

Chapter 4

Post-Translationally Modified Proteins: Glycosylation and Disulfide Bond Formation

Fotini N. Bazoti and Anthony Tsarbopoulos

Abbreviations

AEC	Anion-Exchange Chromatography
BEMAD	β -Elimination followed by Michael Addition with Dithiothreitol
CE	Capillary Electrophoresis
CHO	Chinese Hamster Ovary
CHO IL-4	CHO-derived Interleukin-4
CDG	Congenital Disorders of Glycosylation
CDT	Carbohydrate-Deficient Transferrin
CID	Collision-Induced Dissociation
ConA	Concanavalin A
CFG	Consortium for Functional Glycomics
CZE	Capillary Zone Electrophoresis
CREB	cAMP-responsive Element-Binding Factor
DHB	2,5-Dihydroxybenzoic acid
ECD	Electron-Capture Dissociation
ESI	Electrospray Ionization
ETD	Electron-Transfer Dissociation
FAB	Fast Atom Bombardment
FT ICR	Fourier Transform Ion Cyclotron Resonance
GCC	Graphitized Carbon Columns
HILIC	Hydrophilic Interaction Chromatography
HPA	3-hydroxypicolinic acid
HPLC	High-Performance Liquid Chromatography
HPV	Human Papillomavirus

A. Tsarbopoulos (✉)

Department of Pharmacology, University of Athens Medical School, 75 M. Asias street, Athens 115 27, Greece

e-mail: atsarbop@med.uoa.gr; atsarbop@gnhm.gr

F. N. Bazoti

Bioanalytical Department, The Goulandris Natural History Museum, 13 Levidou street, Kifissia 145 62, Greece

e-mail: fbazoti@gmail.com

HRP	Horseradish Peroxidase
IL-4	Interleukin-4
IL-4R	Interleukin-4 receptor
IL-5R α	Interleukin-5 receptor α -subunit
IA	Immunoaffinity
IM	Ion Mobility
IT	Ion Trap
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAC	Lectin Affinity Chromatography
LC	Liquid Chromatography
LC-MS	Liquid Chromatography-Mass Spectrometry
LSI	Liquid Secondary Ion
mAb	Monoclonal Antibodies
MALDI	Matrix-Assisted Laser Desorption/Ionization
M-LAC	Multi-Lectin Affinity Chromatography
M _r	Molecular Weight
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass-to-charge ratio
nESI	Nano-Electrospray Ionization
NMR	Nuclear Magnetic Resonance
oTOF	Orthogonal Time-of-Flight
PAS	Periodate-acid-Schiff
PD	Plasma Desorption
PGC	Porous Graphitized Carbon Chromatography
PSA	Pisum Sativum Agglutinin
PSD	Post-Source Decay
PTMs	Post-Translational Modifications
rHuEPO	Recombinant Human Erythropoietin
r-RhCG	Recombinant human Chorionic Gonadotrophin
rtPA	Recombinant Tissue Plasminogen Activator
QTOF	Quadrupole Time-of-Flight
sAPP	Secreted Amyloid Precursor Protein
sDHB	2,5-Dihydroxybenzoic acid (DHB) with a 10 % admixture of 2-hydroxy-5-methoxybenzoic acid (super DHB)
SEC	Size-Exclusion Chromatography
SIM	Selected Ion Monitoring
SLAC	Serial Affinity Chromatography
SPE	Solid-Phase Extraction
TAS	Tagging-via-Substrate
Tf	Human Transferrin Glycoprotein
TOF	Time-of-Flight
uPAR	Urokinase-type Plasminogen Activator Receptor
UPLC	Ultra-Performance Liquid Chromatography
WGA	Wheat Germ Agglutinin

4.1 Introduction

Even though most medicines have historically been small molecules, many newly approved drugs over the last two decades have been derived from proteins. For the past few years, protein therapeutics have been enjoying the fastest growth within the global pharmaceutical industry. Protein-based therapeutics, such as insulin, interferons, monoclonal antibodies (mAb), growth hormones, erythropoietins, blood-clotting factors, colony-stimulating factors (CSFs), plasminogen activators, and reproductive hormones, play a significant role in the treatment of many major diseases, and protein therapies have revolutionized the methodology followed by drugs. These therapies exhibit high efficiency due to their targeted approach, which avoids side effects on healthy organs to a great extent. In recent years, the number of protein-based pharmaceuticals reaching the marketplace has increased exponentially, and they provide innovative as well as effective therapies for several chronic diseases which were previously not responsive to treatment. The global market for biologics or biotechnology therapeutics is one of the most prolific and fastest growing markets in the world, representing at least 24 and 22 % of all new chemical entities approved by the US and EU regulatory authorities, respectively [1]. Sales of biotech products in US showed an annual growth rate of 20 % between 2001 and 2006 compared with 6–8 % in the pharmaceutical market [2], and it is expected to grow at annual growth rate of around 13 % during the next three years (2012–2015), with the introduction of new protein therapeutics and enhanced investments contributing to this booming growth of this industry.

For protein therapeutics to be effective, they must be produced in biologically active forms, which require proper folding, and post-translational modifications (PTMs) with the extent of PTMs depending on the nature of the “host” cell and the conditions of the fermentation and recovery processes. Even though only a few biopharmaceutical proteins such as albumin (Recombunin) and insulin (Humulin N and Lispro) undergo simple modifications such that they can be manufactured using yeast or bacteria [3], most of the production platforms used to produce biopharmaceuticals comprise mammalian cells that have the ability to perform complex PTMs. The most prevalent modifications include variable glycosylation, formation of disulfide bonds, cysteine (C) and methionine (M) oxidation, phosphorylation, misfolding and aggregation, deamidation of asparagine (N) and glutamine (Q), and proteolysis at the C- and N-termini. Even though the presence of PTMs is often required for normal biological function or tissue disposition of the protein, in many cases, the role of the modification is as of yet unknown. Therefore, detailed characterization of these modifications is extremely important, because they may alter physical and chemical properties, folding, conformation distribution, stability, activity, which in turn may affect cellular processes, in which the protein is involved [4–6]. Examples of the latter can be regulation of signal transduction and a wide variety of cellular events such as growth, metabolism, proliferation and differentiation in case of protein phosphorylation [7–9], targeting, cell–matrix interaction, as well as pharmacokinetic and pharmacodynamic behavior in case of glycosylation [10, 11]. Therefore, a thorough verification

of the protein's (amino acid) sequence, assessment of the purity and impurities in a recombinant protein drug product along with a detailed characterization of the existing PTMs, is a regulatory requirement prior to its approval for clinical use [12].

Full structural characterization of the existing PTMs in a recombinant protein often poses a considerable analytical challenge owing to their inherent complexity. The presence of PTMs often complicates or even prevents the use of classical tools for protein sequence analysis (e.g., automated Edman degradation). Moreover, the presence of lipid or carbohydrate covalent attachments on proteins can dramatically decrease the accuracy of the molecular weight (M_r) measurement when using sedimentation velocity, gel permeation, or SDS-PAGE analysis. Separation techniques such as high-performance liquid chromatography (HPLC) or capillary electrophoresis (CE) combined with a variety of mass spectrometry (MS) techniques are commonly employed for the profiling and quantitation of PTMs present in recombinant therapeutic proteins. The development of electrospray ionization (ESI) [13, 14] MS coupled with online liquid chromatographic (LC-MS) or electrophoretic separation (CE-MS) [15, 16] and matrix-assisted laser desorption/ionization (MALDI) [17, 18] has established MS as the technology of choice for protein mapping, localization, structure identification, and quantification of existing PTMs [19, 20]. Several MS-based approaches have been developed employing tailored tandem MS scanning methods diagnostic for specific PTMs, such as monitoring precursor/product-ion transitions and neutral loss scan [21–23].

Recently, online LC-MS combined with collision-induced dissociation (CID) and electron-capture dissociation (ECD) [24] or electron-transfer dissociation (ETD) [25] fragmentation has been used to elucidate disulfide linkages and site-specific glycosylation in recombinant therapeutic proteins and glycoproteins [26, 27]. Similarly, MS-based approaches can be employed in the production of a recombinant therapeutic protein in order to ensure the purity, the production yield, and the absence of chemical degradation and/or aggregation products in the protein formulations for clinical and eventually commercial use.

In this chapter, we discuss MS-based methodologies that are employed to detect, identify, and characterize two of the most prevalent PTMs in the production of therapeutic recombinant proteins, glycosylation and disulfide bond formation. These MS-based approaches discussed here are representative of those used for the comprehensive characterization and quantitation of other PTMs encountered in recombinant proteins intended for therapeutic use in humans.

4.2 Glycosylation

Glycosylation process, that is, the covalent attachment of oligosaccharide chains on the protein backbone, is considered as the most important and common PTM of proteins. It is estimated that over 70 % of all human proteins are glycosylated [28] and 90 % of protein therapeutics are glycosylated [10]. The carbohydrate moieties of glycoproteins (glycans) can modulate the biological functions of a

glycoprotein such as circulation, cell-to-cell interactions, receptor binding, molecular and immune recognition, which in turn affect intracellular signaling, fertilization, embryonic development, immune defense, recognition of hormones, cell adhesion, and pathogenicity [4]. In addition, glycan-chain modification can significantly impact their physicochemical properties such as protein folding, solubility, stability, aggregation, and susceptibility to proteolysis [29]. Finally, carbohydrate modifications can also considerably alter protein conformation, which may consequently modulate the functional activity of the protein, especially in its interactions with other proteins or ligands. It has been established that altered glycosylation or variation of a protein's glycosylation pattern is associated with numerous diseases and disorders [30–32]. Therefore, detailed structural studies of the glycosylation and its inherent heterogeneity are also potentially vital toward understanding their function in complex physiopathological processes and establishing glycan profile changes between healthy and disease states [33, 34]. The latter has increased the potential of using glycan biomarkers for the diagnosis of several diseases [35], as well as for the design of new therapeutics [10, 36, 37]. Moreover, carbohydrate modification can be used toward the production of “custom-made” glycoproteins tailored, such as glycoproteins with defined homogeneous glycosylation structure, for specific therapeutic use [38].

Therefore, complete structural analysis of a glycoprotein end product will involve not only the determination of the primary peptide sequence, but also detailed analysis of the glycan structures including information on the individual glycosylation sites, the glycosylation patterns, and the structure elucidation of the attached carbohydrates (glycoproteome) [39–43]. As it has become obvious that many of the changes associated with disease and differentiation are due to the glycans attached to proteins (glycome), a thorough understanding of these glycan structures will be invaluable for gaining insight into their involvement in disease mechanisms and the potential for novel therapeutic interventions [44]. Characterizing the glycoproteome, however, is a challenging and daunting task because the structural heterogeneity of these glycans is vast, necessitating the development of highly sensitive and efficient analytical methods for detection, separation, and structural investigation of glycoproteins.

4.2.1 Intact Glycoprotein Analysis by Mass Spectrometry

An important preliminary step in the quality control and structure characterization of a therapeutic recombinant protein is the M_r determination of the protein product. On the intact glycoprotein level, non-spectrometric techniques such as SDS-PAGE, lectin affinity chromatography (LAC), isoelectric focusing (also in a capillary), or capillary zone electrophoresis (CZE) are generally used. In case of the two-dimensional (2D) gel electrophoresis separation of glycoproteins, characteristic spots reflecting their different isoelectric points and M_r of different glycoforms can be seen. The subsequent detection of the glycosylation pattern of the electroblotted glycoproteins

may be performed by LAC [45, 46], where carbohydrate-specific lectins can be used to probe distinct oligosaccharide structures (motifs). In addition, this affinity purification can be employed as an enrichment method for the glycosylated peptides and proteins (see Sect. 4.2.2.2). Nevertheless, the low solubility of the membrane glycoproteins, resulting in their poor detection, is a significant drawback of the 2D gel electrophoresis approach. An alternative method of higher resolving potential is CZE or CE, where the various glycoforms are detected even though no information on the nature of the attached glycans is revealed [47]. These electrophoretic methods have been successfully used in the separation of sialic acid isoforms of endogenous and recombinant glycoproteins, and they have proved their usefulness in clinical diagnosis and product quality assessment [48].

In the late 1980s, the incorporation of ESI and MALDI MS, along with advances in electrophoretic separations and high-resolution MS, has provided a powerful analytical tool for the analysis and even quantitation of the intact individual glycoforms in glycoproteins [15]. ESI and MALDI MS are the premier methods of choice for M_r measurement and the ensuing protein mapping. In case of ESI MS analysis of therapeutic proteins, spraying of an aqueous protein solution at $\mu\text{L}/\text{min}$ or nL/min flow rates generates multiply protonated signals with reduced mass-to-charge (m/z) ratios, thus making them readily detected by typical mass analyzers with a mass range up to 2,500 Da. This is demonstrated in the ESI mass spectrum of human recombinant interferon α -2b (INTRON A) (Fig. 4.1), which is used in the treatment of certain viral infections, including chronic hepatitis B, C, and D, malignant melanoma, follicular lymphoma, Kaposi's sarcoma caused by AIDS, and infections caused by human papillomavirus (HPV). The ESI mass spectrum

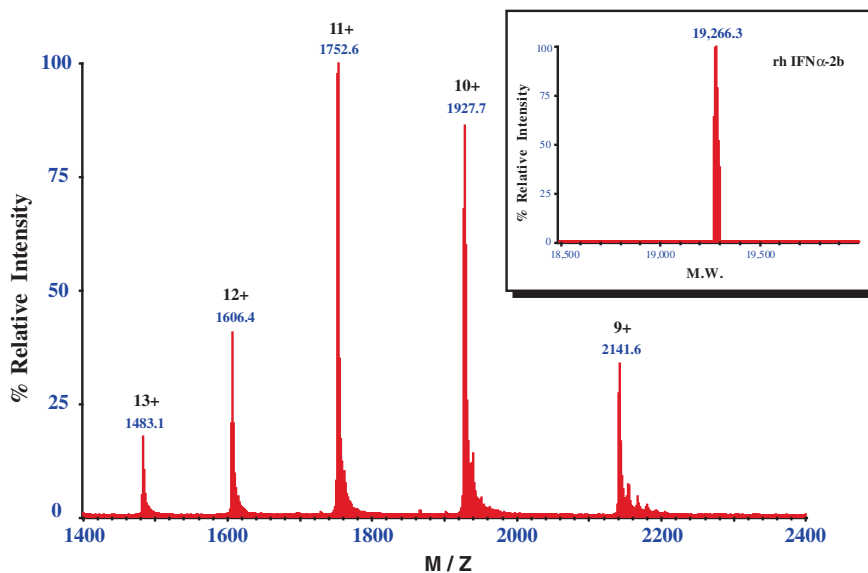


Fig. 4.1 Positive-ion ESI mass spectrum of human recombinant interferon α -2b (INTRON A)

exhibited a bell-shaped distribution of multiply charged ions ranging from the 9+ to the 13+ charge state, and the average M_r value derived from the five multiply charged ions present in the ESI mass spectrum was 19,266.3 (Fig. 4.1, inset). The excellent mass measurement accuracy, which is usually better than 0.01 % for masses up to 100 kDa [49], makes ESI MS an ideal preliminary method for monitoring the integrity of the therapeutic recombinant protein batches. In case of larger proteins, we observe greater charge states, often in the presence of a dilute acid, due to the presence of more available sites to carry the positive charge (i.e., K-, R-, H-, N-terminus). The simultaneous shift of the observed ion envelope distribution to lower m/z values is also accompanied by a decrease in the spacing between adjacent charge states, thus making the identification of the envelope's charge-state components difficult. This becomes more complicated in the analysis of a recombinant glycoprotein sample where the inherent complexity of the carbohydrate structure heterogeneity enhances the aforementioned analytical challenge. This complexity is shown in the ESI mass spectrum of the Chinese hamster ovary (CHO)-derived interleukin-4 (IL-4), a glycoprotein containing two N-linked glycosylation sites (Fig. 4.2) [50].

