Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1764 (2006) 1870–1880
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.elsevier.com/locate/bbapap

Review

Global methods for protein glycosylation analysis by mass spectrometry

Bogdan A. Budnik [a,1], Richard S. Lee [b,1], Judith A.J. Steen [c,*]

[a] Department of Pathology, Children's Hospital Boston and Harvard Medical School, Boston, MA 02115, USA
[b] Department of Urology, Children's Hospital Boston and Harvard Medical School, Boston, MA 02115, USA
[c] Department of Neurology, Children's Hospital Boston and Harvard Medical School, Boston, MA 02115, USA

Received 23 June 2006; received in revised form 1 October 2006; accepted 10 October 2006
Available online 18 October 2006

Abstract

Mass spectrometry has been an analytical tool of choice for glycosylation analysis of individual proteins. Over the last 5 years several previously and newly developed mass spectrometry methods have been extended to global glycoprotein studies. In this review we discuss the importance of these global studies and the advances that have been made in enrichment analyses and fragmentation methods. We also briefly describe relevant sample preparation methods that have been used for the analysis of a single glycoprotein that could be extrapolated to global studies. Finally this review covers aspects of improvements and advances on the instrument front which are important to future global glycoproteomic studies.

Keywords: Global glycosylation analysis; Enrichment analysis; Chemical derivatization; Mass spectrometric methods of analysis; Precursor ion scanning; Electron capture dissociation; Electron transfer dissociation

1. Introduction

A focus on global glycoprotein analysis has recently resulted in technical advances in the field. The application of these techniques could generate basic functional knowledge in glycobiology and yield many candidate biomarkers for several diseases including cancer. The major focus of these global glycosylation studies has been to find biomarkers [1] and to map glycosylation sites [2–7]. The mapping of glycosylation sites and the structures of the glycans may yield valuable information about glycosylation consensus sequences. Recent advances in other "omics" fields have created the potential to identify new candidate biomarkers of disease that may provide the needed information to improve diagnosis and staging/grading of disease; discover new potential therapeutic targets; and provide more accurate prognostic information for patients [8]. Although there have been a high number of publications about candidate markers, there has been a striking decrease in the number of US Food and Drug Administration (FDA) approved biomarkers since 2003 [8].

Refocusing the analysis of a global proteome to sub-proteomes may provide the necessary direction to produce more clinically relevant biomarkers of disease. One particular sub-proteome is the glycoproteome. Studying changes of global proteome may lead to the identification of clinically useful biomarkers and therapeutic targets of disease. Glycoproteins compromise approximately 25% of the currently available cancer biomarkers approved by the FDA [8]. A systems biology approach to the glycoproteome of various diseases or organisms may provide the necessary insight into the variations in glycosylation that are associated with varying degrees of pathogenesis. For instance, specific alterations in the glycans of currently available testicular tumor markers, such as α-fetoprotein or human chorionic gonadotropin-β may result in improved clinical diagnostic and/or prognostic capability and lead to new insights into the pathogenesis of certain grades of the disease.

Overall, glycosylated proteins represent the majority of cell surface markers and secreted proteins [9]. It is estimated that 50–60% of proteins in the human body are modified by

* Corresponding author. Children's Hospital Boston, Division of Neuroscience, Enders 307, 320 Longwood Ave, Boston, MA 02115, USA. Tel.: +1 617 919 2450; fax: +1 617 730 0243.
E-mail address: judith.jebanathirajah@childrens.harvard.edu (J.A.J. Steen).
[1] Both authors contributed equally to this work.

1570-9639/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2006.10.005

glycosylation [9–11]. Glycans are critical in determining a protein's stability, conformation, cellular signaling, and binding affinity for other molecules or glycan binding proteins [12,13]. Additionally, glycosylation has been implicated in numerous biological processes including cell growth and developmental biology, immune response, tumor growth, metastasis, anticoagulation, cell to cell communication, and microbial pathogenesis [14–25].

However, characterizing the glycoproteome is challenging because of the inherent heterogeneous and diverse nature of glycans and the complex nature of this modification [26]. The complete characterization of the glycoproteome requires the identification and analysis of the glycan structure; the protein expressing the glycan; and the protein that binds the glycan [27]. The glycans are not static, they may vary temporally with disease, and may be responsible for modulating as opposed to initiating a biological function as in a direct protein–protein interaction [26]. Interestingly, variations in the concentration of a single glycosyltransferase may change the degree of glycosylation of multiple proteins [28].

