Talanta 85 (2011) 499–505

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Large-scale assignment of *N*-glycosylation sites using complementary enzymatic deglycosylation

Wei Zhang, Hong Wang, Lei Zhang, Jun Yao, Pengyuan Yang*

Department of Chemistry and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 15 January 2011
Received in revised form 4 April 2011
Accepted 10 April 2011
Available online 16 April 2011

Keywords:
Glycoproteomics
Lectin affinity chromatography
PNGase F
Endo H
Mass spectrometry

ABSTRACT

Endoglycosidase is a class of glycosidases that specifically cleaves the glycosidic bond between two proximal residues of GlcNAc in the pentasaccharide core of *N*-glycan, leaving the innermost GlcNAc still attached to its parent protein, which provides a different diagnostic maker for *N*-glycosylation site assignment. This study aims to validate the use of endoglycosidase for high throughput *N*-glycosylation analysis. An endoglycosidase of Endo H and the conventional PNGase F were employed, with a similar accessible procedure, for large-scale assignment of *N*-glycosylation sites and then *N*-glycoproteome for rat liver tissue. ConA affinity chromatography was used to enrich selectively high-mannose and hybrid glycopeptides before enzymatic deglycosylation. As a result, a total of 1063 unique *N*-glycosites were identified by nano liquid chromatography tandem mass spectrometry, of which 53.0% were unknown in the Swiss-Prot database and 47.1% could be assigned only by either of the methods, confirmed the possibility of large-scale glycoproteomics by use of endoglycosidase. In addition, 11 glycosites were assigned with core-fucosylation by Endo H. A comparison between the two enzymatic deglycosylation methods was also investigated. Briefly, Endo H provides a more confident assignment but a smaller dataset compared with PNGase F, showing the complementary nature of the two *N*-glycosite assignment methods.

1. Introduction

Glycosylation is one of the most common, important, and complicated post-translational modifications (PTMs) of proteins [1]. The oligosaccharide chains of glycoproteins have vital roles in protein conformation, activity, and localization, as well as participate in many biological processes, such as molecular and cellular recognition, signaling, and communication [2–4]. As a consequence, many diseases have affinitive relations with aberrant glycosylation [5]. Changes in glycosylation abundance and alterations in the glycan structure of serum and membrane glycoproteins have been shown to correlate evidently with the progression of cancer [6,7]. Thus, glycoproteins and the corresponding glycans may be considered potential disease biomarkers and drug targets [5]. Based on these biological and medical significances, glycoproteomics, the combined field of proteomics and glycomics, is emerging as an essential issue, aimed at comprehensive characterizing and complete understanding of protein glycosylation [8].

N-Glycosylation, where glycans are *N*-linked to asparagine residues in proteins, constitutes the most common types of glycosylation. *N*-Glycosylation sites fall into a consensus sequence of Asn-X-Ser/Thr, where X is any amino acid other than Pro [9,10]. All *N*-linked glycans share a pentasaccharide core ($Man_3GlcNAc_2$) and can be classified into three subgroups based on the antennas attached to the core. Glycans containing only oligomannose antennas attached to the core are high-mannose type glycans; complex-type glycans have antennas that exhibit variable numbers of *N*-acetylglucosamine (GlcNAc), galactose (Gal), sialic acid (Neu5Ac), fucose (Fuc), and rarely some other sugars; and glycans containing both oligomannose antennas and complex antennas are hybrid type-glycans [11]. The carbohydrate binding specificities of different lectins, a group of sugar-binding proteins, offer a powerful approach for specific glycan structure recognition. For example, ConA specifically binds mannosyl (Man) residues of glycans and has affinity to high-mannose and hybrid glycoforms; WGA recognizes GlcNAc residues of glycans and prefers to bind complex glycoforms [12,8]. Therefore, lectin affinity chromatography has been developed and extensively used for glyco-protein/peptide isolation and purification [13–15].

More than half of all proteins in the SWISS-PROT database have been estimated to have at least one glycosylation site, but only a few of them have been reported [16]. Many efforts have been made to elucidate more glycosites using various techniques and approaches [17,18]. The traditional method for *N*-glycosylation site assignment entails the enzymatic removal of glycans by peptide-*N*-glycosidase (PNGase) accompanied by the conversion of Asn (N) to Asp (D)

* Corresponding author. Tel.: +86 21 5423 7961; fax: +86 21 5423 7961.
E-mail address: pyyang@fudan.edu.cn (P. Yang).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.04.019

through deamidation. This deamidation provides an indirect indication of *N*-glycosites, as a resulting mass shift of 0.98 Da from N to D can be detected by mass spectrometry (MS) [19]. However, as deamidation can occur both *in vivo* and *in vitro*, false positive identification of glycosites is inevitable [20,21]. Moreover, accurate determination of 0.98 Da by tandem MS (MS/MS) is also a challenge for high throughput glycosylation analysis [22]. These problems can be partially rectified by performing the PNGase digestion in ^{18}O-water as a result of a larger mass shift of 2.98 Da [14,15,23].

Recently, a novel strategy for the identification of *N*-glycosylation sites involving the use of *endo-β-N*-acetylglucosaminidase (endoglycosidases) has been described [21,24,25]. Endoglycosidase is a class of glycosidases that specifically hydrolyzes the glycosidic bond between two proximal residues of GlcNAc in the pentasaccharide core of *N*-glycoprotein, leaving a terminal GlcNAc still attached to its parent peptide, which provides a mass increase of 203.08 Da [26,27]. Thus, glycosites can be elucidated by MS/MS with less ambiguity due to higher mass increase and avoidance of *in vitro* reaction. Furthermore, the presence or absence of core-fucosylation, an important structure in the innermost GlcNAc of some hybrid and complex glycans, can also be determined by the mass increment (349.14 Da for fucosylated GlcNAc) [28]. However, this strategy is still difficult for large-scale analysis of glycoproteome because only 33 glycosites by Endo M [21] and 90 glycosites at most by an enzyme mixture of two endoglycosidases [24,25] have been identified, respectively, from human serum in previous reports. To facilitate the efficiency of these endoglycosidases, various exoglycosidases are needed to simplify the complexity of glycans. Simultaneously, the complementarity between the two enzymatic deglycosylation methods (PNGase and Endo) for glycosites identification has not been well investigated.