The ESI mass spectrum of CHO IL-4 (Fig. 4.2a) contained three envelopes of multiply charged ions ranging from 8+ to 10+ charge state, with each envelope containing several peaks corresponding to individual glycoforms of the glycoprotein and adducts thereof. This is better depicted in the deconvoluted mass spectrum (Fig. 4.2b), with the mono- and disialylated components (separated by 291 Da) representing the most abundant signals. Other higher M_r components indicated the presence of tri- and tetraantennary glycans containing up to three additional lactosamine units (in-chain mass of 365 Da), whereas satellite signals 98 Da higher were also observed (Fig. 4.2b). These signals probably arise from the attachment of sulfate groups, since sulfate salts were used in the protein isolation process and operating at a higher desolvation potential or using low pH solvents can minimize their formation [51]. Overall, the success of glycoprotein analysis by ESI MS depends on their relative carbohydrate content, with the success decreasing significantly with a relatively high percentage weight of the carbohydrate component. ESI MS analysis of complex glycoproteins by direct infusion often results in broad unresolved signals arising from the large number of different glycoforms and potential salt adducts, along with the ESI multiple charging phenomenon that spreads the signals over a large m/z region. In agreement with that, ESI MS analysis of the 44 kDa ovalbumin containing 4 % carbohydrate was successful [52], whereas glycoproteins with higher carbohydrate content such as the CHO IL-5 (M_r ~ 31 kDa; 15 % carbohydrate) and CHO IL-4 receptor (IL-4R; M_r ~ 38 kDa with 35 % carbohydrate) did not give any ESI signals [53]. Another contribution to the unsuccessful ESI MS analysis is the poor ionization efficiency in the positive ion mode due to the presence of negatively charged glycans, as this is demonstrated in the comparative analysis of recombinant human erythropoietin (rHuEPO) and its asialo counterpart [54]. The use of nano-electrospray ionization (nESI) overcomes this drawback and improves the sensitivity of analysis due to the generation of smaller-sized droplets [55]. Moreover, the interfacing of the nESI source with orthogonal time-of-flight

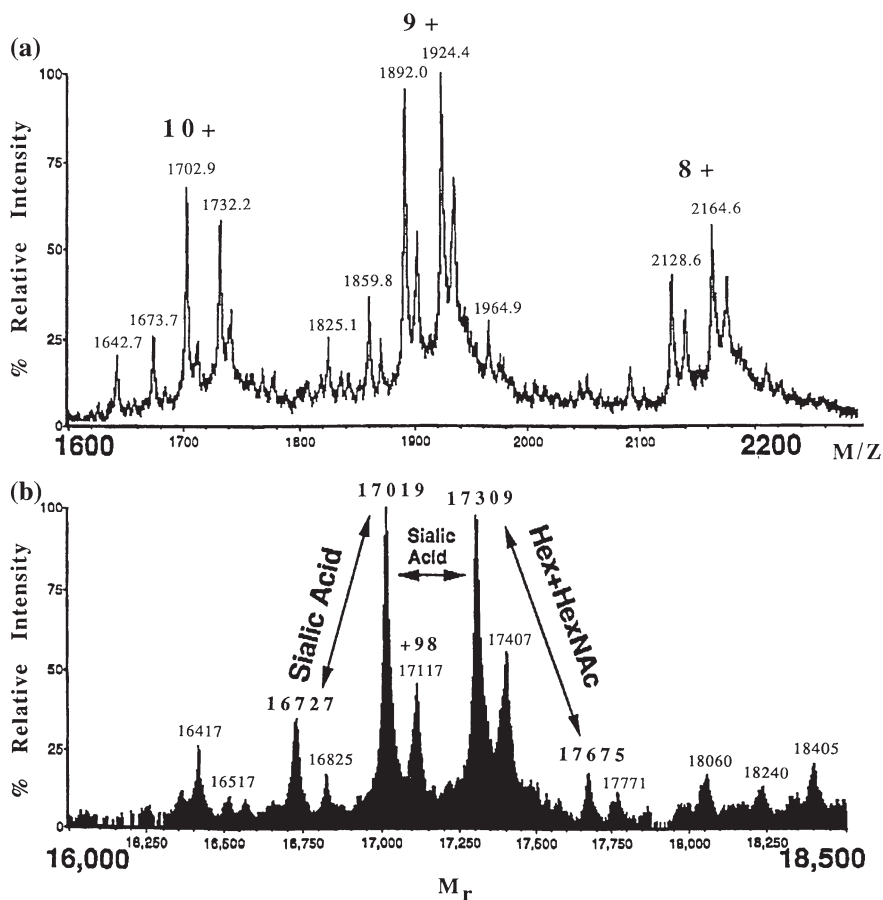


Fig. 4.2 Positive-ion ESI mass spectrum of CHO-derived interleukin-4 (IL-4): Raw spectrum (a) and deconvoluted spectrum (b) showing the individual glycoforms of the glycoprotein. (Reprinted with permission from Wiley [50])

(oTOF) instrumentation [56] has led to better mass measurement accuracy and increased analytical mass range, thus offering new momentum to the ESI MS analysis of glycoproteins. It should be emphasized that the commonly used quadrupole, quadrupole ion trap, and even Orbitrap [57] analyzers have mass range of analysis limited to m/z 2,000 and 4,000 (Orbitrap), which is a significant drawback when larger glycoproteins or non-covalent complexes thereof must be detected; thus, an upper mass limit greater than even m/z 10,000 may be required [58, 59]. This is shown in the nESI mass spectrum of Sf9-derived IL-4R ($M_r \sim 30.2$ kDa) obtained on an oTOF mass spectrometer, where an extensive series of multiply charged ions up to m/z 3,000 corresponding to two sets of high-mannose glycoforms separated by a fucosylated $\text{Man}_3(\text{GlcNAc})_2$ structure (in-chain mass of 1,039 Da) were observed [51]. Therefore, the improved mass resolving power, sensitivity and extended mass

range, has made the oTOF, hybrid quadrupole TOF (QTOF), and recently the ion mobility (IM) [60] TOF as the analyzers of choice for nESI MS analysis of glycoproteins. The use of the IM TOF analyzer is nicely shown in the nESI mass spectrum of the intact therapeutic mAb trastuzumab (Herceptin), which is a humanized monoclonal immunoglobulin γ -1 (IgG1) antibody directed against the HER2/neu receptor, which is over-expressed in about 25 % of all breast cancer patients [61]. In the ESI mass spectrum of trastuzumab obtained on an IM TOF mass spectrometer [62] (Fig. 4.3), an extensive series of multiply charged ions ranging from the 35+ up to the 75+ charge state were observed, and the separation between successive charge states was sufficient to reveal the presence of six glycoform variants. The illustration of these glycoforms is portrayed in the zoomed spectrum for the 53+ charge state (Fig. 4.3b), while their respective assignment is shown in the deconvoluted mass spectrum (Fig. 4.3c). The spectrum clearly reveals the glycoprofile difference between trastuzumab antibodies from different batches (shown in different colors) where the intensity of each glycoform varies.

It should be mentioned that the mass measurement accuracy of the main glycoform is within 1.5–2 Da (~10 ppm) from its theoretical mass value (148,057 Da), an unthinkable achievement prior to the advent of ESI and MALDI MS. The latter is an essential attribute of this method and renders it suitable to distinguish the lot-to-lot heterogeneity in glycosylation profile of the commercially available glycoprotein biopharmaceutical. Glycoprotein heterogeneity can result in an enhancement or loss of the protein's biological activity, as this has been demonstrated in the case of rHuEPO, where desialylation causes complete loss of its hormonal activity *in vivo* [63]. In particular, intravenously administered rHuEPO consisting of highly branched sialylated oligosaccharide structures has been shown to result in a plasma half-life of 5–6 h as compared to desialylated rHuEPO, which is cleared within minutes [64].

On the other hand, glycoprotein analysis by MALDI MS yields signals corresponding to protonated molecules (MH^+) of the individual glycoforms and allows the determination of the heterogeneity for glycoproteins with M_r less than 30 kDa and a relatively low percentage of carbohydrate content. This is clearly shown in the screening of the glycosylation profile for the human soluble urokinase-type plasminogen activator receptor (uPAR) expressed in CHO cells, where the extent and type of glycosylation in its three domains was assessed by MALDI MS [65]. On the contrary, MALDI MS analysis of the Sf9-derived interleukin-5 receptor α -subunit (IL-5R α) [66] and CHO IL-4R [53] with 17 and 35 % carbohydrate content, respectively, did not provide any information on the type of the glycosylation. In addition, the choice of an appropriate MALDI matrix is very important toward achieving the optimum mass resolving power and separation of the individual glycoform signals [67, 68]. This is shown in the MALDI mass spectra of the Sf9-derived IL-5R α (in a reflectron and a linear TOF instrument) using different matrices, where the use of the sDHB matrix (2,5-dihydroxybenzoic acid with a 10 % admixture of 2-hydroxy-5-methoxybenzoic acid) in a linear TOF instrument resulted in a more reliable mass measurement due to minimized metastable fragmentation [69] (Fig. 4.4).

Fig. 4.3 Positive-ion ESI ion mobility (IM) TOF mass spectrum of the intact therapeutic monoclonal antibody trastuzumab (Herceptin) (a); The 53+ charged ion with the signals corresponding to various glycoforms is annotated (b); The glycoform assignments and the glycosylation heterogeneity of the monoclonal antibody arising from variations in the hexose and fucose content are shown in the deconvoluted mass spectrum (c). (Reprinted with permission from Springer [62])

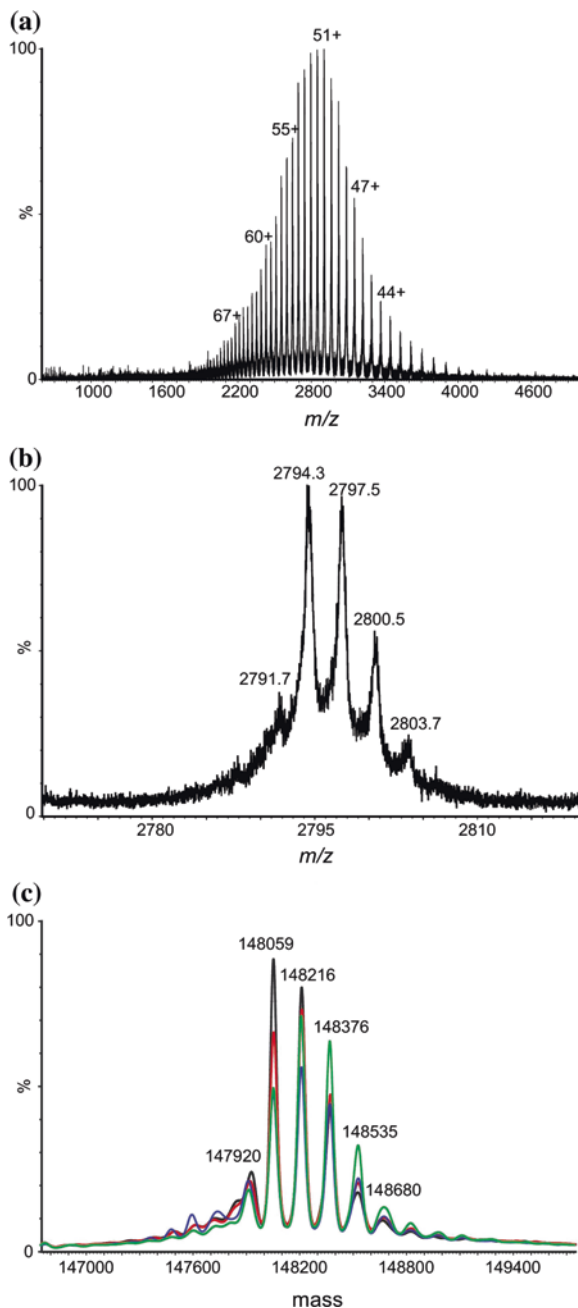
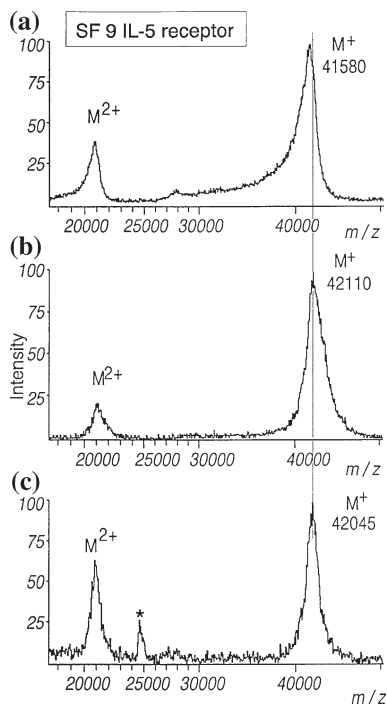


Fig. 4.4 Positive-ion MALDI-TOF mass spectrum of Sf9-derived interleukin-5 receptor α (IL-5R α) obtained with reflectron TOF with sDHB (a) and HPA (b) matrix, and linear TOF instrument using sDHB matrix (c). The *asterisk* denotes an internal calibrant. (Reprinted with permission from Wiley [69])



Overall, ESI MS analysis of intact glycoproteins has better success over MALDI MS for surveying the individual glycoforms in a glycoprotein biopharmaceutical sample and ensuring the homogeneity of the manufacturing batches. Nevertheless, the biggest challenge for the analysis of glycoproteins is their low abundance compared to that of unmodified proteins and the resulting low intensities of the mass spectral signals. This is mainly due to the low ionization efficiency of glycoproteins and the distribution of their signal among the various glycoforms sharing a common peptide sequence, thus rendering their detection an overwhelming task. This can be overcome by performing an enrichment step for the glycoproteins, which eliminates the most abundant unmodified proteins from competing for charge during the ionization process and results in higher ionization efficiencies and increased probability for detecting glycoproteins. The commonly used analytical methods for glycoprotein/glycopeptide enrichment are discussed in Sect. 4.2.2.2. Another promising solution to this problem is coupling of ESI MS with a separation device such as nano-LC [70], CE [71] or CZE [72] that can definitely improve the chances for a successful analysis. This is shown in the analysis of intact rHuEPO and bovine α 1-acid glycoproteins by a developed CZE-ESI MS method without any complicated sample treatment, where characterization of the intact glycoforms was provided along with their relative intensities [73, 74]. In addition to the efficient separation of the intact glycoforms, small glycan modifications such as acetylation, oxidation, and sulfation could be

successfully characterized. Similarly, high-resolution CE-Fourier transform ion cyclotron resonance (FT ICR) MS analysis was used for the profiling of the intact glycoforms of recombinant human chorionic gonadotrophin (r-RhCG) produced in a murine cell line, which resulted in the identification of over 60 different glycoforms with up to nine sialic acids [75]. These studies suggest that CE-MS can be an important tool for rapid assessment of the recombinant product quality either for product release or for in-process control, and even for demonstrating comparability of a glycoprotein therapeutic biosimilar to the innovator product being replicated.

Moreover, the rapid assessment of glycosylation at the molecular level is invaluable in glycoform screening of glycoproteins involved in certain diseases, such as the human transferrin (Tf) model glycoprotein for congenital disorders of glycosylation (CDG) diagnosis. CE-ESI MS was used successfully for carbohydrate-deficient transferrin (CDT) detection and CDG-type characterization [76]. Comparative analysis of serum samples from healthy and CDG patients by CE-ESI MS (Fig. 4.5) provided partial separation of Tf glycoforms and identification of the carbohydrate-deficient Tf glycoforms in the CDG patients' serum. It is clearly shown that the Tf glycoforms in the CDG serum correspond to a disialoform containing one free N-glycosylation site (Fig. 4.5e) and another one occupied by a biantennary instead of a triantennary N-linked sialylated glycan (Fig. 4.5f), thus confirming that the sample belongs to a patient who has CDG of type I [76].

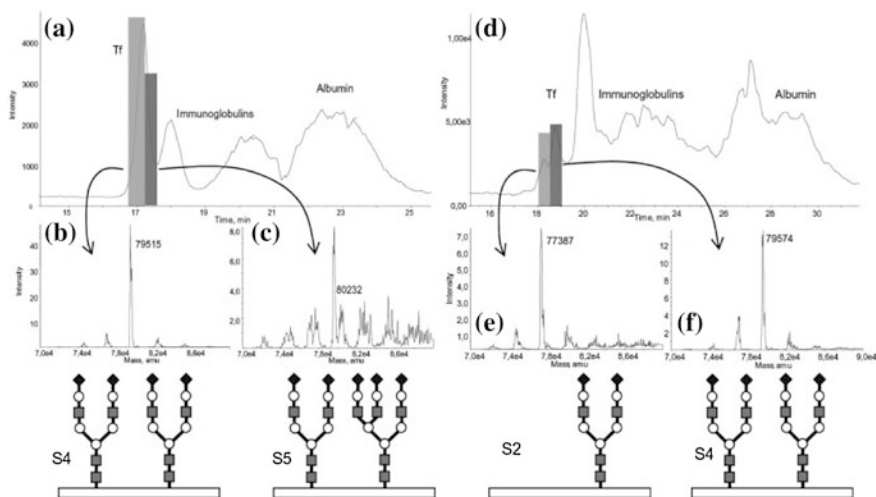


Fig. 4.5 Total ion electropherogram obtained for a serum from a healthy individual (a) and a CDG patient (d) under optimized CE-ESI MS conditions. The deconvoluted spectra obtained from the beginning and the end of the Tf peak are shown in (b) and (c), respectively. In case of the CDG patient, the deconvoluted spectra obtained from the two partial resolved glycoforms of Tf provided identification of the carbohydrate-deficient Tf glycoforms in the CDG patients' serum (e) and (f). The most probable glycan composition is displayed below the deconvoluted mass spectra. (Reprinted with permission from Wiley [76])

4.2.2 Mass Spectrometry and Glycoproteomics

Glycoproteomics involves the study of the glycosylation of proteins, including the structures of the attached oligosaccharide moieties and the identification of the glycosylation sites. There are two distinct classes of protein glycosylation in nature depending on the linkage site. First, the “O-linked” are the ones that are linked to serine (S), threonine (T), or hydroxyproline residues in the protein backbone, and then the “N-linked” which are linked to N residues through an N-acetylglucosamine residue (GlcNAc). Regarding O-glycosylation, a number of monosaccharides attached to S and T have been found, most commonly N-acetylgalactosamine (GalNAc), GlcNAc, xylose, mannose, and fucose [29]. O-glycans are synthesized in a stepwise process that involves single monosaccharide transfer steps, and their biosynthesis takes place after protein N-glycosylation, folding, and oligomerization. O-glycosylation may occur at any S or T residue with no single common core structure or consensus protein sequence. Extended structures from a core GalNAc that are called mucin-type O-glycans are the most frequently occurring [77, 78]. In contrast to O-glycans, N-glycosylation sites can be predicted by the tripeptide sequon Asn-Xaa-Ser/Thr (N-X-S/T, where X is any amino acid except P) [79, 80] (Fig. 4.6). All three types of N-glycans found in mature glycoproteins share a pentasaccharide core (i.e., the trimannosyl core with two N-acetylglucosamine residues ($\text{Man}_3\text{GlcNAc}_2$)) because of a common biosynthetic pathway in the endoplasmic reticulum compartment of the cell. This N-glycan $\text{Man}_3\text{GlcNAc}_2$ core is common to complex, high-mannose, and hybrid structures as shown in Fig. 4.6.

