

Analyst

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Mammalian Protein Glycosylation – Structure versus Function

S. Defaus^{a,b}, P. Gupta^b, D. Andreu^a, R. Gutiérrez-Gallego^{a,b,c*}

^a S. Defaus, D. Andreu, R. Gutiérrez-Gallego, Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona Biomedical Research Park, 08003 Barcelona, Spain

^b S. Defaus, P. Gupta, R. Gutiérrez-Gallego, Bio-analysis group, Neuroscience Research Program, IMIM-Parc Salut Mar, Barcelona Biomedical Research Park, 08003 Barcelona, Spain

^c R. Gutiérrez-Gallego, Anapharm Biotech, 08038 Barcelona, Spain

* Corresponding author: R. Gutiérrez-Gallego E-mail: ricardogutierrez@upf.edu.

Summary

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 carbohydrates 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 function.

Abbreviations

| | |
|-------------|---|
| 2AB | 2-amino benzamide |
| AChE | acetylcholinesterase |
| AD | Alzheimer disease |
| α DG | 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 |

| | | |
|----|--------|--|
| 1 | | |
| 2 | | |
| 3 | ETD | electron transfer dissociation |
| 4 | FAB | fast atom bombardment |
| 5 | | |
| 6 | FASP | filter aided sample preparation |
| 7 | | |
| 8 | FID | flame ionization detector |
| 9 | | |
| 10 | Fuc | fucose |
| 11 | | |
| 12 | GAGs | glycosaminoglycans |
| 13 | | |
| 14 | Gal | galactose |
| 15 | | |
| 16 | GalNAc | <i>N</i> -acetylgalactosamine |
| 17 | | |
| 18 | GC | gas chromatography |
| 19 | | |
| 20 | GCMS | gas chromatography mass spectrometry |
| 21 | | |
| 22 | GH | growth hormone |
| 23 | | |
| 24 | Glc | glucose |
| 25 | | |
| 26 | GlcA | glucuronic acid |
| 27 | | |
| 28 | GlcNAc | <i>N</i> -acetylglucosamine |
| 29 | | |
| 30 | Gln | glutamine |
| 31 | | |
| 32 | Glu | glutamic acid |
| 33 | | |
| 34 | GnT-V | <i>N</i> -acetylglucosaminyl transferase V |
| 35 | | |
| 36 | GT | glycosyltransferase |
| 37 | | |
| 38 | GU | glucose unit |
| 39 | | |
| 40 | HCC | hepatocellular carcinoma |
| 41 | | |
| 42 | HCD | higher-energy collisional dissociation |
| 43 | | |
| 44 | HCl | hydrochloric acid |
| 45 | | |
| 46 | HILIC | hydrophilic interaction liquid chromatography |
| 47 | | |
| 48 | HNK-1 | human natural killer epitope 1 |
| 49 | | |
| 50 | Hp | haptoglobin |
| 51 | | |
| 52 | HPAEC | high performance anion exchange chromatography |
| 53 | | |
| 54 | HPAEC | high pH anion exchange chromatography |
| 55 | | |
| 56 | HPLC | high performance liquid chromatography |
| 57 | | |
| 58 | HS | Heparin sulfate |
| 59 | | |
| 60 | HSA | human serum albumin |
| | | |
| | ICAM | inter cellular adhesion molecule |

| | | |
|----|-----------|--|
| 1 | | |
| 2 | | |
| 3 | IDAWG | isotopic detection of aminosugars with glutamine |
| 4 | IEF | isoelectric focusing |
| 5 | | |
| 6 | INPEG | in-gel non-specific proteolysis for elucidating glycoproteins |
| 7 | | |
| 8 | ITC | isothermal titration calorimetry |
| 9 | | |
| 10 | LC | liquid chromatography |
| 11 | | |
| 12 | LC-FLC | liquid chromatography with post-column fluorescence derivatization |
| 13 | | |
| 14 | Lewis A | Gal β 1-3(Fuca α 1-4)GlcNAc |
| 15 | Lewis B | (Fuca α 1-2)Gal β 1-3(Fuca α 1-4)GlcNAc |
| 16 | | |
| 17 | Lewis X | Gal β 1-4(Fuca α 1-3)GlcNAc |
| 18 | | |
| 19 | Lewis Y | (Fuca α 1-2)Gal β 1-4(Fuca α 1-3)GlcNAc |
| 20 | | |
| 21 | LIF | laser-induced fluorescence |
| 22 | | |
| 23 | LIF | leukemia inhibitory factor |
| 24 | | |
| 25 | Lys | lysine |
| 26 | | |
| 27 | M | molar - mol per liter |
| 28 | | |
| 29 | M6P | mannose-6-phosphate |
| 30 | | |
| 31 | MAA | <i>Maackia amurensis</i> agglutinin |
| 32 | | |
| 33 | MALDI-TOF | matrix assisted laser desorption ionization time of flight |
| 34 | | |
| 35 | Man | mannose |
| 36 | | |
| 37 | ManNAc | <i>N</i> -acetylmannosamine |
| 38 | | |
| 39 | mM | millimolar |
| 40 | | |
| 41 | MRM | multiple reaction monitoring |
| 42 | | |
| 43 | MS | mass spectrometry |
| 44 | | |
| 45 | NCAM | neural cell adhesion molecule |
| 46 | | |
| 47 | NCBI | national center for biotechnology information |
| 48 | | |
| 49 | NeuAc | <i>N</i> -acetylneuraminic acid |
| 50 | | |
| 51 | NIBRT | national institute for bioprocessing and training |
| 52 | | |
| 53 | NK | natural killer |
| 54 | | |
| 55 | nm | nanometer |
| 56 | | |
| 57 | NMR | nuclear magnetic resonance |
| 58 | | |
| 59 | NP | normal-phase |
| 60 | | |
| | OXM | oxyntomodulin |

| | | |
|----|----------------|---|
| 1 | | |
| 2 | | |
| 3 | PAD | pulsed amperometric detection |
| 4 | PDIA | protein disulfide isomerase |
| 5 | | |
| 6 | PECAM | platelet endothelial cell adhesion molecule |
| 7 | | |
| 8 | PFK | phosphofructokinase |
| 9 | | |
| 10 | PGMs | post-glycosylational modifications |
| 11 | PMAA | partially methylated alditol acetate |
| 12 | | |
| 13 | PNGase F | Peptide-N4-(N-acetyl-beta-glucosaminy)asparagine amidase |
| 14 | | |
| 15 | Pro | proline |
| 16 | | |
| 17 | PTC | papillary thyroid carcinoma |
| 18 | | |
| 19 | PTM | post-translational modification |
| 20 | | |
| 21 | Q-TOF | quadrupole time-of-flight |
| 22 | | |
| 23 | RP | reverse-phase |
| 24 | SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| 25 | | |
| 26 | Ser | serine |
| 27 | | |
| 28 | Sia | sialic acid |
| 29 | | |
| 30 | Sialyl-Lewis X | Neu5Ac α 2-3Gal β 1-4(Fuca1-3)GlcNAc |
| 31 | SNA | <i>Sambucus nigra</i> agglutinin |
| 32 | | |
| 33 | SPR | surface plasmon resonance |
| 34 | | |
| 35 | S-Tn | Neu5Ac α 2-6GalNAc α 1-O-Ser/Thr |
| 36 | T antigen | Gal β 1-3-GalNAc α -O-Ser/Thr |
| 37 | | |
| 38 | Thr | threonine |
| 39 | | |
| 40 | Tn antigen | GalNAc α 1-O-Ser/Thr |
| 41 | | |
| 42 | Tyr | tyrosine |
| 43 | | |
| 44 | UC | ulcerative colitis |
| 45 | | |
| 46 | UPLC | ultra performance liquid chromatography |
| 47 | | |
| 48 | VCAM | vascular cell adhesion molecule |
| 49 | | |
| 50 | WAX | weak anion exchange |
| 51 | | |
| 52 | Xyl | xylose |
| 53 | | |
| 54 | ZFN | zinc-finger nuclease |
| 55 | | |
| 56 | | |
| 57 | | |
| 58 | | |
| 59 | | |
| 60 | | |

Content

| | |
|---|-----------|
| 1. Introduction..... | 7 |
| 2. Structural analysis of glycoprotein glycans..... | 7 |
| 2.1. Analyzing glycosylation..... | 11 |
| 2.1.1. Analysis of intact glycoproteins..... | 11 |
| 2.1.2. Analysis of glycoprotein glycans..... | 15 |
| 2.1.3. Analysis of site-specific glycosylation..... | 21 |
| 3. Functional analysis of glycans..... | 22 |
| 3.1. Glycosyltransferases..... | 23 |
| 3.2. Carbohydrate determinants (glycotopes)..... | 23 |
| 3.2.1. Monosaccharide constituents..... | 24 |
| 3.2.2. Oligosaccharides..... | 26 |
| 3.3. Glycosylation site occupancy..... | 27 |
| 3.4. Site-specific glycosylation..... | 29 |
| 3.5. Modifications of carbohydrates..... | 30 |
| 4. Simultaneous structural and functional analysis of glycans..... | 33 |
| 4.1. Analysis through biomolecular interactions..... | 34 |
| 4.2. Glycoengineering (genetic, chemoenzymatic, chemical)..... | 36 |
| 4.3. Evaluating glycosylation within the biological context..... | 37 |
| 5. Outlook..... | 39 |

1. Introduction

The world of carbohydrates is extremely complex, rendering it both fascinating and awesome 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 decades in terms of attention paid, importance attributed, and level of understanding reached. Recently, 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 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 on its relevance. A 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 system¹. 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 B², or rather complex as in the case of CD59 where at least 123 different desialylated glycan variants have been identified at a single site³. 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, linkages⁴, 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 principle⁵. 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.

| LINKAGE | TYPE | SEQUENCE |
|---------------------------|-----------------|---|
| Man- α -Trp | C-mannosylation | W-X-X-W |
| GlcNAc- β -Asn | N-glycosylation | N-X-(S/T) (X \neq 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-P-T-X-(T/X)-X-X |
| Man- α -1-P-Ser | | S rich domains |
| Xyl- β -Ser | | -G-S-G- (near acidic residues) |

X may be any amino acid

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 cells⁶. 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 of how glycans are directly involved in almost every biological process or human disease⁷. 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 cell or tissue type, but also with developmental stage⁸, metabolic state, or changes such as disease⁹, aging^{10,11}, environmental factors¹², 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 information¹⁵. It is therefore of utmost importance to know what carbohydrate structures decorate which glycoproteins under particular conditions.

