Analyst

CRITICAL REVIEW

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2014, **139**, 2944

Mammalian protein glycosylation – structure *versus* function

S. Defaus,[ab] P. Gupta,[b] D. Andreu[a] and R. Gutiérrez-Gallego*[abc]

Received 4th December 2013
Accepted 6th March 2014

DOI: 10.1039/c3an02245e

www.rsc.org/analyst

Carbohydrates fulfil many common as well as extremely important functions in nature. They show a variety of molecular displays – *e.g.*, free mono-, oligo-, and polysaccharides, glycolipids, proteoglycans, glycoproteins, *etc.* – with particular roles and localizations in living organisms. Structure-specific peculiarities are so many and diverse that it becomes virtually impossible to cover them all from an analytical perspective. Hence this manuscript, focused on mammalian glycosylation, rather than a complete list of analytical descriptors or recognized functions for carbohydrate structures, comprehensively reviews three central issues in current glycoscience, namely (i) structural analysis of glycoprotein glycans, covering both classical and novel approaches for teasing out the structural puzzle as well as potential pitfalls of these processes; (ii) an overview of functions attributed to carbohydrates, covering from monosaccharide to complex, well-defined epitopes and full glycans, including post-glycosylational modifications, and (iii) recent technical advances allowing structural identification of glycoprotein glycans with simultaneous assignation of biological functions.

1. Introduction

The world of carbohydrates is extremely complex, rendering it both fascinating and challenging to those facing the task of unraveling their structural features. The term carbohydrate spans many different disciplines from large-scale industrial applications to fine-tuned biomedical uses, and the science of carbohydrates has experienced ups and downs over the last few decades in terms of attention paid, importance attributed, and level of understanding reached. Currently, the field of carbohydrate (bio)chemistry is enjoying renewed interest at both basic and applied (biomedical, pharmaceutical) levels, as clearly evidenced by the >500 reviews on the subject over the past 18 months. Most efforts are devoted to the study of carbohydrate-mediated biomolecular interactions and glycoprotein engineering but the structural analysis of carbohydrates, in all

[a]*Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona Biomedical Research Park, 08003 Barcelona, Spain. E-mail: ricardogutierrez@upf.edu*

[b]*Bio-analysis Group, Neuroscience Research Program, IMIM-Parc Salut Mar, Barcelona Biomedical Research Park, 08003 Barcelona, Spain*

[c]*Anapharm Biotech, 08038 Barcelona, Spain*

Sira Defaus received her B.S. (2008) and M.Sc. (2009) degrees from the Department of Organic Chemistry at the University of Barcelona. Later she joined the Bioanalysis and Analytical Services Research Group, and the Protein Chemistry and Proteomics laboratory at the Barcelona Biomedical Research Park (PRBB) to obtain her PhD degree. During her study she investigated the structure–activity relationships of glycans and glycopeptides and developed analytical tools for protein detection in biological sources. Her research interests are peptide chemistry and glycobiology.

Preeti Gupta obtained her M.Sc. degree in Biotechnology and M.Tech degree in Biochemical Engineering from Banaras Hindu University, India. In 2012 she joined the Bio-analysis Group at the Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona through the Agency for Management of University and Research Grants (AGAUR) offered by the Government of Catalonia, Spain to pursue her PhD. She is interested in characterizing protein glycosylation, glycan structure and glycoprotein identification and quantification in plasma samples using HPLC and MALDI-TOF and LC-MS.

its aspects, remains the basis of nearly all the developments of recent times. The goal of this review is to highlight relevant aspects of structural analysis of carbohydrates with focus on mammalian protein glycosylation and insights into its relevance. The final section deals with recent advances paving the way towards structural analysis within actual biological settings, ideally, without any external interference.

2. Structural analysis of glycoprotein glycans

Glycoproteins are fundamental in most important biological processes including fertilization, immune response, inflammation, viral replication, parasite infection, cell growth, cell–cell adhesion, or glycoproteins clearance. Whereas protein synthesis follows a well-defined, genetically encoded linear process, glycosylation is a non-template-driven, secondary gene event initiated during protein synthesis and involving a large collection of redundant and overlapping enzymes (glycosidases and glycosyltransferases) partially compartmentalized throughout the endoplasmic reticulum (ER) and the Golgi system.[1] Various competing reactions in the processing pathways, plus the need for enzyme, acceptor and substrate concurrence, as well as other physiological factors contribute to glycan microheterogeneity, *i.e.*, glycoprotein isoforms resulting from different glycans at a given site. This heterogeneity may be relatively simple, such as for RNAse B,[2] or rather complex as in the case of CD59 where at least 123 different desialylated glycan variants have been identified at a single site.[3] Thus, carbohydrate diversity and consequent complexity arises from several factors. Firstly, from the structural variety at the monosaccharide level, where multidirectional combinations of different monosaccharide building blocks, linkages,[4] anomericity, and branching generate a vast number of complex glycan structures (polysaccharides) that can be further modified by sulfation, acetylation, methylation, phosphorylation, *etc.*, and linked covalently to aglycones such as peptides (in different ways) or lipids forming the corresponding glycoconjugates (see Table 1). Secondly, from the influence of the peptide sequence in determining potential glycosylation sites, the effect of the 3D protein display in subsequent glycan processing events, and the spatial distribution or multivalent presentation leading to the avidity principle.[5] Thirdly, from microheterogeneity and macroheterogeneity phenomena inherent to carbohydrate chemistry resulting from the fact that in an individual glycoprotein a specific glycosylation site is not always associated with the same glycan structure and that not all *N*-glycan sequons are necessarily glycosylated.

Table 1 Different types of glycosylation. The letters in the sequence correspond to the 1-letter annotation of amino acids[a]

Linkage	Type	Sequence
Man–α-Trp	*C*-Mannosylation	W–X–X–W
GlcNAc–β-Asn	*N*-Glycosylation	N–X–(S/T) (X ≠ P)
		N–X–C, N–G, N–X– (rare)
GalNAc–α-Ser/Thr	*O*-Glycosylation	Various ppGalNAcT act concertedly
GlcNAc–β-Ser/Thr		Any S or T
GlcNAc–α-Thr		T (near P residues)
Glc–α-Tyr		GYG (glycogenin)
Glc–β-Ser		C–X–S–X–P–C
Glc–β-Asn		N–X–(S/T)
Gal–Thr		G–X–T (X = A, R, P, hP, S)
Gal–β-Hyl		X–Hyl–G
Fuc–α-Ser/Thr		C–X–X–G–G–(S/T)–C
		X–X–X–X–(S/T)
Man–α-Ser/Thr		I–X–P–T–(P/X)–T–X–P–
		X–X–X–X–P–T–X–(T/X)–X–X
Man–α-1-P–Ser		S rich domains
Xyl–β-Ser		–G–S–G– (near acidic residues)

[a] X may be any amino acid.

David Andreu studied chemistry at the University of Barcelona, where he obtained his PhD (1981, with E. Giralt). After further training in peptide chemistry during a postdoc with Bruce Merrifield at Rockefeller University (1982–1985), in 1985 he returned to Spain as Associate Professor of organic chemistry at the University of Barcelona. In 2001 he was appointed Professor of Chemistry at the Pompeu Fabra University where he heads the Protein Chemistry and Proteomics laboratory at the Barcelona Biomedical Research Park. In addition to glycan–protein interactions, his other scientific endeavours include cell-penetrating and antimicrobial peptides, and peptide-based synthetic vaccines.

Ricardo Gutiérrez-Gallego studied chemistry at Utrecht University (NL) and obtained a PhD in 2001 (with J. Vliegenthart and J. Kamerling). Subsequently, he joined the IMIM-Parc de Salut Mar where he developed the analytical methodology program for (glyco) protein hormone detection in the context of antidoping policies. Simultaneously, in 2001, he was appointed assistant professor of chemistry at Pompeu Fabra University where he expanded his research on the structure–function relationship of protein glycosylation with special emphasis on glycan-mediated biomolecular interactions. In 2013 he amplified his scope of activities joining Anapharm Biotech to provide analytical support in drug-development programs.

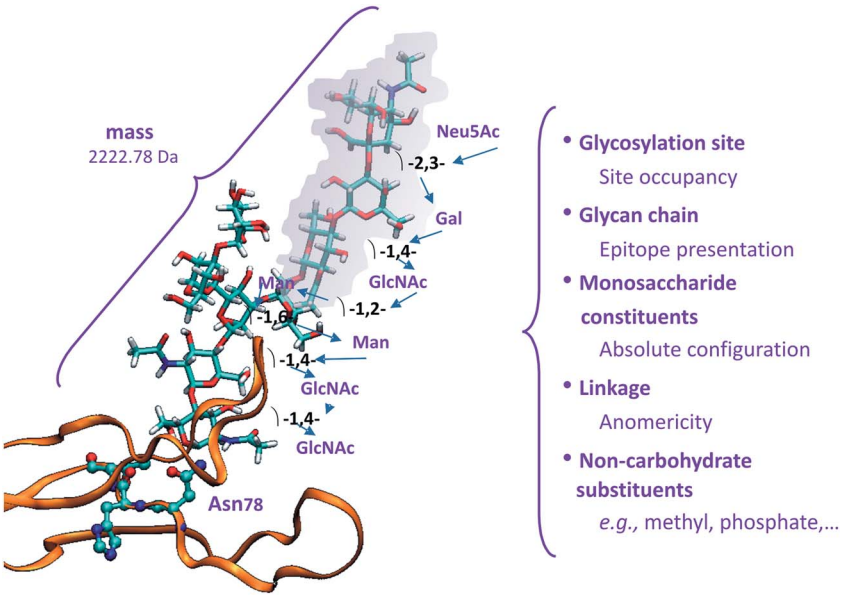


Fig. 1 As an example in the structural elucidation of glycoproteins an *N*-glycan in the human chorionic gonadotropin (hCG) glycoprotein α-chain is shown. Elements to be specified are listed on the right and some of them displayed. The shaded part represents the epitope potentially recognized in a carbohydrate-driven interaction.

Eventually, such diversity gives rise to a set of glycoforms, in both soluble and membrane-anchored forms that are as essential to life as a genetic code, and constitute an evolutionary conserved feature of all living cells.[6] The identification of the number, structure, and function of glycans in a particular biological context, initiated decades before the "omics" boom, was recently termed glycomics, and substantial progress has been made in understanding how glycans are directly involved in almost every biological process or human disease.[7] Still, the glycome is far more complex than the genome, transcriptome, or proteome, due to a much more dynamic character that varies considerably not only with the cell or tissue type, but also with the developmental stage,[8] metabolic state, or changes such as disease,[9] aging,[10,11] environmental factors,[12] or evolution.[13,14] For instance, epigenetic regulation may induce novel glycan structures that make the organism fitter in a specific environment without altering genetic information.[15] It is therefore of utmost importance to know what carbohydrate structures decorate which glycoproteins under particular conditions.

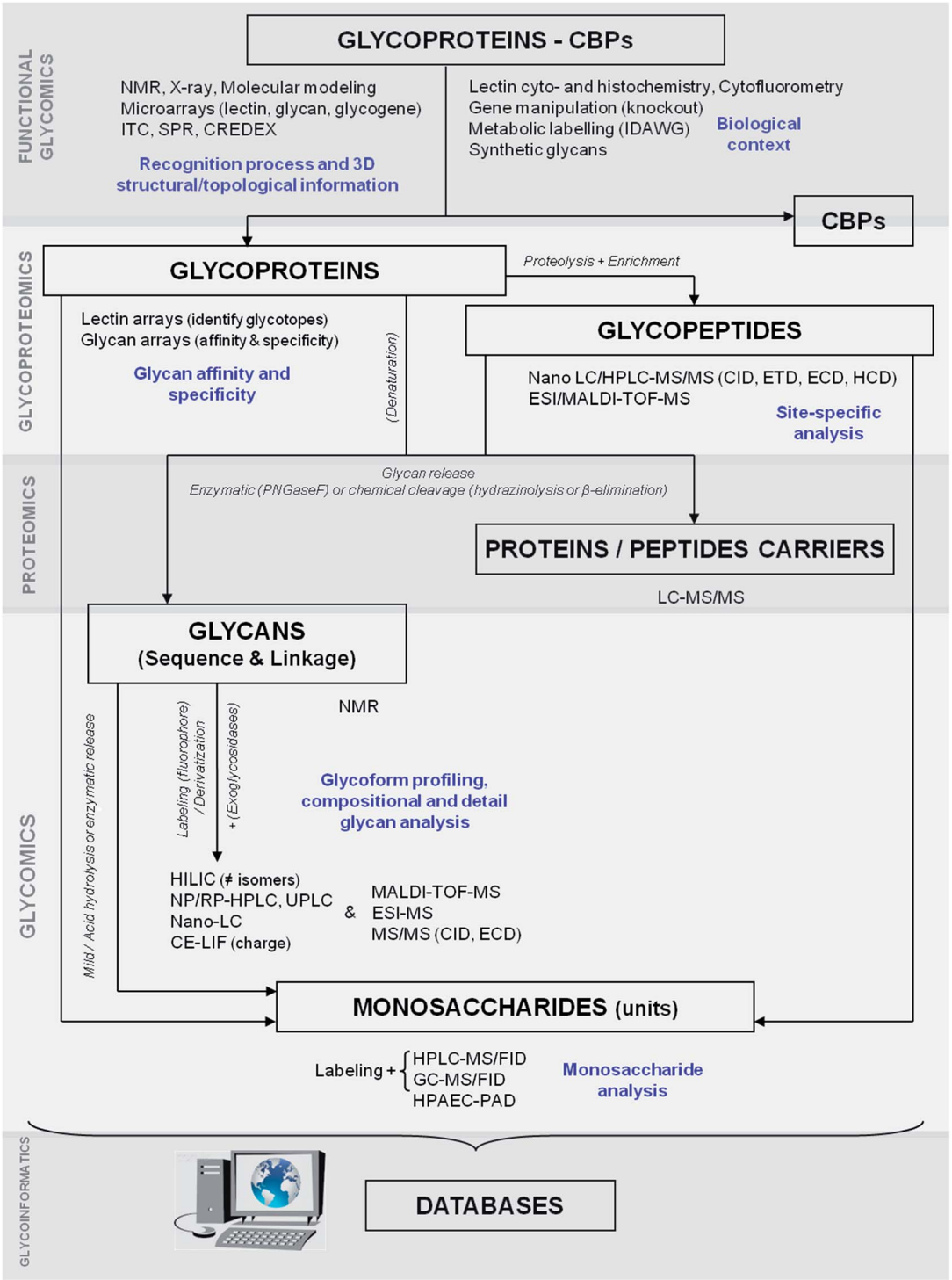
Even for dedicated specialists analysis of protein glycosylation remains an extremely challenging task due to many different physical parameters that must be established before a structural characterization can be considered complete (Fig. 1). As a consequence, there is no single analytical method capable of providing all the necessary information for fast and reliable identification and quantification of a particular structure, let alone to also establish its particular functionality. Rather, a multidimensional approach involving several orthogonal, physical, chemical, and biochemical techniques as depicted in Scheme 1 is required.

In the following pages an overview is provided of the techniques employed in structural analysis of protein glycosylation, their shortcomings and particular virtues, and the latest trends in this field.

2.1. Analyzing glycosylation

Over the last 2–3 decades the continuous refinement of analytical tools has greatly facilitated glycan analysis; numerous reviews[16–21] and papers cover the main technologies routinely used today for *N*- and *O*-linked glycan analysis, including capillary electrophoresis (CE),[22–24] liquid chromatography (LC),[25,26] mass spectrometry (MS)[27–30] and microarray-based[31–35] approaches to glycomics and glycoproteomics.[28,36,37] It is important to stress that in all these techniques a compromise exists between analytical sensitivity and the degree of structural detail provided. None of these tools, or any other for that matter, can single-handedly reveal all the features (see Fig. 1) necessary for full characterization. Hence, an unambiguously structural analysis must be conducted at different levels, namely intact glycoprotein, glycopeptides and released glycans, and in each case the most appropriate technique for deciphering that part of the puzzle must be chosen. This, in turn, entails another compromise between the degree of information obtained *vs.* the amount of (purified) material required.

