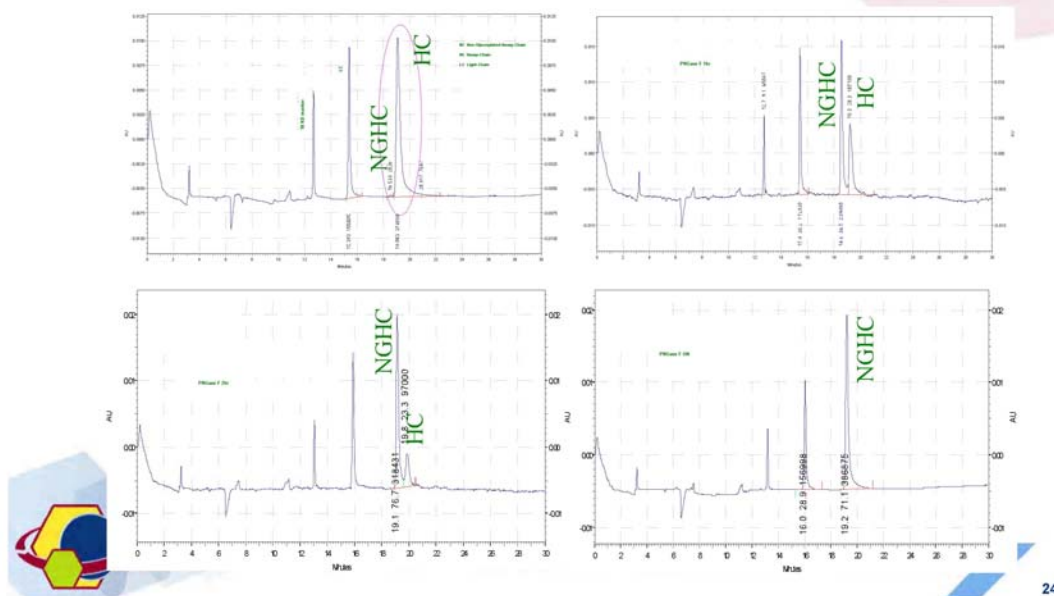


Figure 6: Confirmation of non-glycosylated heavy chain (NGHC) of a mab by capillary gel electrophoresis (CGE)³

Various modes of CE (e.g. zone electrophoresis, isoelectric focusing, isotachopheresis and micellar electrokinetic chromatography) are available and have demonstrated the capacity to characterize the glycosylation of various recombinant proteins, including EPO and interferon- γ [53, 96].

Although CE has a low consumption of expensive reagents and solvents, a drawback of this methodology is the relative high price for the device itself, approximately 100.000 Euro (information by Beckman 2005). Small start-up biotechnology companies are often not able to pay that price for an essential but also very specialized analytical technique.

³ HC = heavy chain, modified from Que, A. H. - Pfizer Global Biologics - presented at Glycomics and Carbohydrates in Drug Development, March 2005, La Jolla - California

2.4.2 Analytical techniques for glycosylation microheterogeneity

The detailed analysis of glycan structures needs much more effort than the macroheterogeneity profiling. In general, there are two approaches for microheterogeneity profiling. The first approach is the analysis of glycopeptides. Therefore the glycoprotein can be proteolytically digested to get glycosylated and non-glycosylated fragments. These glycopeptides can be analyzed by mass spectrometry, e.g. MALDI-TOF-MS. The second approach is the analysis of free glycans. Here, the oligosaccharide moieties are chemically or enzymatically cleaved from the intact peptide backbone and then chromatographic or spectrometric analyses follow.

2.4.2.1 Analysis of free glycans

Most of the used methods require two steps: the preparation of free oligosaccharides followed by their analysis. Preparation of oligosaccharides means in general the cleavage of the carbohydrate moiety from the peptide backbone. The most common way of preparing N-linked glycans is the digestion with the enzyme peptide N-glycosidase F (PNGase F), which cleaves most common mammalian N-linked oligosaccharides at the N-glycosidic linkage [97]. To prepare N- and O-linked glycans of a protein together, the chemical cleavage of carbohydrates is the preferred way. Chemical cleavage of carbohydrates can be achieved by hydrazinolysis or β -elimination [98]. O-linked oligosaccharides can be selectively released by alkaline β -elimination [99, 100].