The high-mannose-type glycoproteins (e.g., ovalbumin) contain two to eight mannose residues added to the pentasaccharide core. Glycoproteins containing complex-type N-structures (e.g., fetuin) exhibit the highest structural variation by having a number of GlcNAc, Gal, Fuc and NeuAc (sialic acid) residues attached to the N-glycan $\text{Man}_3\text{GlcNAc}_2$ core, as well as possible extension and/or branching of the outer chains through lactosamine repeats and sialylation. Finally, the hybrid-type glycoproteins combine features from both high-mannose- and complex-type glycans [79]. At this point, it should be emphasized that for both N- and O-glycosylation, there is an inherent *microheterogeneity* resulting from the array of glycan structures associated with each glycosylation site. Moreover, there is *macroheterogeneity* due to the fact that not all N-glycan sequons or S/T residues present in the glycoproteins are glycosylated. The end result is a diverse degree of occupancy at different O- or N-linked glycosylation sites with a wide array of structurally different oligosaccharides that generate a complex mixture of glycosylated variants (*glycoforms*). The variety of these glycoforms depends not only on the polypeptide backbone and the number of putative glycosylation sites but also on the cell type, in which the glycoprotein is expressed, and its development stage. Therefore, characterizing the glycoproteome is a demanding task because of the inherent macro- and microheterogeneity of glycans along with the complex nature of this modification.

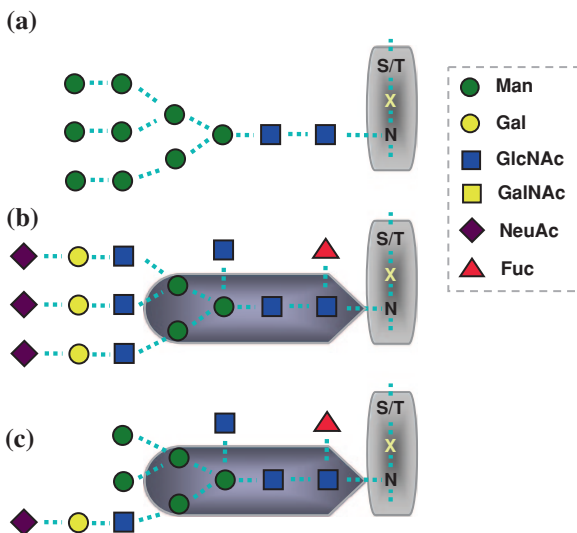


Fig. 4.6 Classes of N-linked carbohydrate structures sharing a common pentasaccharide core structure, that is, the trimannosyl core with two N-acetylglucosamine residues ($\text{Man}_3\text{GlcNAc}_2$). **a** High-mannose-type; **b** Complex-type (triantennary); **c** Hybrid-type. The sugar symbols used throughout this chapter are those adopted by the consortium for functional glycomics (CFG). *Circles* represent hexoses (Hex) [*yellow*: Galactose (Gal), *green*: Mannose], *squares* represent N-acetylhexosamines (HexNAc) [*blue*: N-acetylglucosamine (GlcNAc), *yellow*: N-acetylgalactosamine (GalNAc)], *red triangle*: fucose, *purple diamond*: N-acetylneuraminic acid (NeuAc). (Gal: ●, Man: ●, GlcNAc: ■, GalNAc: ■, Fuc: ▲, NeuAc: ◆)

4.2.2.1 Top-Down and Bottom-Up Analytical Approaches

Complete structural characterization of a glycoprotein includes the following tasks: (1) characterization of glycans in intact glycoproteins (2) determination of the protein primary sequence and the glycosylation attachment sites (3) characterization of glycopeptides, and (4) structural analysis of chemically or enzymatically released glycans. The M_r determination of intact glycoproteins by either ESI or MALDI MS analysis is successful only for glycoproteins up to 20–30 kDa with a relatively low percentage of carbohydrate content as demonstrated above (Sect. 4.2.1). Even though this accurate M_r measurement is valuable for profiling of intact glycoproteins and providing very useful information on the type and extent of glycosylation, there is no information on the nature and the attachment sites of the glycan chains. Therefore, one needs to cleave the protein into smaller fragments before attempting MS analysis. In the *top-down* approach, the intact molecule is introduced into the mass spectrometer where limited fragmentation of the ionized protein is induced (i.e., in the gas phase) and the resulting production mass spectra can provide information on the location of the glycosylation sites (or other PTMs) [81, 82]. Even though there are several mass analyzers capable of

measuring intact proteins and large ionic fragments (such as TOF, QTOF, FT ICR), the unusually high resolving power ($>10^5$) of FT ICR has made possible accurate assignments of ESI charge state and mass, even for MS/MS of intact proteins [83, 84]. Such *top-down* methods have proven especially powerful in stability and formulation studies of intact antibodies with $M_r \sim 150$ kDa used as therapeutics in the biopharmaceutical industry. Nevertheless, FT ICR instruments have not become standard analytical tools for the characterization of recombinant biopharmaceuticals mainly due to the high cost of acquisition and maintenance. For that reason, the most commonly followed MS-based approach for characterization of a recombinant biopharmaceutical involves the enzymatic digestion of the glycoprotein (usually with trypsin or another endoprotease) followed by the separation/analysis of the resulting peptide digests by LC-MS/MS [41, 85, 86] or CE-MS/MS [87] and MALDI MS [88] (*bottom-up* approach). In case of purified proteins or simple mixtures thereof, LC-MS or MALDI MS analysis of the proteolytic mixture provides M_r information on the peptide components. The advantage of MALDI-TOF MS is the simplicity of the spectra, which contain usually intense protonated (MH^+) or sodiated signals corresponding to the enzyme-generated peptides. Further, protein structural information can be deduced by carrying out LC-MS/MS analysis of the enzyme-generated peptides. Peptide identification is performed through a direct search of the M_r measured values and the tandem MS-derived fragment ions (sequence tags) [89] against a protein sequence database (*peptide fingerprinting*). The general experimental workflow comprising the commonly employed approaches in glycoproteomic analysis is shown in Fig. 4.7. Of course, the nature of the glycoprotein sample determines the number of the necessary steps needed in order to determine site-specific glycosylation and heterogeneity.

In general, MS mapping of the enzyme-generated peptide mixtures provides not only confirmation of the expected protein sequence but also identification of any existing modifications, including the glycosylation attachment sites. In addition, unexpected mass spectral signals can provide insights into the glycosylation profile of the protein, taking into consideration the known N-glycan structures (Fig. 4.6). Nevertheless, there are several problems associated with the *bottom-up* approach. The major problem arises from the fact that many glycoproteins are resistant to enzymatic proteolysis (e.g., trypsin or *S. aureus* V8 protease) due to the presence of the attached glycans near the proteolytic site, thus requiring an additional specific enzymatic proteolysis. In addition, the resulting mixture of peptides and glycopeptides could complicate the analysis because glycosylation strongly diminishes the ionization efficiency of the peptide [90, 91], especially when the glycans are terminated with the negatively charged sialic acid moiety [47]. This problem becomes more significant considering that the glycopeptides are in much lower abundance than the peptides from the same glycoprotein, and the glycopeptide signals are distributed over several peaks due to the glycan heterogeneity and multiple adduct ion formation. However, several enrichment methods (either in parallel or sequentially) prior to glycoprotein analysis can be used to compensate for the low abundance of glycopeptides (and glycoproteins) and the presence of multiple glycan structures (heterogeneity) [92]. The use of

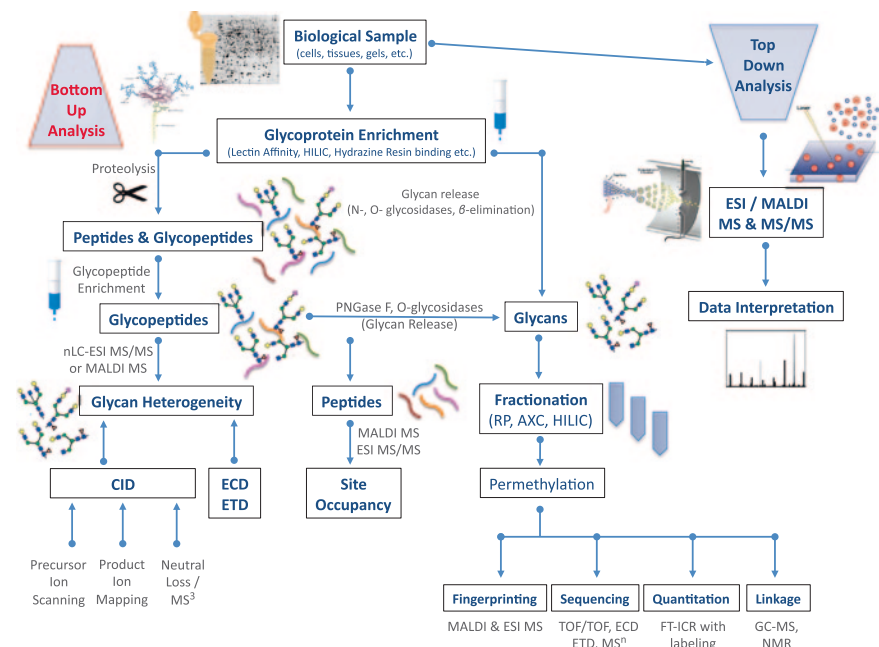


Fig. 4.7 The general experimental workflow comprising the commonly employed approaches in glycoproteomic analysis. The *top-down* approach starts with the analysis of intact glycoproteins, while the *bottom-up* analytical approach begins with proteolytic or chemical cleavage of the glycoprotein, followed by mapping of the generated glycopeptides by an assortment of LC–MS and tandem MS techniques

glycoprotein enrichment methods can bypass the aforementioned obstacles in glycoprotein analysis by achieving exclusion or reduction of the most abundant unmodified peptides from the analysis, thus improving the ionization efficiency of the low-abundance glycopeptides, which do not have to compete for charge during the ionization process with unmodified peptides.

4.2.2.2 Glycopeptides Enrichment Methods

Enrichment of glycoproteins and glycopeptides can be achieved by using the natural affinity of lectins for their glycan “handles” [93], whereas other analytical methodologies based on general physical and chemical properties of glycopeptides have been employed, such as size-exclusion chromatography (SEC) [94], hydrophilic interaction chromatography (HILIC) [95–97] or graphitized carbon columns (GCC) [98, 99]. A rough classification of the commonly used enrichment techniques in glycopeptides analysis can be made into chemical [100, 101] and chromatographic methods (such as affinity chromatography [102–104], LAC [93], immunoaffinity chromatography [105], SEC [94], hydrophilic phases [96, 97] and GCC [99]).

Lectin Affinity Chromatography

Lectins are proteins originating from plants, fungi, bacteria, or animals that express a special affinity toward glycans [106] and thus are used for glycopeptide/glycoprotein isolation from complex mixtures after being immobilized onto various solid supports such as silica [107], agarose [46], resins, magnetic beads, and affinity membranes. These are used in different arrangements, such as columns [108–110], tubes [46], and microfluidic chips [111]. Lectins generally interact with specific motifs in a glycan and demonstrate selectivity for different oligosaccharides and broad range of specificity [112], thus enabling glycoprotein/glycopeptide isolation from a complex protein mixture along with glycoform pre-fractionation. Widely used lectins include concanavalin A (ConA) [113, 114], which binds glycan residues containing mannose and glucose and affords broad selectivity (i.e., high-mannose, hybrid, complex biantennary [115]), wheat germ agglutinin (WGA), which presents selectivity for GlcNAc and NeuAc, and Jacalin (JCA), which expresses affinity against galactose (b1-3) GalNAc and O-linked glycoproteins.

Various analytical strategies have been proposed for the isolation and pre-concentration of glycoproteins/glycopeptides prior to MS analysis [93]. In summary, the sample enrichment using lectin columns can be performed before or after the protein mixture digestion by loading the sample onto the columns under high-ionic-strength buffers to prevent non-specific retention. The same loading buffer containing a displacer (haptene saccharide) is used to elute the captured glycopeptides/glycoproteins, which can then be subjected to MS analysis.

There are two principal enrichment methodologies based on LAC: Serial Affinity Chromatography (SLAC) [116] and Multi-Lectin approach (M-LAC) [117]. The first one uses a serial set of lectin columns with different specificity, thus enabling the sequential selective binding of various glycan moieties of a peptide or protein mixture. SLAC has proven to be a powerful tool for rapid and primary elucidation of glycans' structural features, especially when columns with broad (ConA, WGA, or JCA) and narrow selectivities (also known as “structure-specific affinity selectors,” i.e., *Sambucus nigra* agglutinin, SNA) are combined [118]. The SLAC approach was used in the characterization of a prostate-specific antigen in human prostate cancer [119]. Furthermore, coupling LAC with advances in stable isotopic labeling has been successfully applied for the comparative analysis of sialylated proteins [120], thus providing a valuable tool for exploring the glycosylation sites of the whole proteome as well as an excellent tool for biomarker discovery. On the other hand, the M-LAC approach uses a single column (multi-lectin column) containing various lectins with broad specificity (i.e., ConA, WGA, JCA), thus enabling the comprehensive isolation of glycoproteins/glycopeptides from a complex mixture covering an extended dynamic range. This approach was used in the study of glycoproteins in human serum [117, 121] and plasma [122].

Integrated analytical platforms combining LAC with various separation techniques have been developed lately in order to overcome the low-throughput drawback of the off-line procedures. Such methodologies include a microfluidic

chip [111] containing a polymeric monolithic column with immobilized pisum sativum agglutinin (PSA), an integrated glass microchip [123] for online tryptic digestion of glycoproteins in the first channel, followed by selective enrichment of resulting peptides through ConA in the second channel. The eluted fractions were subjected to CE and nano-LC–MS analysis employing capillary polymethacrylate monolithic columns with immobilized ConA and WGA [109] allowing large volume injection and adequate sensitivity. Similarly, a fully automated LAC system coupled online to ESI MS with silica-based lectin microcolumns [108] demonstrated high-binding capacity and excellent reproducibility, whereas a variation of this platform with SLAC [124] was proved to be superior over the M-LAC approach for the selective enrichment of small volumes of blood serum [115].

Immunoaffinity Chromatography

Immunoaffinity (IA) enrichment protocols for glycoproteins/glycopeptides rely on the unique specificity of the antibody–antigen interaction and enable the highly selective adsorption of a target analyte through the covalent attachment on a properly functionalized solid support containing an affinity ligand [125]. This can be performed either by the covalent attachment of antibody fragments via proper chemistries that provide correct orientation of the fragment, or by immobilization of a secondary binder molecule. The elution of the bound ligands is achieved by lowering the pH of the eluting buffer to pH 1–3, by using chaotropic salts, or by using polarity-reducing agents in order to weaken the antibody–antigen hydrophobic interactions. Although the IA enrichment approach has been mainly used for off-line targeted glycoproteomics [105], an online integration of this technique with CE was employed for the pre-concentration of rHuEPO in diluted solutions [126]. This integrated platform has demonstrated high loading sample capacity and good separation efficiency of the glycoforms.

Porous Graphitized Carbon Chromatography

Porous graphitized carbon chromatography (PGC) has been employed for the separation of oligosaccharides, in their native form as well as after derivatization, based on a retention mechanism driven mainly by hydrophobic and electrostatic stacking interactions. The oligosaccharide analytes are eluted in order of increased size, and structural isomer resolution is often provided [47]. In addition to separation, PGC has been used for the selective enrichment of glycans and glycopeptides. An off-line approach combining solid-phase extraction (SPE) with PGC cartridges was used to concentrate and pre-fractionate pronase glycopeptides and glycans prior to MALDI-TOF MS analysis [127]. An automated variation of the aforementioned approach for glycoprotein analysis has been reported combining digestion, extraction, and separation processes in one analysis [128]. This integrated platform employs a pronase-based chromatographic bioreactor for the *in situ* rapid digestion

of glycoproteins, an online SPE of the produced glycopeptides with a PGC trap column, and separation by LC-MS/MS. This system allowed the direct sequencing of the glycans and peptides along with simultaneous characterization of the glycan composition and localization of the glycosylation site.

Chemical Derivatization Methods

In addition to the affinity techniques described above, that do not change the structure of the modification and the peptides/proteins, several chemical methods specific to the glycan moieties have been used for the detection and the purification of glycosylated proteins. Most of the chemical derivatization strategies use two basic reactions: (1) the Schiff base reaction of aldehydes with a hydrazine [129–131], and (2) a Staudinger ligation between a phosphine and an azide [132, 133]. However, most of these derivatization methods provide peptide/protein identification without much information about the site or the structure of glycosylation, mainly due to inadequate search algorithms and the occasional modification of the glycan structure [112].

One of the strategies using the Schiff base reaction is the O-GlcNAc ketone enrichment method [134], where a chemo-enzymatic approach using an engineered β -1,4-galactosyl transferase is employed to transfer a ketone containing substrate onto O-GlcNAc-modified proteins. A Schiff base reaction was used to biotinylate the ketones with biotin-hydrazine and subsequently the O-GlcNAcylated peptides/proteins were captured on a streptavidin affinity column. This methodology was successfully used for the identification of the cAMP-responsive Element-Binding Factor (CREB), a low-abundance protein with two known O-GlcNAc sites, in a whole cell lysate [135]. Another derivatization enrichment approach for the glycoproteome and especially for N-glycosylation, is the Periodate-acid-Schiff (PAS) reaction using an iminobiotin hydrazide via the Schiff base reaction [129]. The derivatized peptides/proteins are affinity purified on a streptavidin column and analyzed by MS. This reaction exploits the unique vicinal diol functionality of glycans, thus oxidizing these diols to aldehydes without affecting any other amino acid apart from M, which is oxidized to its sulfoxide analog. This approach provides important information regarding N-glycosylation site modifications and has been used for high-throughput quantitative analyses [136]. In addition, it is an extremely versatile process for proteins and peptides, as different coupling agents such as biotin hydrazides and digoxigenin hydrazides can be incorporated. However, the major disadvantage of the PAS strategy is the heterogeneous modification of the glycan structures by an undefined number of hydrazide tags, thus necessitating PNGase cleavage of the glycans in order to sequence the peptide backbone. In this way, all information pertaining to the glycan structure is lost and only N-glycosylation sites can be determined.