2.1.1. Analysis of intact glycoproteins. In the first evaluation of protein glycosylation it is recommended to assess the microheterogeneity at the glycoprotein level as it provides an excellent starting point. Quite often this is done by means of conventional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and comparing the resulting bands to a protein molecular weight standard (Fig. 5). Such evaluation, when conducted with non-specific staining techniques using coomassie, silver, or Pro-Q emerald dyes, should provide an unbiased view of the glycoform distribution. Alternatively, the detection could be performed through specific biomolecular recognition (using lectins, antibodies, *etc.*) at much better sensitivity than the non-specific staining. However, one should bear in mind that such biorecognition may be biased towards



Scheme 1 Different levels of glycan analysis include compositional and detailed glycan structure, glycan affinity and specificity, glycoform profiling, site-specific analysis and 3D structural and topological studies. Moreover, determination of carbohydrate-binding protein (CBP) structures and characterization of glycan–CBP recognition and complex formation are required, particularly in biological contexts. Advanced glyco-informatic resources are essential for analytical data collection, annotation, and analysis of the large-scale data generated.

particular glycoforms because of steric effects or other factors hampering interactions with other glycoforms. For its part, SDS-PAGE does not provide an accurate molecular weight determination as separation is governed by the hydrodynamic volume of the migrating species. One alternative technique is gel-based isoelectric focusing (IEF) (Fig. 2). This technique gained much momentum in the early days of proteomics as part of the two-dimensional gel-electrophoretic sample preparation and provides a rough charge distribution of the glycoprotein. Given the limited number of charged modifications of amino acid

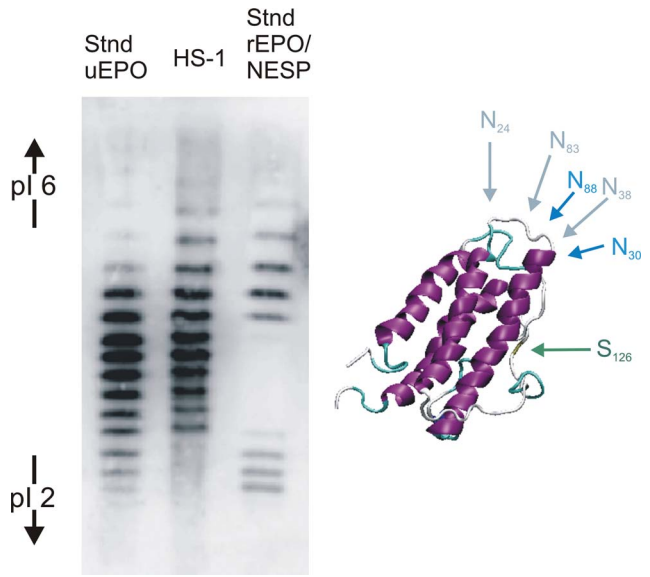


Fig. 2 Isoelectric focusing profiles of an endogenous erythropoietin (EPO) standard (left lane), EPO from a human urinary sample (center lane) and a recombinant erythropoietin mixture (right lane) composed of Eprex (3 *N*-glycans and migrating just below pI 6) and Darbepoetin alpha (5 *N*-glycans and migrating just above pI 2). On the right the crystal structure of erythropoietin (1BUY) and its glycosylation sites are shown.

residues, the IEF profile usually provides a reliable sketch of the degree of sialylation, sulfation, phosphorylation and/or glucuronidation of the glycoprotein.

Similar information can also be obtained from the mass spectrum of the intact glycoprotein. Matrix Assisted Laser Desorption Ionization Time Of Flight (MALDI-TOF) mass spectrometry is particularly suited for this purpose, being capable of handling complex mixtures and fairly tolerant of impurities, aside from detergents, which produce significant ion suppression.[38] Fig. 3 compares the MALDI-TOF spectra of a glycoprotein (rAT-III) and a non-glycosylated protein (rGH) and shows how the peak width provides information on the heterogeneity of the protein, and the peak number on the prevalent glycoforms. Depending on the purity of the glycoprotein, on its structural complexity, and on instrument resolution, information on microheterogeneity can be quite exhaustive.

A subsequent step in the structural interrogation, still at the glycoprotein level, concerns the evaluation of the monosaccharide residues present. One may distinguish different levels of analysis, all requiring the chemical hydrolysis of the glycoprotein. Typically, a first level of analysis addresses sialic acid (Sia) residues. The relevance of Sia was acknowledged nearly six decades ago[39] and, at the time, a specific colorimetric protocol named Bial's reaction and based on orcinol was employed for its detection. With time, the number of residues in the Sia family has increased and currently more than 50 structurally different sialic acid residues,[40] with a variety of associated functions,[41] are known. Even though the analysis of Sia has been pursued through many different approaches, it was selective conjugation of the released α-keto-acids with *ortho*-diamines to form quinoxaline derivatives that allowed both sensitive and specific analysis by liquid chromatography of this family of compounds. In particular, 1,2-diamino-4,5-methylenedioxybenzene (DMB) has found widespread use due to its fluorescent properties (ex: 373 nm, em: 448 nm),[42] which grants the protocol a yet unmatched sensitivity, and also because the mild acid conditions required to release Sia residues do not cause migration of the labile acetyl at C–O7.[43] For example, the detection of *N*-glycolyl neuraminic acid in erythropoietin – found at picomolar concentrations in human specimens – is an unambiguous evidence for a doping violation that could be established by this protocol[44] but not with MS analysis of the same sample (personal communication). Hence, obtaining a complete picture in terms of Sia speciation will usually require the DMB protocol, though care must be exercised as, in addition to the release-related degradation, other α-keto acids or 1,2-diketones in biological samples, *e.g.*, α-ketoglutaric, pyruvic or *p*-hydroxyphenyl-pyruvic acids, can interfere. In such cases, hyphenation of liquid chromatography with fluorescence detection (LC-FLD) to mass spectrometry is possibly the best solution.

The next level of analysis, namely determining all monosaccharides present in a glycoprotein, usually requires a strong acid (*e.g.*, 1 M HCl in methanol, 65 °C, overnight, in the presence of an internal standard) to hydrolyze all glycosidic linkages – except that between the first GlcNAc residue and Asn in *N*-linked glycosylation – and convert glycosidic acetals into the corresponding methylglycosides. The procedure, however, will also irretrievably cleave most post-glycosylational modifications. Following neutralization and evaporation, free hydroxyls are further derivatized with trimethylsilyl (or analogous) functionalities for both qualitative (four characteristic peaks for each monosaccharide) and quantitative evaluation using gas chromatography flame ionization detection (GC-FID) or gas chromatography mass spectrometry (GC-MS) and comparison with a standard monosaccharide mix.[45,46] Alternatively, following acid release, monosaccharides may be separated chromatographically by high performance anion exchange chromatography (HPAEC) (CarboPack PA-100) and detected by pulsed amperometric detection (PAD).[47] The latter protocol offers the advantage of a single peak per monosaccharide and of direct analysis without derivatization, but the basic LC conditions may induce C2-epimerization in GlcNAc to yield ManNAc,[48] or peeling reactions where some monosaccharides are degraded from the reducing end. Altogether, monosaccharide analysis offers the possibility of identifying which type of glycosylation is present; Man and GalNAc being representative of *N*- and *O*-glycosylation, respectively. Furthermore, the stoichiometry of the different sugars allows an educated guess on the type of *N*-linked glycans present by considering the ratio between the distinct monosaccharides with respect to Man. A similar approach can be employed to estimate substitution profiles (*i.e.* α/β1–2,3,4,6) in glycans. In this case, carbohydrates and other functional moieties are permethylated using the Hakomori protocol,[49] subsequently the monosaccharides are released by acid hydrolysis (leaving the methyl-ether bonds intact) and the resulting hydrolyzed monomers are reduced and acetylated to give volatile, partially methylated alditol acetates

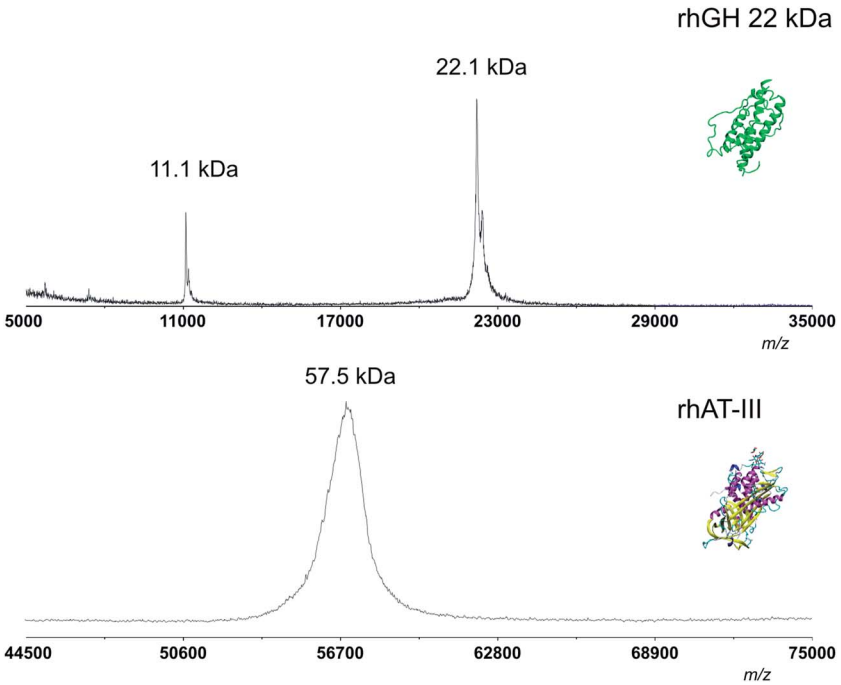


Fig. 3 Top: MALDI-TOF spectrum of recombinant human growth hormone (not glycosylated); bottom: recombinant human antithrombin-III (tri/tetra glycosylated). Microheterogeneity due to glycosylation can be clearly appreciated from the peak width.

(PMAA), again analyzed by GC-MS. This procedure provides unambiguous information on the linkage pattern as well as the ring size of the corresponding sugar, but it is important to emphasize that it is unable to distinguish between a 4-*O*-linked aldopyranose and a 5-*O*-linked aldofuranose.

A more recent development to assess glycosylation at the glycoprotein level consists in the interrogation of a particular glycoprotein or complex biological sample through a lectin array. Even though lectins have long been recognized as tools in the study of glycosylation, their systematic application in array format to detect the glycotopes in a given sample is relatively new.[50–53] Even if the information obtained cannot be compared with a thorough structural analysis (*vide infra*), it has the advantage of analyzing a crude biological sample, *e.g.* the cellular glycome,[54] without too much manipulation, and has demonstrated its value in assessing glycosylation changes in cancer cells on the basis of a direct or an antibody-assisted evanescent-field fluorescence detection scanner.[55,56] In addition, combination with antibodies allows the changes in glycosylation to be pinpointed to specific proteins,[57] adding one more level of specificity to the analysis. Hence, dynamic glycome analysis can be undertaken by means of differentially labeled CBPs.[58] Even so, there are several drawbacks to lectin arrays. For one, while current plant lectin-based arrays[32] cover most monosaccharides in the mammalian glycome, mammalian lectins would obviously provide a more representative glycoprofile. Also, one should not ignore that most lectins are promiscuous to a certain degree and that this behavior, different for each lectin and with different affinities for different sugars, will complicate glycome readout. Ultimately, it appears that lectin–carbohydrate interactions are not always straightforward and that glycoclusters, of either homo- or heterogenic nature, will strongly influence the interaction, and in this case, the analytical data.[5,59] The latter phenomenon appears to be, at least in part, responsible for the fine-tuning of biological communication processes and will as such be very difficult to interpret in terms of precise structural entities.

2.1.2. Analysis of glycoprotein glycans. Evaluation of intact glycans almost inevitably requires their release from the peptide backbone. While some high resolution approaches, in particular those based on MS, are capable of addressing microheterogeneity at the intact glycoprotein level, this is restricted to those entities with a very limited number of glycans and glycoforms such as apolipoprotein C3.[60] For more complex entities, separation of carbohydrate from the protein backbone is needed. Since, in this process, both the site-specificity and the protein origin of the glycans are lost, it is crucial to ensure the maximum degree of protein purity before the procedure is initiated. Deglycosylation can be achieved by either chemical or biological means, each with their respective dis/advantages.

The most widely used chemical method is hydrazinolysis, a procedure that releases the two major classes of glycans (Fig. 4) yet requires highly skilled staff and strict conditions for success, and is invariably accompanied by side-reactions and byproducts. In addition, re-acetylation is necessary to avoid *N*-glycans being lost during the process but may also induce *O*-acetylation. Selective and sequential release of oligosaccharides is achieved by mild hydrazinolysis of *O*-linked oligosaccharides at 60 °C, followed by that of *N*-linked oligosaccharides at 95 °C, but there may be a significant overlap between both processes depending on the protein and the degree of glycosylation. In addition, the procedure will destroy the protein backbone so that if both

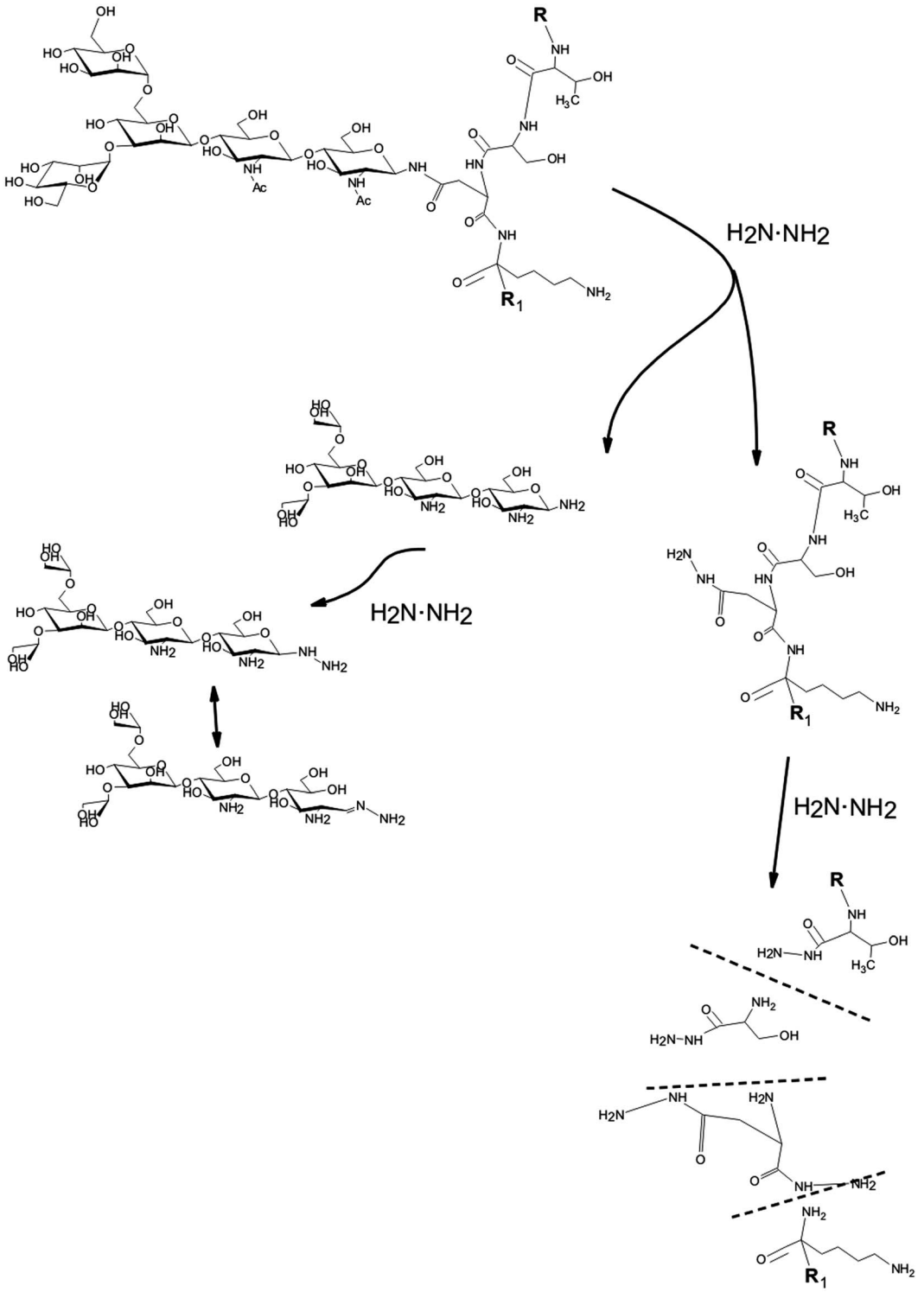


Fig. 4 Release of carbohydrates from the protein backbone following hydrazine treatment. An undesired side effect is protein destruction as indicated in the bottom right.

glycans and protein sequence are to be investigated, hydrazinolysis is not the method of choice. Another chemical procedure, *i.e.*, alkaline β-elimination (0.05 to 0.1 M NaOH or KOH, 60 °C, 12 h), can be applied for *O*-linked carbohydrates attached to Ser or Thr (except those at the carboxy-terminus), but not to Tyr, hydroxy-Pro or hydroxy-Lys. In this case, *N*-linked carbohydrates are unaffected. To prevent base-catalyzed peeling (*vide supra*), sugars must be immediately captured,[61] the alkaline solution carefully prepared,[62] or a reducing agent (*e.g.* 1 M $NaBH_4$) added which forms an alditol that precludes

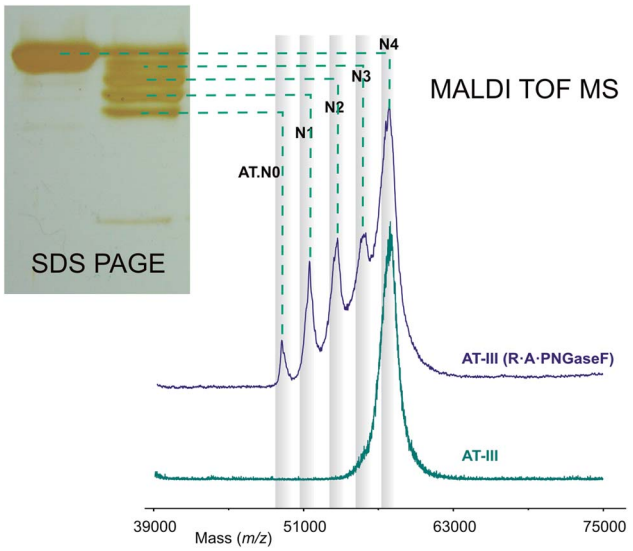


Fig. 5 SDS-PAGE and MALDI-TOF-MS analysis of human antithrombin III (AT-III) after conventional PNGaseF de-*N*-glycosylation. From the number of bands in the SDS-PAGE and peaks in the MALDI-TOF spectrum it is evident that deglycosylation is inefficient and does not reach completion.

reducing-end derivatization. If tagging is intended, it is best performed during release.[63] As with hydrazinolysis, the protein backbone is destroyed in the process.