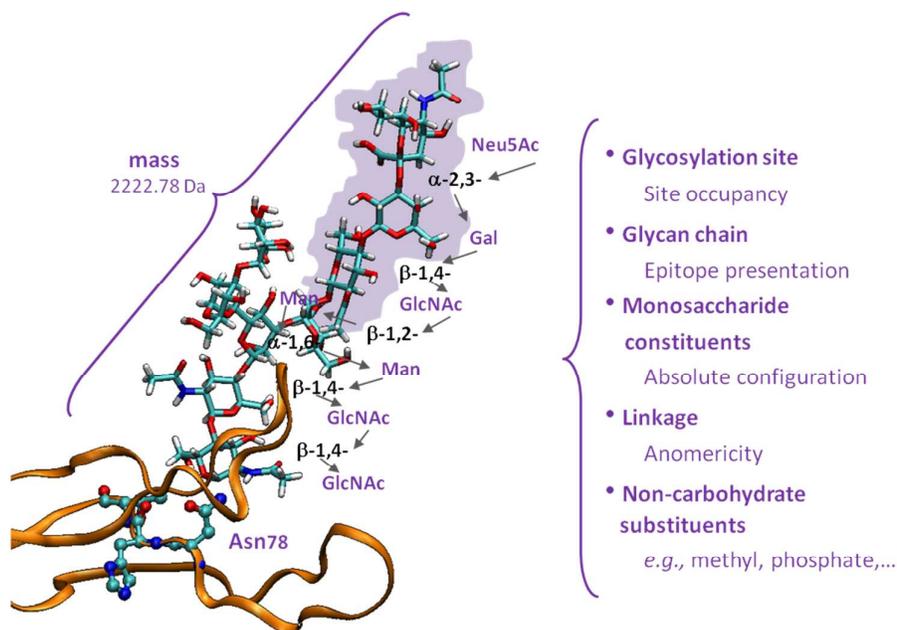
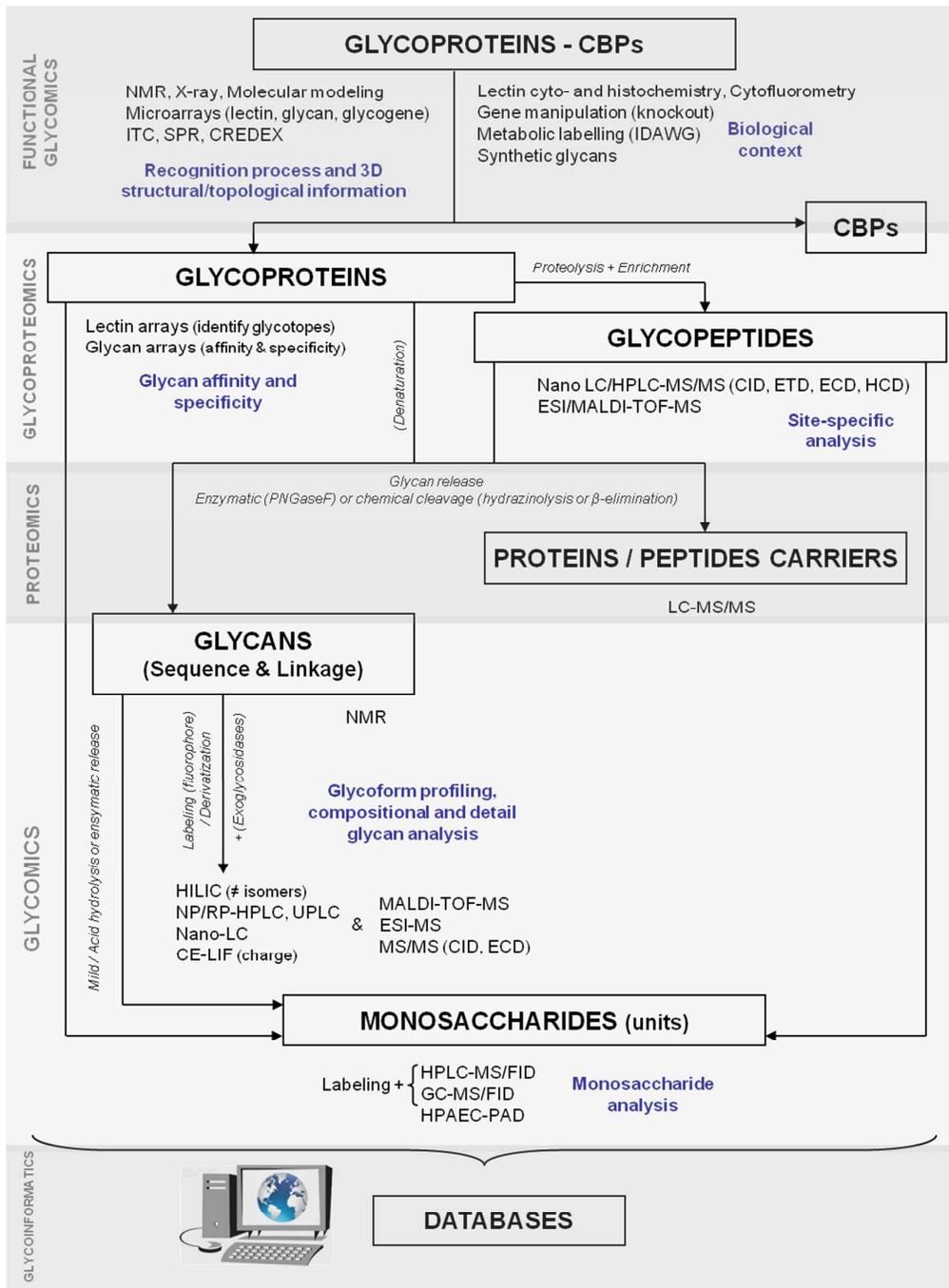


Figure 1. As an example in the structural elucidation of glycoproteins an N-glycan in human chorionic gonadotropin (hCG) glycoprotein α -chain is shown. Elements to be specified are listed on the right and some of them displayed. The shaded part represents the epitope potentially recognized in a carbohydrate-driven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

Even for dedicated specialist 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 (figure 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.

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.



1
2
3 In the following pages an overview is provided of the techniques employed in structural analysis
4 of protein glycosylation, their shortcomings and particular virtues, and the latest trends in this
5 field.
6

7 8 2.1. Analyzing glycosylation 9

10 Over the last 2-3 decades the continuous refinement of analytical tools has greatly facilitated
11 glycan analysis; numerous reviews¹⁶⁻²¹ and papers cover the main technologies routinely used
12 today for N- and O-linked glycan analysis, including capillary electrophoresis (CE)²²⁻²⁴, liquid
13 chromatography (LC)^{25,26}, mass spectrometry (MS)²⁷⁻³⁰ and microarray-based³¹⁻³⁵ approaches to
14 glycomics and glycoproteomics^{28,36,37}. It is important to stress that in all these techniques a
15 compromise exists between analytical sensitivity and the degree of structural detail provided.
16 None of these tools, or any other for that matter, can singlehandedly reveal all the features (see
17 figure 1) necessary for full characterization. Hence, an unambiguously structural analysis must
18 be conducted at different levels, namely intact glycoprotein, glycopeptides and released
19 glycans, and in each case the most appropriate technique for deciphering that part of the puzzle
20 must be chosen. This, in turn, entails another compromise between the degree of information
21 obtained vs the amount of (purified) material required.
22
23
24
25
26

27 28 2.1.1. Analysis of intact glycoproteins 29

30 In the first evaluation of protein glycosylation it is recommended to assess the
31 microheterogeneity at the glycoprotein level as it provides an excellent starting point. Quite
32 often this is done by means of conventional sodium dodecyl sulfate polyacrylamide gel
33 electrophoresis (SDS-PAGE) and comparing the resulting bands to a protein molecular weight
34 standard (figure 5). Such evaluation, when conducted with non-specific staining techniques
35 such as Coomassie, Silver, or Pro-Q emerald dyes, should provide an unbiased view of the
36 glycoform distribution. Alternatively, the detection could be performed through specific
37 biomolecular recognition (using lectins, antibodies, *etc.*) at much better sensitivity than the non-
38 specific staining. However, one should bear in mind that such biorecognition may be biased
39 towards particular glycoforms because of steric effects or other factors hampering interactions
40 with other glycoforms. For its part, SDS-PAGE does not provide an accurate molecular weight
41 determination as separation is governed by the hydrodynamic volume of the migrating species.
42 One alternative technique is gel-based isoelectric focusing (IEF) (figure 2). This technique
43 gained much momentum in the early days of proteomics as part of the two-dimensional gel-
44 electrophoretic sample preparation and provides a rough charge distribution of the glycoprotein.
45 Given the limited number of charged modifications of amino acid residues, the IEF profile
46 usually provides, a reliable sketch of the degree of sialylation, sulfation, phosphorylation and/or
47 glucuronidation of the glycoprotein.
48
49
50
51
52
53
54
55
56
57
58
59
60

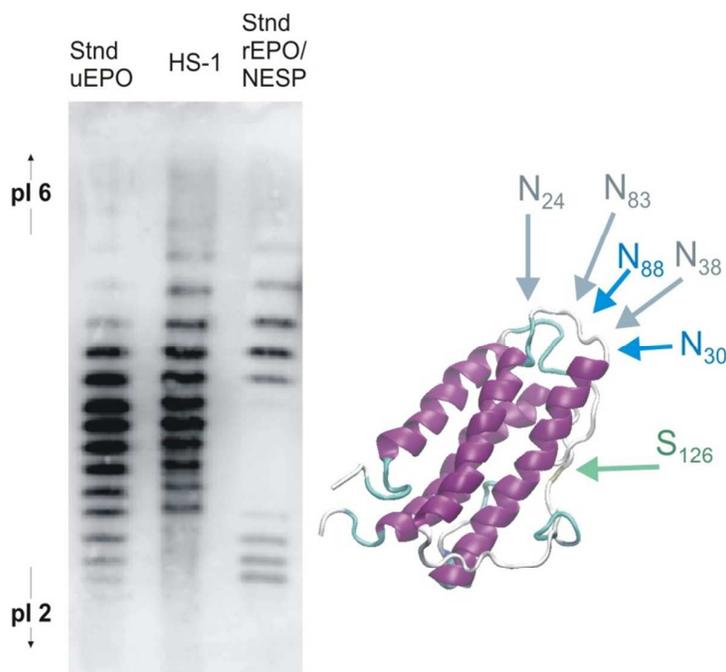


Figure 2. Isoelectric focusing profiles of an endogenous erythropoietin (EPO) standard (left lane), EPO from a human urinary sample (center lane) and a recombinant erythropoietin mixture (right lane) composed of Eprex (3 N-glycans and migrating just below pI 6) and Darbepoietin alpha (5 N glycans and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

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 suppression³⁸. Figure 3 compares the MALDI-TOF spectra of a glycoprotein (rAT-III) and a non-glycosylated protein (rGH) and shows how peak width provides information on the heterogeneity of the protein, and 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.

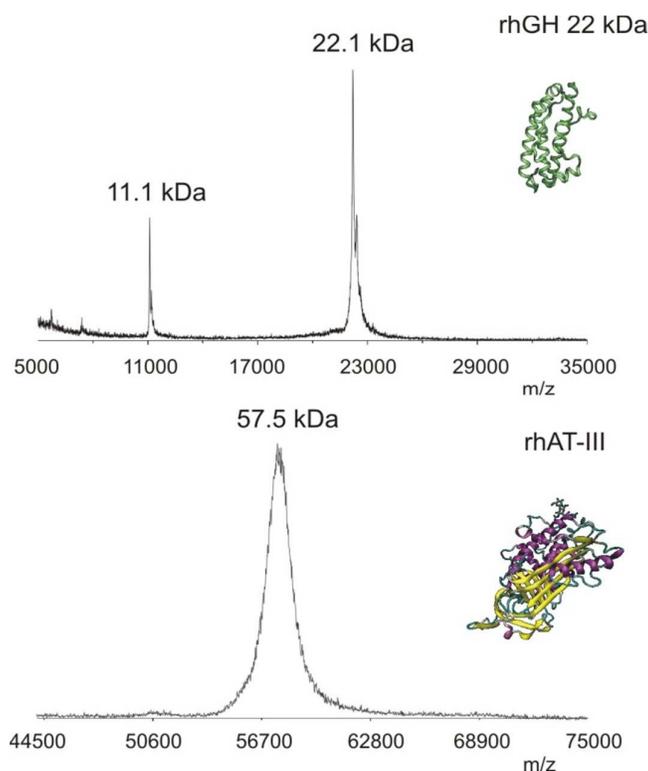


Figure 3. Top: MALDI-TOF spectrum of recombinant human growth hormone (not glycosylated); bottom; recombinant human antithrombin-III (tri/tetra glycosylated). Microheterogeneity due to glycosylation can be clearly appreciated from peak width.