Global methods using mass spectrometry to study glycosylation may be a key technology in unraveling the biological function and intricacies of glycosylation. However, the study of protein glycosylation by mass spectrometry has its challenges. First, glycosylations can be highly labile, especially if the glycosylation is bound to threonine or serine residues, where the glycan is connected to the protein by an acetal bond. This lability of glycan complicates analyses and reproducibility. Second, glycosylated peptides have been described as exhibiting poor ionization efficiencies. Third, glycan modifications larger than a single sugar residue are extremely heterogeneous and are not particularly amenable to current bioinformatics platforms. The heterogeneity of glycans further complicates the analysis, as a single peptide may be modified by several forms of glycans, each having a different structure and mass, and to complicate things further, sometimes a different structure can also have the same mass.

Although there are many obstacles to interrogating the glycoproteome, recent advances in sample preparation and mass spectrometry have demonstrated the ability to isolate and identify both the peptide backbone and glycan. Several groups in the last few years have used various analytical techniques in order to overcome some of the above described problems associated with the analysis of glycoproteins and glycopeptides. In this review we will cover the global approaches used for glycoprotein and glycopeptide analysis including enrichment analysis, chemical derivatization and mass spectrometry instrument based methods that aid the analysis of glycoproteins.

2. Challenges of analyzing glycosylation by mass spectrometry

The biggest challenge faced in the analysis of protein modifications is the dynamic range of protein concentrations. Most mass spectrometers can only analyze samples in which the concentrations of proteins cover three to four orders of magnitude, whereas the protein concentrations of biological fluids span up to 10 to 12 orders of magnitude [29]. Some modified proteins are signaling proteins which are found in low abundance and as such are not selected for analysis in a complex mixture with a dynamic range greater than four orders of magnitude. These problems can be bypassed if the enrichment analysis for glycopeptides is performed. Enriching for glycoproteins by mass spectrometry has three significant advantages: (1) the most abundant unmodified proteins are excluded from the analyses such that low abundance glycoproteins are analyzed; (2) the glycopeptides do not have to compete with unmodified proteins for charge carriers during ionization thereby improving ionization efficiencies and increasing the likelihood of detecting the modified peptide; (3) instrument parameters can be optimized for glycopeptide analysis, such that labile glycans are not lost during ionization and the fragmentation methods and conditions are optimized for glycan analysis. For example by using low voltages in the interface region, glycans from O-glycosylated peptides will remain intact on the peptide backbone during analysis. In addition, if the goal of an experiment is to obtain peptide and glycan sequence information, CID (collision induced dissociation) fragmentation energies for glycopeptides are often at least 10 eV higher that that for unmodified peptides [30]. This is because the primary fragmentation process is glycan fragmentation when using CID for glycopeptides, whereas the peptide backbone fragmentation is a secondary process that has to be induced by increased collision energies. The large glycan structures are especially difficult to fragment as the branched structure has numerous degrees of freedom and much of the energy applied to fragment the glycopeptide will be absorbed by this glycan structure. However, it must be noted that if one attempts to analyze unmodified proteins along with the glycoproteins, these high energy conditions do not necessarily provide good sequence information for unmodified peptides.

3. Enrichment methods

Several recent studies highlight the success of enrichment analysis strategies using affinity chromatography methods to enrich proteins or peptides with a particular modification [31–33]. The glycan moiety can be used as a handle for affinity purification to enrich for glycosylated peptides as they have distinct properties, which can be used to specifically select the modified proteins/peptides using structural aspects (e.g. antibodies and lectins) [34,35] or chemical properties (e.g. derivatization methods) [36,37].

3.1. Lectin affinity purification

Lectin affinity chromatography for glycoprotein isolation has been used widely for several years [38]. Lectins are proteins that specifically interact with carbohydrates without modifying them. They generally interact with specific motifs in a glycan and the structural domain which interacts with the glycan varies between lectins. There are several different commercially available lectins, that can be used to selectively enrich for particular subsets of glycoproteins from a complex protein