The only strategy that preserves both protein and carbohydrate is enzymatic deglycosylation, which has been successfully developed for *N*-linked glycans using several endoglycosidases. For mammalian glycoproteins, peptide-N4-(*N*-acetyl-beta-glucosaminyl)asparagine amidase (PNGase F) is the enzyme of choice; it liberates nearly all *N*-linked carbohydrates under standard conditions (*e.g.*, phosphate buffer, 50 mM pH 7.3, 16 h, 37 °C). Nevertheless, the efficiency of this procedure needs to be checked to ensure a correct assessment of the subsequent analysis. One example where conventional conditions do not result in full release is human antithrombin III (AT-III) (Fig. 5), where only partial release of the four *N*-linked structures of the α-variant is achieved if the procedure is not optimized. For plant or invertebrate glycosylations, PNGase A is the preferred choice. In contrast to PNGase F, this enzyme, although of poorer overall efficiency, is capable of releasing α1–3-fucose-bearing core structures. Other endoglycanases (endo F1 to F3 or endo H) can be employed to release the carbohydrate chains, except for Asn-bound GlcNAc, as the enzymes specifically target chitobiose units. In sum, the full repertoire of *N*-glycans can be released by enzymatic means but caution is still advisable. Co- and post-release glycosylamine modification to functionalities other than C1-hydroxyl, such as urea,[64] glycerol[65] or thiol[66] or incomplete conversion in the presence of ammonium may obscure the final analysis. For *O*-glycans, one single deglycosylation enzyme has been identified thus far,[67] and its activity is restricted to T- and Tn-antigenic structures on Ser or Thr. As such, its main application is in unveiling the presence of either epitope, without providing further evidence on the presence of other *O*-glycans.

Once the release from the protein has been completed, the carbohydrates must be purified from the protein and buffer components prior to analysis. Separation into simpler glycan mixtures, a discipline in itself, can significantly facilitate subsequent analyses, but loss of low-abundance glycans may inevitably bias structural identification. On average, every separation step may cause a 10–50% loss of starting material. Generally, separation of the mixture is done by filtration, CE,[23,68] HPAEC,[26] or high-performance liquid chromatography (HPLC). The latter is one of the most versatile, as separation can be based on charge (weak or strong cation or anion exchange), hydrophobicity[69] or hydrophilic interaction (HILIC), and can be performed on either conventional, micro- or nanosized platforms.[25] Separation is typically performed on normal-phase (NP), but reverse-phase (RP) analysis is also possible after permethylation, as demonstrated recently in a comparative study of RP-LC-electrospray (ESI-MS), RP-LC-MALDI-MS, and MALDI-MS[70] using model proteins as well as human blood serum. This study concluded that, for complex samples such as serum, RP-LC-ESI-MS yielded the confident detection of more and lower-abundance glycans, and also permitted the separation of several structural isomers. Another type of derivatization, *i.e.*, selective incorporation of a reporter group at the reducing end of every glycan, is one major step forward in the field of carbohydrate profiling. Research at the Oxford Glycobiology Institute pioneered this approach for comprehensive glycosylation analysis when the starting material is scarce.[71,72] In this approximation, carbohydrates are labeled with fluorescent 2-amino benzamide (2AB), profiled by both weak anion exchange (WAX) and NP (nowadays HILIC) HPLC, and elution times standardized against a partial acid hydrolysate of 2AB-labeled dextran. The resulting glucose unit (GU) values[73,74] allow a preliminary assignment that can then be corroborated by targeted and sequential exoglycosidase digestions, followed by another round of HPLC profiling. Subsequent glycan trimming is of particular interest as not only does it provide confirmation of structural assignments, it simplifies the glycan pool, ultimately contributing to unveiling epitopes that are obscured in the overall microheterogeneity.[75] Fluorescence labeling at the reducing end is not restricted to 2AB as several other tags are described[76] and not only reduces sample requirement to the low femtomole level of individual structures,[77] it also allows accurate relative and absolute quantitation of the glycans present in a given glycoprotein.[78] It has become one of the standard techniques in carbohydrate profiling,[71,79] which can be amplified with internal standards if a different fluorescent tag is used for dextran and the sample, and can be easily extended with back-end MS evaluation when the material is not required for further exoglycosidase treatment.[80] Despite these advantages, it is a laborious approach that requires considerable care, especially during the 2AB labeling (in 30% acetic acid, 65 °C) to avoid desialylation that may confound structural assignments.[78] Automated sample preparation, *i.e.* both the fluorescence labeling and the post-release and post-labeling purification steps, greatly reduces analysis variability, providing robust and reliable glycomics data.[16,19,81]

Arguably one of the more powerful and versatile analytical techniques for all sorts of compounds, including carbohydrates, MS has become the cutting-edge technology for glycomics, linking mass with composition and providing precise characterization of complex structures. A wide range of MS equipments are available for glycan analysis. The introduction of MALDI-TOF instruments allowed rapid and straightforward evaluation of complex mixtures[82,83] and was a giant leap forward in MS evaluation of carbohydrates, hitherto restricted to the cumbersome, low-sensitivity fast atom bombardment mass spectrometry (FAB-MS). Improved analyses are made possible through combination with well-known derivatization strategies (*e.g.* permethylation or peracetylation) that reduce polarity and improve sensitivity by either MALDI-TOF-MS[84] or LC-ESI-MS.[69] Isotope-based differential derivatization protocols, *e.g.*, using CH_3I and CD_3I, allow exact determination of the number of free hydroxyls in a given structure, from which valuable information on carbohydrate composition can be inferred.[85–87] When glycan sequencing is the goal, analyses must include tandem MS experiments where structure-revealing ions are obtained by a combination of ion activation/fragmentation strategies such as collision-induced dissociation (CID), electron transfer dissociation (ETD) or electron capture dissociation (ECD). For instance, a recent study involving a series of oligosaccharide-derived oxonium fragment ions generated by CID enabled simultaneous characterization of IgG glycoforms at both Fc and Fab glycosylation sites by combining multiple reaction monitoring (MRM) MS with energy-resolved structural analysis.[88] In another study, CID-MS was employed to selectively monitor the generation of a *m*/*z* 284.053 fragment, consistent with GlcNAc phosphorylation in a mouse brain dataset, and unveiled this new post-glycosylational modification.[89] Within the last few years, ECD and ETD have enabled the assignment of *O*-GlcNAc sites at the proteomic scale and greatly facilitated protein-specific studies of single *O*-GlcNAc events. Particularly, ETD has been used to identify *O*-GlcNAc sites and PTMs such as phosphorylation and Arg methylation, on host cell factor C1 (HCF-1), a chromatin-associated protein involved in transcriptional regulation and cell proliferation, and one of the most highly *O*-GlcNAc-rich proteins found in cells.[90] Although MS is clearly indispensable in glycomic analysis, some techniques still present limitations such as susceptibility to salts, difficult assignment of isomeric and isobaric monosaccharides – even though the evolution of ion-mobility strategies are addressing this –,[91] complicated behaviour of acyl groups on glycans, and ionization efficiency dependence. Moreover, interpretation of MS^n fragmentation datasets remains a limiting factor with regard to throughput, user-dependent variability in discrimination and/or interpretation and complete identification of all glycoforms.

When sample complexity is limited to only a few glycan structures, the analytical technique of choice is nuclear magnetic resonance (NMR), the only one providing both qualitative and quantitative information on the glycan without being destructive. While mono-dimensional (^1H, ^{15}N, ^{13}C) experiments readily provide information on structural reporter groups,[92,93] multi-dimensional, both homo- and heteronuclear experiments yield information on the spatial orientation of the glycotope. Moreover, NMR may provide unambiguous information on the presence and position of post-glycosylational modifications such as sulfation, methylation, acetylation or phosphorylation.[94] Ironically for a technique that had proven crucial in the early development of the carbohydrate field, NMR had gradually lost influence due to the often prohibitive amounts of natural material required. Nevertheless, recent developments enabling analysis of picomoles[95] may reinvigorate a technique which in fact has never lost its appeal for the analysis of carbohydrate biological interactions[96,97] or the effect of changes in glycosylation.[98,99]

Altogether glycomics can be addressed through a variety of strategies and technologies that turn out to be orthogonal rather than parallel. While all of them rapidly generate very large amounts of data, differences between platforms can turn data analysis into a complex, time-consuming task requiring bio-informatics tools and databases to facilitate data processing and interpretation. Most of these glycoinformatic tools have particular focuses, *e.g.*, data from HPLC,[74] MS,[100] NMR or microarray[101,102] experiments. Initiatives for cross-linking data from different techniques and integrating multiple datasets are prospering and extremely useful,[103,104] although in the use of database search outputs critical interrogation is advisable. Additionally, the field of glycobiology would greatly benefit from a single glycan structural annotation, easy and of worldwide access, and supported by public agencies such as NCBI or EMBL. Limited public initiatives in this regard (*e.g.* Consortium for Functional Glycomics – http://www.functionalglycomics.org/static/consortium/consortium.shtml) are at risk of being overshadowed by commercial enterprises (*e.g.* Waters & NIBRT – http://www.waters.com/waters/promotionDetail.htm?id=134654015), most likely with ensuing limitations in accessing data, let alone seeking a universal output.

2.1.3. Analysis of site-specific glycosylation. With the increasing awareness of the importance of site specific glycosylation much effort is invested in addressing the glycoforms, enrichment of glycopeptides, and evaluating glycans at their site of attachment.[36,37] The prerequisite of preserving the peptide backbone eliminates the possibility of quantitative glycan profiling through the 2AB protocol, or any other procedure involving tagging of the reducing end. Furthermore, the analytical strategy is limited to MS as the only technique capable of differentiating peptide and carbohydrate sequences. However, a main drawback of glycopeptide MS analysis is that glycosidic bonds are less stable than amide bonds, so that predominant cleavage of the former leads to deglycosylated peptides with no information on the attachment site. The problem has been solved by simply varying the collision energy, so that fragmentation is selectively directed to either carbohydrate or peptide, and information on either part is obtained.[105] Another useful approach, requiring as above no hardware modification, is switching between high and low cone voltage during the LC-MS analysis. Whereas high voltage promotes glycan fragmentation, low voltage produces intact glycopeptides that are identified through

accurate mass measurements and signal intensity.[106] When applied to complex mixtures, deconvolution of the data is of the utmost importance for precise identification and quantification of singular glycopeptides. A significant advancement in the analysis of labile posttranslational modifications, including glycans, has resulted from the implementation of ECD or ETD. In these experiments, electron transfer from a radical anion to the peptide backbone results in preferential cleavage of the N–Cα bond, hence preserving the modification and allowing reliable analysis of both permanent[107] and transient glycosylation.[90] This high-accuracy mass spectrometric characterization combined with a strategy based on "filter aided sample preparation" (FASP) technology and multi-lectin affinity enrichment recently allowed the characterization of more than 5500 new glycosylation sites, confirming 74% of known sites in different mouse tissues and revealing their topological organization.[108] Still another strategy, named "in-gel non-specific proteolysis for elucidating glycoproteins" (INPEG), includes gel-based separation and subsequent digestion with a protease cocktail. With the reduced sample complexity afforded by SDS-PAGE and the help of a software package (Glycopeptide Finder), complex samples such as crude bovine milk or human serum can be evaluated.[109] It seems clear that standardized analysis protocols[79,110] as well as dedicated software applications[100] will be necessary to accurately and reproducibly assess glycosylation at the glycopeptide level, and to extract biologically relevant conclusions, *e.g.*, differentiation between hepatic and liver cell-surface gamma-glutamyl peptidases,[111] site-specific alteration of haptoglobin glycosylation related to hepatocarcinoma and liver cirrhosis,[112] or how a particular congenital glycosylation disorder (CDG-Id) is associated with site-specific glycan deficiencies.[113]

3. Functional analysis of glycans

The chemical and biological diversity of carbohydrates gives rise to a structural complexity that underlies their functional variety. Thus, glycosylation is not only important for protein folding and stability[114,115] but also plays important roles in various biological processes and recognition events (Fig. 6). These roles may be unrelated to the close structural environment where glycosylation occurs or, in contrast, very stringent in terms of glycotope structure and protein localization. Also, the functions exerted are very diverse including: (i) structural, organizational and stabilizing roles, (ii) protective or barrier functions, (iii) provision of specific receptors for microorganisms, toxins or antibodies to attack, shield or lure, (iv) modulation of protein functions in a glycosylation-dependent manner, (v) intra- and intercellular trafficking roles, and (vi) mediation of cell–matrix or cell–cell interactions.[116,117] Therefore, no particular function can or should be attributed to a given oligosaccharide, so that general statements on the subject are practically impossible. The only common general principle emerging from the numerous functions is that glycans generate important functional diversity required for the development, differentiation, and crosstalk in complex organisms as well as for their interactions with other organisms in the environment.

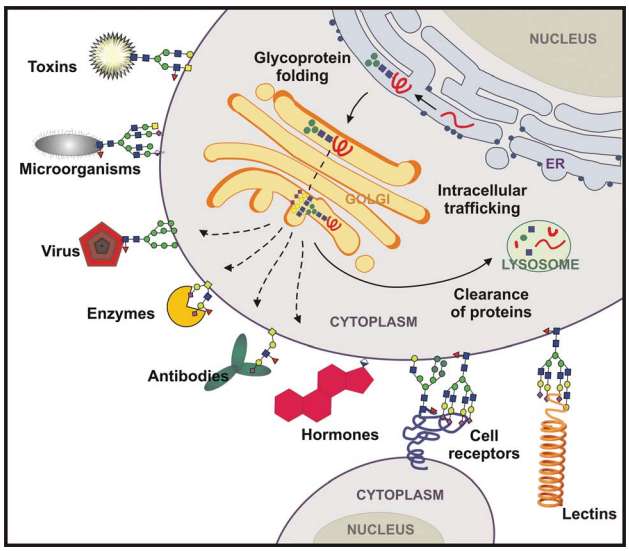


Fig. 6 Glycans participate in multiple mechanisms of cellular regulation. The general functions of glycans span from nascent protein folding and intracellular trafficking to roles in extracellular compartments such as cell–cell communication, providing specific receptors for noxious agents, protecting from microorganisms and antibodies or regulating myriad receptor–ligand interactions.

In the following sections, functions attributed to carbohydrates are reviewed through studies going from the smallest entity to larger glycosidic structures and finally including post-glycosylational modifications (see Fig. 1 and 7).

3.1. Glycosyltransferases

The majority of proteins synthesized in the rough ER undergo glycosylation and the carbohydrate chains attached to these target proteins serve a variety of structural and functional roles in membrane-anchored and secreted proteins. Glycosylation increases proteome diversity, because almost every aspect of glycosylation can be modified, including glycan composition, structure, bond and length.

The cellular glycome assembly *i.e.* the biosynthesis of disaccharides, oligosaccharides and polysaccharides, involves the action of hundreds of different glycosyltransferases (GTs), the enzymes that catalyze the regio- and stereospecific transfer of sugar moieties from activated donor molecules to a variety of acceptor biomolecules including glycans, lipids, peptides, and small molecules forming glycosidic bonds.[118] The complex glycans synthesized by these mammalian GTs are known to play crucial roles in cell–cell, cell–matrix and cell–pathogen interactions, which impact growth and development, infection and immunity, signaling, malignancy, and metabolic disorders. For instance, congenital disorders of glycosylation (CDG) are genetic diseases causing defects in the synthesis or the attachment of the glycan moiety of glycoproteins and glycolipids. Of the more than 40 CDG reported in humans, some 80% affect the nervous system and no effective treatment is known for any of these disorders.

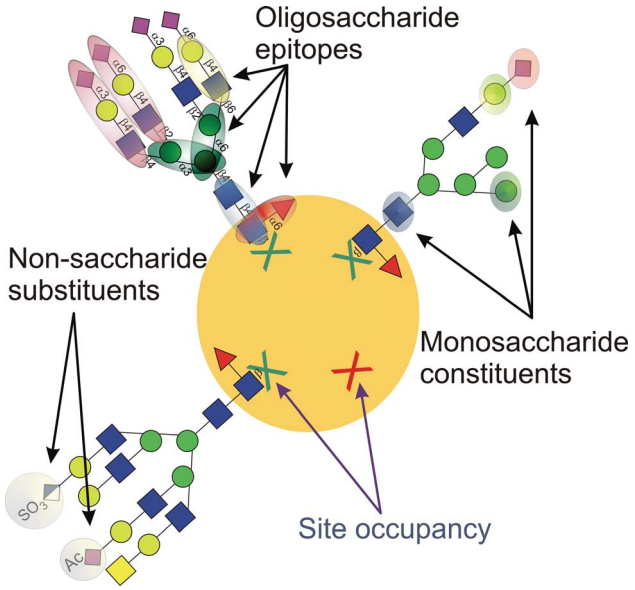


Fig. 7 Different levels at which carbohydrates contribute to glycoconjugate heterogeneity: *i.e.* by occupancy, the monosaccharides that build-up the structure, the specific epitopes composed of the monosaccharides, and ultimately, the non-carbohydrate substituents.