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 ago³⁹ 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 residues⁴⁰, with a variety of associated functions⁴¹, 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-O7⁴³. For example, the detection of *N*-glycolyl neuraminic acid in erythropoietin –present at picomolar concentrations in human specimens– is an unambiguous evidence for a doping violation that could be established by this protocol⁴⁴ but not with MS analysis of the same sample (personal communication). Hence, obtaining a complete picture in

1
2
3 terms of Sia speciation will usually require the DMB protocol, though care must be exercised as,
4 in addition to the release-related degradation, other α -keto acids or 1,2-diketones in biological
5 samples, *e.g.*, α -ketoglutaric, pyruvic or p-hydroxyphenyl-pyruvic acids, can interfere. In such
6 cases, hyphenation of liquid chromatography with fluorescence detection (LC-FLD) to mass
7 spectrometry is possibly the best solution.
8
9

10
11 The next level of analysis, namely determining all monosaccharides present in a glycoprotein,
12 usually requires strong acid (*e.g.*, 1M HCl in methanol, 65°C, overnight, in the presence of an
13 internal standard) to hydrolyze all glycosidic linkages –except that between the first GlcNAc
14 residue and Asn in N-linked glycosylation– and convert glycosidic acetals into the
15 corresponding methylglycosides. The procedure, however, will also irretrievably cleave most
16 post-glycosylational modifications. Following neutralization and evaporation, free hydroxyls are
17 further derivatized with trimethylsilyl (or analogous) functionalities for both qualitative (four
18 characteristic peaks for each monosaccharide) and quantitative evaluation using gas
19 chromatography flame ionization detection (GC-FID) or gas chromatography mass spectrometry
20 (GC-MS) and comparison with a standard monosaccharide mix^{45,46}. Alternatively, following acid
21 release, monosaccharides may be separated chromatographically by high performance anion
22 exchange chromatography (HPAEC) (CarboPack PA-100) and detected by pulsed
23 amperometric detection (PAD)⁴⁷. The latter protocol offers the advantage of a single peak per
24 monosaccharide and of direct analysis without derivatization, but the basic LC conditions may
25 induce C2-epimerization in GlcNAc to yield ManNAc⁴⁸, or peeling reactions where some
26 monosaccharides are degraded from the reducing end. Altogether, monosaccharide analysis
27 offers the possibility of identifying which type of glycosylation is present; Man and GalNAc being
28 representative of N- and O-glycosylation, respectively. Furthermore, the stoichiometry of the
29 different sugars allows an educated guess on the type of N-linked glycans present by
30 considering the ratio between the distinct monosaccharides with respect to Man. A similar
31 approach can be employed to estimate substitution profiles (*i.e.* $\alpha\beta$ 1-2,3,4,6) in glycans. In this
32 case, carbohydrates and other functional moieties are permethylated using the Hakomori
33 protocol⁴⁹, subsequently the monosaccharides are released by acid hydrolysis (leaving the
34 methyl-ether bonds intact) and the resulting hydrolyzed monomers are reduced and acetylated
35 to give volatile, partially methylated alditol acetates (PMAA), again analyzed by GC-MS. This
36 procedure provides unambiguous information on the linkage pattern as well as the ring size of
37 the corresponding sugar, but it is important to emphasize that it is unable to distinguish between
38 a 4-O-linked aldopyranose and a 5-O-linked aldofuranose.
39
40
41
42
43
44
45
46
47
48
49

50
51 A more recent development to assess glycosylation at the glycoprotein level consists in the
52 interrogation of a particular glycoprotein or complex biological sample through a lectin array.
53 Even though lectins have long been recognized as tools in the study of glycosylation, their
54 systematic application in array format to detect the glycotopes in a given sample is relatively
55 new⁵⁰⁻⁵³. Even if the information obtained cannot be compared with a thorough structural
56 analysis (*vide infra*), it has the advantage of analyzing a crude biological sample, *e.g.* the
57
58
59
60

1
2
3 cellular glycome⁵⁴, without too much manipulation, and has demonstrated its value in assessing
4 glycosylation changes in cancer cells on the basis of a direct or an antibody-assisted
5 evanescent-field fluorescence detection scanner^{55,56}. In addition, combination with antibodies
6 allows the changes in glycosylation to be pinpointed to specific proteins⁵⁷, adding one more
7 level of specificity to the analysis. Hence, dynamic glycome analysis can be undertaken by
8 means of differentially labeled CBPs⁵⁸. Even so, there are several drawbacks to lectin arrays.
9 For one, while current plant lectin-based arrays³² cover most monosaccharides in the
10 mammalian glycome, mammalian lectins would obviously provide a more representative
11 glycoprofile. Also, one should not ignore that most lectins are promiscuous to a certain degree
12 and that this behavior, different for each lectin and with different affinities for different sugars,
13 will complicate glycome readout. Ultimately, it appears that lectin-carbohydrate interactions are
14 not always straightforward and that glycoclusters, of either homo- or heterogenic nature, will
15 strongly influence the interaction, and in this case, the analytical data^{5,59}. The latter
16 phenomenon appears to be, at least in part, responsible for the fine-tuning of biological
17 communication processes and will as such be very difficult to interpret in terms of precise
18 structural entities.
19
20
21
22
23
24

25 26 *2.1.2. Analysis of glycoprotein glycans*

27
28 Evaluation of intact glycans almost inevitably requires their release from the peptide backbone.
29 While some high resolution approaches, in particular those based on MS, are capable of
30 addressing microheterogeneity at the intact glycoprotein level, this is restricted to those entities
31 with a very limited number of glycans and glycoforms such as apolipoprotein C3⁶⁰. For more
32 complex entities, separation of carbohydrate from protein backbone is needed. Since, in this
33 process, both the site-specificity and the protein origin of the glycans are lost, it is crucial to
34 ensure the maximum degree of protein purity before the procedure is initiated. Deglycosylation
35 can be achieved by either chemical or biological means, each with their respective
36 dis/advantages.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

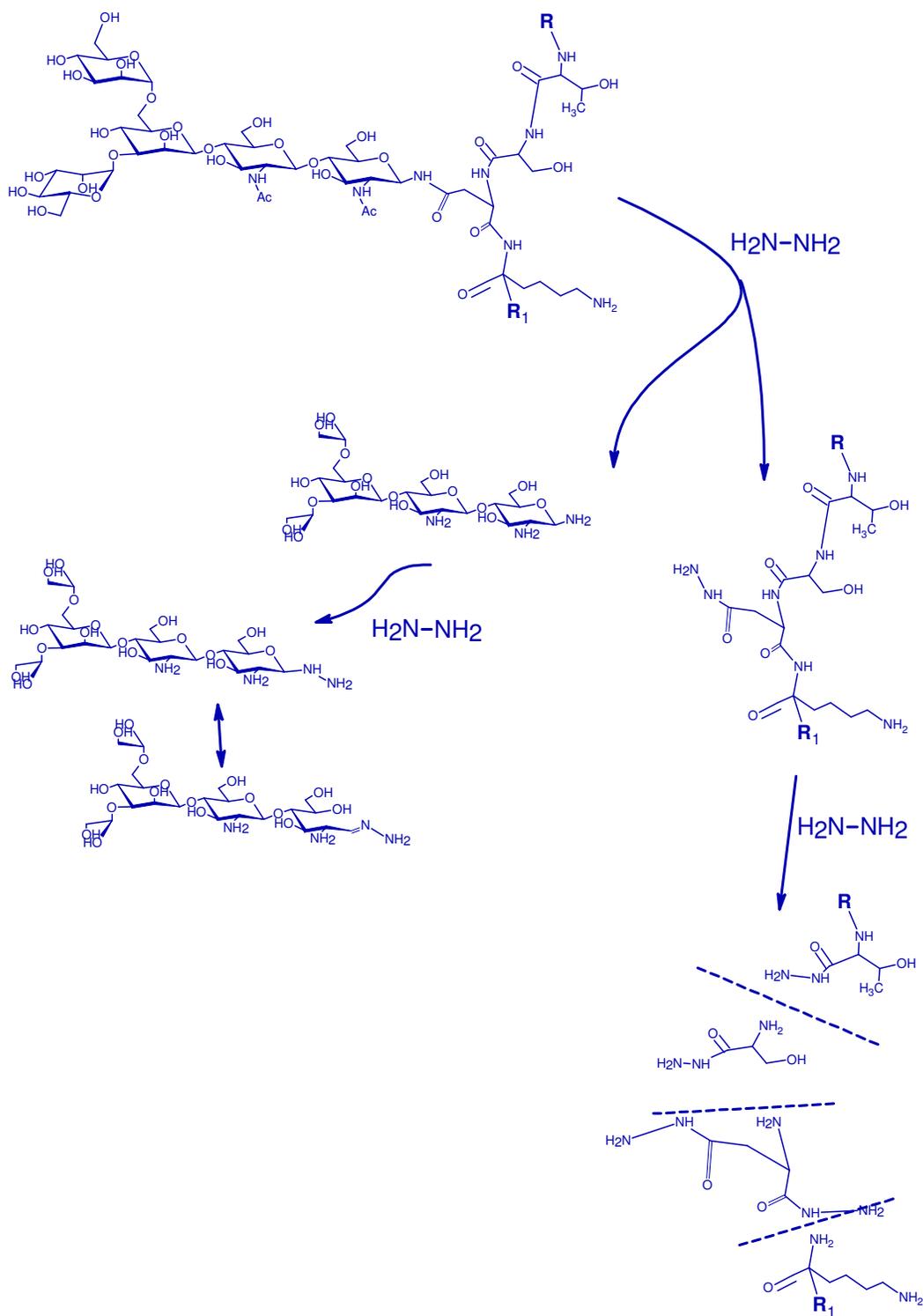


Figure 4. Release of carbohydrates from the protein backbone following hydrazine treatment. An undesired side effect is protein destruction as indicated at right bottom.

The most widely used chemical method is hydrazinolysis, a procedure that releases the two major classes of glycans (figure 4) yet requires highly skilled staff and strict conditions for

1
2
3 success, and is invariably accompanied by side-reactions and byproducts. In addition, re-
4 acetylation is necessary to avoid N-glycans being lost during the process but may also induce
5 O-acetylation. Selective and sequential release of oligosaccharides is achieved by mild
6 hydrazinolysis of O-linked oligosaccharides at 60°C, followed by that of N-linked
7 oligosaccharides at 95°C, but there may be a significant overlap between both processes
8 depending on the protein and the degree of glycosylation. In addition, the procedure will destroy
9 the protein backbone so that if both glycans and protein sequence are to be investigated,
10 hydrazinolysis is not the method of choice. Another chemical procedure, *i.e.*, alkaline β -
11 elimination (0.05 to 0.1 M NaOH or KOH, 60°C, 12 h), can be applied for O-linked
12 carbohydrates attached to Ser or Thr (except those at the carboxy-terminus), but not to Tyr,
13 hydroxy-Pro or hydroxy-Lys. In this case, N-linked carbohydrates are unaffected. To prevent
14 base-catalyzed peeling (*vide supra*), sugars must be immediately captured⁶¹, the alkaline
15 solution carefully prepared⁶², or a reducing agent (*e.g.* 1 M NaBH₄) added which forms an alditol
16 that precludes reducing-end derivatization. If tagging is intended, it is best performed during
17 release⁶³. As with hydrazinolysis, the protein backbone is destroyed in the process.