The analysis of released glycans can be performed in many different ways. Ziad El Rassi gives a very detailed overview in his book "Carbohydrate analysis by modern chromatography and electrophoresis" [101].

The most widespread method for the analysis of oligosaccharides today, is the High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) developed and patented by the Dionex Corporation [102-106]. This method became widely accepted within the biopharmaceutical industry because of its good reliability and easy establishment in the companies. Another advantage is the fact that this

method is relative easy to validate and validation is a necessary requirement for drug release testing. HPAEC-PAD is one of the few methods that are able to profile glycan structures directly without derivatization. Other methods of detection are UV or RI (Refraction Index) but PAD is much more sensitive and specific for oligosaccharides. HPAEC-PAD works with an electrochemical detection at a gold electrode. The carbohydrates are ionized in a strong alkaline milieu and get separated on different ion exchange matrices. However, there are some disadvantages of this method. Regarding quantification, the response factors of the different signals play a major role. Those response factors are not equal for the different oligosaccharide structures [107]. Unless there is no availability of an oligosaccharide reference library, chromatograms showing glycosylation patterns can only be used for comparison aspects, e.g. confirmation of batch-to-batch-consistency, but quantification of the glycans is not possible.

The other major approach for glycosylation analysis is the release of the oligosaccharides and additional derivatization with a fluorophore. This makes a fluorimetric detection possible which is the most sensitive detection method today, going into ranges of attomol as detection limit. Although the preparation of the oligosaccharides is more time- and labor-consuming, this method has one immense advantage: the response factors of all labeled glycans are equal which makes a general relative quantification possible. Besides, an absolute quantification is enabled by use of only one weighable reference standard, giving the opportunity not only to compare different products by their microheterogeneity but also by their macroheterogeneity.

The most widespread separation techniques of labeled glycans are anion exchange chromatography (AEX) and hydrophilic interaction chromatography (HILIC). Anion exchange chromatography is normally used for charged glycans, e.g. sialylated oligosaccharides, and hydrophilic interaction chromatography is preferably used for neutral glycans. The group of sialic acids contains several monosaccharides with one carbonic acid function which gives the molecule one negative charge at pH 7. Neutral glycans are uncharged glycans, e.g. high-mannose and desialylated oligosaccharides. There is very often an analytical regime where the total amount of released

glycans is divided into two groups. One group is analyzed by AEX and the other group is treated with the enzyme sialidase to desialylate the glycans and to get uncharged glycans. This second group is then further analyzed by HILIC.

Anion exchange chromatography of charged oligosaccharides is often performed on weak anion exchange matrices like DEAE-resins (diethyl-amino-ethyl). The resin itself often consists of inert divinyl benzene.

Hydrophilic interaction chromatography (HILIC) is a relative new term for a separation technique that has its earliest roots in 1975 [108-110]. The term HILIC was invented by Alpert in 1990 [111]. Four years later, he published a study focusing on the separation of complex carbohydrates with this technique [112]. Today this technique is the most widely used separation mode for carbohydrate analysis with HPLC using an amino-bonded silica gel column with acetonitrile-water mixture as mobile phase. HILIC needs a hydrophilic stationary phase and a hydrophobic mobile phase. It is a kind of "normal-phase" chromatography where elution is promoted by the use of more polar (often more aqueous) mobile phases. The order of elution is approximately the opposite of that expected for RPC (reversed phase chromatography), a special kind of HIC (hydrophobic interaction chromatography). The retention of carbohydrates on amino-bonded silica is not directly based on the positive charge of the amino-groups but on the hydrophilic character of them. The partitioning mechanism of the carbohydrates is as follows: the stationary phase retains a semi-immobilized layer of mobile phase enriched with water. Chromatography of carbohydrates involves partitioning between this stagnant aqueous layer and the bulk of the (mostly organic) mobile phase. Thus, HILIC is a special case of partition chromatography which distinguishes this variant of normal-phase chromatography from other variants involving adsorption directly on the stationary phase. Addition of salts like ammonium formiate, phosphate or acetate to the water phase can be used to adjust the pH of the mobile phase as well as accelerate the elution of the oligosaccharides. Untreated saccharides tend to react with the amino functions of the stationary phases via their aldehyde groups, resulting in imin formation. After labeling with e.g.