Given their importance in both normal development and pathological conditions, GTs are targets for inhibition and specific small-molecule inhibitors have long been sought to manipulate their activity in cells and to determine the functional roles of glycans. Although recent, structural, kinetic and inhibitor studies have provided important information about the evolution and reaction mechanism of GTs,[119] virtually nothing is known about their donor and acceptor specificity. Therefore, functional characterization remains the greatest challenge in the GT field as there is presently no easy way to assign functions to the many uncharacterized GTs.

3.2. Carbohydrate determinants (glycotopes)

While the complexity and diversity of the totality of glycan structures in an organism is almost impossible to calculate, some 7000 glycan determinants (glycotopes) recognized by CBPs including lectins, receptors, toxins, antibodies, and enzymes have been reckoned for the human glycome.[120] This value is probably underestimated but it provides an idea of the dimension generated by the approximately 700 proteins that make up the mammalian glycan repertoire, and sets the boundaries for glycan–CBP interaction studies[121] where the use of lectins, receptors, antibodies, enzymes, and glycan microarray technologies is crucial for elucidating carbohydrate-specific functions.

3.2.1. Monosaccharide constituents. In this text, "monosaccharide" refers to the simplest form of a sugar, found either as a stand-alone residue or as a terminal or internal part of a polysaccharide. Sialic acids (Sias) are a family of nine-carbon sugars typically attached to the outermost ends of glycoconjugate chains as well as on secreted glycoproteins. The high prevalence of Sias terminating glycan extensions suggests that their predominant function is modulating interactions with the environment. For example, receptor 2B4 of human natural killer (NK) cells has sialic acid residues on both *N*- and *O*-linked glycans. Removal of predicted 2B4 *N*-glycosylation sites decreases binding to its ligand CD48 suggesting that *N*-linked sugars are essential for binding, yet sialylation of 2B4 has a negative impact on ligand binding and 2B4-mediated NK cell cytotoxicity.[122] Similarly, Sias on human corticosteroid-binding globulin (CBG) *N*-glycans were shown to modulate its function, specifically by restricting the binding of CBG to its receptor through steric and/or electrostatic means. Removal of CBG NeuAc residues, or the entire *N*-glycan, increased cAMP production significantly, which was used to evaluate the CBG–receptor interaction.[123]

O-Glycosylation of the Notch extracellular domains in epidermal growth factor (EGF)-like repeats is essential for activity, and tissue-specific alterations in the glycan structures are known to regulate activity. As such, *O*-fucose and *O*-glucose-initiated glycans modulate Notch signaling events critical to cell fate determination and tissue development. More specifically, *O*-fucose-initiated glycans modulate the strength of Notch binding to DSL Notch ligands, while *O*-glucose-initiated glycans facilitate juxta-membrane cleavage, generating the substrate for intramembrane cleavage and Notch activation.[124,125] Moreover, increasing both sialylation and terminal α1–3-linked fucosylation in *N*-glycans could lead to suppression of EGF receptor (EGFR) dimerization and activation in lung cancer cells, which could in turn affect the metastatic ability of cancer cells, EGFR-mediated signaling, and cellular behavior. In particular, the Sia and Fuc residues in the Asn420 *N*-glycan could be critical in inhibiting EGFR dimerization and phosphorylation. In contrast, core fucosylation would promote EGFR dimerization and phosphorylation.[126]

Another prominent example of *O*-glycosylation is the intracellular modification of cytoplasmic and nuclear proteins with *O*-linked-*N*-acetylglucosamine (*O*-GlcNAc) that regulates basic and multiple cellular functions such as transcription and translation, neuronal function, nutrient sensing, cell cycle, and stress. Moreover, it is involved in the etiology of diabetes and neurodegeneration.[127] Indeed, CREB, a central transcription factor in the brain, is highly *O*-GlcNAc monoglycosylated in neurons and influences gene expression by inhibiting both basal and activity-induced CREB-mediated transcription, neuronal function regulation and long-term memory.[128] One of the earliest examples of *O*-GlcNAc modification was found over 25 years ago in nuclear proteins,[129] and since then numerous studies have suggested the existence of dynamic interaction networks, whereby *O*-GlcNAc simultaneously senses and modulates metabolic flow through essential pathways. For instance, histones are modified with *O*-GlcNAc within the nucleosomal core *in vivo*. In particular, histone H2B is GlcNAcylated at S112, and this PTM facilitates K120 monoubiquitination, presumably for transcriptional activation and is responsive to serum glucose levels and/or cellular energy states in certain cell types.[130] Moreover, histone *O*-GlcNAcylation levels change during mitosis and with heat shock showing that *O*-GlcNAc cycles dynamically on histones and can be considered

part of the histone code.[131] This modification is not confined to the nuclear environment as demonstrated by the dynamic induction of *O*-GlcNAc at Ser529 of phosphofructokinase 1 (PFK1) in response to hypoxia. Here the modification inhibits PFK1 activity and redirects the glucose-flux from glycolysis through the pentose phosphate pathway (PPP), thereby conferring a selective growth advantage to cancer cells. This was confirmed by blocking glycosylation of PFK1 in cancer cells resulting in reduced proliferation *in vitro* and impaired tumor formation *in vivo*.[132] Extracellular *O*-GlcNAcylation of secreted and membrane glycoproteins also occurs and mediates cell–cell or cell–matrix interactions at the cell surface.[133] Several recent reviews on *O*-GlcNAcylation have been published providing more details and studies on different aspects of this PTM[134,135] but it is certainly worth mentioning that modulation of these cellular processes by *O*-GlcNAcylation involves a very extensive cross-talk with phosphorylation[136] and that combinations of both, *i.e. O*-GlcNAc-6-phosphate have been proposed recently as a novel PTM of mammalian proteins with a variety of possible cellular functions.[89]

3.2.2. Oligosaccharides. As described above, glycans are mostly constituted by multiple monosaccharides and biological activity may be traced to single building blocks. However, with time, evidence has accumulated that in carbohydrate-mediated interactions larger entities (di to hexasaccharides) add yet another level of complexity. Thus, CBPs may recognize complex and relatively large structures that may be either linear or branched homo- or heteropolymeric in nature. One of the very first examples in this context are the ABO(H) major blood group antigens,[137,138] where the absence (O) or presence of an α-Gal (B) or α-GalNAc (A) on Fuc(α1–2)Gal is of paramount importance. Of similar size is the Sd^a-antigen, comprising a Neu5Ac(α2–3)[GalNAc(β1–4)]Gal(β1–R) trisaccharide, expressed in a donor-specific manner in males, and with no particular function hitherto attributed.[139] More recently, this glycotope has been coined as a potential biomarker for colon cancer and its absence related to downregulation of β-1,4-*N*-acetylgalactosaminyltransferase II (β4GalNAcT-II).[140] In close relationship to Sd^a-downregulation stands upregulation of sialyl Lewis^x expression, as α-1,3-fucosyltransferase activity directly competes with β4GalNAcT-II for the acceptor substrate. The Lewis type carbohydrate sequences (Lewis^a, Lewis^b, Lewis^x, Lewis^y, sulfo-Lewis^a, and pseudo-Lewis^y antigens) are expressed on many human glycoproteins and have been assigned a myriad of functions. Just to cite a recent description, terminal Lewis^x and Lewis^y antigens have been reported to be abundantly expressed on *N*-glycans in human seminal plasma glycoproteins and to bind specifically with the lectin domains of DC-SIGN in both male and female to maintain immune homeostasis.[141] The sialyl Lewis^x moiety is also of utmost importance in the interaction between P-selectin glycoprotein ligand 1 (PSGL-1) and P-selectin during the initial phases of inflammatory response.[142] While this interaction is promoted by the *N*-glycan in PSGL-1, in combination with upstream tyrosine sulfation, P-selectin itself is also functionally glycosylated. On a broader scope, the specific *N*-glycosylation status of a particular endothelial adhesion molecule (P/E-selectins, ICAM-1, VCAM-1, or PECAM-1) may regulate protein function during inflammation, affecting both leukocyte capturing and endothelial signalling functions. Adhesion molecule *N*-glycosylation is a dynamic process regulated during inflammation by mechanisms that operate in parallel, but independent of up-regulation of protein expression, and only under those conditions where the appropriate adhesion molecule protein and the corresponding *N*-glycan are expressed will efficient leukocyte adhesion be achieved.[143] For example, the presence of polysialic acid, long chains of α2–8-linked sialic acid residues, on neural cell adhesion molecules (NCAMs) has been demonstrated to decrease cell adhesion, and it is critical for a variety of processes including brain development; synaptic plasticity; axon guidance and path-finding; neurite outgrowth; and general cell migration.[144] Another unique carbohydrate structure characteristically expressed on a series of cell adhesion molecules (L1, myelin associated glycoprotein, TAG1, P0, *etc.*) is the human natural killer (HNK-1) epitope. Initially targeted by an antibody raised to natural killer cells, the epitope was soon recognized to consist of a sulfated trisaccharide, SO_4-3GlcA(β1–3)Gal(β1–4)GlcNAc(β1–R), that is expressed in a spatio-temporally regulated manner during the development and maintenance of the peripheral nervous system. Particularly, the single glycan moiety contained in P0 plays an important role in cell–cell adhesion.[82]

Finally, in the phenomenon of carbohydrate-mediated biological recognition, an extra level of complexity can be added when the carbohydrate binding event is potentiated by a multivalent expression of glycotopes that result in stronger CBP recognition. This phenomenon has been extensively studied under laboratory conditions,[59,145,146] much less in biological settings. A clear example is a recent study on the requirements for neuronal interactions and subsequent axon growth, where clustered presentation of *N*-glycans with *N*-acetyllactosamine (LacNAc) epitopes at branch ends of neural cell adhesion molecule L1 is required for neuronal galectin-4/L1 binding. Impairing the maturation of these epitopes precludes Gal-4/L1 association resulting in a failure of L1 membrane cluster organization, required for proper axon growth.[147]

The above mentioned examples are merely a glimpse of recent descriptive studies and illustrate the increasing relevance of glycotopes. One hopes that the information flow will grow exponentially to meet the vast challenge posed by glycomics and to establish a comprehensive functional appreciation of the human glycome.

3.3. Glycosylation site occupancy

Glycosylation impacts significantly on the physico-chemical properties of the glycoprotein and may thereby exert influence on its viability and activity. These effects are apparently independent of the structural modification but the modification *per se* is necessary. One of the better known examples in this respect is the folding of the nascent polypeptide chain where the monoglucosylated oligomannose structure serves as an anchor point for the chaperone-assisted event.[1] Persistent failure to fold properly, possibly due to the absence of the carbohydrate chain, ultimately results in lysosomal targeting. The common

approach to site occupancy issues involves studying the biological role after (enzymatic or chemical) glycan removal, or upon inhibition of glycosylation, alteration of oligosaccharide processing, or elimination of specific glycosylation sites. The consequences of altering, decreasing or abrogating glycan site occupancy are variable and unpredictable, ranging from nearly undetectable to decreased protein function, production level, stability, or even complete loss of function; in tune with this, the functional interpretation of the absence/defects in glycosylation is not always straightforward.

Several recent reports stress the need of glycosylation for viability; for example, in human chymotrypsin C (CTRC) it is required for efficient folding and secretion. Elimination of *N*-glycosylation by mutation of the single glycan (N52S) reduced CTRC secretion by about 10-fold.[148] Similarly, OATP1B1, an organic anion transporting polypeptide expressed in the human liver and containing three *N*-linked glycans, underwent dramatically decreased expression and was retained within the endoplasmic reticulum when all three sites were mutated to Gln.[149] A further example is BRI2, a type-II transmembrane protein where inhibition of its single *N*-glycosylation reduced cell surface trafficking and led to intracellular accumulation, although the mutation did not affect cleavage by furin or ADAM10.[150]

For glycoproteins whose glycosylation is relevant for functionality, different functional levels can be attained, as shown by numerous reports in recent literature. For instance, blocking the glycosylation of the hepatocyte growth factor receptor (c-Met), a transmembrane tyrosine kinase, attenuates c-Met function through inhibiting its cell membrane targeting.[151] CREB-H, a liver-abundant bZIP transcription factor, requires *N*-glycosylation at three sites in its luminal C-terminal domain for optimal activation.[152] Another example is glycoprotein KCC4, a K^+Cl^- co-transporter isoform involved in maintaining protein stability, regulation of cell volume, anchorage-independent cell growth, tumor formation, and lung colonization by tumors. Deglycosylated KCC4 forms decrease tumor formation and lung colonization in mice. Also, site-directed mutagenesis on the four putative *N*-glycosylation sites established that KCC4 localization to the cell surface depends on the central N331 and N344 sites.[153] This example serves to introduce a next level of functionality, namely when glycosylation of one or more, but not all, sites is required for proper functioning. While this type of study is much more informative, it is also harder to perform as multiple mutant strains must be produced or, alternatively, selective deglycosylation must be achieved. An example can be found in human acetylcholinesterase ($AChE_T$), with three putative *N*-glycosylation sites that are very important for maintaining the catalytically active conformation. Mutants $AChE_T^{N381Q}$, $AChE_T^{N495Q}$ and $AChE_T^{N296Q/N381Q/N495Q}$, particularly the former, showed a dramatic decrease in enzymatic activity compared with $AChE_T^{WT}$. In contrast, glycan removal did not change the sedimentation properties or proportions of AChE, indicating that *N*-linked glycosylation does not affect oligomerization.[154] Similarly, human serum carnosinase CN-1, involved in diabetic nephropathy, contains three potential *N*-glycosylation sites which, if deleted, result in impaired protein secretion; enzyme activity, for its part, is already reduced when two sites are deleted.[155] Finally, myeloperoxidase, a lysosomal protein of neutrophils with five *N*-glycans (N323, N355, N391, N483, and N729), undergoes significant loss of activity upon deglycosylation at N355.[156]

The *N*-glycosylation cases described above need to be completed with a few equally important examples of *O*-glycosylation. In addition to the well-established protective role of *O*-glycosylation in mucins, several more specific functions have been recently discovered. Thus, both regulated and aberrant glycosylation modulate the electrical signaling of the I_{Ks} channel, a macromolecular complex composed of a pore-forming α-(KCNQ1)-subunit and a modulatory β-(KCNE1)-subunit that is crucial for repolarization of the cardiac action potential.[157] Moreover, *O*-glycosylation at Thr-7 in the KCNE1 subunit is essential for proper biosynthesis and trafficking of the complex.[158] Similar examples are two-pore-domain potassium (K(2P)) channels, where disruption of glycosylation reduced current through decreasing the number of channels on the cell surface and hence influencing cellular depolarization.[159]

3.4. Site-specific glycosylation

As described in the previous section, the sole occupation of one or multiple glycosylation sites may affect glycoprotein functionality. It is being increasingly recognized that the selective and specific glycosylation of particular domains, among multiple potential sites, may be key in the regulation of the protein function. Many examples can already be found where glycosylation is impeded through mutations or eliminated after expression yielding a change in functioning. However, examples where glycosylation has been characterized at the structural level and the function studied are still scarce, and it is even more difficult to find structural studies at the site-specific level. In the following section, we review the state of the knowledge with several examples.