24 The only strategy that preserves both protein and carbohydrate is enzymatic deglycosylation,
25 which has been successfully developed for N-linked glycans using several endoglycosidases.
26 For mammalian glycoproteins, peptide-N4-(N-acetyl-beta-glucosaminy)asparagine amidase
27 (PNGase F) is the enzyme of choice; it liberates nearly all N-linked carbohydrates under
28 standard conditions (*e.g.*, phosphate buffer, 50 mM pH 7.3, 16h, 37°C). Nevertheless, the
29 efficiency of this procedure needs to be checked to ensure a correct assessment of the
30 subsequent analysis. One example where conventional conditions do not result in full release is
31 human antithrombin III (AT-III) (figure 5), where only partial release of the four N-linked
32 structures of the α -variant is achieved if the procedure is not optimized. For plant or invertebrate
33 glycosylations, PNGase A is the preferred choice. In contrast to PNGase F, this enzyme,
34 although of poorer overall efficiency, is capable of releasing α 1-3-fucose-bearing core
35 structures. Other endoglycanases (endo F1 to F3 or endo H) can be employed to release the
36 carbohydrate chains, except for Asn-bound GlcNAc, as the enzymes specifically target
37 chitobiose units. In sum, the full repertoire of N-glycans can be released by enzymatic means
38 but caution is still advisable. Co- and post-release glycosylamine modification to functionalities
39 other than C1-hydroxyl, such as urea⁶⁴, glycerol⁶⁵ or thiol⁶⁶ or incomplete conversion in the
40 presence of ammonium may obscure the final analysis. For O-glycans, one single
41 deglycosylation enzyme has been identified thus far⁶⁷, and its activity is restricted to T- and Tn-
42 antigenic structures on Ser or Thr. As such, its main application is in unveiling the presence of
43 either epitope, without providing further evidence on the presence of other O-glycans.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

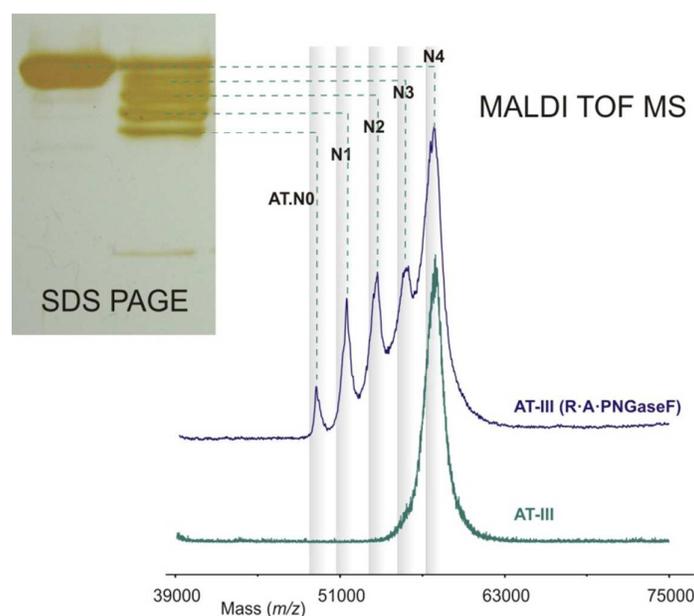


Figure 5. SDS-PAGE and MALDI-TOF-MS analysis of human antithrombin III (AT-III) after conventional PNGaseF de-N-glycosylation. From the number of bands in the SDS-PAGE and peaks in the MALDI-TOF spectrum it is evident that deglycosylation is inefficient and does not reach completion.

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}, HPAEC²⁶, 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), hydrophobicity⁶⁹ or hydrophilic interaction (HILIC), and can be performed on either conventional, micro- or nanosized platforms²⁵. 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-MS⁷⁰ 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 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

1
2
3 unit (GU) values^{73,74} allow a preliminary assignment that can then be corroborated by targeted
4 and sequential exoglycosidase digestions, followed by another round of HPLC profiling.
5 Subsequent glycan trimming is of particular interest as not only does it provide confirmation of
6 structural assignments, it simplifies the glycan pool, ultimately contributing to unveiling epitopes
7 that are obscured in the overall microheterogeneity⁷⁵. Fluorescent labeling at the reducing end
8 is not restricted to 2AB as several other tags are described⁷⁶ and not only reduces sample
9 requirement to the low femtomole level of individual structures⁷⁷, it also allows accurate relative
10 and absolute quantitation of the glycans present in a given glycoprotein⁷⁸. It has become one of
11 the standard techniques in carbohydrate profiling^{71,79}, which can be amplified with internal
12 standards if a different fluorescent tag is used for dextran and the sample, and can be easily
13 extended with back-end MS evaluation when the material is not required for further
14 exoglycosidase treatment⁸⁰. Despite these advantages, it is a laborious approach that requires
15 considerable care, especially during the 2AB labeling (in 30% acetic acid, 65°C) to avoid
16 desialylation that may confound structural assignments⁷⁸. Automated sample preparation, *i.e.*
17 both the fluorescent labeling and the post-release and post-labeling purification steps, greatly
18 reduces analysis variability, providing robust and reliable glycomics data^{16,19,81}.

19
20
21 Arguably one of the more powerful and versatile analytical techniques for all sorts of
22 compounds, including carbohydrates, MS has become the cutting-edge technology for
23 glycomics, linking mass with composition and providing precise characterization of complex
24 structures. A wide range of MS equipments are available for glycan analysis. The introduction of
25 MALDI-TOF instruments allowed rapid and straightforward evaluation of complex mixtures^{82,83}
26 and was a giant leap forward in MS evaluation of carbohydrates, hitherto restricted to the
27 cumbersome, low-sensitivity fast atom bombardment mass spectrometry (FAB-MS). Improved
28 analyses are made possible through combination with well-known derivatization strategies (*e.g.*
29 permethylation or peracetylation) that reduce polarity and improve sensitivity by either MALDI-
30 TOF-MS⁸⁴ or LC-ESI-MS⁶⁹. Isotope-based differential derivatization protocols, *e.g.* using CH₃I
31 and CD₃I, allow exact determination of the number of free hydroxyls in a given structure, from
32 which valuable information on carbohydrate composition can be inferred⁸⁵⁻⁸⁷. When glycan
33 sequencing is the goal, analyses must include tandem MS experiments where structure-
34 revealing ions are obtained by a combination of ion activation/fragmentation strategies such as
35 collision-induced dissociation (CID), electron transfer dissociation (ETD) or electron capture
36 dissociation (ECD). For instance, a recent study involving a series of oligosaccharide-derived
37 oxonium fragment ions generated by CID enabled simultaneous characterization of IgG
38 glycoforms at both Fc and Fab glycosylation sites by combining multiple reaction monitoring
39 (MRM) MS with energy-resolved structural analysis⁸⁸. In another study, CID-MS was employed
40 to selectively monitor the generation of a *m/z* 284.053 fragment, consistent with GlcNAc
41 phosphorylation in a mouse brain dataset, and unveiled this new post-glycosylational
42 modification⁸⁹. Within the last years, ECD and ETD have enabled the assignment of *O*-GlcNAc
43 sites at the proteomic scale and greatly facilitated protein-specific studies of single *O*-GlcNAc
44 events. Particularly, ETD has been used to identify *O*-GlcNAc sites and PTMs such as

1
2
3 phosphorylation and Arg methylation, on host cell factor C1 (HCF-1), a chromatin-associated
4 protein involved in transcriptional regulation and cell proliferation, and one of the most highly *O*-
5 GlcNAc-rich proteins found in cells⁹⁰. Although MS is clearly indispensable in glycomic analysis,
6 some techniques still present limitations such as susceptibility to salts, difficult assignment of
7 isomeric and isobaric monosaccharides –even though the evolution of ion-mobility strategies
8 are addressing this⁹¹–, complicated behaviour of acyl groups on glycans, and ionization
9 efficiency dependence. Moreover, interpretation of MSⁿ fragmentation datasets remains a
10 limiting factor with regard to throughput, user-dependent variability in discrimination and/or
11 interpretation and complete identification of all glycoforms.
12
13
14
15

16 When sample complexity is limited to only a few glycan structures, the analytical technique of
17 choice is nuclear magnetic resonance (NMR), the only one providing both qualitative and
18 quantitative information on the glycan without being destructive. While mono-dimensional (¹H,
19 ¹⁵N, ¹³C) experiments readily provide information on structural reporter groups^{92,93}, multi-
20 dimensional, both homo- and heteronuclear experiments yield information on the spatial
21 orientation of the glycotope. Moreover, NMR may provide unambiguous information on the
22 presence and position of post-glycosylational modifications such as sulfation, methylation,
23 acetylation or phosphorylation⁹⁴. Ironically for a technique that had proven crucial in the early
24 development of the carbohydrate field, NMR had gradually lost influence due to the often
25 prohibitive amounts of natural material required. Nevertheless, recent developments enabling
26 analysis at picomoles⁹⁵ may reinvigorate a technique which in fact has never lost its appeal for
27 the analysis of carbohydrate biological interactions^{96,97} or the effect of changes in
28 glycosylation^{98,99}.
29
30
31
32
33
34

35 Altogether glycomics can be addressed through a variety of strategies and technologies that
36 turn out to be orthogonal rather than parallel. While all of them rapidly generate very large
37 amounts of data, differences between platforms can turn data analysis into a complex, time-
38 consuming task requiring bio-informatics tools and databases to facilitate data processing and
39 interpretation. Most of these glycoinformatic tools have particular focuses, *e.g.*, data from
40 HPLC⁷⁴, MS¹⁰⁰, NMR or microarray^{101,102} experiments. Initiatives for cross-linking data from
41 different techniques and integrating multiple data sets are prospering and extremely
42 useful^{103,104}, although in the use of database search outputs critical interrogation is advisable.
43 Additionally, the field of glycobiology would greatly benefit from a single glycan structural
44 annotation, easy and of worldwide access, and support by public agencies such as NCBI or
45 EMBL. Limited public initiatives in this regard (*e.g.* Consortium for Functional Glycomics -
46 <http://www.functionalglycomics.org/static/consortium/consortium.shtml>) are in risk of being
47 overshadowed by commercial enterprises (*e.g.* Waters & NIBRT -
48 <http://www.waters.com/waters/promotionDetail.htm?id=134654015>), most likely with ensuing
49 limitations in accessing data, let alone seeking a universal output.
50
51
52
53
54
55
56
57
58
59
60

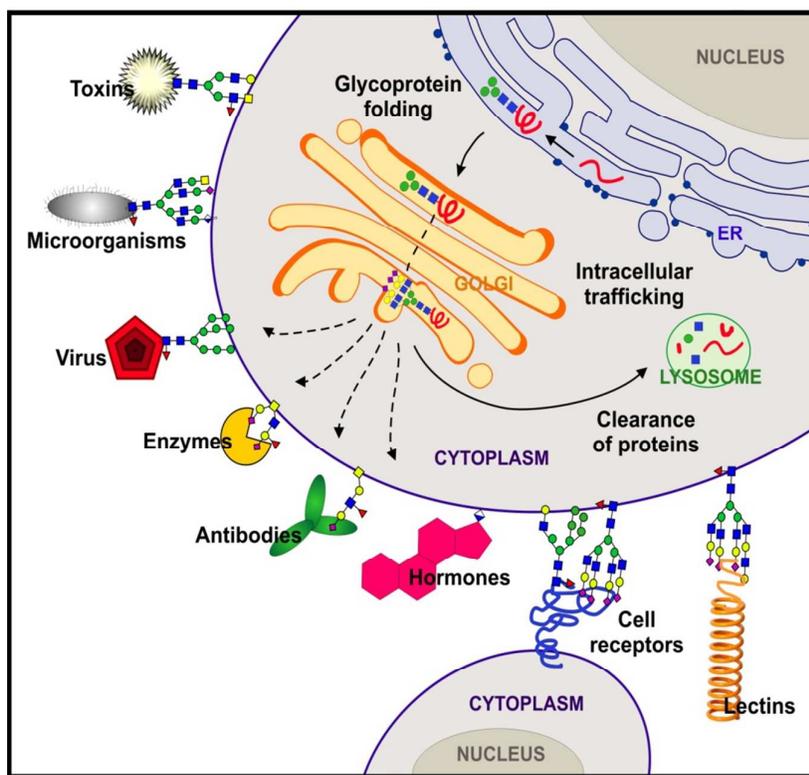
2.1.3. Analysis of site-specific glycosylation

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 obtained¹⁰⁵. 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 intensity¹⁰⁶. 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-C α bond, hence preserving the modification and allowing reliable analysis of both permanent¹⁰⁷ and transient glycosylation⁹⁰. These 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 organization¹⁰⁸. 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 evaluated¹⁰⁹. It seems clear that standardized analysis protocols^{79,110} as well as dedicated software applications¹⁰⁰ 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 peptidases¹¹¹, site-specific alteration of haptoglobin glycosylation related to hepatocarcinoma and liver cirrhosis¹¹², or how a particular congenital glycosylation disorder (CDG-Id) is associated with site-specific glycan deficiencies¹¹³.