In a modification [133] to the standard Staudinger reaction (a reaction of an azide with a phosphine), the intermediate aza-ylide formed in the standard Staudinger reaction reacts with an electrophilic trap to form an amide bond with a compound

that is biotin tagged. This reaction is biologically unique as neither phosphines nor azides occur in biomolecules and also offers the possibility to design phosphines in order to incorporate a wide variety of tags, such as fluorescent probes and affinity tags [137, 138]. A tagging-via-substrate (TAS) strategy based on a tag attached to the modification substrate was used for the identification of O-GlcNAc glycosylated proteins [117, 124], as well as for the detection and isolation of other PTMs in proteins, such as farnesylation [139]. Another derivatization method that has been used for the enrichment of O-linked β -GlcNAc is β -elimination followed by Michael addition with dithiothreitol (BEMAD) [140]. BEMAD has also been used to quantify both O-glycosylated and O-phosphorylated peptides [141].

4.2.3 Determination of Site-Specific Glycosylation and Heterogeneity

The complete characterization of a glycoprotein biopharmaceutical involves the analysis of the glycan structures that are expressed on the glycoprotein of a given organism or cell line, the identification of the proteins that express these glycans, as well as the individual glycosylation sites on each protein [39, 41]. MS and tandem MS analysis of glycopeptides usually after chromatographic or electrophoretic separation, either online or off-line, holds a central role in all the strategies for glycoproteomic analysis [142] (Fig. 4.7). The most commonly followed experimental approach for providing a detailed glycoprotein mapping involves the analysis of enzymatically derived glycopeptides by fast atom bombardment (FAB) [22, 143], MALDI [144] or LC-ESI MS and MS/MS [50, 85]. This MS mapping identifies most of the expected peptide signals (*peptide fingerprinting*), whereas any new, unexpected mass spectral signals may correspond to glycopeptides. In a similar off-line strategy, the isolated fractions are mapped by ESI or MALDI MS and MS/MS approaches. Nevertheless, there are several issues related to these MS approaches, such as the potential deglycosylation of glycopeptides in the gas phase combined with the low ionization efficiency and low abundance of the glycopeptides compared with the peptides derived from the proteolytic digestion. One of the remedies to ensure the appearance of glycoproteomic information within the copious proteomic data is enrichment of glycoproteins and/or glycopeptides prior to analysis (as discussed above). Another way to overcome this difficulty is the carbohydrate removal from the glycoprotein by base-catalyzed β -elimination for O-linked glycans or digestion with PNGase F (N-Glycanase) for N-linked glycans. The former leads to the conversion of S and T residues to A and α -aminobutyric acid sequences, respectively (i.e., loss of 16 Da), whereas the latter converts the glycosylated N residues to D (i.e., increase of 1 Da). In the MS mapping of the enzyme-generated peptide mixture of the deglycosylated protein, the former O- and N-glycosylated peptides can be readily identified by the appearance of new mass spectral signals at lower m/z (for O-linked sugars) or higher m/z (for N-linked sugars) than those of the unglycosylated peptides [145]. This mass

difference can be magnified by carrying out the N-Glycanase reaction in fully or partially (50 %) ^{18}O -labeled glycosylated N residues, which results in characteristic doublets separated by 2 Da [146]. These doublets can be used to locate the modification site and to determine the degree of occupancy at each N-linked glycosylation site. Another approach for N-linked glycans involves the release of the high-mannose- and hybrid-type oligosaccharides by digestion of the glycoprotein with endoglycosidase H, leaving a GlcNAc residue attached to the peptide's N residue. That results in the detection of peptides having M_r values 203 Da higher than that of the respective unglycosylated peptides. Glycosylation sites containing complex-type glycans are unaffected by the endoglycosidase H treatment. This approach was employed in the FAB carbohydrate mapping of the major envelope glycoprotein gp120 of HIV type 1 [147] and recombinant tissue plasminogen activator (rtPA) [148]. On a similar approach, glycoprotein mapping of CHO rHuEPO was facilitated by removal of terminal NeuAc residues with neuraminidase followed by LC-ESI MS analysis of the enzyme-generated peptide fragments of asialo CHO rHuEPO [54]. rHuEPO contains three N-glycosylation sites at N-24, N-38, and N-83 and a single O-glycosylation site at S-126; the glycans account for up to 40 % of the total molecular mass. This LC-MS mapping provided information on the microheterogeneity of the carbohydrate structures, which is associated with the presence or absence of lactosamine extensions and varying levels of O-acetylated NeuAc residues. Similarly, comparative LC-ESI MS tryptic mapping of untreated and neuraminidase-treated rtPA allowed the identification of the attachment site of two hybrid-type carbohydrates on one of the tryptic peptides [149]. The same analytical protocol was applied in the characterization of a rtPA mutant with an additional glycosylation site (T103N), where two new complex-type carbohydrate chains have been observed [149]. An analogous LC-ESI MS/MS approach combined with a multi-enzymatic digestion strategy was employed for the characterization of the glycosylation occupancy in the generic variant of rtPA (TNK-tPA), which was approved for treatments of acute myocardial infarction and ischemic stroke [150]. TNK-tPA has the same amino acid sequence as natural human tPA except for the three substitutions: T103N, N117Q, and AAAA for KHRR (296–299) which lead to longer half-life and higher fibrin activity than those of tPA. Nevertheless, differences in the glycosylation occupancy at N184 along with different extents of deamidation at N184 and oxidation at M207 have been observed between the therapeutic biosimilar and the innovator product, thus raising concerns as to its bioequivalence.

In the case of CHO IL-4, comparative LC-MS tryptic and V8 protease mapping of CHO IL-4 and its N-Glycanase-treated protein revealed that the N residue in the sequon N^{38}TT was glycosylated rather than the other potential site at N^{105}QS . We should point out that the presence of carbohydrate often provides shielding of a neighboring proteolytic site, thus leading to the incorporation of the adjacent peptide fragment, as demonstrated by the incorporation of the T5 tryptic glycopeptide into the adjacent disulfide-linked peptide T4–T10 of CHO IL-4 [50]. When ESI MS/MS approaches are incorporated in the analysis of the LC- or CE-separated enzymatic fragments of a glycoprotein, the identification of

glycopeptide-containing chromatographic fractions is facilitated by the appearance of several diagnostic fragment ions. CID product-ion spectra of ESI generated glycopeptides in a variety of instruments, such as triple quadrupole, ion trap (IT), and QTOF, are dominated by fragmentation of glycosidic linkages thereby revealing predominantly information on the composition and sequence of the glycan moiety. Glycopeptide marker ions under CID conditions are low-mass sugar-specific oxonium ions (B-type fragmentation in the Domon and Costello nomenclature [151]) of m/z 162 for Hex⁺, m/z 204 for HexNAc⁺, m/z 274 and 292 for NeuAc⁺, m/z 366 for Hex-HexNAc⁺, and m/z 657 for NeuAc-Hex-HexNAc⁺. Scanning for these diagnostic fragment ions in the “precursor ion” mode on triple-quadrupole mass spectrometers can selectively identify the glycopeptides within the enzymatic digest mixture, whereas screening of constant neutral losses of terminal monosaccharides could also pinpoint the glycopeptides. Selected ion monitoring (SIM) experiments can also be performed for glycopeptide identification with IT and QTOF mass analyzers. In cases, where MS/MS is not available, these low-mass glycopeptides marker ions can be generated by either “in-source” fragmentation of ESI-produced ions [50, 149, 152] or post-source decay (PSD) of MALDI-produced ions [153]. In the former, increasing the source entrance potential into the mass spectrometer, which controls the collision excitation and the extent of fragmentation, induces the fragmentation. This online LC-MS “in-source” CID mapping of glycopeptides utilizes both low and high source potentials and monitoring of the resulting sugar-specific oxonium ions. In case of complex/hybrid or high-mannose structures, monitoring of the oxonium ions at m/z 204, 274/292, 366 and 657 has allowed the fast glycan profiling in the LC-ESI MS analysis of the trypsin-treated CHO rTPA [154] and CHO IL-4 [50] without having to search each individual mass spectrum for glycopeptide-characteristic patterns. In the case of rTPA, this method allowed the identification of a low-level novel N-glycosylation at N142, which is part of an atypical N-Y-C consensus motif. Although this site is only 1 % occupied by predominantly biantennary hybrid structures, it was readily detected by this sensitive LC-ESI MS tryptic mapping approach. In the case of CHO IL-4, the observation of the glycopeptides marker ions at m/z 274, 366 and 657 revealed the presence of sialylated complex-type N-glycans in the specific chromatographic fraction. In addition, the mass separation of the signals within the triply and quadruply multiply charged ion envelopes revealed the presence of mono- and di-sialylated glycoforms (291 Da apart) along with higher M_r components containing additional lactosamine units (365 Da apart) owing to the presence of extended arms or branching. Similarly, this rapid glycopeptide screening approach was applied to other mammalian-cell-derived proteins, such as the Sf9-derived IL-5R α , where this low/high “in-source” fragmentation allowed the identification of all glycopeptide-containing fractions in the LC-ESI MS tryptic peptide map of Sf9 IL-5R α (Fig. 4.8). This method allowed the identification of four glycosylation sites in Sf9 IL-5R α out of the six potential sites fulfilling the N-glycosylation consensus sequence [66].

The ESI mass spectrum of one glycopeptide-containing fraction (Fig. 4.8, peak 10) showed signals corresponding to doubly and triply charged tryptic glycopeptides containing a Man₉(GlcNAc)₂ high-mannose carbohydrate (Fig. 4.9). All these

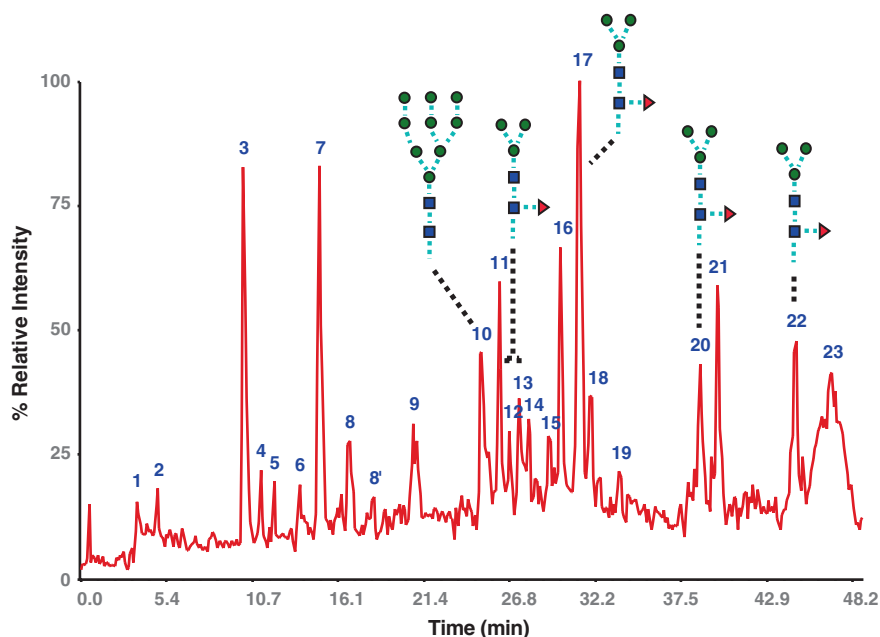


Fig. 4.8 Total ion current (TIC) chromatogram obtained from the reversed-phase LC-ESI MS analysis of the Sf9 IL-5R α tryptic digest using “in-source” fragmentation. The tryptic fractions containing glycopeptides are labeled, as inferred by the detection of the low-mass diagnostic ions at m/z 162, 204, 274, 366, and 657

glycopeptides contain the N196 glycosylation site and the M_r values of the respective glycoforms differ by 162 Da due to an extensive heterogeneity in the Man (●) content, as shown in the deconvoluted mass spectrum (Fig. 4.9, inset).

The assignment of the putative glycan structures to the experimental masses with a high degree of confidence is made possible by the excellent mass measurement accuracy provided by ESI MS analysis. Corroborative information on the composition and sequence of the attached glycans can be attained from MS/MS analysis of the glycopeptides, because CID tandem mass spectra of glycopeptides contain mainly fragments arising from glycosidic bond cleavage [155]. In the analysis of the therapeutic glycoprotein BRP 3 EPO by a combined anion-exchange chromatography (AEC)—ultra-performance liquid chromatography (UPLC) MS/MS approach, tetra-antennary glycans with up to four NeuAc and up to five poly-N-acetyl lactosamine extensions were observed at the glycosylation sites N24 and N83, whereas biantennary glycans were the major structures at N38 [156]. The presence of these large repeating glycan motifs although at low levels may infer additional functional interactions for EPO and may be beneficial in terms of immunogenicity. A more detailed characterization of N-glycopeptides, especially in terms of the peptide sequence, can be obtained by an alternative approach combining MS/MS and MS³ experiments in an IT MS [142]. The glycopeptide ion is

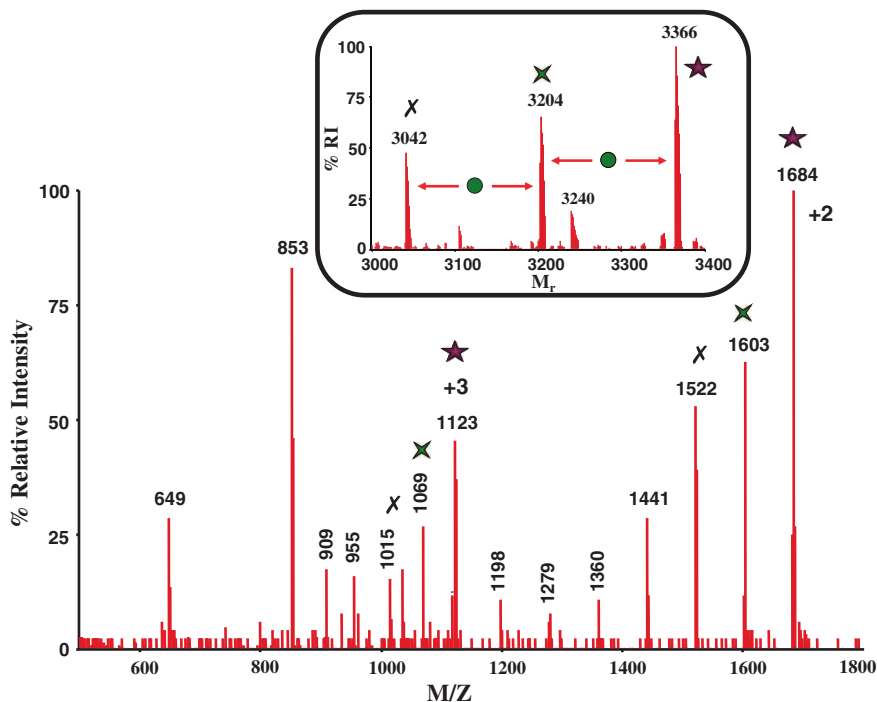


Fig. 4.9 Positive-ion ESI mass spectrum of Sf9 IL-5R α tryptic glycopeptide component (Fig. 4.8, TIC peak 10). The deconvoluted mass spectrum (shown in the *inset*) indicates the presence of a high-mannose carbohydrate component with M_r value of 3366. The *asterisk* and *cross*-denoted peaks in the doubly and triply charged ESI ion envelopes correspond to two glycoforms with M_r values 3204 and 3042, respectively, arising from glycoform heterogeneity due to variations in the mannose (●) content

selected and fragmented, and the peptide ion carrying a single GlcNAc (which is often the most abundant ion) is subjected to a second fragmentation cycle resulting in extended fragmentation of the peptide moiety into *b*- and *y*-series ions, thus allowing the deduction of the glycan attachment site. MS/MS analysis of N-glycopeptides with QTOF mass analyzers at low collision energy exhibited mostly cleavages of glycosidic linkages providing information on the glycan moiety [157]. Nevertheless, CID mass spectra at elevated collision energies resulted in a significant level of *b*-type and *y*-type peptide fragmentation, thus allowing identification of the glycosylation site. The potential of the nESI QTOF MS/MS in the characterization of O-glycopeptides has also been demonstrated in the analysis of mucin-type glycopeptides with S- or T-linked O-glycans [88, 158] where information on the structure and the attachment site of the O-glycan has been provided based on the *b*-type and *y*-type peptide ions comprising the glycan attachment site.