2-aminobenzamide these aldehyde functions are blocked. This is another important reason for precolumn-derivatization of glycans, to inactivate their very reactive aldehyde groups.

Earlier methods in glycosylation analysis included gel permeation chromatography [61] and paper electrophoresis [113], but are in most cases obsolete today.

However, one major drawback of the analysis of released free oligosaccharides is that the glycans are separated from the glycoprotein as a pool, site-specific glycosylation information is generally lost by this approach for all glycoproteins containing more than one glycosylation site.

2.4.2.2 Analysis of glycopeptides

The analysis of glycopeptides has the advantage that site-specific glycan information can be produced. Many products are shown to exhibit glycan differences between individual glycosylation sites on the same protein [35].

To be able to identify the different glycosylation sites of a protein, peptide maps obtained by proteolytic digestion (e.g. trypsin) of the intact glycoprotein are compared to those observed when the protein has been deglycosylated previously (e.g. with glycosidases like PNGase F) [114]. PNGase F is frequently used for this purpose due to its broad substrate specificity for the hydrolysis of all commonly encountered N-linked glycans.

For the determination of specific attachment sites, the glycosylated and the deglycosylated peptides are typically analyzed by mass spectrometry. The peptide masses give information about the glycosylation sites and by calculating the mass differences of the corresponding peptides, it is possible to get information about the attached glycans concerning composition, heterogeneity and sequence information.

2.4.2.3 Identification of oligosaccharides and glycopeptides by mass spectrometry

Mass spectrometric (MS) techniques have begun to play a dominant role in the structure determination of carbohydrates and peptides of glycoproteins. MS techniques are mainly divided into several groups by their ionization methods. One of the earliest MS technique for biomolecules is the fast atom bombardment (FAB) mass spectrometry. Later, electrospray ionisation (ESI) and matrix-assisted laser desorption ionization (MALDI) MS techniques followed. FAB-MS offers the highest mass accuracy of these methods and has been extensively used to deduce the glycan structures of several glycoproteins [115]. However, the drawbacks of this method are the high instrument price as well as the relative large sample amount which is needed. Although ESI-MS and MALDI-MS do not offer quite the mass accuracy of FAB-MS, they exhibit greater mass ranges than FAB-MS and require much smaller sample amounts for analysis. ESI-MS has also the advantage that it can be coupled directly to HPLC- and CE-devices so that online detection is possible [116-118]. Newer ESI-MS-devices are connected to an ion trap so that fragmentation analysis of carbohydrates gets feasible (ESI-IT-MSⁿ-analysis). The specific fragmentation pattern of an oligosaccharide does not only give information about its mass, but also about the linkages between its monosaccharides [119-122]. With this method, an exoglycosidase sequencing for linkage analysis is unnecessary. But those devices are expensive and this is especially a problem for small biotech companies with limited budgets.

For these companies, MALDI-MS is an useful alternative. It is the least expensive MS-method and it is more robust regarding buffer salts in the analytes than ESI-MS. Newer MALDI-MS-techniques even allow ion trap fragmentation experiments [123]. MALDI-MS is also the preferred MS-technique for the analysis of large biomolecules like peptides and proteins with molecular weights up to 200.000 Da, e.g. IgG with 150.000 Da.

In MALDI-MS, a low concentration of analyte molecules that exhibit only a moderate absorption at the emission wavelength of a laser, is embedded in either a solid or liquid matrix consisting of a small, highly-absorbing species.

In theory, the matrix serves two major functions, absorption of energy from the laser light and isolation of the analyte molecules from each other. The energy of the laser is transmitted by the matrix to the analyte molecules and leads to a desorption of matrix and analyte molecules from the sample plate. The desorbed molecules get softly ionized in vacuum, resulting in molecules with one positive or one negative charge (Figure 7), depending on the operation mode.

