FEBS Letters 583 (2009) 1728–1735

journal homepage: www.FEBSLetters.org

Minireview

Glycoproteomics: Past, present and future

Bérangère Tissot [a], Simon J. North [a], Alessio Ceroni [a], Poh-Choo Pang [a], Maria Panico [a], Floriana Rosati [b], Antonietta Capone [b], Stuart M. Haslam [a], Anne Dell [a,*], Howard R. Morris [a,c,*]

[a] Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, United Kingdom
[b] Department of Evolutionary Biology, University of Siena, Via A. Moro 2, 53100 Siena, Italy
[c] M-SCAN Ltd., Millars Business Centre, Wokingham, Berks, RG41 2TZ, United Kingdom

ARTICLE INFO

Article history:
Received 16 March 2009
Revised 20 March 2009
Accepted 20 March 2009
Available online 27 March 2009

Edited by Miguel De la Rosa

Keywords:
Carbohydrate sequencing
Glycodelin
Glycoproteomics
gp273
Mass spectrometry
Peptide sequencing
GlycoWorkBench

ABSTRACT

This invited paper reviews the study of protein glycosylation, commonly known as *glycoproteomics*, beginning with the origins of the subject area in the early 1970s shortly after mass spectrometry was first applied to protein sequencing. We go on to describe current analytical approaches to glycoproteomic analyses, with exemplar projects presented in the form of the complex story of human glycodelin and the characterisation of blood group H eptitopes on the O-glycans of gp273 from *Unio elongatulus*. Finally, we present an update on the latest progress in the field of automated and semi-automated interpretation and annotation of these data in the form of *GlycoWorkBench*, a powerful informatics tool that provides valuable assistance in unravelling the complexities of glycoproteomic studies.

1. Introduction and historical perspective

Glycoproteomics, as distinct from proteomics or glycomics, is the study of the glycosylation of proteins, a covalent modification which confers altered physico-chemical properties and functional activity on the nascent protein chain. There are two broad classes of protein glycosylation in nature, those 'O-linked' to Serine or Threonine residues in the protein backbone, and those 'N-linked' to Asparagine residues. Mass spectrometry (MS) has played a key and irreplaceable role in defining the structures of glycoproteins over the past 30 years [1,2] using methods developed from the earlier studies on Antarctic fish blood "antifreeze" and prothrombin glycoproteins [3,4] together with the general 'mass mapping' strategy [5] of determining and screening the masses of peptides/glycopeptides produced from specific proteolytic or chemical digests, which itself evolved from earlier 'mixture analysis' approaches to protein and glycoprotein sequencing [6,7]. The concept of mapping (sometimes called fingerprinting) derives from the realisation in the late 1970s that the data set comprised of peptide molecular ions $M^{+\cdot}$ or quasi-molecular ions $[M+H]^+$ produced by digesting any given protein is likely to be unique (especially if more than one digest is used), and therefore it provides a reasonable diagnostic for characterisation or identification of the protein, distinguishing it from others, importantly without the need for sequencing. From 1981, early research applications of mass mapping ranged from the screening of recombinant proteins and glycoproteins for the Biotech industry, detecting errors of translation or confirming mass matching and thus identity [8] using a software mass-search aid ProtMap, through to assisting in the structural characterisation of new peptide hormones [9] and the detection and characterisation of glycosylation in human Interleukin 2 [10]. With the later advent of comprehensive computerised protein databases, the peptide maps could then be used to interrogate those databases for matches to, and thus identification of

Abbreviations: conA, concanavalin A; dHex, deoxyhexose; ES, electrospray; Fuc, fucose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GC, gas chromatography; Glc, glucose; GlcNAc, *N*-acetylglucosamine; Hex, hexose; HexNAc, *N*-acetylhexosamine; LC, liquid chromatography; Man, mannose; MALDI, matrix-assisted laser desorption ionisation; NeuAc, *N*-acetylneuraminic acid; MS, mass spectrometry; MS/MS, tandem mass spectrometry; NMR, nuclear magnetic resonance; POMC, pro-opiomelanocortin; PTM, post-translational modification; Q-TOF, quadrupole orthogonal acceleration time of flight; TOF, time of flight; TIC, total ion chromatogram; VVL, *Vicia villosa* lectin

* Corresponding authors. Address: Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, United Kingdom (H.R. Morris).

E-mail addresses: a.dell@imperial.ac.uk (A. Dell), h.morris@imperial.ac.uk (H.R. Morris).

0014-5793/$36.00
doi:10.1016/j.febslet.2009.03.049

unknown proteins [11], which in turn has stimulated the general development of the field of proteomics.

Two unique strengths of mass mapping were, and still are, the ability to map (and therefore visualise) the N-terminal and C-terminal domains of a protein with equal probability, and most importantly, the ability to discover post-translational modifications (PTMs) including *glycosylation* by detecting mass shifts in component peptides in the mass map or by locking on to sugar mass differences in the map, created by facile glycosidic bond cleavage. Once detection is achieved in this way, a whole battery of techniques including tandem mass spectrometry (MS/MS) can then be applied to determine even the most complex structures, and this laboratory has reported many such novel glycosylation studies over the past 20 years including defining the glycosylation of tissue plasminogen activator [12], of pro-opiomelanocortin (POMC) [13] of glycodelins A [14] and S [15], cytoplasmic glycosylation of Skp 1 [16,17], multiple 'O-linked' glycosylation of CD8 [18] and an unexpected novel 'N-linked' glycosylation in *Campylobacter jejuni* glycoproteins [19].