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 (figure 6). These roles may be unrelated to the close structural environment where glycosylation occurs or, to the contrary, 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.

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 figure 1 and 7).



1
2
3 **Figure 6.** Glycans participate in multiple mechanisms of cellular regulation. The general
4 functions of glycans span from nascent protein folding and intracellular trafficking to roles in
5 extracellular compartments such as cell-cell communication, providing specific receptors for
6 noxious agents, protect from microorganisms and antibodies or regulate myriad receptor-ligand
7 interactions.
8
9

10 3.1. Glycosyltransferases

11
12 The majority of proteins synthesized in the rough ER undergo glycosylation and the
13 carbohydrate chains attached to these target proteins serve a variety of structural and functional
14 roles in membrane-anchored and secreted proteins. Glycosylation increases proteome diversity,
15 because almost every aspect of glycosylation can be modified, including glycan composition,
16 structure, bond and length.
17
18
19

20 The cellular glycome assembly *i.e.* the biosynthesis of disaccharides, oligosaccharides and
21 polysaccharides, involves the action of hundreds different glycosyltransferases (GTs), the
22 enzymes that catalyze the regio- and stereospecific transfer of sugar moieties from activated
23 donor molecules to a variety of acceptor biomolecules including glycans, lipids, peptides, and
24 small molecules forming glycosidic bonds¹¹⁸. The complex glycans synthesized by these
25 mammalian GTs are known to play crucial roles in cell–cell, cell–matrix and cell–pathogen
26 interactions, which impact growth and development, infection and immunity, signaling,
27 malignancy, and metabolic disorders. For instance, congenital disorders of glycosylation (CDG)
28 are genetic diseases causing defects in the synthesis or the attachment of the glycan moiety of
29 glycoproteins and glycolipids. Of the more than 40 CDG reported in humans, some 80% affect
30 the nervous system and no effective treatment is known for any of these disorders.
31
32
33
34
35

36 Given their importance in both normal development and pathological conditions, GTs are
37 targets for inhibition and specific small-molecules inhibitors have long been sought to
38 manipulate their activity in cells and to determine the functional roles of glycans. Although
39 recent, structural, kinetic and inhibitor studies have provided important information about the
40 evolution and reaction mechanism of GTs¹¹⁹, virtually nothing is known about their donor and
41 acceptor specificity. Therefore, functional characterization remains the greatest challenge in the
42 GT field as there is presently no easy way to assign functions to the many uncharacterized GTs.
43
44
45
46

47 3.2. Carbohydrate determinants (glycotopes)

48
49 While the complexity and diversity of the totality of glycan structures in an organism is almost
50 impossible to calculate, some 7,000 glycan determinants (glycotopes) recognized by CBPs
51 including lectins, receptors, toxins, antibodies, and enzymes has been reckoned for the human
52 glycome¹²⁰. This value is probably underestimated but it provides an idea of the dimension
53 generated by the approximately 700 proteins that make up the mammalian glycan repertoire,
54 and sets the boundaries for glycan-CBP interaction studies¹²¹ where the use of lectins,
55
56
57
58
59
60

1
2
3 receptors, antibodies, enzymes, and glycan microarray technologies is crucial for elucidating
4 carbohydrate-specific functions.
5

6 7 3.2.1. Monosaccharide constituents

8
9 In this text, “monosaccharide” refers to the simplest form of a sugar, found either as a stand-
10 alone residue or as a terminal or internal part of a polysaccharide. Sialic acids (Sias) are a
11 family of nine-carbon sugars typically attached to the outermost ends of glycoconjugate chains
12 as well as on secreted glycoproteins. The high prevalence of Sias terminating glycan extensions
13 suggests that their predominant function is modulating interactions with the environment. For
14 example, receptor 2B4 of human natural killer (NK) cells has sialic acid residues on both N- and
15 O-linked glycans. Removal of predicted 2B4 N-glycosylation sites decreases binding to its
16 ligand CD48 suggesting that N-linked sugars are essential for binding, yet sialylation of 2B4 has
17 a negative impact on ligand binding and 2B4-mediated NK cell cytotoxicity¹²². Similarly, Sias on
18 human corticosteroid-binding globulin (CBG) N-glycans were shown to modulate its function,
19 specifically by restricting the binding of CBG to its receptor through steric and/or electrostatic
20 means. Removal of CBG NeuAc residues, or the entire N-glycan, increased cAMP production
21 significantly, which was used to evaluate the CBG-receptor interaction¹²³.
22
23

24
25
26
27
28 O-glycosylation of the Notch extracellular domains in epidermal growth factor (EGF)-like repeats
29 is essential for activity, and tissue-specific alterations in the glycan structures are known to
30 regulate activity. As such, O-fucose and O-glucose-initiated glycans modulate Notch signaling
31 events critical to cell fate determination and tissue development. More specifically, O-fucose-
32 initiated glycans modulate the strength of Notch binding to DSL Notch ligands, while O-glucose-
33 initiated glycans facilitate juxta-membrane cleavage, generating the substrate for
34 intramembrane cleavage and Notch activation^{124,125}. Moreover, increasing both sialylation and
35 terminal α 1-3-linked fucosylation in N-glycans could lead to suppression of EGF receptor
36 (EGFR) dimerization and activation in lung cancer cells, which could in turn affect the metastatic
37 ability of cancer cells, EGFR-mediated signaling, and cellular behavior. In particular, the Sia and
38 Fuc residues in the Asn420 N-glycan could be critical in inhibiting EGFR dimerization and
39 phosphorylation. In contrast, core fucosylation would promote EGFR dimerization and
40 phosphorylation¹²⁶.
41
42

43
44
45
46
47 Another prominent example of O-glycosylation is the intracellular modification of cytoplasmic
48 and nuclear proteins with O-linked-N-acetylglucosamine (O-GlcNAc) that regulates basic and
49 multiple cellular functions such as transcription and translation, neuronal function, nutrient
50 sensing, cell cycle, and stress. Moreover, it is involved in the etiology of diabetes and
51 neurodegeneration¹²⁷. Indeed, CREB, a central transcription factor in the brain, is highly O-
52 GlcNAc monoglycosylated in neurons and influences gene expression by inhibiting both basal
53 and activity-induced CREB-mediated transcription, neuronal function regulation and long-term
54 memory¹²⁸. One of the earliest examples of O-GlcNAc modification was found over 25 years
55 ago in nuclear proteins¹²⁹, and since then numerous studies have suggested the existence of
56
57
58
59
60

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 types¹³⁰. 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 code¹³¹. 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 surface¹³³. 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 phosphorylation¹³⁶ and that combinations of both, *i.e.* *O*-GlcNAc-6-phosphate has been proposed recently as a novel PTM of mammalian proteins with a variety of possible cellular functions⁸⁹.

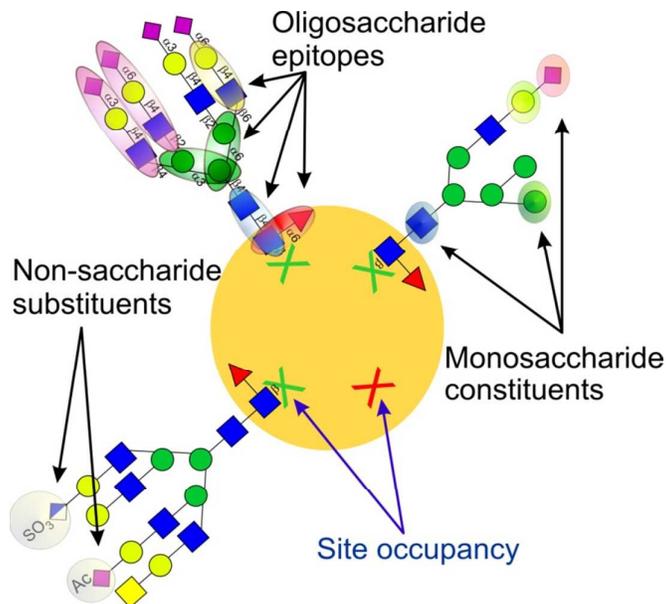


Figure 7. Different levels at which carbohydrates contribute to glycoconjugate heterogeneity: *i.e.* by occupancy, the monosaccharides that build-up the structure, the specific epitopes composed by the monosaccharides, and ultimately, the non-carbohydrate substituents.

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(α 1-2)Gal is of paramount importance. Of similar size is the Sd^a-antigen, comprising a Neu5Ac(α 2-3)[GalNAc(β 1-4)]Gal(β 1-R) trisaccharide, expressed in a donor-specific manner in males, and with no particular function hitherto attributed¹³⁹. 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 (β 4GalNAcT-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 β 4GalNAcT-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 homeostasis¹⁴¹. 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 response¹⁴². 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 corresponding N-glycan are expressed will efficient leukocyte adhesion be achieved¹⁴³. For example, the presence of polysialic acid, long chains of α 2-8-linked sialic acid residues, on neural cell adhesion molecule (NCAM) 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 migration¹⁴⁴. 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, SO₄-3GlcA(β 1-3)Gal(β 1-4)GlcNAc(β 1-R), that is expressed in a spatio-temporally

1
2
3 regulated manner during the development and maintenance of the peripheral nervous system.
4 Particularly, the single glycan moiety contained in P0 plays an important role in cell-cell
5 adhesion⁸².
6

7
8 Finally, in the phenomenon of carbohydrate-mediated biological recognition, an extra level of
9 complexity can be added when the carbohydrate binding event is potentiated by a multivalent
10 expression of glycotopes that result in stronger CBP recognition. This phenomenon has been
11 extensively studied under laboratory conditions^{59,145,146}, much less in biological settings. A clear
12 example is a recent study on the requirements for neuronal interactions and subsequent axon
13 growth, where clustered presentation of N-glycans with *N*-acetyllactosamine (LacNAc) epitopes
14 at branch ends of neural cell adhesion molecule L1 is required for neuronal galectin-4/L1
15 binding. Impairing the maturation of these epitopes precludes Gal-4/L1 association resulting in a
16 failure of L1 membrane cluster organization, required for proper axon growth¹⁴⁷.
17
18
19

20
21 The above mentioned examples are merely a glimpse of recent descriptive studies and illustrate
22 the increasing relevance of glycotopes. One hopes that the information flow will grow
23 exponentially to meet the vast challenge posed by glycomics and to establish a comprehensive
24 functional appreciation of the human glycome.
25
26