The human protein disulfide isomerase family A member 2 (PDIA2), an ER enzyme involved in protein folding and maturation, contains three *N*-glycans, one of which modulates PDIA2 homodimer formation and subsequent chaperoning activity. When devoid of carbohydrate, dimerization was highly efficient and *vice versa*.[160] The precise glycan structure of PDIA2 has not yet been elucidated but it is plausible that a decrease in glycan complexity accelerates protein folding as required. Similarly, upon investigating the role of glycosylation in E-cadherin (four *N*-glycans) and cancer, the N633 glycan is shown to be required for proper folding, trafficking, and expression whereas other glycans are related to stability of adheren junctions. Furthermore, the presence of (α1–6)-fucosylation on Asn-linked GlcNAc promotes cell–cell adhesion in both cancer and downstream signaling pathways.[161] Another study involved the melanocortin 1 receptor, the main determinant of skin pigmentation and phototype, which is *N*-glycosylated at N15 and N29. Mutagenesis and proteolytic studies showed that the N15-bound glycan was not essential while the N29-linked counterpart was crucially involved in ligand binding and normal cell surface expression.[162] In a recent paper on the tumor-associated antigen

Table 2 Examples of glycosylation changes in disease context

Protein/substrate	Alteration	Related disease	Ref.
AMPA receptor GluR2 subunit	Altered *N*-linked glycosylation suggests abnormal trafficking of AMPA receptors from the ER to the synaptic membrane	Schizophrenia	175
Amyloid-beta (Abeta) peptides	The sulfated galactose moiety of sulfatides is essential for Abeta peptide clearance. A deficiency of sulfatides in conjunction with ceramide elevation is associated with AD pathology and is present by the very earliest clinical stage of AD	Alzheimer's disease (AD)	176 and 177
Haptoglobin (Hp)	Unusual hyper-fucosylated site specific glycoforms of Hp	Liver cirrhosis and hepatocellular carcinoma (HCC)	112
Heparan sulfate (HS)	*N*-Sulfation and 2-*O*-sulfation *vs.* lipoprotein binding. Binding and uptake of lipoproteins depends on the degree of sulfation of the chains. Clearance appears to depend on *N*-sulfation based on loss of inhibitory activity of *N*-desulfated	Hepatic clearance of triglyceride-rich lipoproteins	178
Human serum and cerebrospinal fluid (CSF) proteins	Tetraantennary tetrasialylated glycan with a polylactosamine extension shows a 2-fold increase in patient sera. Triantennary trisialylated glycan containing the sLe^x epitope is significantly increased. Levels of bisecting and sialylated glycans in the cerebrospinal fluid show a general downregulation	Schizophrenia	179
Leukemia inhibitory factor (LIF)	Mannose phosphorylation of LIF mediates its internalization thereby reducing extracellular levels and stimulating embryonic stem cell differentiation	Leukemia	180
Lymphoblasts, glycoproteins and gangliosides	Enhanced expression of 9-*O*-acetylated sialoglycoproteins and 9-*O*-acetylated disialoganglioside on lymphoblasts	Childhood acute lymphoblastic leukemia (ALL)	181
Mucosal addressin cell adhesion molecule 1 (MAdCAM-1)	Sulfation of MAdCAM-1 protein with L-selectin ligand carbohydrates (6-sulfo sialyl Lewis X-capped *O*-glycans) regulates UC disease activity	Ulcerative colitis (UC) disease	182
Sialyl-Le(x)-positive mucins	Decrease of *O*-acetylation contributes to colon carcinoma-associated overexpression of sialyl-Le^x	Colorectal carcinoma	183
Sulfated mucins	Cystic fibrosis mucins contain a higher proportion of sialylated and sulfated *O*-glycans compared with non-pathogenic mucins	Cystic fibrosis (CF)	83
Thyroglobulin antibody (TgAb)	HT patients have significantly lower core fucose content on TgAb. Increasing trend of sialylation was found in PTC sera. In all patients, the sialic acid content and TgAb IgG levels showed negative correlation	Thyroid diseases: Hashimoto's thyroiditis (HT), Graves' disease (GD), papillary thyroid carcinoma (PTC), and PTC with histological lymphocytic thyroiditis (PTC-T)	184
α-Dystroglycan (α-DG)	*O*-Mannosyl phosphorylation of α-DG plays critical roles in the pathogenesis of dystroglycanopathy and is a key determinant of α-DG functional expression as a laminin receptor in normal tissues and cells. T192 → M mutation caused deficiencies in α-DG glycosylation and a marked reduction in its ability to bind extracellular-matrix components	Limb-girdle and congenital muscular dystrophy; and muscle–eye–brain disease	94,185 and 186

CD147 (*N*-glycans at N44, N152, and N186), enzymatic deglycosylation and permethylation followed by high-resolution MS analysis revealed the presence of Man_3 to Man_7 structures and barely processed bi-antennary *N*-glycans in which core-fucosylated Man_3 accounted for ~30% of the structures. All glycans were found to stabilize tertiary and quaternary structures and to maintain the active conformation essential for CD147 activity.[163] In addition, N152 was crucial for cell-surface

expression and (β1–6)-GlcNAc (~14%) residues were crucial for translocation to the plasma membrane. These same authors speculate that elevated core-fucosylation, as in E-cadherin, combined with metastasis-associated GnT-V overexpression, could potentiate the role of CD147 in hepatocellular carcinoma cells.

Slightly more dated literature on the subject includes myelin P0 protein, involved in myelin sheet formation, which is glycosylated at a single site and whose microheterogeneity has been fully elucidated and in which the sulfated HNK-1 epitope, crucial for homophilic binding, is only a minor component.[82] Another example, thoroughly studied from both structural and functional perspectives, is human chorionic gonadotropin (hCG),[164] a heterodimeric, cysteine-knot-type glycoprotein that was the first of its kind produced for medical purposes. Both hCG subunits contain two *N*-glycans each (αN52 and αN78; βN13 and βN30) in addition to several *O*-glycans. Oligosaccharides comprise (α2–3)-monosialylated di/monoantennary complex type structures with partial core fucosylation, as well as (α2–3)-monosialylated hybrid type structures. Core fucosylation is found only in the β-subunit and both *N*-glycans are of diantennary complex type, while in the α-subunit N78 lacks hybrid type structures but are instead predominant at N52. This site-specific glycosylation is required for efficient recombination of both α and β subunits to form the active hormone.[165] One should also mention the laborious work on the glycosylation of Tamm–Horsfall glycoprotein,[166] where only one (N14) out of eight potential sites was non-occupied, and the rest exhibited remarkable diversity: the N489 site included di- and tri-charged oligosaccharides exposing, among others, the $4HSO_3$-GalNAc(β1–4)GlcNAc epitope; N251 contained only oligomannose-type chains ranging from $Man_5GlcNAc_2$ to $Man_8GlcNAc_2$, while N208 was quite heterogeneous, with multiply charged complex glycan structures terminated by sulfate groups, Sia residues, and/or the Sd^a-determinant. A final example is human erythropoietin (EPO), possibly the most extensively studied cytokine, for which full glycan profiling (of the endogenous form) is, after nearly 50 years of research effort, still incomplete. The recombinant version used as a pharmaceutical is less negatively charged despite being fully sialylated,[167] its three N-sites contain complex type tetra-antennary *N*-acetyllactosamine repeats with acetylated Sia residues.[78]

3.5. Modifications of carbohydrates

While glycosylation is unrivalled as PTM in terms of abundance, complexity, and relevance, carbohydrates themselves may be subject to yet another level of structural multifariousness. Post-glycosylational modifications (PGMs) of specific sites (mostly hydroxyl or amino groups) within the glycan chain occur after the oligomer has been assembled and include sulfation, acylation, phosphorylation, methylation or epimerization that may modulate the biological function of the carbohydrate and as such play a critical role in many normal and pathological processes.[168]

Several examples of PGM have already been mentioned above such as the sulfation of the HNK-epitope, or sulfo-Sd^a variant, which may be more common than anticipated. One prime candidate here is EPO, for which sulfation may explain the difference between the fully sialylated recombinant version and the even more charged endogenous variant,[167] but unambiguous evidence is hard to collect in view of endogenous EPO levels and current analytical sensitivity. Sulfation is not only important in glycoprotein glycans; a prominent group of carbohydrates bearing this modification are glycosaminoglycans (GAGs), unbranched polysaccharides made up of repeating disaccharide units of hexosamine and uronic acid, found on the extracellular matrix of cell surfaces and classified in different types depending on sulfation patterns. These molecules participate directly or indirectly in many different physiological processes ranging from the balance between morphogenetic protein and fibroblast growth factor signaling to maintain cartilage homeostasis[169] to axon growth inhibition after central nervous system injury by specific chondroitin sulfate-E (CS-E) sulfation motif within chondroitin sulfate proteoglycans (CSPGs);[170] activation of the antithrombin–thrombin complex by heparin to promote fibrinogen cleavage;[171] and many others.[172] Another PGM is hydroxyl acylation, particularly, *O*-acetylation of sialic acids in positions 4, 7, 8, and/or 9 that gives rise to many different variants hence communicational possibilities. For instance, 9-*O*-acetylation of Sia regulates the function of CD22 (Siglec-2) *in vivo* as an inhibitor of B cell receptor signaling. Enzymatic acetylation and deacetylation of cell surface α2–6-linked Sia residues controls B cell development, signaling, and immunological tolerance.[173] Phosphorylation, arguably one of the best studied PTMs transiently affecting protein charge, is also a PGM with particular functions in glycans. For instance, mannose-6-phosphate (M6P) is the key targeting signal for acid hydrolase precursor proteins destined for lysosome transport. The M6P tag enables recognition by the M6P receptor, and NMR analysis has revealed the role of phosphodiester-containing lysosomal enzymes in the process.[174] As shown above, the effects of altered oligosaccharides on glycoconjugate functions are highly variable and quite unpredictable, and the resulting aberrant glycome composition is often associated with specific diseases. As an illustration, Table 2 summarizes the diversity of pathological states in which altered glycosylation has been implicated.

Hitherto in this section the relevance of glycosylation as a general phenomenon, its occurrence at a specific position, the site-specific presentation of a particular epitope, and the importance of glycosylation modifications have been presented. From this, it easily follows the considerable interest in the identification of glycan profiles of particular glycoproteins, body fluids or tissues under healthy or disease conditions. For instance, since in most cancers fucosylation and sialylation levels are significantly modified, such aberrant glycan structures can become useful glyco-biomarkers.[187–191] High-throughput discovery and new analytical approaches, including those addressed to PGMs,[192] are becoming essential for unraveling the biological significance of carbohydrate modification and for developing candidate biomarkers for particular conditions. One of the driving forces in the current biomarker research is "single protein-omics", namely, elucidating the

association between disease and site-specific glycoform variants of a protein rather than full-proteome coverage. The following section summarizes the state of the art in this field.

4. Simultaneous structural and functional analysis of glycans

As described in the previous part, glycoproteins are fundamental in many important biological processes and it is quite clear that no single function can be attributed to a particular oligosaccharide. Likewise, there is no single method that routinely provides all the information required for fast and reliable identification and quantification of a particular structure, let alone its particular functionality. One must also bear in mind that, from a biological point of view, identifying the carbohydrate binding entity is as important as deciphering the cognate sugar epitope. The vast majority of structural studies conducted today are performed within the constraints dictated by either physiological or technical boundaries. In an ideal situation, the analysis of biological interactions with glycoprotein participation and directly triggering a physiological response would be performed *in situ*, in real time, and without external intervention (Fig. 8).

However, this goal is as yet unattainable and state-of-the-art approaches still require the use of chemical and/or biological labeling strategies or the analysis under *in vitro* conditions where the biological context is greatly reduced to the cellular level. In the following sections, the latest scientific achievements, focused on the biological functionality, are summarized, often also referring to analytical innovations not included hitherto in this review. Thus, henceforth we will review approaches to (i) investigating glycosylation diversity under physiological conditions, (ii) biological interactions promoted by a particular glycotope, or (iii) introducing well-defined glycosylation by biological (*vs.* chemical) means to explore its functionality.

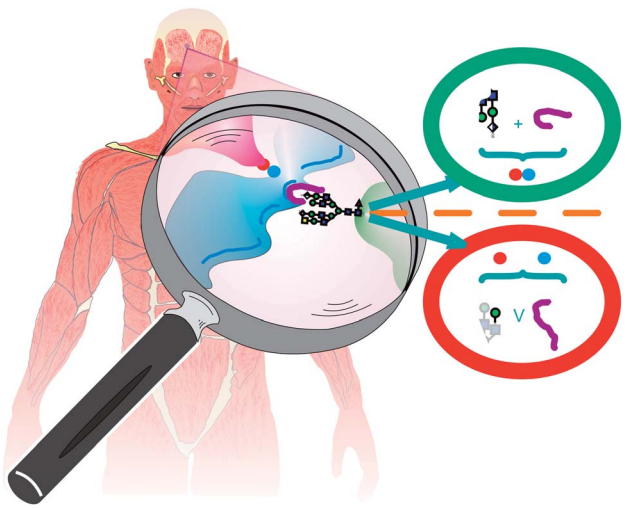


Fig. 8 An ideal analytical setup for monitoring carbohydrate-driven biological functionality. In a productive scenario (green circle), specific binding of a particular carbohydrate epitope and its cognate CBP (purple) triggers further interaction (blue and red circles). In an unproductive scenario (red circle), in contrast, the lack of a sulfated GlcA in the carbohydrate epitope or a non-matching CBP structure precludes binding and subsequent interaction.

4.1. Analysis through biomolecular interactions

There seems to be little doubt that MS-based applications, with their excellent trade-off between analytical capacity, information flow, and sensitivity, will remain a key tool in glycoproteomics studies.[28,36,37] However, these approaches have as an inherent downside the non-natural environment in which analysis takes place, and the fact that comparative analysis (*i.e.* glycosylation *vs.* pathological state or glycan interactions with different binding partners) always requires independent runs. In this regard, array-type experiments, either with a single specimen analyzed over a panel of glycans or lectins, or the reverse format where a panel of glycans or lectins is run over a single sample, constitute interesting alternatives for functional glycomics studies.

One of the classical approaches, *i.e.* use of lectin histochemistry to map the presence and localization of reactive glycan epitopes, as well as detect subtle glycosylation alterations that attend both transformation to malignancy and tumor progression in cells and tissues, has regained prominence of late.[193,194] This technique relies on the readout of biomolecular interactions with surface-exposed carbohydrates, can be employed directly on complex tissue specimens, and is of particular value in extracellular explorations. As such, it has been used exhaustively in the evaluation of gametes. For instance, experiments on fixed sections of the adult murine testis and epididymis revealed that Leydig cells react specifically with SNA or CD22 lectins, both recognizing α2–6-linked Sia; and that the same sugar, but α2–3-linked, allowed differentiation between basal (no Sia) and apical (MAA lectin staining) cells of the epididymis.[195] Also, the application of human galectins as tools for glycophenotyping has been demonstrated by a detailed comparison of their staining properties in the different layers of the *zona pellucida* extracellular matrix using confocal laser scanning microscopy.[196] In the reverse situation, carbohydrate binding molecules are fixed to a solid support and samples are interrogated.[32,54] One noticeable contribution in this context is the development of a dual-color ratiometric readout, similar to that used in gene microarrays. Briefly, two differentially labeled samples mixed in equal amounts compete for lectin binding, allowing the detection of subtle differences in glycosylation expression among many samples by comparing them with a common reference. The versatility, consistency, reproducibility and sensitivity of this approach is nicely illustrated by its application in the comparison of whole mammalian glycomes and the examination of dynamic glycosylation changes upon cell differentiation.[58] Another level of specificity can be added to lectin microarrays by combining them with protein specific antibodies in a sandwich-type approach.[57] This has been successfully applied to establish the prevalence and carriers of particular glycosylation patterns in pancreatic cancer.[56] In these assays, issues such as orientation, native multimeric quaternary

structure, clustering, and metal ion requirements influence the recognition and binding ability of lectins and should not be ignored. Other aspects to consider when working with lectin microarrays are overlapping substrate specificities, which complicate relative quantification, or the detrimental effect of washing steps on weak-to-moderate interactions. Current challenges in the lectin microarray field are expanding the lectin repertoire to include non-plant (and novel plant) lectins, developing recombinant lectins, as well as introducing affinity rather than just specificity parameters in the read-out.

Attention must also be paid to CBPs from a functional perspective. In many situations, the binding partners for particular glycans are not known and, for simplicity, are regarded as constants rather than variables in most studies.[102] In any case, CBP study has become a discipline on its own, mainly fueled by efforts from the Consortium for Functional Glycomics (CFG) (http://www.functionalglycomics.org/static/consortium/consortium.shtml). The generation of mammalian glycan arrays (currently with more than 650 structures) has greatly enabled the systematic study of carbohydrate–protein-binding interactions, despite the intrinsic limitations discussed below. The CFG has developed databases containing not only structures but also experimental data generated with their microarrays. Although microarrays do not reveal glycosylation patterns, they constitute a powerful technology to characterize CBP binding specificities.[31] As an example, both carbohydrate microarray and computational modeling approaches have been used for the rapid screening of glycosaminoglycan (GAG) interactions with proteins and multimeric protein complexes. Novel interactions between a specific sulfated epitope, CS-E, and the neurotrophin family of growth factors have been identified with these methods, providing unique molecular-level insights into the diverse biological functions of GAGs.[197] In another recent study, glycosphingolipids (GSLs) extracted from bovine brain gangliosides and fluorescently labeled were bound to a microarray for subsequent interrogation by biologically relevant CBPs like cholera toxin, antibodies and sera from individuals with Lyme disease.[198] More recently, glyconanoparticles (GNPs) in microarray format have been used to study glycan–lectin interactions. The GNPs were made by conjugating carbohydrate ligands on silica nanoparticles and microarrays were generated by conventional photocoupling chemistry. They were then probed with fluorescein-labeled lectins and with fluorescein-doped silica nanoparticles.[199] The above studies are representative of glycofunctional approaches but several aspects of glycan array design and biomolecular interaction assay should not be overlooked.[34] In addition to the limitations in oligosaccharide synthesis[200] or in glycan isolation from natural sources, aspects such as oligosaccharide density,[201] spacing, and orientation achieved upon immobilization, as well as the nature, flexibility and length of the linker are key parameters in optimizing array strategies. Indeed, since many CBPs achieve their specificity and affinity through multivalent interactions with glycans,[202] glycoarrays should aim at faithful replication of multivalent sugar display, and at capturing the physiological avidity of such interactions in as native-like fashion as possible. For instance, a bead-modified surface providing multivalency (*i.e.* the cluster effect) was used for probing carbohydrate–protein interactions mimicking a cellular environment.[203] These clusters can be of identical or mixed sugar composition to assess both valency and hetero-cluster effects.[5] Also, a new class of end-functionalized polymers mimicking the multivalent architecture of chondroitin sulfate (CS) proteoglycans have been designed, providing insights into how multivalency within and between GAG chains enhances the avidity, specificity and cooperativity of GAG–protein interactions.[204]

Carbohydrates immobilized to solid surfaces have also been employed in slightly different formats and with various other objectives, such as affinity-based systems to detect specific CBP structures or to ascertain other interaction characteristics. For instance, a novel glycan-affinity method combining proteolytic digestion of protein–glycan complexes and mass spectrometry (CREDEX-MS, "Carbohydrate REcognition Domain EXcision Mass Spectrometry") has proven useful in the structural definition of CBPs of two human galectins with lactose as the binding partner.[205]