Alternatively, the development of the complementary mass spectrometric fragmentation techniques of electron-capture dissociation (ECD) [24, 159] and

electron-transfer dissociation (ETD) [25] has expanded the analytical options for mapping the modification sites of both N-glycosylation and O-glycosylation. In the ECD technique, which is mainly restricted to FT ICR analyzers, multiply protonated peptide ions are irradiated with low-energy electrons (<0.2 eV) and undergo fragmentation. On the other hand, ETD can be combined with IT, QIT, and Orbitrap analyzers and peptide fragmentation is generated through gas-phase electron-transfer reactions from singly charged anions (e.g., anions of fluoranthene, sulfur dioxide) to a multiply charged peptide/glycopeptide. Unlike the traditional MS/MS techniques, both ECD and ETD appear to retain labile PTMs and induce fragmentation of the peptide backbone with minimal loss of the glycan moiety. ECD and ETD of glycopeptides result in the cleavage of the amine backbone (N-C $_{\alpha}$) to generate preferentially c' and z• fragments ions (nomenclature of Zubarev and co-workers [160]). The intact oligosaccharide moieties are retained in the fragment ions containing the site of glycosylation. Consequently, ECD and ETD represent excellent tools for the localization of modification sites in post-translationally modified proteins [161–163], and there have been few reports of using these techniques in the characterization of N-linked [142, 164] and O-linked glycopeptides [162, 165]. This is nicely shown in the ESI tandem MS analysis of a tryptic glycopeptide (S295-R313) from horseradish peroxidase (HRP) containing a core-fucosylated and core-xylosylated trimannosyl N-glycan attached to the N298 residue (Fig. 4.10) [142]. The [M+3H] $^{3+}$ ion at m/z 1119 was subjected to CID fragmentation which led to preferential cleavage of glycosidic linkages rather than polypeptide bonds (Fig. 4.10a), thus providing information primarily on the composition and sequence of the glycan moiety. On the contrary, ETD ion activation of the [M+3H] $^{3+}$ ion yielded the cleavage of the peptide backbone with no loss of the glycan moiety, thus leaving the N-glycan modification on the N298 residue intact and providing complete peptide backbone sequence through the observed c'- and z•-ion series (Fig. 4.10b).

Therefore, the use of both CID and ETD ion activation in the LC-MS analysis of glycopeptides has allowed the characterization of both glycan structure (CID-MS/MS) and peptide sequence/site attachment (ETD-MS/MS) within the same LC-MS run. Similarly, use of LC-MS and the ETD and CID fragmentation techniques allowed the identification of two distinct O-glycopeptide structures and three glycosylation sites from the secreted amyloid precursor protein (sAPP695) expressed in CHO cells [166]. This de novo characterization of unknown O-glycosylation sites was extremely challenging due to the large number of S and T residues (27 S and 39 T residues) contained in the protein sequence of the APP fragment. In a modified strategy, LC-MS combined with CID, ETD, and CID of an isolated charge-reduced species derived from ETD was employed to determine the peptide backbone sequence and the site of modification for an O-linked glycosylated peptide fragment of rtPA at the low femtomol level [167].

In case of glycoprotein mapping by MALDI MS, the intense protonated (MH $^{+}$) glycopeptides signals are much more stable in CID than the multiply protonated glycopeptide species obtained by ESI. Although PSD, as well as CID, is used for MS/MS of glycopeptides, precise analysis of fragment ion peaks often seems to

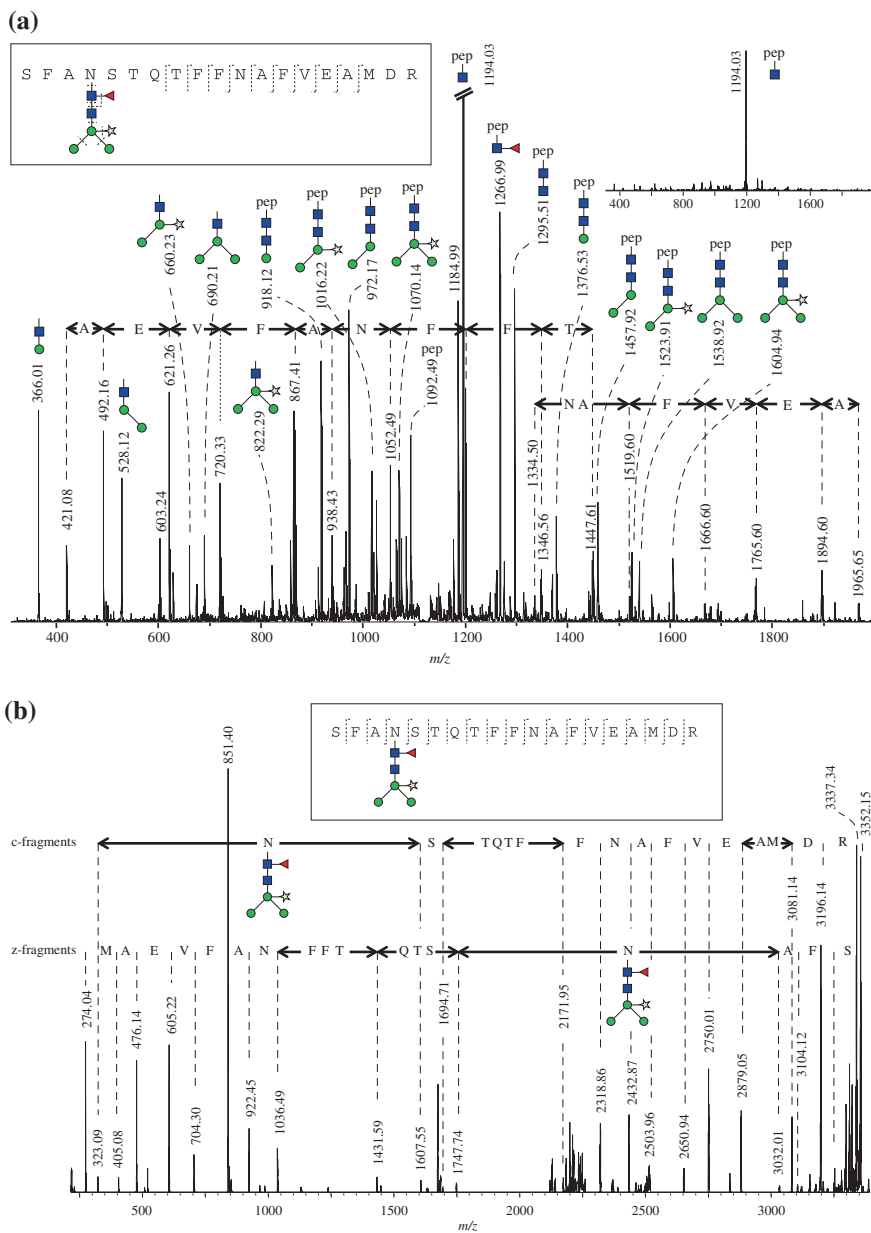


Fig. 4.10 Positive-ion nESI tandem mass spectrum of the RP-HPLC-purified tryptic glycopeptide Ser295-Arg313 from horseradish peroxidase (HRP) containing a core-fucosylated and core-xylosylated trimannosyl N-glycan attached to N298 residue. The $[M+3H]^{3+}$ ion at m/z 1119 was subjected to CID **(a)** and ETD **(b)** fragmentation which led to cleavage of the glycosidic bonds and the peptide backbone, respectively. (Reprinted with permission from Elsevier [142])

be difficult because of preferential and fast deglycosylation, and the limited peptide sequence information [168]. Therefore, fragmentation of these glycopeptide ions by metastable dissociation in a MALDI-TOF/TOF MS or by CID in a MALDI QTOF instrument is performed at higher energies. MALDI-TOF/TOF MS of N-glycopeptides results in a set of cleavages at or near the innermost GlcNAc residue, with the peptide moiety retained in all the fragment ions. In addition, peptide bond cleavages next to the fragmentation of glycosidic bonds are observed (predominantly *b*-type and *y*-type ions), which provide useful peptide sequence tags [169, 170]. All these fragments comprising the N-glycosylation site retain the attached glycan, thus confirming the glycan attachment site. Similarly, MALDI-TOF/TOF MS of O-glycopeptides generate fragmentation patterns from the glycopeptides precursor ions (*b*- and *y*-series ions), which can be used for identification of O-glycosylation sites as it was demonstrated in the case of mucin-type glycopeptide derivatives [171].

At this point, we should point out that parallel *glycomic* analyses for providing information on the linkage, branching points, and configuration of the constituent monosaccharides (microheterogeneity) are also essential in the whole glycoproteomic strategy. In general, the glycans are released by enzymatic or chemical digestion of the glycoprotein or the glycopeptide mixture, undergo permethylation and then subjected to a range of techniques, selected upon the level of analysis to be carried out, that is, fingerprinting, linear sequencing, linkage, branching, or quantitation of monosaccharides [172, 173]. In one of the followed approaches, the permethylated glycans are subjected to LC-MS analysis and the supplied mass spectral information on the specific glycans and their relative amounts can be compared and matched with data at the glycopeptide and overall glycoprotein levels (Fig. 4.7). Incorporation of MALDI-TOF and ESI tandem MS can definitely enhance the analytical potential for tackling complex glycobiology structural issues [43]. Further information on the carbohydrate secondary structures can be provided by well-established methods in structural glycobiology such as X-ray crystallography and especially 2D nuclear magnetic resonance (NMR) analysis [174, 175], albeit the requirement for highly purified glycans and large amounts of sample.

4.2.4 Bioinformatics Tools for Glycoprotein Analysis

Because of the extreme glycan heterogeneity, interpretation of the data produced from the aforementioned glycoproteomic approaches and glycopeptide identification through a comprehensive large-scale data analysis is a challenging task. The development and use of informatics tools and databases for glycobiology research has increased considerably in recent years [176], even though the progress of these tools for glycobiology and glycomics is still in its infancy compared to those already used in genomics and proteomics. Even though, the automated identification of proteins from MS and MS/MS spectra is now almost routine by

using informatics tools such as Mascot (<http://www.matrixscience.com/>), there is lack of rapid and accurate automated tools for retrieving structural information from MS data in case of glycoproteomics. The MS and MS/MS-derived information should be searched for putative glycopeptides predicted by comparison with other glycoconjugate structures derived through the same biosynthetic machinery in other closely related organism, cell line, or tissue. Nevertheless, the inherent complexity of the glycan structures combined with the wide range of techniques employed in their study renders the development of similar automated computational tools a formidable task [177]. In addition, the lack of libraries of glycan sequences similar to the SWISS-PROT protein databank makes matters more challenging. It should be emphasized that more than half of all proteins are glycosylated, based on the analysis of well-characterized proteins deposited in the SWISS-PROT databank [28].

In case of proteomics, bioinformatics tools essentially utilize sequences of the building blocks of proteins (20 amino acids), which are always linked in a predicted linear way in order to provide automated protein identification from MS and tandem MS data. On the contrary, carbohydrates are structurally diverse as their building blocks, the monosaccharides, may be connected in various ways to form branched structures, thus complicating their digital encoding. Moreover, in contrast to protein expression, glycosylation is a non-template-driven synthetic process where multiple enzymes are involved and the final glycoprotein product depends on the type of enzymes expressed in the cell that synthesizes the glycoprotein. The development of bioinformatics methods has mainly found applications in glycosylation analysis, glycomics, glycan structure analysis, glycan biomarker prediction, and glycan structure mining (e.g., using lectins that recognize a certain glycan [178]). In the glycosylation analysis and the prediction of glycosylation binding sites on proteins, the first step is the selective search of protein databases for proteins containing only the consensus sequence for N-linked glycosylation. Several software platforms have been developed for the identification of intact N-linked glycopeptides, such as *GlycoMod* [179], *GlycoPep DB* [180], *Cartoonist* [181], *Peptonist* [182], and *Glyco-Miner* [183]. These methods can be used mainly for glycopeptides generated from specific enzymes, for example, trypsin or endoprotease Glu-C, whereas *GlycoX* [184] can be used for interpretation of mass spectra obtained from non-specific proteases, such as protease K. *Cartoonist* is one of the earlier developed glycomic MS interpretation approaches containing a library of several hundred archetype glycans derived from information about biosynthetic pathways and employing a set of rules to modify these structures. *Cartoonist* incorporates the same assumptions used by human expert in the annotation of MS data, and it is used to automatically annotate N-glycans in MALDI mass spectra with diagrams or cartoons of the most possible glycans consistent with the observed mass values. *Peptonist* [182] uses MS/MS data to identify glycosylated peptides in LC-ESI MS runs of enzymatically digested glycoproteins and MS data to identify the N-glycans present on each of those peptides. On the other hand, the *GlyDB* [185] approach has been developed to address the need for structure annotation of N-linked glycopeptides in the LC-ESI MS analysis of glycoprotein proteolytic

digests. The annotation of low-resolution tandem MS spectra of N-linked glycopeptides arising from low-energy CID, where cleavage along the glycosidic bonds occurs preferentially, is based on matching experimental spectra to theoretical spectra generated by a linearized database of glycan structures using the established search engine SEQUEST. Similarly, *GlycoPep ID* [186] is a web-based tool used to identify the peptide moiety of either sialylated, sulfated, or both sialylated and sulfated glycopeptides, by correlating the product ions of suspected glycopeptides to a peptide composition. Following the identification of the peptide portion, the mass of the remaining segment can be attributed to the carbohydrate component.

Even though the development and use of informatics tools and databases for glycobiology and glycomics research has increased significantly in recent years, it has lagged behind the development of similar tools for genomics and proteomics. This drawback arises from the lack of comprehensive and well-organized compilations of glycan sequences and efficient automatic assignment procedures for high-throughput analysis of glycans. Most of the aforementioned library-based sequencing and N-glycopeptide identification tools for MS data interpretation are not publicly available; they have their own standards, databases and/or run on a special hardware platform. Moreover, the independently developed database with their own format and language along with the absence of publicly available databases with carefully assigned MS spectra of glycans hinders the development of efficient scoring algorithms. Therefore, rules should be established for the standardization of the structural description of glycans and the deposit of glycan structures and the associated glyco-related data in databases of complex glycan structures. In addition, the deposit of complex glycan structures and glyco-related data in generally accepted databases should be maintained by well-recognized international institutions such as NCBI (www.ncbi.nlm.nih.gov) and European Bioinformatics Institute (EMBL-EBI, www.ebi.ac.uk), which house genome sequencing data (GenBank) and protein related databases, respectively. It is also essential to ensure the intercompatibility of the related data formats, in order to facilitate data exchange between different databases and efficient cross-linking and referencing thereof between various projects.

Toward this direction, the EU FP6-funded EUROCarbDB project (<http://www.ebi.ac.uk/eurocarb/home.action>) was an initiative to create the technical framework where interested research groups could feed in their complex glycan structural data, which would be archived and maintained at the EMBL-EBI. Other most prominent publically available glycan-related databases are the Consortium for Functional Glycomics (CFG) relational database (<http://www.functionalglycomics.org/glycomics/common/jsp/firstpage.jsp>), the Kyoto Encyclopedia of Genes and Genomes glycome informatics resource (KEGG GLYCAN) (<http://www.genome.jp/kegg/glycan/>), and Glycosciences.de (<http://www.dkfz.de/spec/glycosciences.de/sweetdb/index.php>). Finally, genomic/proteomic findings need to be integrated with biomedical studies where glycan structures can serve as biomarkers for specific diseases or malfunctions [187], like the ones provided by the KEGG resources [188–190].

4.3 Disulfide Bond Formation

4.3.1 MS Determination of Disulfide Bonds

Even though glycosylation enjoys more popularity in the PTM literature, disulfide bond formation is one of the most common PTMs playing a critical role in establishing and stabilizing the three-dimensional structure of proteins [6, 191]. The physiological and pathological relevance of disulfide bonds to diseases has been recognized in several cases, such as tumor immunity [192], neurodegenerative diseases [193], and G-protein receptors [194]. These cross-linkages between the sulfhydryl groups of two C residues can be either *intramolecular* or *intermolecular*. The former stabilize the tertiary structures of proteins, while the latter are involved in stabilizing quaternary structures of proteins [195, 196]. For protein therapeutics, the generation of correctly folded recombinant proteins is of paramount importance. Difficulties in folding recombinant protein products are common from *E.coli* cell line, thus resulting in loss of specific activity compared to the native material. Similarly, over-expression of proteins in CHO cell line leads to disulfide scrambling. Therefore, there are significant efforts to develop reliable methods for mapping disulfide bonds in therapeutic proteins, thus ensuring drug quality. The determination of disulfide bond arrangements of proteins not only provides insights into protein activity relationships but also guides further structural determination by NMR and X-ray crystallography. The first step in disulfide mapping is the determination of the number of disulfides in a given protein, which can be readily deduced by a simple MS analysis before and after protein reduction. This is nicely illustrated in the ESI MS analysis of recombinant interferon α -2b and GM-CSF, where reduction resulted in a 4 Da shift in the measured M_r , thus indicating the presence of two disulfide bonds [197]. In case of GM-CSF, the ESI mass spectrum prior to and after treatment with β -mercaptoethanol clearly showed a 4 Da shift in the measured M_r (Fig. 4.11 insets), hence confirming the presence of two disulfide bonds in the recombinant protein product.