27 3.3. Glycosylation site occupancy

28
29 Glycosylation impacts significantly on the physico-chemical properties of the glycoprotein and
30 may thereby exert influence on its viability and activity. These effects are apparently
31 independent of the structural modification but the modification per se is necessary. One of the
32 better known examples in this respect is the folding of the nascent polypeptide chain where the
33 mono-glucosylated oligomannose structure serves as anchor point for the chaperone-assisted
34 event¹. Persistent failure to fold properly, possibly due to the absence of the carbohydrate chain,
35 ultimately results in lysosomal targeting. The common approach to site occupancy issues
36 involves studying the biological role after (enzymatic or chemical) glycan removal, or upon
37 inhibition of glycosylation, alteration of oligosaccharide processing, or elimination of specific
38 glycosylation sites. The consequences of altering, decreasing or abrogating glycan site
39 occupancy are variable and unpredictable, ranging from nearly undetectable to decreased
40 protein function, production level, stability, or even complete loss of function; in tune with this,
41 the functional interpretation of the absence/defects in glycosylation is not always
42 straightforward.
43
44
45
46
47
48

49
50 Several recent reports stress the need of glycosylation for viability; for example, in human
51 chymotrypsin C (CTRC) it is required for efficient folding and secretion. Elimination of N-
52 glycosylation by mutation of the single glycan (N52S) reduced CTCRC secretion about 10-fold¹⁴⁸.
53 Similarly, OATP1B1, an organic anion transporting polypeptide expressed in the human liver
54 and containing three N-linked glycans, underwent dramatically decreased expression and was
55 retained within the endoplasmic reticulum when all three sites were mutated to Gln¹⁴⁹. A further
56 example is BRI2, a type-II transmembrane protein where inhibition of its single N-glycosylation
57
58
59
60

1
2
3 reduced cell surface trafficking and led to intracellular accumulation, although the mutation did
4 not affect cleavage by furin or ADAM10¹⁵⁰.
5

6 For glycoproteins whose glycosylation is relevant for functionality, different functional levels can
7 be attained, as shown by numerous reports in recent literature. For instance, blocking the
8 glycosylation of the hepatocyte growth factor receptor (c-Met), a transmembrane tyrosine
9 kinase, attenuates c-Met function through inhibiting its cell membrane targeting¹⁵¹. CREB-H, a
10 liver-abundant bZIP transcription factor, requires N-glycosylation at three sites in its luminal C-
11 terminal domain for optimal activation¹⁵². Another example is glycoprotein KCC4, a K⁺Cl⁻ co-
12 transporter isoform involved in maintaining protein stability, regulation of cell volume,
13 anchorage-independent cell growth, tumor formation, and lung colonization by tumors.
14 Deglycosylated KCC4 forms decrease tumor formation and lung colonization in mice. Also, site-
15 directed mutagenesis on the four putative N-glycosylation sites established that KCC4
16 localization to the cell surface depends on the central N331 and N344 sites¹⁵³. This example
17 serves to introduce a next level of functionality, namely when glycosylation of one or more, but
18 not all, sites is required for proper functioning. While this type of studies is much more
19 informative, it is also harder to perform as multiple mutant strains must be produced or,
20 alternatively, selective deglycosylation must be achieved. An example can be found in human
21 acetylcholinesterase (AChE_T), with three putative N-glycosylation sites that are very important
22 for maintaining the catalytically active conformation. Mutants AChE_T^{N381Q}, AChE_T^{N495Q} and
23 AChE_T^{N296Q/N381Q/N495Q}, particularly the former, showed a dramatic decrease in enzymatic activity
24 compared with AChE_T^{WT}. In contrast, glycan removal did not change the sedimentation
25 properties or proportions of AChE, indicating that N-linked glycosylation does not affect
26 oligomerization¹⁵⁴. Similarly, human serum carnosinase CN-1, involved in diabetic nephropathy,
27 contains three potential N-glycosylation sites which, if deleted, result in impaired protein
28 secretion; enzyme activity, for its part, is already reduced when two sites are deleted¹⁵⁵. Finally,
29 myeloperoxidase, a lysosomal protein of neutrophils with five N-glycans (N323, N355, N391,
30 N483, and N729), undergoes significant loss of activity upon deglycosylation at N355¹⁵⁶.
31
32
33
34
35
36
37
38
39
40
41

42 The N-glycosylation cases described above need to be completed with a few equally important
43 examples of O-glycosylation. In addition to the well-established protective role of O-
44 glycosylation in mucins, several more specific functions have been recently discovered. Thus,
45 both regulated and aberrant glycosylation modulate the electrical signaling of the I_{Ks} channel, a
46 macromolecular complex composed of pore-forming α-(KCNQ1)-subunit and modulatory β-
47 (KCNE1)-subunit that is crucial for repolarization of the cardiac action potential¹⁵⁷. Moreover, O-
48 glycosylation at Thr-7 in the KCNE1 subunit is essential for proper biosynthesis and trafficking
49 of the complex¹⁵⁸. A similar example are two-pore-domain potassium (K(2P)) channels, where
50 disruption of glycosylation reduced current through decreasing the number of channels on the
51 cell surface and hence influencing cellular depolarization¹⁵⁹.
52
53
54
55
56
57
58
59
60

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 versa¹⁶⁰. 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 adherens junctions. Furthermore, the presence of (α 1-6)-fucosylation on Asn-linked GlcNAc promotes cell-cell adhesion in both cancer and downstream signaling pathways¹⁶¹. Another study involved 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 expression¹⁶². In a recent paper on the tumor-associated antigen CD147 (N-glycans at N44, N152, and N186), enzymatic deglycosylation and permethylation followed by high-resolution MS analysis revealed the presence of Man₃ to Man₇ structures and barely processed bi-antennary N-glycans in which core-fucosylated Man₃ accounted for ~30% of the structures. All glycans were found to stabilize tertiary and quaternary structure and to maintain the active conformation essential for CD147 activity¹⁶³. In addition, N152 was crucial for cell-surface expression and (β 1-6)-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 component⁸². 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

1
2
3 contain two N-glycans each (α N52 and α N78; β N13 and β N30) in addition to several O-glycans.
4 Oligosaccharides comprise (α 2-3)-monosialylated di/monoantennary complex type structures
5 with partial core fucosylation, as well as (α 2-3)-monosialylated hybrid type structures. Core
6 fucosylation is found only in the β -subunit and both N-glycans are diantennary complex type,
7 while in the α -subunit N78 lacks hybrid type structures but are instead predominant at N52. This
8 site-specific glycosylation is required for efficient recombination of both α and β subunits to form
9 the active hormone¹⁶⁵. One should also mention the laborious work on the glycosylation of
10 Tamm-Horsfall glycoprotein¹⁶⁶, where only one (N14) out of eight potential sites was non-
11 occupied, and the rest exhibited remarkable diversity: the N489 site included di- and tri-charged
12 oligosaccharides exposing, among others, the 4HSO₃-GalNAc(β 1-4)GlcNAc epitope; N251
13 contained only oligomannose-type chains ranging from Man₅GlcNAc₂ to Man₈GlcNAc₂, while
14 N208 was quite heterogeneous, with multiply charged complex glycan structures terminated by
15 sulfate groups, Sia residues, and/or the Sda-determinant. A final example is human
16 erythropoietin (EPO), possibly the most extensively studied cytokine, for which full glycan
17 profiling (of the endogenous form) is, after nearly 50 years of research effort, still incomplete.
18 The recombinant version used as a pharmaceutical is less negatively charged despite being
19 fully sialylated¹⁶⁷, its three N-sites contain complex type tetra-antennary *N*-acetylglucosamine
20 repeats with acetylated Sia residues⁷⁸.

21 3.5. Modifications of carbohydrates

22 While glycosylation is unrivalled as PTM in terms of abundance, complexity, and relevance,
23 carbohydrates themselves may be subject to yet another level of structural multifariousness.
24 Post-glycosylational modifications (PGMs) of specific sites (mostly hydroxyl or amino groups)
25 within the glycan chain occur after the oligomer has been assembled and include sulfation,
26 acylation, phosphorylation, methylation or epimerization that may modulate the biological
27 function of the carbohydrate and as such play a critical role in many normal and pathological
28 processes¹⁶⁸.

29 Several examples of PGM have already been mentioned above such as the sulfation of the
30 HNK-epitope, or sulfo-Sd^a variant, which may be more common than anticipated. One prime
31 candidate here is EPO, for which sulfation may explain the difference between the fully
32 sialylated recombinant version and the even more charged endogenous variant¹⁶⁷, but
33 unambiguous evidence is hard to collect in view of endogenous EPO levels and current
34 analytical sensitivity. Sulfation is not only important in glycoprotein glycans; a prominent group
35 of carbohydrates bearing this modification are glycosaminoglycans (GAGs), unbranched
36 polysaccharides made up of repeating disaccharides units of hexosamine and uronic acid,
37 found on the extracellular matrix of cell surfaces and classified in different types depending on
38 sulfation patterns. These molecules participate directly or indirectly in many different
39 physiological processes ranging from the balance between morphogenetic protein and fibroblast
40 growth factor signaling to maintain cartilage homeostasis¹⁶⁹; to axon growth inhibition after
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 cleavage¹⁷¹; and many others¹⁷². 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 α 2–6-linked Sia residues controls B cell development, signaling, and immunological tolerance¹⁷³. 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 process¹⁷⁴. 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.

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) | [176,177] |
| Haptoglobin (Hp) | Unusual hyper-fucosylated site specific glycoforms of Hp. | Liver cirrhosis and hepatocellular carcinoma (HCC) | [112] |
| Heparan sulfate (HS) | <i>N</i> -sulfation and 2- <i>O</i> -sulfation vs lipoprotein binding. Binding and uptake of lipoproteins depends on the degree of sulfation of the chains. Clearance appears to depend on <i>N</i> -sulfation based on loss of inhibitory activity of <i>N</i> -desulfated. | Hepatic clearance of triglyceride-rich lipoproteins | [178] |
| Human serum and cerebrospinal fluid (CSF) proteins | Tetraantennary tetrasialylated glycan with a polyactosamine 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- <i>O</i> -acetylated sialoglycoproteins and 9- <i>O</i> -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) regulate 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, 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 | [94,185,186] |

Hitherto in this section the relevance of glycosylation as a general phenomenon, its presence 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-biomarkers¹⁸⁷⁻¹⁹¹. High-throughput discovery and new analytical approaches, including those addressed to PGMs¹⁹², 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 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 of 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 (figure 8).

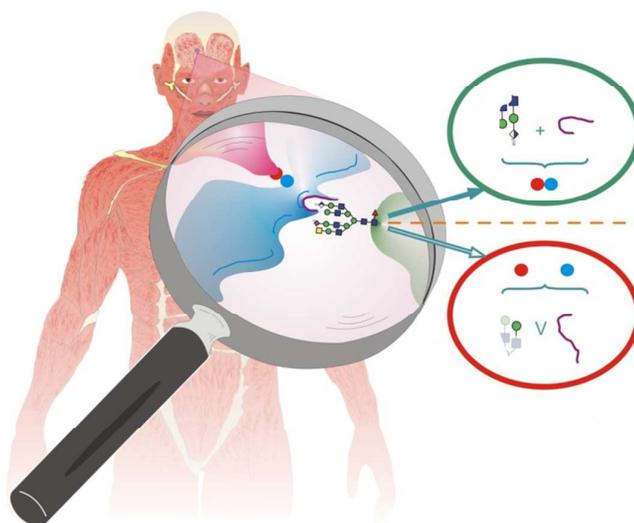


Figure 8. An ideal analytical setup for monitoring carbohydrate-driven biological functionality. In a productive scenario (green circle), specific binding of a particular carbohydrate epitope and its cognate CBP (purple) triggers further interaction (blue and red circles). In an unproductive scenario (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.