An extremely valuable biophysical tool for carbohydrate–protein interaction studies is surface plasmon resonance (SPR). In studies aimed at detecting multiple sclerosis autoantibodies in sera, the glycopeptide antigen CSF114(Glc) was immobilized at the 3D-sensor surface and real-time specific autoantibody detection was achieved.[206] SPR is particularly valuable because it allows simultaneous monitoring of several surfaces (a limited form of arraying) and because non-specific binding can be discarded by using a non-glycosylated peptide as a reference surface. A refined version of SPR, Au nano-island based localized surface plasmon resonance, has been used to characterize the specific recognition between concavalin A (Con A) and mannose.[207] The exquisite potential of SPR for analyzing carbohydrate binding proteins was also highlighted in a study of galactose-specific *Erythrina crista-galli* agglutinin binding to several galactoside-epitopes exposed in a well-defined manner[86] at the 3D-sensor surface.[208] Results complemented those of CREDEX-MS, demonstrating that the combination of both techniques can provide good insights into CBPs in various settings. Also, apparent differences in the binding preference of carbohydrate ligands have been observed by quantitative SPR analysis, suggesting that glycan presentation and the conformational space it occupies plays an important role in binding, regardless of affinity.[209]

4.2. Glycoengineering (genetic, chemoenzymatic, chemical)

One alternative approach to assess the importance of glycosylation is selective modification of the carbohydrate decoration and study of its effects. In this context, gene targeting uses homologous recombination to change an endogenous gene. The method can be used to delete a gene, remove exons, add a gene, and introduce point mutations. Indeed, gene targeting has been widely used in glycomics research by removing ("knocking out") or adding ("knocking in") specific mutations of interest to a variety of models. The regulation of genes whose protein products are involved in glycan synthesis and glycan–

protein interaction provides insights into glycan structural diversity and function in complex biological systems.[210] Recently, a strategy for developing cell lines that produce simplified homogenous *O*-glycan structures and thus interrogating the human *O*-glycoproteome has been presented. Named 'Simple Cell', it uses zinc-finger nuclease (ZFN)-based gene targeting of COSMC gene to glycoengineer stable human cell lines displaying only truncated Tn and STn *O*-glycans. More than 100 *O*-glycoproteins and up to 350 glycosylation sites, including a previously unidentified linkage to tyrosine, were elucidated by this approach.[211] The strategy has been used to analyze the function of a single GalNAc-transferase (GalNAc-T) isoform and its role in congenital diseases and disorders.[212]

An attractive approach toward predefined glycoforms is *in vitro* chemoenzymatic glycosylation, *i.e.*, remodeling natural or recombinant glycoproteins by addition of sugar units through sequential glycosyltransferase-catalyzed reactions, or by endoglycosidase-catalyzed transglycosylation and *en bloc* transfer of pre-assembled large oligosaccharides to the protein in a single step under the catalysis of an endo-β-*N*-acetylglucosaminidase (ENGase).[213] Alternatively, site-specific glycosylation can be achieved by chemoselective ligation of proteins to appropriately tagged glycans. For instance, cysteine residues in the protein can be reacted with a thiol-reactive group pre-installed in the sugar moiety to give a disulfide or thioether-linked glycoconjugate. Other strategies involve ligation (oxime, hydrazone) between amino and carbonyl groups, or azide–alkyne cycloadditions under mild, bio-compatible conditions. For instance, using recently developed chemoenzymatic strategies, *N*-glycans containing core-fucose substitution and/or bisecting GlcNAc with otherwise ordinary complex-type antennae terminated in α2–3- or α2–6-linked sialic acid were synthesized and neoglycoproteins produced. With these ultra-defined entities *in vivo* bio-distribution was assessed showing that core substitutions alter glycan ligand properties through conformational changes which act as molecular switches for target affinity and influence glycoprotein-mediated cell binding and serum clearance.[214,215]

4.3. Evaluating glycosylation within the biological context

As mentioned above, evaluating glycosylation within a given biological context and without external manipulation is complicated; hence strategies with minimal impact on the system are actively sought. One approach in this direction is selective targeting, under physiological conditions, of particular glycans that can be subsequently evaluated. A recent study employs an engineered β-1,4-galactosyltransferase to specifically transfer a keto-Gal functionality to *O*-GlcNAc-modified proteins. The ketone moiety was subsequently reacted with various aminooxy-functionalized polyethylene glycol tags of defined mass and the resulting samples were analyzed by gel-based methods. In this way, a direct read-out of *O*-GlcNAc stoichiometry *vs.* state (*e.g.* mono-, di-, tri-, *etc.*) was possible, with insights into the complex interplay between *O*-GlcNAc glycosylation and phosphorylation.[216] Another chemoenzymatic strategy enabled rapid, sensitive and selective detection of the (Fuc(α1–2)Gal) disaccharide motif involved in processes such as learning and memory, inflammation, asthma, and tumorigenesis. By exploiting the restricted substrate tolerance of a blood group A GalNAc-transferase, the disaccharide is targeted with azido-functionalized UDP–GalNAc that is later captured from the complex sample mixture. This labeling strategy provides a variety of different enrichment strategies and imaging read-outs for a variety of Fuc(α1–2)Gal motifs.[217] Broader applications would of course require a supply of such restricted enzymes, currently unavailable at a larger scale. In another example of azide–alkyne chemoselective ("click chemistry") conjugation,[218] by introducing tetraacetylated *N*-azidoacetyl-D-mannosamine in the Sia biosynthetic pathway; mature glycoproteins containing azido-Sia were produced and targeted at the cellular level by capture with a biotinylated alkyne reagent and subsequent MS evaluation. Sias have also been targeted by periodate oxidation, in which vicinal hydroxyls in a *cis* configuration (present only in terminal Man, Gal(NAc) or non *O*-acetylated Sia residues) are converted to aldehydes. Subsequent oxime ligation with aminooxy-biotin labels glycoprotein subpopulations with high efficiency and cell viability, after which samples can be evaluated by MS.[219] The authors use ultra-mild conditions to assess only Sia and target terminal GalNAc through an enzymatic protocol to allow differentiation between Sia-containing and deficient cells.

Another elegant example of glycome comparison, based on stable isotope labeling with amino acids in cell culture (SILAC), labels amine-containing monosaccharides in cells using ^{14}N or ^{15}N glutamine as the sole nitrogen source. Named Isotopic Detection of Aminosugars With Glutamine (IDAWG),[220] the technique shows great promise for analyzing glycome dynamics under different conditions. However, interpretation of the data may not be straightforward as the protocol targets ManNAc, GlcNAc, GalNAc and NeuAc simultaneously, setting an equation with at least four variables added to the intrinsic microheterogeneity of glycosylation. On the backflip is the fact that the proteome may be targeted in a synchronized fashion. A rather different approach was chosen in a recent study where cell surface amino groups were reacted with unsaturated aldehyde yielding dihydropyridines products without affecting cell viability and simultaneously introducing the Hilyte Fluor 750 tag to perform noninvasive whole body fluorescence imaging. Examples included labeling of colon and gastric cancer cell lines in BALB/c nude mice to monitor tumor metastasis.[221]

5. Outlook

In this review we have not aimed at an exhaustive examination of all aspects related to a greater or lesser extent to protein glycosylation. Rather, we sought to provide a taste of some of the disciplines involving glycoscience that will landmark the future. One of the major challenges that glycoscience faced since its very beginning remains: handling the glycoproteome at the endogenous level, addressing its complexity in an automated high-throughput mode, and analyzing glycoproteins in complex samples with simultaneous characterization of both the glycan moieties and the corresponding protein carriers. Novel instrumental developments, such as ion mobility mass

spectrometry, to name only one, or the intelligent hyphenation of orthogonal existing techniques such as combining front-end biomolecular interaction analysis with in-line mass spectrometric evaluation, will be required to meet this challenge which will always constitute the first step in understanding the biological function of a glycoconjugate. In this respect, integration of glycomics with other –omics fields such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics[222] will certainly rank glycomics according to its merits. Current efforts towards systems glycobiology modelling, *i.e.* coupling biochemical knowledge and mathematics into *in silico* models of the cellular glycosylation system, will no doubt be decisive in this respect.[223] Evidently, a broad picture of how glycosylation is regulated through omics-data acquisition and systematic integration will be an enormously valuable asset to gain understanding of glycan functions as well as to develop clinical diagnostics and glyco-biomarker discoveries.[190] Such systems-level studies will help establish novel quantitative and mechanistic links between gene expression, protein expression, enzyme activity, carbohydrate structure and glycoconjugate function.

Abbreviations

2AB	2-Amino benzamide
AChE	Acetylcholinesterase
AD	Alzheimer disease
αDG	Alpha dystroglycan
ALL	Acute lymphoblastic leukemia
Arg	Arginine
Asn	Asparagine
AT	Antithrombin
CBPs	Carbohydrate binding proteins
CD147	Cluster of differentiation 147
CD22	Cluster of differentiation 22
CDG	Congenital disorder of glycosylation
CE	Capillary electrophoresis
CF	Cystic fibrosis
CFG	Consortium for Functional Glycomics
CID	Collision-induced dissociation
CREDEX	Carbohydrate REcognition Domain EXcision
CS-E	Chondroitin sulfate E
CSF	Cerebrospinal fluid
CSPG	Chondroitin sulfate proteoglycans
CTRC	Chymotrypsin C
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DMB	1,2-Diamino-4,5-methylenedioxybenzene
DTT	Dithiothreitol
ECD	Electron capture dissociation
EGFR	Epidermal growth factor receptor
EMBL	European Molecular Biology Laboratory
EPO	Erythropoietin
ER	Endoplasmic reticulum
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FAB	Fast atom bombardment
FASP	Filter aided sample preparation
FID	Flame ionization detector
Fuc	Fucose
GAGs	Glycosaminoglycans
Gal	Galactose
GalNAc	*N*-Acetylgalactosamine
GC	Gas chromatography
GCMS	Gas chromatography mass spectrometry
GH	Growth hormone
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	*N*-Acetylglucosamine
Gln	Glutamine
Glu	Glutamic acid
GnT-V	*N*-Acetylglucosaminyl transferase V
GT	Glycosyltransferase
GU	Glucose unit
HCC	Hepatocellular carcinoma
HCD	Higher-energy collisional dissociation
HCl	Hydrochloric acid
HILIC	Hydrophilic interaction liquid chromatography
HNK-1	Human natural killer epitope 1
Hp	Haptoglobin
HPAEC	High performance anion exchange chromatography
HPAEC	High pH anion exchange chromatography
HPLC	High performance liquid chromatography
HS	Heparin sulfate
HSA	Human serum albumin
ICAM	Inter cellular adhesion molecule
IDAWG	Isotopic detection of aminosugars with glutamine
IEF	Isoelectric focusing
INPEG	In-gel non-specific proteolysis for elucidating glycoproteins
ITC	Isothermal titration calorimetry
LC	Liquid chromatography
LC-FLC	Liquid chromatography with post-column fluorescence derivatization
Lewis A	Galβ1–3(Fucα1–4)GlcNAc
Lewis B	(Fucα1–2)Galβ1–3(Fucα1–4)GlcNAc
Lewis X	Galβ1–4(Fucα1–3)GlcNAc
Lewis Y	(Fucα1–2)Galβ1–4(Fucα1–3)GlcNAc
LIF	Laser-induced fluorescence
LIF	Leukemia inhibitory factor
Lys	Lysine
M	Molar – mole per liter
M6P	Mannose-6-phosphate
MAA	*Maackia amurensis* agglutinin
MALDI-TOF	Matrix assisted laser desorption ionization time of flight
Man	Mannose
ManNAc	*N*-Acetylmannosamine
mM	Millimolar
MRM	Multiple reaction monitoring
MS	Mass spectrometry
NCAM	Neural cell adhesion molecule
NCBI	National Center for Biotechnology Information
NeuAc	*N*-Acetylneuraminic acid
NIBRT	National institute for bioprocessing and training

NK	Natural killer
nm	Nanometer
NMR	Nuclear magnetic resonance
NP	Normal-phase
OXM	Oxyntomodulin
PAD	Pulsed amperometric detection
PDIA	Protein disulfide isomerase
PECAM	Platelet endothelial cell adhesion molecule
PFK	Phosphofructokinase
PGMs	Post-glycosylational modifications
PMAA	Partially methylated alditol acetate
PNGase F	Peptide-N4-(*N*-acetyl-beta-glucosaminyl) asparagine amidase
Pro	Proline
PTC	Papillary thyroid carcinoma
PTM	Post-translational modification
Q-TOF	Quadrupole time-of-flight
RP	Reverse-phase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
Sia	Sialic acid
Sialyl-Lewis X	Neu5Acα2–3Galβ1–4(Fucα1–3)GlcNAc
SNA	*Sambucus nigra* agglutinin
SPR	Surface plasmon resonance
S-Tn	Neu5Acα2–6GalNAcα1-*O*-Ser/Thr
T antigen	Galβ1–3-GalNAcα-*O*-Ser/Thr
Thr	Threonine
Tn antigen	GalNAcα1-*O*-Ser/Thr
Tyr	Tyrosine
UC	Ulcerative colitis
UPLC	Ultra performance liquid chromatography
VCAM	Vascular cell adhesion molecule
WAX	Weak anion exchange
Xyl	Xylose
ZFN	Zinc-finger nuclease

References

1 K. W. Moremen, M. Tiemeyer and A. V. Nairn, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 448–462.
2 H. C. Joao and R. A. Dwek, *Eur. J. Biochem.*, 1993, **218**, 239–244.
3 P. M. Rudd, B. P. Morgan, M. R. Wormald, D. J. Harvey, C. W. van den Berg, S. J. Davis, M. A. Ferguson and R. A. Dwek, *J. Biol. Chem.*, 1997, **272**, 7229–7244.
4 R. G. Spiro, *Glycobiology*, 2002, **12**, 43R–56R.
5 M. Gómez-García, J. M. Benito, D. Rodríguez-Lucena, J. X. Yu, K. Chmurski, M. C. Ortiz, R. Gutiérrez-Gallego, A. Maestre, J. Defaye and J. M. García Fernández, *J. Am. Chem. Soc.*, 2005, **127**, 7970–7971.
6 A. Varki, *Cold Spring Harbor Perspect. Biol.*, 2011, **3**, 1–14.
7 G. W. Hart and R. J. Copeland, *Cell*, 2010, **143**, 672–676.
8 R. S. Haltiwanger and J. B. Lowe, *Annu. Rev. Biochem.*, 2004, **73**, 491–537.
9 K. Ohtsubo and J. D. Marth, *Cell*, 2006, **126**, 855–867.
10 L. R. Ruhaak, H. W. Uh, M. Beekman, C. A. Koeleman, C. H. Hokke, R. G. Westendorp, M. Wuhrer, J. J. Houwing-Duistermaat, P. E. Slagboom and A. M. Deelder, *PLoS One*, 2010, **5**, e12566.
11 M. Pucic, A. Muzinic, M. Novokmet, M. Skledar, N. Pivac, G. Lauc and O. Gornik, *Glycobiology*, 2012, **22**, 975–982.
12 J. Wang, C. I. Balog, K. Stavenhagen, C. A. Koeleman, H. U. Scherer, M. H. Selman, A. M. Deelder, T. W. Huizinga, R. E. Toes and M. Wuhrer, *Mol. Cell. Proteomics*, 2011, **10**, M110.
13 S. A. Springer and P. Gagneux, *J. Biol. Chem.*, 2013, **288**, 6904–6911.
14 G. Lauc and V. Zoldos, *Mol. BioSyst.*, 2010, **6**, 2373–2379.
15 V. Zoldos, M. Novokmet, I. Beceheli and G. Lauc, *Glycoconjugate J.*, 2013, **30**, 41–50.
16 T. Tharmalingam, B. Adamczyk, M. A. Doherty, L. Royle and P. M. Rudd, *Glycoconjugate J.*, 2013, **30**, 137–146.
17 J. F. Rakus and L. K. Mahal, *Annu. Rev. Anal. Chem.*, 2011, **4**, 367–392.
18 D. Vanderschaeghe, N. Festjens, J. Delanghe and N. Callewaert, *Biol. Chem.*, 2010, **391**, 149–161.
19 K. Mariño, J. Bones, J. J. Kattla and P. M. Rudd, *Nat. Chem. Biol.*, 2010, **6**, 713–723.
20 M. Pabst and F. Altmann, *Proteomics*, 2011, **11**, 631–643.
21 K. Yamada and K. Kakehi, *J. Pharm. Biomed. Anal.*, 2011, **55**, 702–727.
22 J. Zaia, *Methods Mol. Biol.*, 2013, **984**, 13–25.
23 Y. Mechref, *Electrophoresis*, 2011, **32**, 3467–3481.
24 D. Vanderschaeghe, A. Guttman and N. Callewaert, *Methods Mol. Biol.*, 2013, **919**, 87–96.
25 S. Hua, C. Lebrilla and H. J. An, *Bioanalysis*, 2011, **3**, 2573–2585.
26 J. L. Behan and K. D. Smith, *Biomed. Chromatogr.*, 2011, **25**, 39–46.
27 M. Wuhrer, *Glycoconjugate J.*, 2013, **30**, 11–22.
28 I. M. Lazar, W. Lee and A. C. Lazar, *Electrophoresis*, 2013, **34**, 113–125.
29 J. Zaia, *OMICS: J. Integr. Biol.*, 2010, **14**, 401–418.
30 N. Leymarie and J. Zaia, *Anal. Chem.*, 2012, **84**, 3040–3048.
31 D. F. Smith and R. D. Cummings, *Mol. Cell. Proteomics*, 2013, **12**, 902–912.
32 G. Gupta, A. Surolia and S. G. Sampathkumar, *OMICS: J. Integr. Biol.*, 2010, **14**, 419–436.
33 D. F. Smith, X. Song and R. D. Cummings, *Methods Enzymol.*, 2010, **480**, 417–444.
34 S. Park, J. C. Gildersleeve, O. Blixt and I. Shin, *Chem. Soc. Rev.*, 2012, **42**, 4310–4326.
35 C. D. Rillahan and J. C. Paulson, *Annu. Rev. Biochem.*, 2011, **80**, 797–823.
36 Y. Pasing, A. Sickmann and U. Lewandrowski, *Biol. Chem.*, 2012, **393**, 249–258.
37 J. Nilsson, A. Halim, A. Grahn and G. Larson, *Glycoconjugate J.*, 2013, **30**, 119–136.
38 C. Reichel, *Drug Test. Anal.*, 2010, **2**, 603–619.
39 H. J. Müller-Eberhard and H. G. Kunkel, *J. Exp. Med.*, 1956, **104**, 253–269.