4.3.2 Disulfide Mapping

Following the determination of the number of disulfide linkages, mapping of the protein's primary sequence by proteolytic cleavage of the protein between half-cystine residues to produce disulfide-linked peptides and MS analysis of the resulting peptide fragments allows the identification of the existing disulfide arrangement [198]. The potential of MS in this disulfide mapping approach was first realized with the implementation of soft ionization techniques, such as FAB/liquid secondary ion (LSI) [199–201], plasma desorption (PD) [202, 203], and later by the more sensitive method of MALDI [18, 204]. That was nicely illustrated in the disulfide mapping of several therapeutic proteins, such as recombinant human interferon

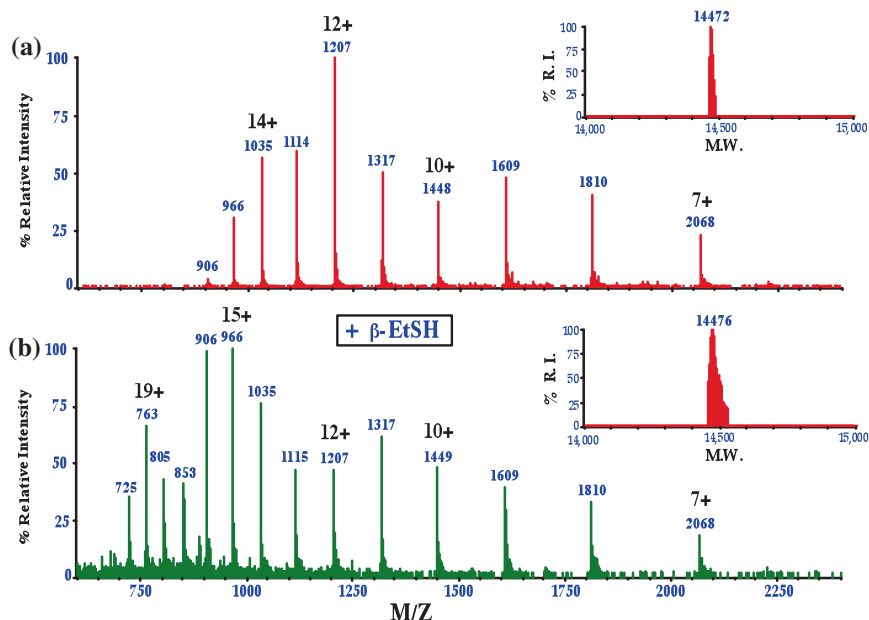


Fig. 4.11 Positive-ion ESI mass spectrum of recombinant human granulocyte-macrophage colony-stimulating factor (rhGMCSF) in 1 % HCOOH (a) and after treatment with β -mercaptoethanol (b). The deconvoluted spectra are shown in the *insets*. (Reprinted with permission from Wiley [197])

α -2b (INTRON A) [205, 206], human growth hormone [203] and IL-4 [207] by FAB, PD, and MALDI mapping. It should be noted that weak ion signals corresponding to the MH^+ of the constituent C-containing peptides were also present in the FAB, LSI, PD, and MALDI mass spectra arising from fragmentation of disulfide-linked peptides during the ionization process [208]. This is shown in the LSI mass spectrum of the disulfide-linked tryptic core peptide of rhGM-CSF (expected M_r 7,613) (Fig. 4.12), where additional signals at 5665.2 and 4412.4 Da were also observed due to the presence of the partially reduced peptides T₅-S-S-T₁₁ and T₁₁-S-S-T₁₃, respectively (Fig. 4.12, inset) [197].

Even though the disulfide-linked peptides yield unique mass spectral signals, the protein fragmentation should be carefully controlled to avoid rearrangement of disulfide bonds (disulfide scrambling), which can take place at neutral and alkaline pH [209]. Therefore, protein cleavage methods performed in aqueous solvents at acidic pH are preferred, such as cyanogen bromide [210] and pepsin [200]. This acidic pH is also optimum for disrupting the protein conformation and making the cleavage sites between half-cystine residues more accessible. That was nicely illustrated in the first report on the disulfide mapping of insulin where FAB MS of peptic digest peptides was combined with *Edman* analysis for disulfide bond analysis [200]. The intramolecularly linked peptides are identified by the 2 Da increase

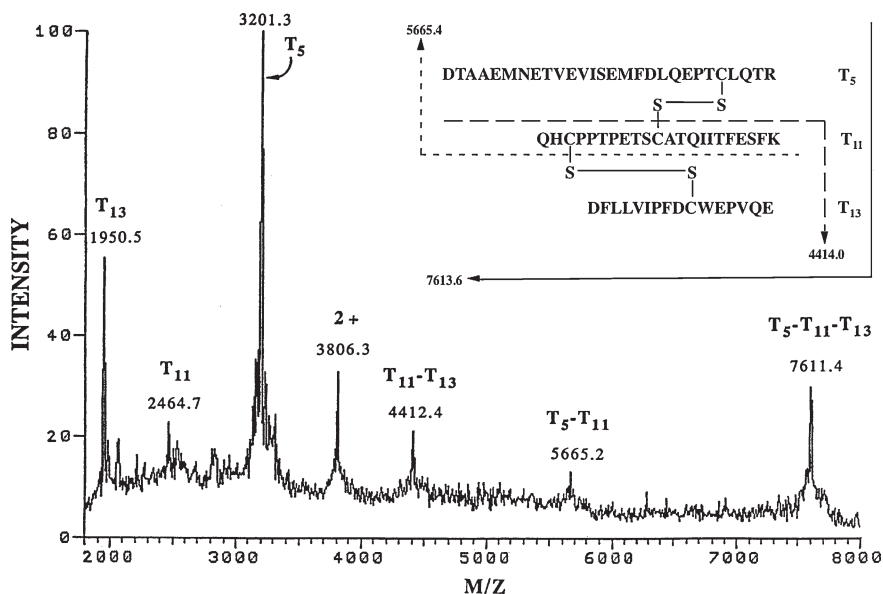


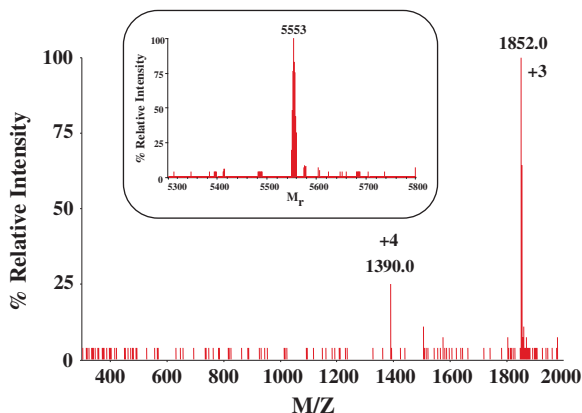
Fig. 4.12 Positive-ion Cs^+ LSI mass spectrum of the HPLC-isolated fraction containing the disulfide-linked tryptic core peptide $\text{T}_5\text{-T}_{11}\text{-T}_{13}$ of rhGM-CSF. (Reprinted with permission from Wiley [197])

upon reduction in their constituent half-cystines with β -mercaptoethanol or dithiothreitol, whereas intermolecularly bridged peptides yield protonated MH^+ signals of the constituent half-cystine-containing peptide fragments.

The advent of ESI [13] has made LC-ESI MS the favorite approach for analyzing the enzyme-generated protein fragments and mapping disulfide linkages in recombinant proteins [198, 206]. Analysis of the peptide mixtures before and after reduction generally allows the identification of the C residues involved in disulfide bonding, taking all aforementioned precautions to minimize disulfide scrambling. It should be noted that ESI MS analysis of disulfide-linked peptides is not conducive to peptide signals arising from partial disulfide bond reduction, as shown in the ESI mass spectrum of the disulfide-linked tryptic peptide $\text{T}_{20}\text{-T}_{25,26}$ of IL-5R α (Fig. 4.13).

When protein chains are disulfide-linked and proteolysis between half-cystine residues is not possible, identification of the exact location of the disulfide linkage often requires (1) successive proteolytic digestions, such as the ones demonstrated for interleukin-13 (chymotrypsin plus *S. aureus* V8 protease) [211] and rtPA (Lys-C plus trypsin) [26] or (2) chromatographic separation of the enzyme-derived protein fragments coupled with online MS/MS analysis (e.g., LC-ESI MS/MS), and/or off-line MS/MS analysis and *Edman* sequencing [212, 213]. This is essential for proteins where three proteolytic fragments are linked by intermolecular disulfides or where two peptide chains contained an intramolecular disulfide

Fig. 4.13 Positive-ion ESI mass spectrum of Sf9 IL-5R α tryptic fraction (Fig. 4.8, TIC peak 19) containing the disulfide-bonded peptides T₂₀ and T_{25,26} with M_r value of 5553



and no further proteolysis is possible. The existence of disulfide bonds is usually confirmed by fragmentation of putatively disulfide-linked peptides by MS/MS analysis following ionization by FAB [214], ESI [215], or MALDI PSD [216]. In the MALDI PSD approach, the characteristic ion triplet separated by 33 Da, arising from cleavage at the C–S bond with a concomitant proton transfer [168], can be used as a diagnostic tool for the location and identification of disulfide-paired peptides, even from complex digest mixtures of proteins.

The LC-ESI MS and tandem MS approach is especially valuable in the disulfide mapping of protein receptors and therapeutic proteins having high M_r, such as rtPA and mAb. In mAb, the inter- and intrachain disulfides are responsible for maintaining the characteristic three-dimensional antibody structure, which allows the highly specific antigen binding. Therefore, complete disulfide mapping in mAb is critical for ensuring its therapeutic activity, because incomplete disulfide linkages and/or free sulfhydryl groups can lead to antibody fragments with no antigen-binding activity [217]. In case of the anti-HER2 mAb (Herceptin) that interferes with the HER2/neu receptor and used for the treatment of early-stage breast cancer, the disulfides were completely mapped by LC-ESI MS with the combination of ETD and CID fragmentation [218]. Using ETD cleaves preferentially the disulfides into two polypeptides while CID generates mainly peptide backbone cleavage (with the disulfides intact). This approach was successful in mapping a total of 16 disulfides, 12 intra- and 4 intermolecular, in anti-HER2 mAb and a similar therapeutic mAb. This ETD fragmentation strategy can be further enhanced by CID-MS³ on the dissociated peptides (after ETD) in order to provide corroborating information on the linkage assignment. The same multi-fragmentation approach in combination with multi-enzyme digestion scheme (Lys-C followed by trypsin and Glu-C) was employed in the mapping of the 17 disulfide linkages in human growth hormone [26] and rtPA, as well as for the identification of the unpaired C residue in rtPA [219]. The ETD-MS² spectrum of the disulfide-linked tryptic peptide T₇-T₈-T₉ clearly showed that the unassigned C residue (C83) was found to be paired with either a glutathione or C molecule, which

could shed light into the activation or signaling pathway of rtPA. A novel approach based on IM MS was also employed for the rapid characterization of disulfide variants in intact IgG2 mAb [220]. IM MS revealed two to three gas-phase conformer populations for IgG2, compared to only one conformer for IgG1 mAb and a C232S mutant of IgG2, thus indicating that the observed conformers are apparently related to disulfide variants. Therefore, IM MS is a new powerful tool for the characterization of intact mAb and may be useful for fingerprinting higher-order structures of these protein therapeutics.

Finally, disulfide mapping combined with *stable isotope-labeling of peptides with* ^{18}O greatly facilitated the identification and characterization of disulfide-linked peptides [221]. Isotope profiles of enzymatically generated peptides produced in 50 % H_2^{18}O (v/v) in H_2^{16}O would produce unique doublets separated by 2 Da, whereas the disulfide-linked peptides should be distinctly different than single-chain peptides [222]. Therefore, the disulfide-linked peptides could be identified in complex peptic digests or chromatographic fractions thereof by MS analysis, and especially MALDI-TOF MS. This procedure is ideally performed in acidic solutions (e.g., peptic digestion) in order to preclude disulfide rearrangement and it may also be used to aid the interpretation of product-ion spectra of disulfide-linked peptides.

4.4 Future Prospects and Challenges

In the past two decades, recombinant protein therapeutics have changed the face of modern medicine as they provide innovative and effective therapies for numerous previously incurable diseases. Protein therapeutics have already a significant role in almost every field of medicine, even though this role is still only in its infancy. The number of recombinant proteins in clinical trials for new and existing therapeutic targets continues to increase annually, as does the total number of protein-based pharmaceuticals reaching the marketplace. The acceptance of the various protein therapies can be attributed to the increasing prevalence of chronic diseases, such as cancer, diabetes, cardiovascular diseases, and neurological/neurodegenerative disorders. In addition, the rising penetration of medical insurance industry has made protein therapeutics available to a wider population. The global protein therapeutics market is expected to grow at an annual rate of 13 % during 2012–2015, arising from the introduction of new protein therapeutics in the major sectors of protein therapeutics market, which include mAb, insulin, interferons, G-CSF, tPA, EPO, coagulation factors, etc.

Recombinant therapeutic proteins for human use must be characterized thoroughly prior to clinical development in order to satisfy the rigorous regulatory requirements (ICH Q6B guidance) [12]. In addition, the manufactured final product should be comparable to that used in preclinical and clinical studies, and its purity, potency, safety, stability, and batch-to-batch consistency should be established. Advances in MS techniques, especially MALDI and ESI, have made MS-based mapping approaches

powerful and essential analytical tools for structure characterization of therapeutic proteins and evaluation of recombinant protein heterogeneity including identification of PTMs, sequence variants, and degradation products in recombinant proteins. Structure characterization of all PTMs in a protein is of a great concern for regulatory agencies, such as glycosylation and disulfide linkages. Glycosylation, the most common form of PTM, plays a crucial role in the stability and therapeutic potency of the glycoprotein, as it was demonstrated for rHuEPO. Moreover, changes in levels and types of glycosylation can be associated with certain diseases, such as aggressive breast cancer [223], thus making glycoprotein screening invaluable, not only for diagnostic purposes, but also for design of novel therapeutic drugs. In addition, glycan profiling of normal and diseased forms of a glycoprotein has provided new insights into future research in rheumatoid arthritis, prostate cancer, and congenital disorders of glycosylation [224–226].

In general, LC–MS and tandem MS peptide mapping is the standard and well-accepted approach by the regulatory agencies (FDA, EMA) for identifying PTMs and establishing the recombinant product purity. Nevertheless, a variety of tandem MS experiments should be performed in order to provide insights into the glycan structure (low-energy CID) and peptide backbone sequence/site attachment (ETD and/or high-energy CID) within the same LC–MS run. These MS fragmentation approaches are ideally suited with higher-resolution mass spectrometers, for example, QTOF, IM TOF, and Orbitrap analyzers. The interpretation of the complex and abundant data generated from these experiments undoubtedly requires the support of the growing resources of bioinformatics tools for automated search and identification of glycopeptides and the attached glycans. The advantages of this multi-fragmentation approach (ETD, CID) combined with these high-resolution mass analyzers are also essential in the mapping of disulfide linkages in recombinant protein therapeutics. Even though disulfide linkages are assigned in the initial development stage of the protein, they often need to be reassigned in large-scale production or when the cell production conditions change. Therefore, confirmation of disulfide linkages and identification of any unpaired C location needs to be provided by the aforementioned mapping approach, thus ensuring the proper folding and biological activity of the protein therapeutic product. The latter is especially critical in case of developing innovative treatments using mAbs, which are expected to top the global market in protein therapeutics in the near future. Fast growth in protein therapeutics will also strengthen the emerging segment of biogenerics (biosimilars), which is a key future growth sector due to patent expirations of the branded innovator products. In that case, a thorough characterization of the biosimilar product in terms of glycosylation occupancy and identification of disulfide linkages will be essential for evaluating the comparability between the innovator and biosimilar products. In case of a generic variant of rtPA (TNK-tPA) [150], the analysis strategy was focused on regions that could impact the clot lysis activity such as the glycosylation occupancy at the N184 site and the different extent of oxidation at several M sites. Finally, the advent of more accurate and sensitive instrumentation will enable the development of novel methodologies for the structural characterization of recombinant protein therapeutics and shed some

light into the role of specific carbohydrates in many complex biological interactions. That, in turn, will incite the development of novel glycosylated therapeutics for treating infectious, chronic, and other diseases, as well as the improvement of the immunogenicity and pharmacokinetic profiles of existing protein therapeutics.

Acknowledgments We acknowledge the kind permission of the Schering-Plough Research Institute to reproduce previously reported, but unpublished data regarding the IL-5R α .