Endo H is an endoglycosidase that specifically hydrolyzes high-mannose and hybrid-type glycans [27,29]. The complexity of glycans does not present an obstacle for the enzymatic efficiency of Endo H, and thus the reduction of the complexity of glycans by exoglycosidases is avoided in Endo H digestion. Combined usage of Endo H and PNGase F has been becoming a useful tool in biological and biochemical analyses [30,31]. Izquierdo et al. recently utilized the combination of Endo H and PNGase F digestion to discriminate the distinct substrate specificities of different oligosaccharyltransferase isoforms from *Trypanosoma brucei* [32].

In this study, Endo H was employed for the large-scale analysis of high-mannose and hybrid-type *N*-glycosylation sites as well as *N*-glycoproteome from rat liver tissue. ConA affinity chromatography was used to enrich high-mannose and hybrid glycopeptides selectively. The enriched sample was divided into two aliquots treated by Endo H and traditional PNGase F with a same procedure, respectively, and then analyzed by nano liquid chromatography-electrospray ionization MS/MS (LC-ESI-MS/MS). As a result, a total of 622 and 1003 unique glycosites were assigned by Endo H and PNGase F, respectively, and 560 sites were identified by both Endo H and PNGase F (52.9% overlap), suggesting the efficiency and the complementary nature of the two methods. These results validated the application of endoglycosidase, as similar process as that of PNGase, for large-scale glycoproteomic analysis. Furthermore, the merits of each method were also analyzed and discussed in detail based on the obtained results.

2. Materials and methods

2.1. Materials and chemicals

ConA (agarose conjugate) was purchased from Vector Laboratories (Burlingame, CA). Porcine trypsin (sequence grade) was purchased from Promega (Madison, WI). PNGase F (glycerol free) and Endo H were purchased from New England Biolabs (Ipswich, MA). Complete mini protease inhibitor cocktail tablets were purchased from Roche (Basel, Switzerland). Sep-Pak C18 columns were purchased from Waters (Milford, MA). All other chemicals were purchased from Sigma–Aldrich (St. Louis, MO). Water used for all experiments was produced by a Milli-Q Plus system from Millipore (Bedford, MA), with resistance ≥18.2 MΩ/cm. Rat liver tissues were obtained from healthy adult rats.

2.2. Sample preparation

Total protein was extracted from rat liver tissues using an ice-cold homogenization PBS buffer (pH 7.4) containing 8 M urea, 2 M thiourea, 0.25 M dithiothreitol (DTT), 1 mM phenylmethanesulfonyl fluoride (PMSF), and a mixture of protease inhibitor (complete mini protease inhibitor cocktail tablets, 1 tablet for 10 mL homogenization buffer). The tissues were homogenized by Qiagen TissueRuptor with 10 mL of the homogenization buffer per 2 g of tissue. The suspension was homogenized at 4 °C for 40 min and then centrifuged (25 000 × *g*) at 4 °C for 40 min. The supernatant was collected and stored at −80 °C. The protein concentration was determined by the Bradford method.

2.3. Protein digestion

A sample containing 2 mg protein was reduced by 10 mM DTT at 56 °C for 30 min, alkylated in the dark by 50 mM iodoacetamide (IAA) at room temperature for 40 min, and precipitated with 6 volumes of acetone at −20 °C for at least 3 h. Afterwards, the precipitates were resuspended in 50 mM NH_4HCO_3 buffer and digested with 40 μg modified trypsin overnight at 37 °C. The reaction was quenched by heating to 95 °C for 10 min. The sample was finally dried through vacuum centrifugation.

2.4. ConA-affinity enrichment of glycopeptides

ConA-agarose settled gel (1:1, v/v slurry) was packed into a 4.6 mm inner diameter × 50 mm long column. The packed column was equilibrated sequentially with 3 column volumes of 1.0 M NaCl, 3 column volumes of 0.1 M NaCl, and 10 column volumes of an equilibration buffer (50 mM Tris base, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, and pH of 7.4 with HCl), at a flow rate of 0.15 mL/min.

The digest was redissolved in 1 column volume of the equilibration buffer and loaded onto the ConA column. After gentle rotation overnight at 4 °C for incubation, the column was washed with 15 column volumes of the equilibration buffer at a flow rate of 0.15 mL/min. Then, 1 column volume of an elution buffer (50 mM Tris base, 500 mM methyl-α-D-mannopyranoside, 150 mM NaCl, and pH of 7.4 with HCl) was applied onto the column, which was rotated for 3 h at 4 °C for elution. Finally, the captured glycopeptides were eluted with 3 column volumes of the elution buffer. The sugar and salts in the eluted solution were removed using a Sep-Pak C18 column.

2.5. Deglycosylation of glycopeptides

The glycopeptides were divided into two identical aliquots, dried through vacuum centrifugation, and redissolved respectively in 50 mM sodium citrate (pH 5.5 for Endo H digestion) and 50 mM NH_4HCO_3 buffer (for PNGase F digestion). In total, 0.5 μL Endo H (500 units/μL) and 0.5 μL PNGase F (500 units/μL) were respectively added to the two aliquots, after which the samples were incubated overnight at 37 °C for deglycosylation. The deglycosylated samples were dried again through vacuum centrifugation.