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.

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 presence and localization of reactive glycan epitopes, as well as detecting 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 adult murine testis and epididymis revealed that Leydig cells react specifically with SNA or CD22 lectins, both recognizing α 2-6-linked Sia; and that the same sugar, but α 2-3-linked, allowed differentiation between basal (no Sia) and apical (MAA lectin staining) cells of the epididymis¹⁹⁵. 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 microscopy¹⁹⁶. 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 differentiation⁵⁸. Another level of specificity can be added to lectin microarrays by combining them with protein specific antibodies in a sandwich-type approach⁵⁷. This has been successfully applied to establish the prevalence and carriers of particular glycosylation patterns in pancreatic cancer⁵⁶. 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

1
2
3 repertoire to include non-plant (and novel plant) lectins, developing recombinant lectins, as well
4 as introducing affinity rather than just specificity parameters in the read-out.
5

6
7 Attention must also be paid to CBPs from a functional perspective. In many situations, the
8 binding partners for particular glycans are not known and, for simplicity, are regarded as
9 constants rather than variables in most studies¹⁰². In any case, CBP study has become a
10 discipline on its own, mainly fueled by efforts from the Consortium for Functional Glycomics
11 (CFG) (<http://www.functionalglycomics.org/static/consortium/consortium.shtml>). The generation
12 of mammalian glycan arrays (currently with more than 650 structures) has greatly enabled the
13 systematic study of carbohydrate-protein-binding interactions, despite the intrinsic limitations
14 discussed below. The CFG has developed databases containing not only structures but also
15 experimental data generated with their microarrays. Although microarrays do not reveal
16 glycosylation patterns, they constitute a powerful technology to characterize CBPs binding
17 specificities³¹. As an example, both carbohydrate microarray and computational modeling
18 approaches have been used for the rapid screening of glycosaminoglycan (GAG) interactions
19 with proteins and multimeric protein complexes. Novel interactions between a specific sulfated
20 epitope, CS-E, and the neurotrophin family of growth factors have been identified with these
21 methods, providing unique molecular-level insights into the diverse biological functions of
22 GAGs¹⁹⁷. In another recent study, glycosphingolipids (GSLs) extracted from bovine brain
23 gangliosides and fluorescently labeled were bound to a microarray for subsequent interrogation
24 by biologically relevant CBPs like cholera toxin, antibodies and sera from individuals with Lyme
25 disease¹⁹⁸. More recently, glyconanoparticles (GNPs) in microarray format have been used to
26 study glycan-lectin interactions. The GNPs were made by conjugating carbohydrate ligands on
27 silica nanoparticles and microarrays were generated by conventional photocoupling chemistry.
28 They were then probed with fluorescein-labeled lectins and with fluorescein-doped silica
29 nanoparticles¹⁹⁹. The above studies are representative of glycofunctional approaches but
30 several aspects of glycan array design and biomolecular interaction assay should not be
31 overlooked³⁴. In addition to the limitations in oligosaccharide synthesis²⁰⁰ or in glycan isolation
32 from natural sources, aspects such as oligosaccharide density²⁰¹, spacing, and orientation
33 achieved upon immobilization, as well as the nature, flexibility and length of the linker are key
34 parameters in optimizing array strategies. Indeed, since many CBPs achieve their specificity
35 and affinity through multivalent interactions with glycans²⁰², glycoarrays should aim at faithful
36 replication of multivalent sugar display, and at capturing the physiological avidity of such
37 interactions in as native-like fashion as possible. For instance, a bead-modified surface
38 providing multivalency (*i.e.* the cluster effect) was used for probing carbohydrate-protein
39 interactions mimicking a cellular environment²⁰³. These clusters can be of identical or mixed
40 sugar composition to assess both valency and heterocluster effects⁵. Also, a new class of end-
41 functionalized polymers mimicking the multivalent architecture of chondroitin sulfate (CS)
42 proteoglycans have been designed, providing insights on how multivalency within and between
43 GAG chains enhances the avidity, specificity and cooperativity of GAG-protein interactions²⁰⁴.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Carbohydrates immobilized to solid surfaces have also been employed in slightly different
4 formats and with various other objectives, such as affinity-based systems to detect specific CBP
5 structures or to ascertain other interaction characteristics. For instance, a novel glycan-affinity
6 method combining proteolytic digestion of protein-glycan complexes and mass spectrometry
7 (CREDEX-MS, "Carbohydrate REcognition Domain EXcision Mass Spectrometry") has proven
8 useful in the structural definition of CBPs of two human galectins with lactose as binding
9 partner²⁰⁵.

10
11
12
13 An extremely valuable biophysical tool for carbohydrate-protein interaction studies is surface
14 plasmon resonance (SPR). In studies aimed at detecting multiple sclerosis autoantibodies in
15 sera, the glycopeptide antigen CSF114(Glc) was immobilized at the 3D-sensor surface and
16 real-time specific autoantibody detection was achieved²⁰⁶. SPR is particularly valuable because
17 it allows simultaneous monitoring of several surfaces (a limited form of arraying) and because
18 non-specific binding can be discarded by using a non-glycosylated peptide as a reference
19 surface. A refined version of SPR, Au nano-island based localized surface plasmon resonance,
20 has been used to characterize the specific recognition between *Concavalin A* (Con A) and
21 mannose²⁰⁷. The exquisite potential of SPR for analyzing carbohydrate binding proteins was
22 also highlighted in a study of galactose-specific *Erythrina cristagalli* agglutinin binding to several
23 galactoside-epitopes exposed in a well-defined manner⁸⁶ at the 3D-sensor surface²⁰⁸. Results
24 complemented those of CREDEX-MS, demonstrating that the combination of both techniques
25 can provide good insights into CBPs in various settings. Also, apparent differences in binding
26 preference of carbohydrate ligands have been observed by quantitative SPR analysis,
27 suggesting that glycan presentation and the conformational space it occupies plays an
28 important role in binding, regardless of affinity²⁰⁹.

36 4.2. Glycoengineering (genetic, chemoenzymatic, chemical)

37
38 One alternative approach to assess the importance of glycosylation is selective modification of
39 the carbohydrate decoration and study of its effects. In this context, gene targeting uses
40 homologous recombination to change an endogenous gene. The method can be used to delete
41 a gene, remove exons, add a gene, and introduce point mutations. Indeed, gene targeting has
42 been widely used in glycomics research by removing ("knocking out") or adding ("knocking in")
43 specific mutations of interest to a variety of models. The regulation of genes whose protein
44 products are involved in glycan synthesis and glycan-protein interaction provides insights into
45 glycan structural diversity and function in complex biological systems²¹⁰. Recently, a strategy for
46 developing cell lines that produce simplified homogenous O-glycan structures and thus
47 interrogating the human O-glycoproteome has been presented. Named 'SimpleCell', it uses
48 zinc-finger nuclease (ZFN)-based gene targeting of *COSMC* gene to glycoengineer stable
49 human cell lines displaying only truncated Tn and STn O-glycans. More than 100 O-
50 glycoproteins and up to 350 glycosylation sites, including a previously unidentified linkage to
51 tyrosine, were elucidated by this approach²¹¹. The strategy has been used to analyze the
52
53
54
55
56
57
58
59
60

1
2
3 function of a single GalNAc-transferase (GalNAc-T) isoform and its role in congenital diseases
4 and disorders²¹².

5
6 An attractive approach toward predefined glycoforms is *in vitro* chemoenzymatic glycosylation,
7 *i.e.*, remodeling natural or recombinant glycoproteins by addition of sugar units through
8 sequential glycosyltransferase-catalyzed reactions, or by endoglycosidase-catalyzed
9 transglycosylation and *en bloc* transfer of a pre-assembled large oligosaccharides to the protein
10 in a single step under the catalysis of an endo- β -*N*-acetylglucosaminidase (ENGase)²¹³.
11 Alternatively, site-specific glycosylation can be achieved by chemoselective ligation of proteins
12 to appropriately tagged glycans. For instance, cysteine residues in the protein can be reacted
13 with a thiol-reactive group pre-installed in the sugar moiety to give a disulfide or thioether-linked
14 glycoconjugate. Other strategies involve ligation (oxime, hydrazone) between amino and
15 carbonyl groups, or azide-alkyne cycloadditions under mild, bio-compatible conditions. For
16 instance, using recently developed chemoenzymatic strategies, N-glycans containing core-
17 fucose substitution and/or bisecting GlcNAc with otherwise ordinary complex-type antennae
18 terminated in α 2-3- or α 2-6-linked sialic acid were synthesized and neo-glycoproteins produced.
19 With these ultra-defined entities *in vivo* bio-distribution was assessed showing that core
20 substitutions alter glycan ligand properties through conformational changes which act as
21 molecular switches for target affinity and influence glycoprotein-mediated cell binding and
22 serum clearance^{214,215}.

30 31 4.3. Evaluating glycosylation within the biological context

32
33 As mentioned above, evaluating glycosylation within a given biological context and without
34 external manipulation is complicated; hence strategies with minimal impact on the system are
35 actively sought. One approach in this direction is selective targeting, under physiological
36 conditions, of particular glycans that can be subsequently evaluated. A recent study employs an
37 engineered β -1,4-galactosyltransferase to specifically transfer a keto-Gal functionality to *O*-
38 GlcNAc-modified proteins. The ketone moiety was subsequently reacted with various aminoxy-
39 functionalized polyethylene glycol tags of defined mass and the resulting samples were
40 analyzed by gel-based methods. In this way, a direct read-out of *O*-GlcNAc stoichiometry *vs*
41 state (*e.g.* mono-, di-, tri-, *etc.*) was possible, with insights into the complex interplay between *O*-
42 GlcNAc glycosylation and phosphorylation²¹⁶. Another chemoenzymatic strategy enabled rapid,
43 sensitive and selective detection of the (Fuc(α 1-2)Gal) disaccharide motif involved in processes
44 such as learning and memory, inflammation, asthma, and tumorigenesis. By exploiting the
45 restricted substrate tolerance of a blood group A GalNAc-transferase, the disaccharide is
46 targeted with azido-functionalized UDP-GalNAc that is later captured from the complex sample
47 mixture. This labeling strategy provides a variety of different enrichment strategies and imaging
48 read-outs for a variety of Fuc(α 1-2)Gal motifs²¹⁷. Broader applications would of course require a
49 supply of such restricted enzymes, currently unavailable at a larger scale. In another example of
50 azide-alkyne chemoselective ("click chemistry") conjugation²¹⁸, by introducing tetraacetylated *N*-
51 azidoacetyl-D-mannosamine in the Sia biosynthetic pathway; mature glycoproteins containing
52
53
54
55
56
57
58
59
60

1
2
3 azido-Sia were produced and targeted at the cellular level by capture with a biotinylated alkyne
4 reagent and subsequent MS evaluation. Sias have also been targeted by periodate oxidation, in
5 which vicinal hydroxyls in a cis configuration (present only in terminal Man, Gal(NAc) or non *O*-
6 acetylated Sia residues) are converted to aldehydes. Subsequent oxime ligation with aminoxy-
7 biotin labels glycoprotein subpopulations with high efficiency and cell viability, after which
8 samples can be evaluated by MS²¹⁹. The authors use ultra-mild conditions to assess only Sia
9 and target terminal GalNAc through an enzymatic protocol to allow differentiation between Sia-
10 containing and deficient cells.
11
12
13

14
15 Another elegant example of glycome comparison, based on stable isotope labeling with amino
16 acids in cell culture (SILAC), labels amine-containing monosaccharides in cells using ¹⁴N or ¹⁵N
17 glutamine as the sole nitrogen source. Named Isotopic Detection of Aminosugars With
18 Glutamine (IDAWG)²²⁰, the technique shows great promise for analyzing glycome dynamics
19 under different conditions. However, interpretation of the data may not be straightforward as the
20 protocol targets ManNAc, GlcNAc, GalNAc and NeuAc simultaneously, setting an equation with
21 at least four variables added to the intrinsic microheterogeneity of glycosylation. On the backflip
22 is the fact that the proteome may be targeted in a synchronized fashion. A rather different
23 approach was chosen in a recent study where cell surface amino groups were reacted with
24 unsaturated aldehyde yielding dihydropyridines products without affecting cell viability and
25 simultaneously introducing the Hilyte Fluor 750 tag to perform noninvasive whole body
26 fluorescence imaging. Examples included labeling of colon and gastric cancer cell lines in
27 BALB/c nude mice to monitor tumor metastasis²²¹.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

5. Outlook

In this review we have not aimed at an exhaustive examination of all aspects related to a greater or lesser extent related to protein glycosylation. Rather, we sought to provide a flavor 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, 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 metabolomics²²² 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 respect²²³. 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 discoveries¹⁹⁰. Such systems-level studies will help establish novel quantitative and mechanistic links between gene expression, protein expression, enzyme activity, carbohydrate structure and glycoconjugate function.