40 H. Nie, Y. Li and X. L. Sun, *J. Proteomics*, 2012, **75**, 3098–3112.
41 A. Varki and P. Gagneux, *Ann. N. Y. Acad. Sci.*, 2012, **1253**, 16–36.
42 K. Hayakawa, C. De Felice, T. Watanabe, T. Tanaka, K. Iinuma, K. Nihei, S. Higuchi, T. Ezoe, I. Hibi and K. Kurosawa, *J. Chromatogr.*, 1993, **620**, 25–31.
43 R. Schauer, G. V. Srinivasan, D. Wipfler, B. Kniep and R. Schwartz-Albiez, *Adv. Exp. Med. Biol.*, 2011, **705**, 525–548.
44 J. Mallorquí, E. Llop, C. de Bolòs, R. Gutiérrez-Gallego, J. Segura and J. A. Pascual, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2010, **878**, 2117–2122.
45 J. Kamerling and J. Vliegenthart, *Carbohydrates*, Walter de Gruyter, Berlin, 1989, pp. 176–263.
46 O. M. Grace, A. Dzajic, A. K. Jäger, N. T. Nyberg, A. Onder and N. Rønsted, *Phytochemistry*, 2013, **93**, 79–87.
47 G. Palmieri, M. Balestrieri, J. Peter-Katalinic, G. Pohlentz, M. Rossi, I. Fiume and G. Pocsfalvi, *J. Proteome Res.*, 2013, **12**, 2779–2790.
48 T. Toida, I. R. Vlahov, A. E. Smith, R. E. Hileman and R. J. Linhardt, *J. Carbohydr. Chem.*, 1996, **15**, 351–360.
49 S. Hakomori, *J. Biochem.*, 1964, **55**, 205–208.
50 R. Rosenfeld, H. Bangio, G. J. Gerwig, R. Rosenberg, R. Aloni, Y. Cohen, Y. Amor, I. Plaschkes, J. P. Kamerling and R. B. Maya, *J. Biochem. Biophys. Methods*, 2007, **70**, 415–426.
51 J. Hirabayashi, M. Yamada, A. Kuno and H. Tateno, *Chem. Soc. Rev.*, 2013, **42**, 4443–4458.
52 A. Kuno, A. Matsuda, Y. Ikehara, H. Narimatsu and J. Hirabayashi, *Methods Enzymol.*, 2010, **478**, 165–179.
53 A. Kuno, N. Uchiyama, S. Koseki-Kuno, Y. Ebe, S. Takashima, M. Yamada and J. Hirabayashi, *Nat. Methods*, 2005, **2**, 851–856.
54 H. Tateno, A. Kuno, Y. Itakura and J. Hirabayashi, *Methods Enzymol.*, 2010, **478**, 181–195.
55 H. Narimatsu, H. Sawaki, A. Kuno, H. Kaji, H. Ito and Y. Ikehara, *FEBS J.*, 2010, **277**, 95–105.
56 T. Yue, I. J. Goldstein, M. A. Hollingsworth, K. Kaul, R. E. Brand and B. B. Haab, *Mol. Cell. Proteomics*, 2009, **8**, 1697–1707.
57 B. B. Haab, *Expert Rev. Proteomics*, 2010, **7**, 9–11.
58 K. T. Pilobello, D. E. Slawek and L. K. Mahal, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11534–11539.
59 J. L. Jiménez Blanco, M. C. Ortiz and J. M. García Fernández, *Chem. Soc. Rev.*, 2013, **42**, 4518–4531.
60 W. Jian, R. W. Edom, D. Wang, N. Weng and S. W. Zhang, *Anal. Chem.*, 2013, **85**, 2867–2874.
61 N. G. Karlsson and N. H. Packer, *Anal. Biochem.*, 2002, **305**, 173–185.
62 R. P. Kozak, L. Royle, R. A. Gardner, D. L. Fernandes and M. Wuhrer, *Anal. Biochem.*, 2012, **423**, 119–128.
63 A. H. Merry, D. C. Neville, L. Royle, B. Matthews, D. J. Harvey, R. A. Dwek and P. M. Rudd, *Anal. Biochem.*, 2002, **304**, 91–99.
64 L. A. Omtvedt, L. Royle, G. Husby, K. Sletten, C. Radcliffe, R. A. Dwek, P. M. Rudd and D. J. Harvey, *Rapid Commun. Mass Spectrom.*, 2004, **18**, 2357–2359.
65 R. B. Trimble, P. H. Atkinson, A. L. Tarentino, T. H. Plummer, Jr, F. Maley and K. B. Tomer, *J. Biol. Chem.*, 1986, **261**, 12000–12005.
66 D. J. Harvey and P. M. Rudd, *J. Mass Spectrom.*, 2010, **45**, 815–819.
67 L. V. Johnson and G. S. Hageman, *Exp. Eye Res.*, 1987, **44**, 553–565.
68 J. Schwarzer, E. Rapp, R. Hennig, Y. Genzel, I. Jordan, V. Sandig and U. Reichl, *Vaccine*, 2009, **27**, 4325–4336.
69 I. Ritamo, J. Räbinä, S. Natunen and L. Valmu, *Anal. Bioanal. Chem.*, 2013, **405**, 2469–2480.
70 Y. Hu and Y. Mechref, *Electrophoresis*, 2012, **33**, 1768–1777.
71 P. M. Rudd, G. R. Guile, B. Küster, D. J. Harvey, G. Opdenakker and R. A. Dwek, *Nature*, 1997, **388**, 205–207.
72 G. R. Guile, P. M. Rudd, D. R. Wing, S. B. Prime and R. A. Dwek, *Anal. Biochem.*, 1996, **240**, 210–226.
73 M. P. Campbell, L. Royle, C. M. Radcliffe, R. A. Dwek and P. M. Rudd, *Bioinformatics*, 2008, **24**, 1214–1216.
74 N. V. Artemenko, M. P. Campbell and P. M. Rudd, *J. Proteome Res.*, 2010, **9**, 2037–2041.
75 C. Reichel, *Drug Test. Anal.*, 2011, **3**, 883–891.
76 K. R. Anumula, *Anal. Biochem.*, 2006, **350**, 1–23.
77 R. Gutiérrez-Gallego, S. R. Haseley, V. F. van Miegem, J. F. Vliegenthart and J. P. Kamerling, *Glycobiology*, 2004, **14**, 373–386.
78 E. Llop, R. Gutiérrez-Gallego, V. Belalcazar, G. J. Gerwig, J. P. Kamerling, J. Segura and J. A. Pascual, *Proteomics*, 2007, **7**, 4278–4291.
79 P. H. Jensen, N. G. Karlsson, D. Kolarich and N. H. Packer, *Nat. Protoc.*, 2012, **7**, 1299–1310.
80 A. R. de Boer, C. H. Hokke, A. M. Deelder and M. Wuhrer, *Glycoconjugate J.*, 2008, **25**, 75–84.
81 M. Pucic, A. Knezevic, J. Vidic, B. Adamczyk, M. Novokmet, O. Polasek, O. Gornik, S. Supraha-Goreta, M. R. Wormald, I. Redzic, H. Campbell, A. Wright, N. D. Hastie, J. F. Wilson, I. Rudan, M. Wuhrer, P. M. Rudd, D. Josic and G. Lauc, *Mol. Cell. Proteomics*, 2011, **10**, M111.
82 R. Gutiérrez-Gallego, J. L. Blanco, C. W. Thijssen-van Zuylen, C. H. Gotfredsen, H. Voshol, J. O. Duus, M. Schachner and J. F. Vliegenthart, *J. Biol. Chem.*, 2001, **276**, 30834–30844.
83 B. Xia, J. A. Royall, G. Damera, G. P. Sachdev and R. D. Cummings, *Glycobiology*, 2005, **15**, 747–775.
84 Z. Lin and D. M. Lubman, *Methods Mol. Biol.*, 2013, **1007**, 289–300.
85 F. Tousi, J. Bones, W. S. Hancock and M. Hincapie, *Anal. Chem.*, 2013, **85**, 8421–8428.
86 C. Jiménez-Castells, B. de la Torre, D. Andreu and R. Gutiérrez-Gallego, *Glycoconjugate J.*, 2008, **25**, 879–887.
87 I. Breloy, *Methods Mol. Biol.*, 2012, **842**, 165–177.
88 A. Toyama, H. Nakagawa, K. Matsuda, T. A. Sato, Y. Nakamura and K. Ueda, *Anal. Chem.*, 2012, **84**, 9655–9662.
89 H. Hahne and B. Kuster, *Mol. Cell. Proteomics*, 2012, **11**, 1063–1069.
90 S. A. Myers, S. Daou, E. B. Affar and A. L. Burlingame, *Proteomics*, 2013, **13**, 982–991.

91 K. Pagel and D. J. Harvey, *Anal. Chem.*, 2013, **85**, 5138–5145.
92 V. H. Pomin, *Anal. Bioanal. Chem.*, 2013, **405**, 3035–3048.
93 J. F. Vliegenthart and J. P. Kamerling, *^{1}H NMR Structural-Reporter-Group Concepts in Carbohydrate Analysis*, Elsevier, 2007, pp. 133–191.
94 T. Yoshida-Moriguchi, L. Yu, S. H. Stalnaker, S. Davis, S. Kunz, M. Madson, M. B. Oldstone, H. Schachter, L. Wells and K. P. Campbell, *Science*, 2010, **327**, 88–92.
95 M. Fellenberg, A. Coksezen and B. Meyer, *Angew. Chem., Int. Ed.*, 2010, **49**, 2630–2633.
96 A. Arda, P. Blasco, S. D. Varon, V. Schubert, S. Andre, M. Bruix, F. J. Canada, H. J. Gabius, C. Unverzagt and J. Jiménez-Barbero, *J. Am. Chem. Soc.*, 2013, **135**, 2667–2675.
97 R. E. Lehotzky, C. L. Partch, S. Mukherjee, H. L. Cash, W. E. Goldman, K. H. Gardner and L. V. Hooper, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 7722–7727.
98 H. Ghasriani, P. J. Belcourt, S. Sauve, D. J. Hodgson, D. Brochu, M. Gilbert and Y. Aubin, *J. Biol. Chem.*, 2013, **288**, 247–254.
99 A. W. Barb and J. H. Prestegard, *Nat. Chem. Biol.*, 2011, **7**, 147–153.
100 C. L. Woodin, M. Maxon and H. Desaire, *Analyst*, 2013, **138**, 2793–2803.
101 A. Porter, T. Yue, L. Heeringa, S. Day, E. Suh and B. B. Haab, *Glycobiology*, 2010, **20**, 369–380.
102 D. Kletter, S. Singh, M. Bern and B. B. Haab, *Mol. Cell. Proteomics*, 2013, **12**, 1026–1035.
103 S. V. Bennun, K. J. Yarema, M. J. Betenbaugh and F. J. Krambeck, *PLoS Comput. Biol.*, 2013, **9**, e1002813.
104 C. W. der Lieth, A. A. Freire, D. Blank, M. P. Campbell, A. Ceroni, D. R. Damerell, A. Dell, R. A. Dwek, B. Ernst, R. Fogh, M. Frank, H. Geyer, R. Geyer, M. J. Harrison, K. Henrick, S. Herget, W. E. Hull, J. Ionides, H. J. Joshi, J. P. Kamerling, B. R. Leeflang, T. Lütteke, M. Lundborg, K. Maass, A. Merry, R. Ranzinger, J. Rosen, L. Royle, P. M. Rudd, S. Schloissnig, R. Stenutz, W. F. Vranken, G. Widmalm and S. M. Haslam, *Glycobiology*, 2011, **21**, 493–502.
105 M. Vila-Perelló, R. Gutiérrez-Gallego and D. Andreu, *ChemBioChem*, 2005, **6**, 1831–1838.
106 M. M. Ivancic, H. S. Gadgil, H. B. Halsall and M. J. Treuheit, *Anal. Biochem.*, 2010, **400**, 25–32.
107 W. R. Alley, Jr, Y. Mechref and M. V. Novotny, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 161–170.
108 D. F. Zielinska, F. Gnad, J. R. Wisniewski and M. Mann, *Cell*, 2010, **141**, 897–907.
109 C. C. Nwosu, J. Huang, D. L. Aldredge, J. S. Strum, S. Hua, R. R. Seipert and C. B. Lebrilla, *Anal. Chem.*, 2013, **85**, 956–963.
110 D. Kolarich, P. H. Jensen, F. Altmann and N. H. Packer, *Nat. Protoc.*, 2012, **7**, 1285–1298.
111 M. B. West, Z. M. Segu, C. L. Feasley, P. Kang, I. Klouckova, C. Li, M. V. Novotny, C. M. West, Y. Mechref and M. H. Hanigan, *J. Biol. Chem.*, 2010, **285**, 29511–29524.
112 P. Pompach, Z. Brnakova, M. Sanda, J. Wu, N. Edwards and R. Goldman, *Mol. Cell. Proteomics*, 2013, **12**, 1281–1293.
113 U. M. Bailey, M. F. Jamaluddin and B. L. Schulz, *J. Proteome Res.*, 2012, **11**, 5376–5383.
114 M. M. Chen, A. I. Bartlett, P. S. Nerenberg, C. T. Friel, C. P. Hackenberger, C. M. Stultz, S. E. Radford and B. Imperiali, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 22528–22533.
115 D. D. Banks, *J. Mol. Biol.*, 2011, **412**, 536–550.
116 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
117 R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683–720.
118 L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, *Annu. Rev. Biochem.*, 2008, **77**, 521–555.
119 C. Breton, S. Fournel-Gigleux and M. M. Palcic, *Curr. Opin. Struct. Biol.*, 2012, **22**, 540–549.
120 R. D. Cummings, *Mol. BioSyst.*, 2009, **5**, 1087–1104.
121 J. C. Paulson, O. Blixt and B. E. Collins, *Nat. Chem. Biol.*, 2006, **2**, 238–248.
122 S. Margraf-Schönfeld, C. Böhm and C. Watzl, *J. Biol. Chem.*, 2011, **286**, 24142–24149.
123 Z. Sumer-Bayraktar, D. Kolarich, M. P. Campbell, S. Ali, N. H. Packer and M. Thaysen-Andersen, *Mol. Cell. Proteomics*, 2011, **10**, M111.
124 P. Stanley and T. Okajima, *Curr. Top. Dev. Biol.*, 2010, **92**, 131–164.
125 N. A. Rana and R. S. Haltiwanger, *Curr. Opin. Struct. Biol.*, 2011, **21**, 583–589.
126 Y. C. Liu, H. Y. Yen, C. Y. Chen, C. H. Chen, P. F. Cheng, Y. H. Juan, C. H. Chen, K. H. Khoo, C. J. Yu, P. C. Yang, T. L. Hsu and C. H. Wong, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11332–11337.
127 G. W. Hart, M. P. Housley and C. Slawson, *Nature*, 2007, **446**, 1017–1022.
128 J. E. Rexach, P. M. Clark, D. E. Mason, R. L. Neve, E. C. Peters and L. C. Hsieh-Wilson, *Nat. Chem. Biol.*, 2012, **8**, 253–261.
129 G. D. Holt, C. M. Snow, A. Senior, R. S. Haltiwanger, L. Gerace and G. W. Hart, *J. Cell Biol.*, 1987, **104**, 1157–1164.
130 R. Fujiki, W. Hashiba, H. Sekine, A. Yokoyama, T. Chikanishi, S. Ito, Y. Imai, J. Kim, H. H. He, K. Igarashi, J. Kanno, F. Ohtake, H. Kitagawa, R. G. Roeder, M. Brown and S. Kato, *Nature*, 2011, **480**, 557–560.
131 K. Sakabe, Z. Wang and G. W. Hart, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 19915–19920.
132 W. Yi, P. M. Clark, D. E. Mason, M. C. Keenan, C. Hill, W. A. Goddard, III, E. C. Peters, E. M. Driggers and L. C. Hsieh-Wilson, *Science*, 2012, **337**, 975–980.
133 Y. Sakaidani, T. Nomura, A. Matsuura, M. Ito, E. Suzuki, K. Murakami, D. Nadano, T. Matsuda, K. Furukawa and T. Okajima, *Nat. Commun.*, 2011, **2**, 583.
134 K. T. Schjoldager and H. Clausen, *Biochim. Biophys. Acta*, 2012, **1820**, 2079–2094.
135 J. A. Hanover, M. W. Krause and D. C. Love, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 312–321.
136 G. W. Hart, C. Slawson, G. Ramirez-Correa and O. Lagerlof, *Annu. Rev. Biochem.*, 2011, **80**, 825–858.
137 W. M. Watkins and W. T. Morgan, *Nature*, 1952, **169**, 825–826.