References

1. Walsh G (2006) Biopharmaceutical benchmarks 2006. *Nat Biotechnol* 24(7):769–776
2. Aggarwal S (2007) What's fueling the biotech engine? *Nat Biotechnol* 25(10):1097–1104
3. Roach P, Woodworth JR (2002) Clinical pharmacokinetics and pharmacodynamics of insulin lispro mixtures. *Clin Pharmacokinet* 41:1043–1057
4. Dwek RA (1996) Glycobiology: toward understanding the function of sugars. *Chem Rev* 96:683–720
5. Collins MO, Yu L, Choudhary JS (2007) Analysis of protein phosphorylation on a proteome-scale. *Proteomics* 7:2751–2768
6. Wedemeyer WJ, Welker E, Narayan M et al (2000) Disulfide bonds and protein folding. *Biochemistry* 39:4207–4216
7. Graves JD, Krebs EG (1999) Protein phosphorylation and signal transduction. *Pharmacol Ther* 82:111–121
8. Hunter T (2000) Signaling-2000 and beyond. *Cell* 100:113–127
9. Cohen P (2002) The origins of protein phosphorylation. *Nat Cell Biol* 4:E127–E130
10. Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* 24:1241–1252
11. Li H, d'Anjou M (2009) Pharmacological significance of glycosylation in therapeutic proteins. *Curr Opin Biotechnol* 20:678–684
12. CPMP/ICH harmonised tripartite guideline Q6B (1999) Specifications: test procedures and acceptance criteria for biotechnological/biological products. March 1999 and EMA guideline (2010) requirements for quality documentation concerning biological investigational medicinal products in clinical trials. February 2010
13. Whitehouse CM, Dreyer RN, Yamashita M et al (1985) Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* 57:675–679
14. Fenn JB, Mann M, Meng CK et al (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64–71
15. Smith RD, Udseth H (1988) Capillary zone electrophoresis-MS. *Nature* 331:639–640
16. Kelly JF, Locke SJ, Ramaley L et al (1996) Development of electrophoretic conditions for the characterization of protein glycoforms by capillary electrophoresis-electrospray mass spectrometry. *J Chromatogr A* 720:409–427
17. Karas M, Bachmann D, Bahr U et al (1987) Matrix-assisted ultraviolet-laser desorption of nonvolatile compounds. *Int J Mass Spectrom Ion Process* 78:53–68
18. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60:2299–2301
19. Hancock WS, Wu SL, Shieh P (2002) The challenges of developing a sound proteomics strategy. *Proteomics* 2:352–359
20. Larsen MR, Trelle MB, Thingholm TE et al (2006) Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* 40:790–798
21. Covey T, Shushan B, Bonner R, Schröder W, Hucho F (1991) Methods in protein sequence analysis. In: Jörnvall H, Höög JO, Gustavsson AM (eds) *LC/MS and LC/MS/MS screening of the sites of posttranslational modification in proteins*. Birkhäuser Press, Basel

22. Dell A, Morris HR (2001) Glycoprotein structure determination by mass spectrometry. *Science* 291:2351–2356
23. Bateman RH, Carruthers R, Hoye JB et al (2002) A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J Am Soc Mass Spectrom* 13:792–803
24. Zubarev RA, Kelleher NL, McLafferty FW (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc* 120:3265–3266
25. Syka JE, Coon JJ, Schroeder MJ et al (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 101:9528–9533
26. Wu SL, Jiang H, Lu Q et al (2009) Mass spectrometric determination of disulfide linkages in recombinant therapeutic proteins using online LC-MS with electron-transfer dissociation. *Anal Chem* 81:112–122
27. Wang D, Hincapie M, Rejtar T et al (2011) Ultrasensitive characterization of site-specific glycosylation of affinity-purified haptoglobin from lung cancer patient plasma using 10 μm i.d. porous layer open tubular liquid chromatography-linear ion trap collision-induced dissociation/electron transfer dissociation mass spectrometry. *Anal Chem* 83(6):2029–2037
28. Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473:4–8
29. Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptides bonds. *Glycobiology* 12:43R–56R
30. Schachter H (2001) The clinical relevance of glycobiology. *J Clin Invest* 108:1579–1582
31. Dwek MV, Brooks SA (2004) Harnessing changes in cellular glycosylation in new cancer treatment strategies. *Curr Cancer Drug Targets* 4:425–442
32. Wührer M (2007) Glycosylation profiling in clinical proteomics: heading for glycan biomarkers. *Expert Rev Proteomics* 4:135–136
33. Dube DH, Bertozzi CR (2005) Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nat Rev Drug Discov* 4:477–488
34. Fuster MM, Esko JD (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer* 5:526–542
35. An HJ, Kronewitter SR, de Leoz ML et al (2009) Glycomics and disease markers. *Curr Opin Chem Biol* 13:601–607
36. Niwa T (2006) Mass spectrometry for the study of protein glycation in disease. *Mass Spectrom Rev* 25:713–723
37. Morelle W, Canis K, Chirat F et al (2006) The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* 6:3993–4015
38. Bennett CS, Dean SM, Payne RJ et al (2008) Sugar-assisted glycopeptide ligation with complex oligosaccharides: scope and limitations. *J Am Chem Soc* 130:11945–11952
39. Novotny MV, Mechref Y (2005) New hyphenated methodologies in high sensitivity glycoprotein analysis. *J Sep Sci* 28:1956–1968
40. Wührer M, Deedler AM, Hokke CH (2005) Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr B* 825:124–133
41. Geyer H, Geyer R (2006) Strategies for analysis of glycoprotein glycosylation. *Biochim Biophys Acta* 1764:1853–1869
42. Mariño K, Bones J, Kattla JJ et al (2010) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol*: 713–723
43. North SJ, Hitchen PG, Haslam SM et al (2009) Mass spectrometry in the analysis of N-linked and O-linked glycans. *Curr Opin Struct Biol* 19:498–506
44. Axford J (2001) The impact of glycobiology on medicine. *Trends Immunol* 22:237–239
45. Mortz E, Sareneva T, Haebel S et al (1996) Mass spectrometric characterization of glycosylated interferon-gamma variants separated by gel electrophoresis. *Electrophoresis* 17:925–931
46. Nawarak J, Phutrakul S, Chen ST (2004) Analysis of lectin-bound glycoproteins in snake venom from the elapidae and viperidae families. *J Proteom Res* 3:383–392
47. Mechref Y, Novotny MV (2002) Structural investigations of glycoconjugates at high sensitivity. *Chem Rev* 102:321–369

48. Ramdani B, Nuyens V, Codden T et al (2003) Analyte comigrating with trisialotransferrin during capillary zone electrophoresis of sera from patients with cancer. *Clin Chem* 49:1854–1864
49. Smith RD, Loo JA, Edmonds CG et al (1990) New developments in biochemical mass spectrometry: electrospray ionization. *Anal Chem* 62:882–899
50. Tzarbopoulos A, Pramanik BN, Nagabhushan TL et al (1995) Structural analysis of the CHO-derived interleukin-4 by liquid-chromatography/electrospray ionization mass spectrometry. *J Mass Spectrom* 30:1752–1763
51. Tzarbopoulos A, Bahr U, Karas M, Pramanik BN (2002) Structural analysis of glycoproteins by electrospray ionization mass spectrometry. In: Pramanik BN, Ganguly AK, Gross ML (eds) *Applied electrospray mass spectrometry*. Marcel Dekker, New York
52. Duffin KL, Welply JK, Huang E et al (1992) Characterization of N-linked oligosaccharides by electrospray and tandem mass spectrometry. *Anal Chem* 64:1440–1448
53. Rajan N, Tzarbopoulos A, Kumarasamy R et al (1995) Characterization of recombinant human interleukin-4 receptor from CHO cells: Role of N-linked oligosaccharides. *Biochem Biophys Res Commun* 206:694–702
54. Rush RS, Derby PL, Smith DM et al (1995) Microheterogeneity of erythropoietin carbohydrate structure. *Anal Chem* 67:1442–1452
55. Wilm M, Mann M (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem* 68:1–8
56. Verentchikov AN, Ens W, Standing KG (1994) Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction. *Anal Chem* 66:99–107
57. Makarov A (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 72:1156–1162
58. Olivova P, Chen W, Chakraborty AB et al (2008) Determination of N-glycosylation sites and site heterogeneity in a monoclonal antibody by electrospray quadrupole ion-mobility time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 22:29–40
59. Benesch JLP, Robinson CV (2006) Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Current Opin Struct Biol* 16:245–251
60. Clemmer DE, Jarrold MF (1997) Ion mobility measurements and their applications to clusters and biomolecules. *J Mass Spectrom* 32:577–592
61. Carter P, Presta L, Gorman CM et al (1992) Humanization of an Anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci USA* 89:4285–4289
62. Damen CWN, Chen W, Chakraborty AB et al (2009) Electrospray ionization quadrupole ion-mobility time-of-flight mass spectrometry as a tool to distinguish the lot-to-lot heterogeneity in N-Glycosylation profile of the therapeutic monoclonal antibody Trastuzumab. *J Amer Soc Mass Spectrom* 20:2021–2033
63. Dube S, Fisher JW, Powell JS (1988) Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion and biological function. *J Biol Chem* 263:17516–17521
64. Delorme E, Lorenzini T, Giffin J et al (1992) Role of glycosylation on the secretion and biological activity of erythropoietin. *Biochemistry* 31:9871–9876
65. Ploug M, Rahbek-Nielsen H, Nielsen PF et al (1998) Glycosylation profile of a recombinant urokinase-type plasminogen activator receptor expressed in Chinese hamster ovary cells. *J Biol Chem* 273(22):13933–13943
66. Tzarbopoulos A, Prongay A, Baldwin S et al (1996) Mass spectrometric analysis of the Sf9 cell-derived interleukin-5 Receptor. In: *Proceedings of the 44th ASMS conference on mass spectrometry and allied topics, Portland: 12–16 May*
67. Karas M, Bahr U, Strupat K et al (1995) Matrix dependence of metastable fragmentation of glycoproteins in MALDI TOF mass spectrometry. *Anal Chem* 67:675–679
68. Giménez E, Benavente F, Barbosa J et al (2007) Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 21:2555–2563
69. Tzarbopoulos A, Pramanik BN, Karas M et al (1995) Factors affecting the choice of matrix in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of glycoproteins. *J Mass Spectrom*: S207–S209

70. Liu CL, Bowers LD (1997) Mass spectrometric characterization of the β -subunit of human chorionic gonadotropin. *J Mass Spectrom* 32:33–42
71. Neusüb C, Demelbauer U, Pelzing M (2005) Glycoform characterization of intact erythropoietin by capillary electrophoresis-electrospray-time of flight-mass spectrometry. *Electrophoresis* 26:1442–1450
72. Demelbauer UM, Plematl A, Kremser L et al (2004) Characterization of glyco isoforms in plasma-derived human antithrombin by on-line capillary zone electrophoresis-electrospray ionization-quadrupole ion trap-mass spectrometry of the intact glycoproteins. *Electrophoresis* 25:2026–2032
73. Balaguer E, Demelbauer U, Pelzing M et al (2006) Glycoform characterization of erythropoietin combining glycan and intact protein analysis by capillary electrophoresis–electrospray–time-of-flight mass spectrometry. *Electrophoresis* 27:2638–2650
74. Balaguer E, Neuss C (2006) Glycoprotein characterization combining intact protein and glycan analysis by capillary electrophoresis-electrospray ionization-mass spectrometry. *Anal Chem* 78:5384–5393
75. Thakur D, Rejtar T, Karger BL et al (2009) Profiling the glycoforms of the intact α subunit of recombinant human chorionic gonadotropin by high-resolution capillary electrophoresis-mass spectrometry. *Anal Chem* 81:8900–8907
76. Sanz-Nebot V, Balaguer E, Benavente F et al (2007) Characterization of transferrin glycoforms in human serum by CE-UV and CE-ESI-MS. *Electrophoresis* 28:1949–1957
77. Hang HC, Bertozzi CR (2005) The chemistry and biology of mucin-type Olinked glycosylation. *Bioorg Med Chem* 13:5021–5034
78. Wopereis S, Lefeber DJ, Morava E et al (2006) Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin Chem* 52:574–600
79. Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 54:631–664
80. Vance BA, Wu W, Ribaldo RK et al (1997) Multiple dimeric forms of human CD69 result from differential addition of N-glycans to typical (Asn–X–Ser/Thr) and atypical (Asn–X–Cys) glycosylation motifs. *J Biol Chem* 272:23117–23122
81. Kelleher NL, Lin H, Valaskovic G et al (1999) Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J Am Chem Soc* 121:806–812
82. Kelleher NL (2004) Top-down proteomics. *Anal Chem* 76:196A–203A
83. Reid GE, McLuckey SA (2002) ‘Top down’ protein characterization via tandem mass spectrometry. *J Mass Spectrom* 37:663–675
84. Siuti N, Kelleher NL (2007) Decoding protein modifications using top-down mass spectrometry. *Nat Methods* 4:817–821
85. Ling V, Guzzetta AW, Canova-Davis E et al (1991) Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry. *Anal Chem* 63:2909–2915
86. Huddleston MJ, Bean MF, Carr SA (1993) Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal Chem* 65:877–884
87. Amon S, Alina D, Zamfir AD et al (2008) Glycosylation analysis of glycoproteins and proteoglycans using capillary electrophoresis-mass spectrometry strategies. *Electrophoresis* 29:2485–2507
88. Alving K, Körner R, Paulsen H et al (1998) Nanospray-ESI low-energy CID and MALDI post-source decay for determination of O-glycosylation sites in MUC4 peptides. *J Mass Spectrom* 33:1124–1133
89. Hunt DF, Shabanowitz J, Yates JR et al (1986) Tandem quadrupole Fourier-transform mass spectrometry of oligopeptides and small proteins. *Proc Natl Acad Sci USA* 83:6233–6237
90. Mechref Y, Madera M, Novotny MV (2009) Assigning glycosylation sites and microheterogeneities in glycoproteins by liquid chromatography/tandem mass spectrometry. *Methods Mol Biol* 492:161–180

91. Annesley TM (2003) Ion suppression in mass spectrometry. *Clin Chem* 49:1041–1044
92. Temporini C, Calleri E, Massolini G et al (2008) Integrated analytical strategies for the study of phosphorylation and glycosylation in proteins. *Mass Spectrom Rev* 27:207–236
93. Drake RR, Schwegler EE, Malik G et al (2006) Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol Cell Proteomics* 5:1957–1967
94. Alvarez-Manilla G, Atwood J III, Guo Y et al (2006) Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *J Proteome Res* 5:701–708
95. Tajiri M, Yoshida S, Wada Y (2005) Differential analysis of site-specific glycans on plasma and cellular fibronectins: Application of a hydrophilic affinity method for glycopeptides enrichment. *Glycobiology* 15(12):1332–1340
96. Wada Y, Tajiri M, Yoshida S (2004) Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal Chem* 76:6560–6565
97. Häggglund P, Bunkenborg J, Elortza F et al (2004) A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* 3:556–566
98. Liu X, Li X, Chan K et al (2007) One-pot methylation in glycomics application: esterification of sialic acids and permanent charge construction. *Anal Chem* 79:3894–3900
99. Larsen MR, Højrup P, Roepstorff P (2005) Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. *Mol Cell Proteomics* 4:107–119
100. Brittain SM, Ficarro SB, Brock A et al (2005) Enrichment analysis of peptide subsets using fluorour affinity tags and mass spectrometry. *Nat Biotechnol* 23:463–468
101. Mirzaei H, Regnier F (2005) Affinity chromatographic selection of carbonylated proteins followed by identification of oxidation sites using tandem mass spectrometry. *Anal Chem* 77:2386–2392
102. Zhang W, Zhou G, Zhao Y et al (2003) Affinity enrichment of plasma membrane for proteomics analysis. *Electrophoresis* 24:2855–2863
103. Zhang H, Yi EC, Li XJ (2005) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol Cell Proteomics* 4:144–155
104. Zhao Y, Zhang W, Kho Y et al (2004) Proteomic analysis of integral plasma membrane proteins. *Anal Chem* 76:1817–1823
105. Bailey MJ, Hooker AD, Adams CS et al (2005) A platform for high-throughput molecular characterization of recombinant monoclonal antibodies. *J Chromatogr B* 826:177–187
106. Bundy JL, Fenselau C (2001) Lectin and carbohydrate affinity surfaces for mass spectrometric analysis of microorganisms. *Anal Chem* 73:751–757
107. Xiong L, Andrews D, Regnier F (2003) Comparative proteomics of glycoproteins based on lectin selection and isotope coding. *J Proteome Res* 2:618–625
108. Madera M, Mechref Y, Novotny MV (2005) Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal Chem* 77:4081–4090
109. Bedair M, El Rassi Z (2005) Affinity chromatography with monolithic capillary columns II. Polymethacrylate monoliths with immobilized lectins for the separation of glycoconjugates by nano-liquid affinity chromatography. *J Chromatogr A* 1079:236–245
110. Okanda FM, El Rassi Z (2006) Affinity chromatography with monolithic capillary columns for glycomics/proteomics: 1. polymethacrylate monoliths with immobilized lectins for glycoprotein separation by affinity capillary electrochromatography and affinity nano-liquid chromatography in either a single column or columns coupled in series. *Electrophoresis* 27:1020–1030
111. Mao X, Luo Y, Dai Z et al (2004) Integrated lectin affinity microfluidic chip for glycoform separation. *Anal Chem* 76:6941–6947

112. Budnik BA, Lee RS, Steen JA (2006) Review Global methods for protein glycosylation analysis by mass spectrometry. *Biochim Biophys Acta* 1764:1870–1880
113. Wang L, Li F, Sun W et al (2006) Concanavalin A-captured glycoproteins in healthy human urine. *Mol Cell Proteomics* 5:560–562
114. Kaji H, Saito H, Yamauchi Y et al (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* 21:667–672
115. Madera M, Mechref Y, Klouckova I et al (2007) High-sensitivity profiling of glycoproteins from human blood serum through multiple-lectin affinity chromatography and liquid chromatography/tandem mass spectrometry. *J Chromatogr B* 845:121–137
116. Cummings RD, Kornfeld S (1984) The distribution of repeating [Gal beta 1, 4GlcNAc beta 1, 3] sequences in asparagine-linked oligosaccharides of the mouse lymphoma cell lines BW5147 and PHAR 2.1. *J Biol Chem* 259:6253–6260
117. Yang Z, Hancock WS (2004) Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J Chromatogr A* 1053:79–88
118. Qiu R, Regnier FE (2005) Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal Chem* 77:2802–2809
119. Sumi S, Arai K, Kitahara S et al (1999) Serial lectin affinity chromatography demonstrates altered asparagine-linked sugar-chain structures of prostate-specific antigen in human prostate carcinoma. *J Chromatogr B* 727:9–14
120. Xiong L, Regnier FE (2002) Use of a lectin affinity selector in the search for unusual glycosylation in proteomics. *J Chromatogr, B: Anal Technol Biomed Life Sci* 782:405–418
121. Yang Z, Hancock WS (2005) Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *J Chromatogr A* 1070:57–64
122. Wang Y, Wu S, Hancock WS (2006) Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap Fourier transform mass spectrometry. *Glycobiology* 16:514–523
123. Yue GE, Roper MG, Balchunas C et al (2006) Protein digestion and phosphopeptides enrichment on glass microchip. *Anal Chim Acta* 564:116–122
124. Madera M, Mechref Y, Klouckova I et al (2006) Semiautomated high-sensitivity profiling of human blood serum glycoproteins through lectin preconcentration and multidimensional chromatography/tandem mass spectrometry. *J Proteome Res* 5:2348–2363
125. Guzman NA, Phillips TM (2005) Immunoaffinity CE for proteomics studies. *Anal Chem* 77:60A–67A
126. Benavente F, Hernández E, Guzman NA et al (2007) Determination of human erythropoietin by on-line immunoaffinity capillary electrophoresis: a preliminary report. *Anal Bioanal Chem* 387:2633–2639
127. An HJ, Peavy TR, Hedrick JL et al (2003) Determination of N-glycosylation sites and site heterogeneity in glycoproteins. *Anal Chem* 75:5628–5637
128. Temporini C, Perani E, Calleri E et al (2007) Pronase-immobilized enzyme reactor: an approach for automation in glycoprotein analysis by LC/LC-ESI/MSn. *Anal Chem* 79:355–363
129. Jebanathirajah J, Steen H, Roepstorff P (2003) Using optimized collision energies and high resolution, high accuracy fragment ion selection to improve glycopeptide detection by precursor ion scanning. *J Am Soc Mass Spectrom* 14:777–784
130. Zhang H, Li XJ, Martin DB et al (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21:660–666
131. Khidekel N, Arndt S, Lamarre-Vincent N et al (2003) A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J Am Chem Soc* 125:16162–16163
132. Sprung R, Nandi A, Chen Y et al (2005) Tagging-via-substrate strategy for probing O-GlcNAc modified proteins. *J Proteome Res* 4:950–957
133. Saxon E, Bertozzi CR (2000) Cell surface engineering by a modified Staudinger reaction. *Science* 287:2007–2010