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 Harb. Perspect. Biol.*, 2011, **3**.
- 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, *Glycoconj. J.*, 2013, **30**, 41-50.
- 16 T. Tharmalingam, B. Adamczyk, M. A. Doherty, L. Royle and P. M. Rudd, *Glycoconj. J.*, 2013, **30**, 137-146.
- 17 J. F. Rakus and L. K. Mahal, *Annual Review of Analytical Chemistry*, 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, *Journal of Pharmaceutical and Biomedical Analysis*, 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, *Glycoconj. J.*, 2013, **30**, 11-22.
- 28 I. M. Lazar, W. Lee and A. C. Lazar, *Electrophoresis*, 2013, **34**, 113-125.
- 29 J. Zaia, *OMICS: A Journal of Integrative Biology*, 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: A Journal of Integrative Biology*, 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, *Glycoconj. J.*, 2013, **30**, 119-136.
- 38 C. Reichel, *Drug Test. Anal.*, 2010, **2**, 603-619.
- 39 Müller-Eberhard H.J and Kunkel H.G, *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 Analyt. Technol. Biomed. Life Sci.*, 2010, **878**, 2117-2122.
- 45 J. Kamerling and J. Vliegthart, *Carbohydrates*, Walter de Gruyter, Berlin, 1989, 176-263.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- 46 O. M. Grace, A. Dzajic, A. K. Jäger, N. T. Nyberg, A. Onder and Rønsted N, *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, *Journal of Carbohydrate Chemistry*, 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. Råbinä, 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. Protocols*, 2012, **7**, 1299-1310.
- 80 A. R. de Boer, C. H. Hokke, A. M. Deelder and M. Wuhrer, *Glycoconj. 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.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- 82 R. G. Gallego, J. L. Blanco, C. W. Thijssen-van Zuylen, C. H. Gotfredsen, H. Voshol, J. O. Duus, M. Schachner and J. F. Vliegthart, *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 Journal*, 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. Vliegthart and J. P. Kamerling, *1H NMR Structural-Reporter-Group Concepts in Carbohydrate Analysis*, Elsevier, 2007, 133-191.
- 94 T. Yoshida-Moriguchi, L. Yu, S. H. Stalaker, 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. Engl.*, 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, *Molecular BioSystems*, 2009, **5**, 1087-1104.
- 121 J. C. Paulson, O. Blixt and B. E. Collins, *Nat. Chem. Biol.*, 2006, **2**, 238-248.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- 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, *Glycoconj. J.*, 2000, **17**, 501-530.
139 J. J. van Rooijen, J. P. Kamerling and J. F. Vliegthart, *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. Drobni, 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-García, 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. Ethimiopoulos, *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.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- 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, *Pigment 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. Vliegthart, *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, de la Morena-Barrio ME, 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, *Inflamm. 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. Many, 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.

- 1
2
3 190 T. Angata, R. Fujinawa, A. Kurimoto, K. Nakajima, M. Kato, S. Takamatsu, H. Korekane, C. X. Gao,
4 K. Ohtsubo, S. Kitazume and N. Taniguchi, *Ann. N. Y. Acad. Sci.*, 2012, **1253**, 159-169.
5 191 H. J. An and C. B. Lebrilla, *Methods Mol. Biol.*, 2010, **600**, 199-213.
6 192 H. Yu and X. Chen, *Org. Biomol. Chem.*, 2007, **5**, 865-872.
7 193 H. J. Gabius, S. André, J. Jiménez-Barbero, A. Romero and D. Solís, *Trends Biochem. Sci.*, 2011,
8 **36**, 298-313.
9 194 S. A. Brooks and D. M. Hall, *Methods Mol. Biol.*, 2012, **878**, 31-50.
10 195 M. Lohr, H. Kaltner, R. Schwartz-Albiez, F. Sinowatz and H. J. Gabius, *Anat. Histol. Embryol.*, 2010,
11 **39**, 481-493.
12 196 F. A. Habermann, S. André, H. Kaltner, D. Kübler, F. Sinowatz and H. J. Gabius, *Histochem. Cell*
13 *Biol.*, 2011, **135**, 539-552.
14 197 C. J. Rogers, P. M. Clark, S. E. Tully, R. Abrol, K. C. Garcia, W. A. Goddard, III and L. C. Hsieh-
15 Wilson, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 9747-9752.
16 198 X. Song, Y. Lasanajak, B. Xia, J. Heimbürg-Molinario, J. M. Rhea, H. Ju, C. Zhao, R. J. Molinaro, R.
17 D. Cummings and D. F. Smith, *Nat. Methods*, 2011, **8**, 85-90.
18 199 Q. Tong, X. Wang, H. Wang, T. Kubo and M. Yan, *Anal. Chem.*, 2012, **84**, 3049-3052.
19 200 B. Lepenies, J. Yin and P. H. Seeberger, *Curr. Opin. Chem. Biol.*, 2010, **14**, 404-411.
20 201 T. K. Dam and C. F. Brewer, *Glycobiology*, 2010, **20**, 270-279.
21 202 T. K. Dam and C. F. Brewer, *Adv. Carbohydr. Chem. Biochem.*, 2010, **63**, 139-164.
22 203 G. Simone, P. Neuzil, G. Perozziello, M. Francardi, N. Malara, E. Di Fabrizio and A. Manz, *Lab.*
23 *Chip.*, 2012, **12**, 1500-1507.
24 204 S. G. Lee, J. M. Brown, C. J. Rogers, J. B. Matson, C. Krishnamurthy, M. Rawat and L. C. Hsieh-
25 Wilson, *Chem. Sci.*, 2010, **1**, 322-325.
26 205 A. Moise, S. André, F. Eggers, M. Krzeminski, M. Przybylski and H. J. Gabius, *J. Am. Chem. Soc.*,
27 2011, **133**, 14844-14847.
28 206 F. Real-Fernández, I. Passalacqua, E. Peroni, M. Chelli, F. Lolli, A. M. Papini and P. Rovero,
29 *Sensors. (Basel)*, 2012, **12**, 5596-5607.
30 207 G. Bellapadrona, A. B. Tesler, D. Grünstein, L. H. Hossain, R. Kikkeri, P. H. Seeberger, A. Vaskevich
31 and I. Rubinstein, *Anal. Chem.*, 2012, **84**, 232-240.
32 208 C. Jiménez-Castells, S. Defaus, A. Moise, M. Przybylski, D. Andreu and R. Gutiérrez-Gallego, *Anal.*
33 *Chem.*, 2012, **84**, 6515-6520.
34 209 S. H. Millen, D. M. Lewallen, A. B. Herr, S. S. Iyer and A. A. Weiss, *Biochemistry*, 2010, **49**, 5954-
35 5967.
36 210 J. L. Baker, E. Celik and M. P. Delisa, *Trends Biotechnol.*, 2013, **31**, 313-323.
37 211 C. Steentoft, S. Y. Vakhrushev, M. B. Vester-Christensen, K. T. Schjoldager, Y. Kong, E. P. Bennett,
38 U. Mandel, H. Wandall, S. B. Levery and H. Clausen, *Nat. Methods*, 2011, **8**, 977-982.
39 212 K. T. Schjoldager, S. Y. Vakhrushev, Y. Kong, C. Steentoft, A. S. Nudelman, N. B. Pedersen, H. H.
40 Wandall, U. Mandel, E. P. Bennett, S. B. Levery and H. Clausen, *Proc. Natl. Acad. Sci. U. S. A.*, 2012,
41 **109**, 9893-9898.
42 213 L. X. Wang and J. V. Lomino, *ACS Chem. Biol.*, 2012, **7**, 110-122.
43 214 S. André, T. Kozár, R. Schuberth, C. Unverzagt, S. Kojima and H. J. Gabius, *Biochemistry*, 2007, **46**,
44 6984-6995.
45 215 S. André, T. Kozár, S. Kojima, C. Unverzagt and H. J. Gabius, *Biol. Chem.*, 2009, **390**, 557-565.
46 216 J. E. Rexach, C. J. Rogers, S. H. Yu, J. Tao, Y. E. Sun and L. C. Hsieh-Wilson, *Nat. Chem. Biol.*,
47 2010, **6**, 645-651.
48 217 J. L. Chaubard, C. Krishnamurthy, W. Yi, D. F. Smith and L. C. Hsieh-Wilson, *J. Am. Chem. Soc.*,
49 2012, **134**, 4489-4492.
50 218 L. Yang, J. O. Nyalwidhe, S. Guo, R. R. Drake and O. J. Semmes, *Mol. Cell Proteomics.*, 2011, **10**,
51 M110.
52 219 T. N. Ramya, E. Weerapana, B. F. Cravatt and J. C. Paulson, *Glycobiology*, 2013, **23**, 211-221.
53 220 R. Orlando, J. M. Lim, J. A. Atwood, III, P. M. Angel, M. Fang, K. Aoki, G. Alvarez-Manilla, K. W.
54 Moremen, W. S. York, M. Tiemeyer, M. Pierce, S. Dalton and L. Wells, *J. Proteome. Res.*, 2009, **8**,
55 3816-3823.
56 221 K. Tanaka, K. Moriwaki, S. Yokoi, K. Koyama, E. Miyoshi and K. Fukase, *Bioorg. Med. Chem.*, 2013,
57 **21**, 1074-1077.
58 222 V. Zoldos, T. Horvat and G. Lauc, *Curr. Opin. Chem. Biol.*, 2012, **17**, 34-40.
59 223 S. Neelamegham and G. Liu, *Glycobiology*, 2011, **21**, 1541-1553.
60

Mammalian Protein Glycosylation – Structure versus Function

S. Defaus^{a,b}, P. Gupta^b, D. Andreu^a, R. Gutiérrez-Gallego^{a,b,c*}

^a S. Defaus, D. Andreu, R. Gutiérrez-Gallego, Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona Biomedical Research Park, 08003 Barcelona, Spain

^b S. Defaus, P. Gupta, R. Gutiérrez-Gallego, Bio-analysis group, Neuroscience Research Program, IMIM-Parc Salut Mar, Barcelona Biomedical Research Park, 08003 Barcelona, Spain

^c R. Gutiérrez-Gallego, Anapharm Biotech, 08038 Barcelona, Spain

* Corresponding author: R. Gutiérrez-Gallego E-mail: ricardogutierrez@upf.edu.

This review highlights relevant aspects of structural analysis of carbohydrates, attributed functions at different levels, and current discoveries on the relevance of mammalian protein glycosylation through structural assessment in a biological environment.