138 W. T. Morgan and W. M. Watkins, *Glycoconjugate J.*, 2000, **17**, 501–530.
139 J. J. van Rooijen, J. P. Kamerling and J. F. Vliegenthart, *Glycobiology*, 1998, **8**, 1065–1075.
140 N. Malagolini, D. Santini, M. Chiricolo and F. Dall'Olio, *Glycobiology*, 2007, **17**, 688–697.
141 G. F. Clark, P. Grassi, P. C. Pang, M. Panico, D. Lafrenz, E. Z. Drobnis, M. R. Baldwin, H. R. Morris, S. M. Haslam, S. Schedin-Weiss, W. Sun and A. Dell, *Mol. Cell. Proteomics*, 2012, **11**, M111.
142 T. Pouyani and B. Seed, *Cell*, 1995, **83**, 333–343.
143 D. W. Scott and R. P. Patel, *Glycobiology*, 2013, **23**, 622–633.
144 K. J. Colley, *Adv. Exp. Med. Biol.*, 2010, **663**, 111–126.
145 M. Gómez-Garcia, J. M. Benito, R. Gutiérrez-Gallego, A. Maestre, C. O. Mellet, J. M. Fernández and J. L. Blanco, *Org. Biomol. Chem.*, 2010, **8**, 1849–1860.
146 P. V. Murphy, S. André and H. J. Gabius, *Molecules*, 2013, **18**, 4026–4053.
147 S. Velasco, N. Díez-Revuelta, T. Hernández-Iglesias, H. Kaltner, S. André, H. J. Gabius and J. Abad-Rodríguez, *J. Neurochem.*, 2013, **125**, 49–62.
148 M. Bence and M. Sahin-Tóth, *FEBS J.*, 2011, **278**, 4338–4350.
149 J. Yao, W. Hong, J. Huang, K. Zhan, H. Huang and M. Hong, *PLoS One*, 2012, **7**, e52563.
150 M. Tsachaki, D. Serlidaki, A. Fetani, V. Zarkou, I. Rozani, J. Ghiso and S. Efthimiopoulos, *Glycobiology*, 2011, **21**, 1382–1388.
151 R. Chen, J. Li, C. Feng, S. Chen, Y. Liu, C. Duan, H. Li, X. Xia, T. He, M. Wei and R. Dai, *J. Cell. Biochem.*, 2012, **114**, 816–822.
152 C. P. Chan, T. Y. Mak, K. T. Chin, I. O. Ng and D. Y. Jin, *J. Cell Sci.*, 2010, **123**, 1438–1448.
153 T. Y. Weng, W. T. Chiu, H. S. Liu, H. C. Cheng, M. R. Shen, D. B. Mount and C. Y. Chou, *Biochim. Biophys. Acta*, 2013, **1833**, 1133–1146.
154 V. P. Chen, R. C. Choi, W. K. Chan, K. W. Leung, A. J. Guo, G. K. Chan, W. K. Luk and K. W. Tsim, *J. Biol. Chem.*, 2011, **286**, 32948–32961.
155 E. Riedl, H. Koeppel, F. Pfister, V. Peters, S. Sauerhoefer, P. Sternik, P. Brinkkoetter, H. Zentgraf, G. Navis, R. H. Henning, B. J. Van Den, S. J. Bakker, B. Janssen, F. J. van der Woude and B. A. Yard, *Diabetes*, 2010, **59**, 1984–1990.
156 P. Van Antwerpen, M. C. Slomianny, K. Z. Boudjeltia, C. Delporte, V. Faid, D. Calay, A. Rousseau, N. Moguilevsky, M. Raes, L. Vanhamme, P. G. Furtmüller, C. Obinger, M. Vanhaeverbeek, J. Nève and J. C. Michalski, *J. Biol. Chem.*, 2010, **285**, 16351–16359.
157 M. L. Montpetit, P. J. Stocker, T. A. Schwetz, J. M. Harper, S. A. Norring, L. Schaffer, S. J. North, J. Jang-Lee, T. Gilmartin, S. R. Head, S. M. Haslam, A. Dell, J. D. Marth and E. S. Bennett, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 16517–16522.
158 K. D. Chandrasekhar, A. Lvov, C. Terrenoire, G. Y. Gao, R. S. Kass and W. R. Kobertz, *J. Physiol.*, 2011, **589**, 3721–3730.
159 A. Mant, S. Williams, L. Roncoroni, E. Lowry, D. Johnson and I. O'Kelly, *J. Biol. Chem.*, 2013, **288**, 3251–3264.
160 A. K. Walker, K. Y. Soo, V. Levina, G. H. Talbo and J. D. Atkin, *FEBS J.*, 2013, **280**, 233–243.
161 S. S. Pinho, R. Seruca, F. Gärtner, Y. Yamaguchi, J. Gu, N. Taniguchi and C. A. Reis, *Cell. Mol. Life Sci.*, 2011, **68**, 1011–1020.
162 C. Herraiz, B. L. Sánchez-Laorden, C. Jiménez-Cervantes and J. C. García-Borrón, *Pigm. Cell Melanoma Res.*, 2011, **24**, 479–489.
163 W. Huang, W. J. Luo, P. Zhu, J. Tang, X. L. Yu, H. Y. Cui, B. Wang, Y. Zhang, J. L. Jiang and Z. N. Chen, *Biochem. J.*, 2013, **449**, 437–448.
164 L. A. Cole, *Reprod. Biol. Endocrinol.*, 2009, **7**, 8.
165 D. L. Blithe and R. K. Iles, *Endocrinology*, 1995, **136**, 903–910.
166 J. J. van Rooijen, A. F. Voskamp, J. P. Kamerling and J. F. Vliegenthart, *Glycobiology*, 1999, **9**, 21–30.
167 V. Belalcazar, R. Gutiérrez-Gallego, E. Llop, J. Segura and J. A. Pascual, *Electrophoresis*, 2006, **27**, 4387–4395.
168 S. M. Muthana, C. T. Campbell and J. C. Gildersleeve, *ACS Chem. Biol.*, 2012, **7**, 31–43.
169 S. Otsuki, S. R. Hanson, S. Miyaki, S. P. Grogan, M. Kinoshita, H. Asahara, C. H. Wong and M. K. Lotz, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 10202–10207.
170 J. M. Brown, J. Xia, B. Zhuang, K. S. Cho, C. J. Rogers, C. I. Gama, M. Rawat, S. E. Tully, N. Uetani, D. E. Mason, M. L. Tremblay, E. C. Peters, O. Habuchi, D. F. Chen and L. C. Hsieh-Wilson, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4768–4773.
171 I. Martínez-Martínez, J. Navarro-Fernández, A. Ostergaard, R. Gutiérrez-Gallego, J. Padilla, N. Bohdan, A. Miñano, C. Pascual, C. Martínez, M. E. de la Morena-Barrio, S. Aguila, S. Pedersen, S. R. Kristensen, V. Vicente and J. Corral, *Blood*, 2012, **120**, 900–904.
172 J. R. Bishop, M. Schuksz and J. D. Esko, *Nature*, 2007, **446**, 1030–1037.
173 A. Cariappa, H. Takematsu, H. Liu, S. Diaz, K. Haider, C. Boboila, G. Kalloo, M. Connole, H. N. Shi, N. Varki, A. Varki and S. Pillai, *J. Exp. Med.*, 2009, **206**, 125–138.
174 L. J. Olson, F. C. Peterson, A. Castonguay, R. N. Bohnsack, M. Kudo, R. R. Gotschall, W. M. Canfield, B. F. Volkman and N. M. Dahms, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 12493–12498.
175 J. Tucholski, M. S. Simmons, A. L. Pinner, V. Haroutunian, R. E. McCullumsmith and J. H. Meador-Woodruff, *Schizophr. Res.*, 2013, **146**, 177–183.
176 Y. Zeng and X. Han, *J. Neurochem.*, 2008, **106**, 1275–1286.
177 X. Han, M. Holtzman, D. W. McKeel, Jr, J. Kelley and J. C. Morris, *J. Neurochem.*, 2002, **82**, 809–818.
178 K. I. Stanford, L. Wang, J. Castagnola, D. Song, J. R. Bishop, J. R. Brown, R. Lawrence, X. Bai, H. Habuchi, M. Tanaka, W. V. Cardoso, K. Kimata and J. D. Esko, *J. Biol. Chem.*, 2010, **285**, 286–294.
179 J. L. Stanta, R. Saldova, W. B. Struwe, J. C. Byrne, F. M. Leweke, M. Rothermund, H. Rahmoune, Y. Levin, P. C. Guest, S. Bahn and P. M. Rudd, *J. Proteome Res.*, 2010, **9**, 4476–4489.

180 J. Barnes, J. M. Lim, A. Godard, F. Blanchard, L. Wells and R. Steet, *J. Biol. Chem.*, 2011, **286**, 24855–24864.
181 C. Mandal, C. Mandal, S. Chandra, R. Schauer and C. Mandal, *Glycobiology*, 2012, **22**, 70–83.
182 M. Kobayashi, H. Hoshino, J. Masumoto, M. Fukushima, K. Suzawa, S. Kageyama, M. Suzuki, H. Ohtani, M. Fukuda and J. Nakayama, *Inflammatory Bowel Dis.*, 2009, **15**, 697–706.
183 B. Mann, E. Klussmann, V. Vandamme-Feldhaus, M. Iwersen, M. L. Hanski, E. O. Riecken, H. J. Buhr, R. Schauer, Y. S. Kim and C. Hanski, *Int. J. Cancer*, 1997, **72**, 258–264.
184 L. Zhao, M. Liu, Y. Gao, Y. Huang, G. Lu, Y. Gao, X. Guo and B. Shi, *Eur. J. Endocrinol.*, 2013, **168**, 585–592.
185 A. Kuga, M. Kanagawa, A. Sudo, Y. M. Chan, M. Tajiri, H. Manya, Y. Kikkawa, M. Nomizu, K. Kobayashi, T. Endo, Q. L. Lu, Y. Wada and T. Toda, *J. Biol. Chem.*, 2012, **287**, 9560–9567.
186 Y. Hara, B. Balci-Hayta, T. Yoshida-Moriguchi, M. Kanagawa, D. B. Beltrán-Valero, H. Gündesli, T. Willer, J. S. Satz, R. W. Crawford, S. J. Burden, S. Kunz, M. B. Oldstone, A. Accardi, B. Talim, F. Muntoni, H. Topaloglu, P. Dincer and K. P. Campbell, *N. Engl. J. Med.*, 2011, **364**, 939–946.
187 B. Adamczyk, T. Tharmalingam and P. M. Rudd, *Biochim. Biophys. Acta*, 2012, **1820**, 1347–1353.
188 K. B. Chandler and R. Goldman, *Mol. Cell. Proteomics*, 2013, **12**, 836–845.
189 Y. Mechref, Y. Hu, A. Garcia and A. Hussein, *Electrophoresis*, 2012, **33**, 1755–1767.
190 T. Angata, R. Fujinawa, A. Kurimoto, K. Nakajima, M. Kato, S. Takamatsu, H. Korekane, C. X. Gao, K. Ohtsubo, S. Kitazume and N. Taniguchi, *Ann. N. Y. Acad. Sci.*, 2012, **1253**, 159–169.
191 H. J. An and C. B. Lebrilla, *Methods Mol. Biol.*, 2010, **600**, 199–213.
192 H. Yu and X. Chen, *Org. Biomol. Chem.*, 2007, **5**, 865–872.
193 H. J. Gabius, S. André, J. Jiménez-Barbero, A. Romero and D. Solís, *Trends Biochem. Sci.*, 2011, **36**, 298–313.
194 S. A. Brooks and D. M. Hall, *Methods Mol. Biol.*, 2012, **878**, 31–50.
195 M. Lohr, H. Kaltner, R. Schwartz-Albiez, F. Sinowatz and H. J. Gabius, *Anat., Histol., Embryol.*, 2010, **39**, 481–493.
196 F. A. Habermann, S. André, H. Kaltner, D. Kübler, F. Sinowatz and H. J. Gabius, *Histochem. Cell Biol.*, 2011, **135**, 539–552.
197 C. J. Rogers, P. M. Clark, S. E. Tully, R. Abrol, K. C. Garcia, W. A. Goddard, III and L. C. Hsieh-Wilson, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 9747–9752.
198 X. Song, Y. Lasanajak, B. Xia, J. Heimburg-Molinaro, J. M. Rhea, H. Ju, C. Zhao, R. J. Molinaro, R. D. Cummings and D. F. Smith, *Nat. Methods*, 2011, **8**, 85–90.
199 Q. Tong, X. Wang, H. Wang, T. Kubo and M. Yan, *Anal. Chem.*, 2012, **84**, 3049–3052.
200 B. Lepenies, J. Yin and P. H. Seeberger, *Curr. Opin. Chem. Biol.*, 2010, **14**, 404–411.
201 T. K. Dam and C. F. Brewer, *Glycobiology*, 2010, **20**, 270–279.
202 T. K. Dam and C. F. Brewer, *Adv. Carbohydr. Chem. Biochem.*, 2010, **63**, 139–164.
203 G. Simone, P. Neuzil, G. Perozziello, M. Francardi, N. Malara, E. Di Fabrizio and A. Manz, *Lab Chip*, 2012, **12**, 1500–1507.
204 S. G. Lee, J. M. Brown, C. J. Rogers, J. B. Matson, C. Krishnamurthy, M. Rawat and L. C. Hsieh-Wilson, *Chem. Sci.*, 2010, **1**, 322–325.
205 A. Moise, S. André, F. Eggers, M. Krzeminski, M. Przybylski and H. J. Gabius, *J. Am. Chem. Soc.*, 2011, **133**, 14844–14847.
206 F. Real-Fernández, I. Passalacqua, E. Peroni, M. Chelli, F. Lolli, A. M. Papini and P. Rovero, *Sensors*, 2012, **12**, 5596–5607.
207 G. Bellapadrona, A. B. Tesler, D. Grünstein, L. H. Hossain, R. Kikkeri, P. H. Seeberger, A. Vaskevich and I. Rubinstein, *Anal. Chem.*, 2012, **84**, 232–240.
208 C. Jiménez-Castells, S. Defaus, A. Moise, M. Przbylski, D. Andreu and R. Gutiérrez-Gallego, *Anal. Chem.*, 2012, **84**, 6515–6520.
209 S. H. Millen, D. M. Lewallen, A. B. Herr, S. S. Iyer and A. A. Weiss, *Biochemistry*, 2010, **49**, 5954–5967.
210 J. L. Baker, E. Celik and M. P. Delisa, *Trends Biotechnol.*, 2013, **31**, 313–323.
211 C. Steentoft, S. Y. Vakhrushev, M. B. Vester-Christensen, K. T. Schjoldager, Y. Kong, E. P. Bennett, U. Mandel, H. Wandall, S. B. Levery and H. Clausen, *Nat. Methods*, 2011, **8**, 977–982.
212 K. T. Schjoldager, S. Y. Vakhrushev, Y. Kong, C. Steentoft, A. S. Nudelman, N. B. Pedersen, H. H. Wandall, U. Mandel, E. P. Bennett, S. B. Levery and H. Clausen, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 9893–9898.
213 L. X. Wang and J. V. Lomino, *ACS Chem. Biol.*, 2012, **7**, 110–122.
214 S. André, T. Kozár, R. Schuberth, C. Unverzagt, S. Kojima and H. J. Gabius, *Biochemistry*, 2007, **46**, 6984–6995.
215 S. André, T. Kozár, S. Kojima, C. Unverzagt and H. J. Gabius, *Biol. Chem.*, 2009, **390**, 557–565.
216 J. E. Rexach, C. J. Rogers, S. H. Yu, J. Tao, Y. E. Sun and L. C. Hsieh-Wilson, *Nat. Chem. Biol.*, 2010, **6**, 645–651.
217 J. L. Chaubard, C. Krishnamurthy, W. Yi, D. F. Smith and L. C. Hsieh-Wilson, *J. Am. Chem. Soc.*, 2012, **134**, 4489–4492.
218 L. Yang, J. O. Nyalwidhe, S. Guo, R. R. Drake and O. J. Semmes, *Mol. Cell. Proteomics*, 2011, **10**, M110.
219 T. N. Ramya, E. Weerapana, B. F. Cravatt and J. C. Paulson, *Glycobiology*, 2013, **23**, 211–221.
220 R. Orlando, J. M. Lim, J. A. Atwood, III, P. M. Angel, M. Fang, K. Aoki, G. Alvarez-Manilla, K. W. Moremen, W. S. York, M. Tiemeyer, M. Pierce, S. Dalton and L. Wells, *J. Proteome Res.*, 2009, **8**, 3816–3823.
221 K. Tanaka, K. Moriwaki, S. Yokoi, K. Koyama, E. Miyoshi and K. Fukase, *Bioorg. Med. Chem.*, 2013, **21**, 1074–1077.
222 V. Zoldos, T. Horvat and G. Lauc, *Curr. Opin. Chem. Biol.*, 2012, **17**, 34–40.
223 S. Neelamegham and G. Liu, *Glycobiology*, 2011, **21**, 1541–1553.