134. Khidekel N, Ficarro SB, Peters EC et al (2004) Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci USA* 101:13132–13137
135. Lamarre-Vincent N, Hsieh-Wilson LC (2003) Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J Am Chem Soc* 125:6612–6613
136. Zhang Y, Wolf-Yadlin A, Ross PL et al (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 4:1240–1250
137. Vocadlo DJ, Hang HC, Kim EJ et al (2003) A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci USA* 100:9116–9121
138. Prescher JA, Dube DH, Bertozzi CR (2004) Chemical remodelling of cell surfaces in living animals. *Nature* 430:873–877
139. Kho Y, Kim SC, Jiang C et al (2004) A tagging-via- substrate technology for detection and proteomics of farnesylated proteins. *Proc Natl Acad Sci USA* 101:12479–12484
140. Wells L, Vosseller K, Cole RN et al (2002) Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* 1:791–804
141. Vosseller K, Hansen KC, Chalkley RJ et al (2005) Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* 5:388–398
142. Wührer M, Catalina MI, Deelder AM et al (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci* 849:115–128
143. Carr SA, Hemling ME, Bean MF et al (1991) Integration of mass spectrometry in analytical biotechnology. *Anal Chem* 63:2802–2824
144. Burlingame AL (1996) Characterization of protein glycosylation by mass spectrometry. *Curr Opin Biotechnol* 7:4–10
145. Carr SA, Roberts GD (1986) Carbohydrate mapping by mass spectrometry: a novel method for identifying attachment sites of Asn-linked sugars in glycoproteins. *Anal Biochem* 157:396–406
146. Küster B, Mann M (1999) 18O-labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal Chem* 71:1431–1440
147. Leonard CK, Spellman MW, Riddle L et al (1990) Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* 265:10373–10382
148. Carr SA, Roberts GD, Jurewicz A et al (1998) Structural fingerprinting of Asn-linked carbohydrates from specific attachment sites in glycoproteins by mass spectrometry: application to tissue plasminogen activator. *Biochimie* 70:1445–1454
149. Guzzetta AW, Basa LJ, Hancock WS et al (1993) Identification of carbohydrate structures in glycoprotein peptide maps by the use of LC/MS with selected ion extraction with special reference to tissue plasminogen activator and a glycosylation variant produced by site directed mutagenesis. *Anal Chem* 65:2953–2962
150. Jiang H, Wu SL, Karger BL et al (2010) Characterization of the glycosylation occupancy and the active site in the follow-on protein therapeutic: TNK-tissue plasminogen activator. *Anal Chem* 82:6154–6162
151. Domon B, Costello CE (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj J* 5:397–409
152. Carr SA, Huddleston MJ, Bean MF (1993) Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci* 2:183–196
153. Harvey DJ, Bateman RH, Bordoli RS et al (2000) Ionization and fragmentation of complex glycans with a quadrupole time-of-flight mass spectrometer fitted with a matrix-assisted laser desorption/ionization ion source. *Rapid Commun Mass Spectrom* 14:2135–2142

154. Borisov OV, Field M, Ling VT et al (2009) Characterization of Oligosaccharides in recombinant tissue plasminogen activator produced in Chinese hamster ovary cells: Two decades of analytical technology development. *Anal Chem* 81:9744–9754
155. Demelbauer UM, Zehl M, Plematl A et al (2004) Determination of glycopeptide structures by multistage mass spectrometry with low-energy collision-induced dissociation: comparison of electrospray ionization quadrupole ion trap and matrix-assisted laser desorption/ionization quadrupole ion trap reflectron time-of-flight approaches. *Rapid Commun Mass Spectrom* 18(14):1575–1582
156. Bones J, McLoughlin N, Hilliard M et al (2011) 2D-LC Analysis of BRP 3 Erythropoietin N-Glycosylation using anion exchange fractionation and hydrophilic interaction UPLC reveals long Poly-N-Acetyl lactosamine extensions. *Anal Chem* 83:4154–4162
157. Harazono A, Kawasaki N, Itoh S et al (2006) Site-specific N-glycosylation analysis of human plasma ceruloplasmin using liquid chromatography with electrospray ionization tandem mass spectrometry. *Anal Biochem* 348:259–268
158. Schmitt S, Glebe D, Alving K et al (1999) Analysis of the Pre-S2 N- and O-Linked Glycans of the M surface protein from human hepatitis B virus. *J Biol Chem* 274:11945–11957
159. Zubarev RA, Horn DM, Fridriksson EK et al (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* 72:563–573
160. Kjeldsen F, Haselmann KF, Budnik BA et al (2002) Dissociative capture of hot (3–13 eV) electrons by polypeptide polycations: an efficient process accompanied by secondary fragmentation. *Chem Phys Lett* 356:201–206
161. Kjeldsen F, Haselmann KF, Budnik BA et al (2003) Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. *Anal Chem* 75(10):2355–2361
162. Mikesh LM, Ueberheide B, Chi A et al (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* 1764(12):1811–1822
163. Schroeder MJ, Webb DJ, Shabanowitz J et al (2005) Methods for the detection of paxillin post-translational modifications and interacting proteins by mass spectrometry. *J Proteome Res* 4(5):1832–1841
164. Hogan JM, Pitteri SJ, Chrisman PA et al (2005) Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J Proteome Res* 4(2):628–632
165. Mirgorodskaya E, Roepstorff P, Zubarev RA (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier Transform mass spectrometer. *Anal Chem* 71:4431–4436
166. Perdivara I, Petrovich R, Allinquant B et al (2009) Elucidation of O-Glycosylation structures of the β -Amyloid precursor protein by liquid chromatography-mass spectrometry using electron transfer dissociation and collision-induced dissociation. *J Proteom Res* 8:631–642
167. Wu SL, Huhmer AF, Hao Z et al (2007) On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with posttranslational modifications. *J Proteome Res* 6(11):4230–4244
168. Tzarbopoulos A, Bahr U, Pramanik BN et al (1997) Glycoprotein Analysis by Delayed extraction and post-source decay MALDI TOF MS. *Int J Mass Spectrom Ion Process* 169(170):251–261
169. Wührer M, Hokke CH, Deelder AM (2004) Glycopeptide analysis by matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry reveals novel features of horseradish peroxidase glycosylation. *Rapid Commun Mass Spectrom* 18:1741–1748
170. Bykova NV, Rampitsch C, Krokhin O et al (2006) Determination and characterization of site-specific N-Glycosylation using MALDI-Qq-TOF tandem mass spectrometry: case study with a plant protease. *Anal Chem* 78:1093–1103
171. Kuroguchi M, Matsushita T, Nishimura SI (2004) Post-translational modifications on proteins: facile and efficient procedure for the identification of O-Glycosylation sites by MALDI-LIFT-TOF/TOF mass spectrometry. *Angew Chem Int Ed Engl* 43:4071–4075

172. Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev* 18:349–451
173. Zaia J (2010) Mass spectrometry and glycomics. *OMICS* 14(4):401–418
174. Wormald MR, Petrescu AJ, Pao Y-L et al (2002) Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modeling. *Chem Rev* 102:371–386
175. Koerner TA, Yu RK, Scarsdale JN et al (1988) Analysis of complex carbohydrate primary and secondary structure via two-dimensional proton nuclear magnetic resonance spectroscopy. *Adv Exp Med Biol* 228:759–784
176. Perez S, Mulloy B (2005) Prospects for glycoinformatics. *Curr Opin Struct Biol* 15:517–524
177. Aoki-Kinoshita KF (2008) An introduction to bioinformatics for glycomics research. *PLoS Comput Biol*. doi:[10.1371/journal.pcbi.1000075](https://doi.org/10.1371/journal.pcbi.1000075)
178. von der Lieth CW, Lütke T, Frank M (2006) The role of informatics in glycobiology research with special emphasis on automatic interpretation of MS spectra. *Biochim Biophys Acta* 1760:568–577
179. Cooper CA, Gasteiger E, Packer NH (2001) GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* 1:340–349
180. Go EP, Rebecchi KR, Dalpathado DS et al (2007) GlycoPep DB: a tool for glycopeptide analysis using a “smart search”. *Anal Chem* 79:1708–1713
181. Goldberg D, Sutton-Smith M, Paulson J et al (2005) Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. *Proteomics* 5:865–875
182. Goldberg D, Bern M, Parry S et al (2007) Automated N-glycopeptide identification using a combination of single- and tandem-MS. *J Proteome Res* 6:3995–4005
183. Ozohanic O, Krenyacz J, Ludanyi K et al (2008) GlycoMiner: a new software tool to elucidate glycopeptide composition. *Rapid Commun Mass Spectrom* 22:3245–3254
184. An HJ, Tillinghast JS, Woodruff DL et al (2006) A new computer program (GlycoX) to determine simultaneously the glycosylation sites and oligosaccharide heterogeneity of glycoproteins. *J Proteome Res* 5:2800–2808
185. Ren JM, Rejtar T, Li L et al (2007) N-Glycan structure annotation of glycopeptides using a linearized glycan structure database (GlyDB). *J Proteome Res* 6:3162–3173
186. Irungu J, Go EP, Dalpathado DS et al (2007) Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal Chem* 79:3065–3074
187. Hizukuri Y, Yamanishi Y, Nakamura O et al (2005) Extraction of leukemia specific glycan motifs in humans by computational glycomics. *Carbohydr Res* 340:2270–2278
188. Aoki K, Yamaguchi A, Ueda N et al (2004) KCaM (KEGG Carbohydrate Matcher): a software tool for analyzing the structures of carbohydrate sugar chains. *Nucleic Acids Res* 32:W267–W272
189. Aoki K, Mamitsuka H, Akutsu T et al (2005) A score matrix to reveal the hidden links in glycans. *Bioinformatics* 21:1457–1463
190. Hashimoto K, Goto S, Kawano S et al (2006) KEGG as a glycome informatics resource. *Glycobiology* 16:63R–70R
191. Creighton TE (1984) Disulfide bond formation in proteins. In: Wold F, Moldave K (eds) *Methods in enzymology*, vol 107. Academic Press, San Diego, p 305
192. Dranoff G (2009) Targets of protective tumor immunity. *Ann NY Acad Sci* 1174:74–80
193. Nakamura T, Lipton SA (2009) Cell death: protein misfolding and neurodegenerative diseases. *Apoptosis* 14:455–468
194. Wess J, Han SJ, Kim SK et al (2008) Conformational changes involved in G-protein-coupled-receptor activation. *Trends Pharmacol Sci* 29:616–625
195. Thornton JM (1981) Disulphide bridges in globular proteins. *J Mol Biol* 151:261–287
196. Welker E, Raymond LD, Scheraga HA et al (2002) Intramolecular versus intermolecular disulfide bonds in prion proteins. *J Biol Chem* 277:33477–33481
197. Tsarbopoulos A, Pramanik B, Labdon J et al (1993) Isolation and characterization of a resistant core peptide of recombinant human Granulocyte-Macrophage colony-stimulating factor (GM-CSF); confirmation of the GM-CSF amino acid sequence by mass spectrometry. *Protein Sci* 2:1948–1958

198. Gorman JJ, Wallis TP, Pitt JJ (2002) Protein disulfide bond determination by mass spectrometry. *Mass Spectrom Rev* 21:183–216
199. Barber M, Bordoli RS, Sedgwick RD et al (1981) Fast atom bombardment of solids (FAB): A new ion source for mass spectrometry. *J Chem Soc, Chem Commun* 7:325–327
200. Morris HR, Pucci P (1985) A new method for rapid assignment of S-S bridges in proteins. *Biochem Biophys Res Commun* 126:1122–1128
201. Smith DL, Zhou Z (1990) Strategies for locating disulfide bonds in proteins. In: McCloskey JA (ed) *Methods in enzymology*, vol 193. Academic Press, New York, p 374
202. Sundqvist B, Roepstorff P, Fohlman J et al (1984) Molecular weight determination of proteins by californium plasma desorption mass spectrometry. *Science* 226:696–698
203. Tsbopopoulos A, Becker GW, Occolowitz JL et al (1988) Peptide and protein mapping by ²⁵²Cf-Plasma desorption mass spectrometry. *Anal Biochem* 171:113–123
204. Robertson JG, Adams GW, Medzihradsky KF et al (1994) Complete assignment of disulfide bonds in bovine dopamine beta-hydroxylase. *Biochemistry* 33:11563–11575
205. Pramanik BN, Tsbopopoulos A, Labdon JE et al (1991) Structural analysis of biologically active peptides and recombinant proteins and their modified counterparts by mass spectrometry. *J Chromatogr* 562:377–389
206. Chen G, Liu YH, Pramanik BN (2007) LC/MS analysis of proteins and peptides in drug discovery. In: Kazakevich Y, LoBrutto R (eds) *HPLC for pharmaceutical scientists*. Wiley, New York
207. Tsbopopoulos A, Karas M, Strupat K et al (1994) Comparative mapping of recombinant proteins and glycoproteins by plasma desorption and matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* 66:2062–2070
208. Patterson SD, Katta V (1994) Prompt fragmentation of disulfide-linked peptides during matrix-assisted laser desorption ionization mass spectrometry. *Anal Chem* 66:3727–3732
209. Sanger F (1953) A disulphide interchange reaction. *Nature* 171:1025–1026
210. Yazdanparast R, Andrews PC, Smith DL et al (1987) Assignment of disulfide bonds in proteins by fast atom bombardment mass spectrometry. *J Biol Chem* 262:2507–2513
211. Tsbopopoulos A, Varnerin J, Cannon-Carlson S et al (2000) Mass spectrometric mapping of disulfide bonds in recombinant human Interleukin-13. *J Mass Spectrom* 35:446–453
212. Sun Y, Bauer MD, Keough TW et al (1996) Disulfide bond location in proteins. *Methods Mol Biol* 61:181–210
213. Bauer M, Sun Y, Degenhardt C et al (1993) Assignment of all four disulfide bridges in echistatin. *J Prot Chem* 12:759–764
214. Bean MF, Carr SA (1992) Characterization of disulfide positions in proteins and sequence analysis of cystine-bridged peptides by tandem mass spectrometry. *Anal Biochem* 201:216–226
215. Pitt JJ, Da Silva E, Gorman JJ (2000) Determination of the disulfide bond arrangement of new castle disease virus hemagglutinin neuraminidase, correlation with a beta-sheet propeller structural fold predicted for paramyxoviridae attachment proteins. *J Biol Chem* 275:6469–6478
216. Gorman JJ, Ferguson BL, Speelman D et al (1997) Determination of the disulfide bond arrangement of human respiratory syncytial virus attachment (G) protein by matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Protein Sci* 6:1308–1315
217. Angal S, King DJ, Bodmer MW et al (1993) A single amino acid substitution abolishes the heterogeneity of chimeric mouse/human (IgG4) antibody. *Mol Immunol* 30:105–108
218. Wang Y, Lu Q, Wu SL et al (2011) Characterization and comparison of disulfide linkages and scrambling patterns in therapeutic monoclonal antibodies: using LC-MS with electron transfer dissociation. *Anal Chem* 83:3133–3140
219. Wu SL, Jiang H, Hancock WS et al (2010) Identification of the unpaired cysteine status and complete mapping of the 17 disulfides of recombinant tissue plasminogen activator using LC-MS with Electron transfer dissociation/collision induced dissociation. *Anal Chem* 82:5296–5303
220. Bagal D, Valliere-Douglass JF, Balland A et al (2010) Resolving disulfide structural isoforms of IgG2 monoclonal antibodies by ion mobility mass spectrometry. *Anal Chem* 82:6751–6755

221. Wallis TP, Pitt JJ, Gorman JJ (2001) Identification of disulfide-linked peptides by isotope profiles produced by peptic digestion of proteins in 50 % (18) O water. *Protein Sci* 10:2251–2271
222. Rose K, Savoy LA, Simona MG et al (1988) C-terminal peptide identification by fast atom bombardment mass spectrometry. *Biochem J* 250:253–259
223. Dwek MV, Ross HA, Leatham AJ (2001) Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. *Proteomics* 1:756–762
224. Rudd PM, Elliott T, Cresswell P et al (2001) Glycosylation and the immune system. *Science* 291:2370–2376
225. Peracaula R, Tabares G, Royle L et al (2003) Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* 13:457–470
226. Butler M, Quelhas D, Critchley AJ et al (2003) Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan processing step and provides an insight into pathogenesis. *Glycobiology* 13:601–622