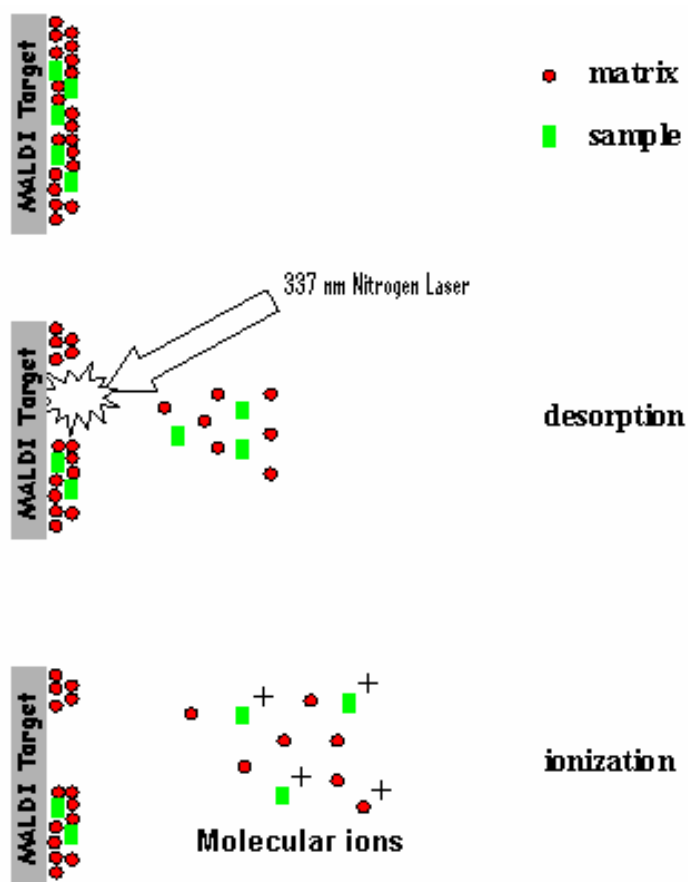


Figure 7: Principle of laser desorption ionisation⁴

The resulting metastable ions can now be accelerated in an electrical field by applying a high voltage to the molecules. Ion detection is generally accomplished by a time-of-flight (TOF) detector (Figure 8).

⁴ www.srsmaldi.com

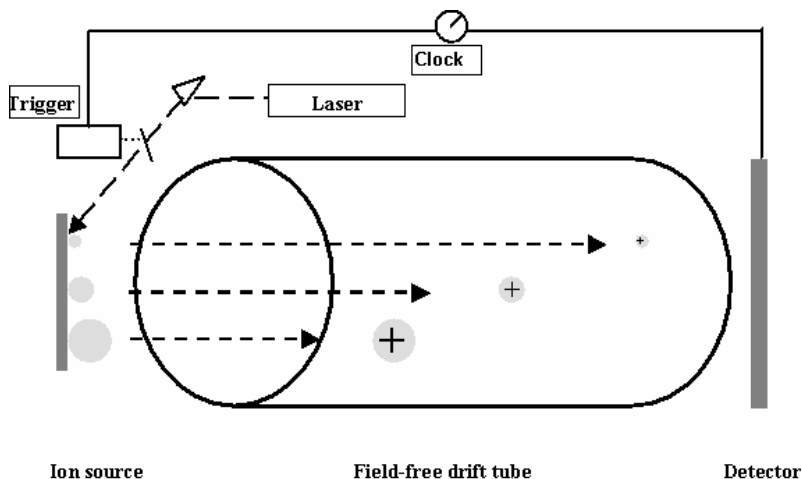


Figure 8: MALDI-TOF mass spectrometer operating in the linear mode⁵

The arrival time of an individual ion at the detector is proportional to $(m/z)^{1/2}$ of the particular species. If the ion beam approaches the detector in a straight line, the method works in the linear mode. If the ion beam is reflected by an ion mirror before it reaches the detector, the reflector mode is applied to the TOF-detector (Figure 9).

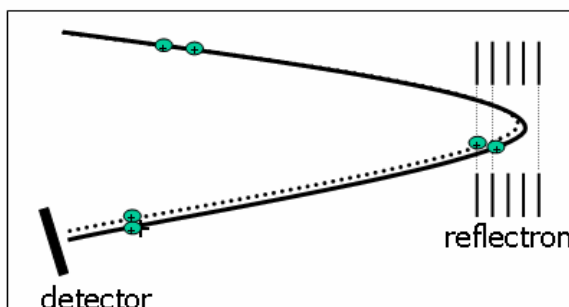


Figure 9: MALDI-TOF mass spectrometer operating in the reflectron mode⁶

The simplified mathematical basics for the TOF-principle are listed beneath. The ions are accelerated by the source electric field (E_S) and enter the drift region with the same kinetic energy (E_K).