Despite those advances, the field of glycoproteomics remains a difficult one to enter for the new researcher, largely due to the sheer complexity and variability in the protein glycosylation we observe in most areas of biological research. In this paper, we attempt to demonstrate the further refinement of the strategies outlined above, with the aim of defining a generic approach to glycoproteomics, illustrated with advanced studies in which the interactive informatic tools which we are currently using to assist in detailed interpretation of MS and MS/MS data are described.

2. Glycoproteomic strategies

A number of reviews have been published which document the historical perspectives, principles and practice of glycoproteomic analysis [1,2,20–25]. The aims of this section are to highlight general issues and to suggest where efforts need to be focused to enable glycoproteomics research to be carried out more effectively. Firstly, of course, a key basic requirement is a well-found laboratory, for example at Imperial this includes three electrospray (ES) quadrupole orthogonal acceleration time of flight (Q-TOF) type instruments (including a Q-Star), 2 matrix-assisted laser desorption ionisation (MALDI) 4800 time of flight (TOF)–TOFs plus a range of ancillary equipment, such as gas chromatography (GC)–MS for composition and linkage analysis and nano-liquid chromatography (LC) for sample presentation both to the Q-TOFs and TOF–TOFs, in the latter case via a Probot auto-spotter. Broadly speaking, the majority of laboratories engaged in glycoproteomic analyses employ all or part of what has become a generic workflow as illustrated in Fig. 1, with specific methodologies being dictated by available infrastructure, instrumentation and expertise. The type of sample being analysed will also influence the choice of methodology. For example, although not always applicable to large and/or highly heterogeneous glycoproteins, molecular weight profiling of intact glycoproteins (purple arrows in Fig. 1) can sometimes provide very useful information on the type and extent of glycosylation. Such "top-down" methods have proven especially powerful, in bacterial glycoproteomics where novel glycans are frequently observed [26–28], and in studies on intact antibodies for the Biopharmaceutical industry, where M-Scan routinely screens intact masses at around 150 kDa by both MALDI-TOF and ES-Q-TOF to give confirmatory total mass analysis when reconstructing the detailed protein and carbohydrate profiles from mass mapping studies.

Central to all general glycoproteomic strategies is the mass spectrometric analysis of glycopeptides, usually after chromatographic separation, either on-line (red arrows) or off-line (blue arrows), to simplify the maps produced. Glycopeptides are normally obtained by specific proteolytic or chemical digestion of

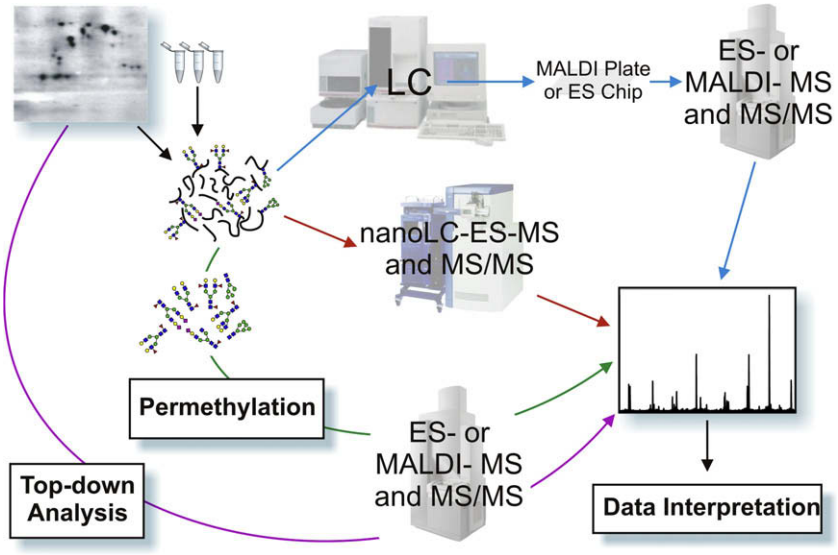

Fig. 1. A simplified glycoproteomic experimental workflow is shown, illustrating common approaches to glycoproteomic analysis. Samples take the form of slices or spots excised from single or multi-dimensional polyacrylamide gels, or batches of cells, fluids, immunoprecipitates or tissue extracts. Analytical approaches can broadly be categorised as "top-down" or "bottom-up". The former, illustrated by a *purple* arrow, begins with work on purified samples of glycoproteins in an attempt to identify the intact molecular weight profile by direct MALDI-TOF MS or by ES-MS. By subtracting the known or inferred mass of the protein component, the type and extent of glycosylation may then be deduced. In "bottom-up" approaches, which incidentally are essential for describing the detailed glycosylation profile of any protein, the glycoprotein is digested enzymatically and/or chemically, ideally with high-specificity procedures, and the resulting peptide/glycopeptide mixture is mapped mass spectrometrically either by on-line (*red* arrows) LC-ES-MS followed by MS/MS analysis of signals of interest, or by off-line (*blue* arrows) strategies involving ES- or MALDI-MS and MS/MS approaches. Prior to these separation and mapping procedures, various strategies for enrichment of glycopeptides may be introduced, including lectin binding (see text). Parallel glycomic analyses (illustrated by the *green* arrow) are an invaluable feature of the "bottom-up" approach involving enzymatic or chemical release of the glycans followed by MALDI-TOF MS or ES-MS mapping of the glycan populations, usually as permethyl derivatives (see text), providing information about specific glycans and their relative amounts, which can then be compared and matched with data at the glycopeptide and overall glycoprotein levels.