⁵ Ions are separated according to their mass depending velocities. www.srsmaldi.com

⁶ The initial velocity distribution of ions of same mass can be corrected. www.srsmaldi.com

$$E_K = zeE_s = \frac{1}{2}mv^2$$

Time required to reach the detector is determined by the mass (m) and the charge (z) of the molecule.

$$t = \left(\frac{m}{2zeE_s} \right)^{\frac{1}{2}}$$

The measured times can be converted to mass/charge (m/z) quotients which are plotted in the mass spectrum.

$$\frac{m}{z} = 2eE_s \left(\frac{t}{D} \right)^{\frac{1}{2}}$$

t = time

e = electron volt

v = velocity

D = detector constant

Table 1 shows the most common MALDI matrices and their applications (Table 1).

Table 1: MALDI matrices and their applications

<u>Matrix</u>	<u>Applications</u>
sinapinic acid	proteins
α-cyano-4-hydroxycinnamic acid	peptides
2,5-dihydroxybenzoic acid	oligosaccharides

For carbohydrate analysis, Huberty et al found out that in the linear mode carbohydrate composition could be attained by digestion of the corresponding glycopeptides with mixtures of glycosidases. They also found

out that this procedure can be simplified by using the reflector mode mass spectrometry which is able to identify the carbohydrate components without requiring digestion by glycosidases [124].

Site-specific glycoform microheterogeneity is best obtained by analyzing glycopeptides resulting from proteolysis of glycoproteins [125, 126]. If each glycosylation site is located on a separate proteolytic fragment, the fragments can be fractionated by reversed-phase HPLC and afterwards be analyzed by MALDI-TOF-MS. Pools of microheterogeneous glycopeptides representing each glycosylation site give several masses within the spectrum. It is now possible to subtract the resulting masses of one glycopeptide pool by the known mass of its peptide portion. Due to the capability of MALDI-TOF to analyze mixtures, the mass shift of each glycopeptide peak to the pure peptide mass can be correlated to a site-specific oligosaccharide structure.

It is also possible to analyze free oligosaccharides by MALDI-TOF-MS. Therefore the glycans are solved in an organic solution of 2,5-dihydroxybenzoic acid. After pipetting a small amount of this solution on a MALDI-target, the high volatile solvent evaporates and the oligosaccharides are embedded in a solid matrix. The detected mass range of the glycans comprises $m/z = 1000 - 4000$. Within this range, most of all possible glycan structures can be found.

2.4.3 Screening methods: Glycoanalytical Fingerprinting

Although MALDI-TOF analysis is very rapid, sample preparation steps often require a lot of time. In the biopharmaceutical industry, glycosylation analysis is usually done with a purified glycoprotein within quality control. The whole purification (downstream process) of the protein from culture supernatant, cell debris and other relicts of the cell culture procedure (upstream process) has been performed before analysis. The sample preparation within the glycoanalytical procedure including buffer exchange, proteolysis of intact glycoprotein and isolation of pools of glycopeptides or even further preparation of free glycans can require several days or even weeks, including a lot of pipetting, purification and concentration steps. Due to the lack of

automation, these methods are not suitable as broad screening tools. Therefore, glycoform microheterogeneity is typically assessed only following termination of cell culture. Since it has been reported that glycosylation microheterogeneity changes during the course of cell culture, the development of rapid and sensitive analytical methods to monitor changes in glycosylation patterns is needed.

However, there are new approaches to make online glycosylation analysis possible within cell culture. These methods are often based on micro-array technology and have been implemented in applications for glycomics [127-133]. A lot of these arrays work with lectins which naturally have a high affinity to carbohydrates [134, 135]. The company Procognia® (www.procognia.com) has developed a technology for automated glycoanalysis which does not need sample purification and sample cleavage or separation and which gives complete quantitative data in less than four hours (manufacturers' instructions). If this technique will become widely accepted in the future, remains to be seen. Until now, the reproducibility of the quantitative analysis with this technique is not as good as the conventional methods (relative standard deviation: 10%) [136].