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.

2.6. Nano-LC-ESI-MS/MS analysis

The samples were redissolved in 5% acetonitrile (ACN) aqueous solution containing 0.1% formic acid (phase A). Analysis of the samples was carried out by an LC-20AB system (Shimadzu, Tokyo, Japan) connected to an LTQ-Orbitrap XL MS (ThermoFisher, San Jose, CA) interfaced with an online nano-electrospray ion source (Michrom Bioresources, Auburn, CA).

The peptide sample was loaded onto a 100 μm inner diameter × 15 mm long Magic C18 AQ column (Michrom Bioresources, Auburn, CA) with a flow of 25 μL/min and eluted directly into the ESI source of the MS with a 60–90 min gradient from 0 to 45% phase B (95% ACN aqueous solution containing 0.1% formic acid) at a constant flow rate of 500 nL/min. The electrospray voltage was used at 1.8 kV. The MS was operated in the data-dependent mode, selecting the 10 most intense parent ions in the Orbitrap for collision-induced dissociation in the LTQ trap per full MS scan. Normalized collision energy was 35.0%. Dynamic exclusion was set to initiate a 60 s duration for ions analyzed twice within a 10 s interval. The scan range was set from m/z 400 to m/z 2000. The samples deglycosylated by Endo H and PNGase F were operated under the same LC–MS condition.

2.7. Database searching and data analysis

All MS/MS data were searched against the rat International Protein Index (IPI) database (IPI rat v3.64 fasta with 39 871 entries) using the SEQUEST algorithm incorporated into the Bioworks software (Version 3.3.1). The search parameters were set as follows: enzyme, partial trypsin; missed cleavage sites allowed, two; fixed modification, carboxyamidomethylation (Cys+ 57.02 Da); variable modifications, oxidation (Met+ 15.99 Da), deamidation (Asn+ 0.98 Da, for PNGase F sample), *N*-GlcNAc (Asn+ 203.08 Da, for Endo H sample), and fucosylated *N*-GlcNAc (Asn+ 349.14 Da, for Endo H sample); peptide tolerance, 10 ppm; and fragment tolerance, 1.0 Da.

Database search results were statistically evaluated using PeptideProphet [33]. A minimum PeptideProphet probability score (P) of 0.95 was selected to remove low-probability results. The Asn modification that did not occur in N-X-S/T motif (X ≠ P) was also removed to reduce the false positive rate of the identified glycosylation sites.

3. Results and discussion

3.1. Overall strategy for large-scale assignment of glycosylation sites

Shotgun glycoproteomics has been used extensively for the comprehensive characterization of glycoproteins in biological samples. Its experiments rely on the digestion of complex protein mixtures into hundreds of thousands of peptides by proteases. Glycoprotein/glycopeptide enrichment is necessary before/after enzymatic digestion to reduce the complexity. These peptides are deglycosylated by traditional PNGase, separated by LC and ultimately analyzed by MS and MS/MS.

Recently, the use of endoglycosidase has provided a new approach for shotgun glycoproteomics because it could potentially provide new glycosylation sites [21,24,25]. Different subtypes of *N*-glycans play variant roles in protein activities and biological processes. Alterations in the degree of branching of *N*-glycans have been reported as a consequence of diseases [34]. For example, recent study shows that glycoproteins exhibit increased branching of *N*-glycans during liver disease [35]. Endoglycosidases have remarkable enzymatic specificities on peptide substrates for certain oligosaccharide structures: High-mannose and most hybrid structures can be released by Endo H, whereas complex structures can be released by Endo F_2 (biantennary complex) or Endo F_3 (triantennary and core-fucosylated complex). Thus, these enzymes for glycoproteomics provide not only glycosites but also the glycan subtypes of each glycosite, and this information is important for further glycobiological investigations.

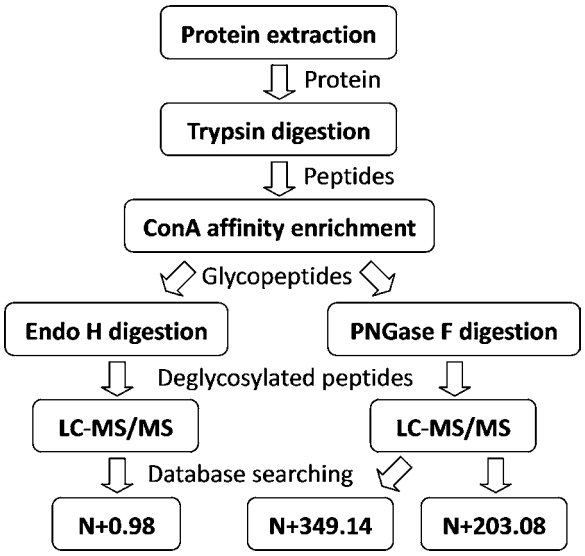

Fig. 1. Overview of the procedure used for the large-scale assignment of glycosylation sites in rat liver proteins.

The flowchart of the analysis of high-mannose and hybrid-type glycosites in rat liver proteins by Endo H and PNGase F is shown in Fig. 1. Briefly, ConA affinity chromatography was used to isolate glycopeptides selectively after protein extraction and trypsin digestion. The bound glycopeptides were competitively eluted by mannopyranoside contained buffer, desalted by a Sep-Pak column, and divided into two identical aliquots. The two aliquots were then treated by Endo H and PNGase F to prompt a mass increase in the peptide backbone possessing the glycosylation sites corresponding to 0.98 and 203.08 Da (349.14 Da if core-fucosylation occurs), respectively. As a result, this molecular weight change can be discernible by database searching, allowing the large-scale assignment of *N*-glycosylation sites.

The use of PNGase F is the most effective method for removing virtually all *N*-linked glycans from glycopeptides, except glycans containing a fucose α(1-3)-linked to the *N*-linked GlcNAc, which are commonly found in glycoproteins from plants or parasitic worms [27]. Conversely, Endo H only removes high-mannose and hybrid structures due to the limited specificities of endoglycosidases, which are remarkably similar to the specificity of ConA lectin. Therefore, lectin affinity enrichment of glycopeptides may be an ideal choice for the endoglycosidase method due to the similar specificities of glycopeptides. Furthermore, endoglycosidase is also suitable for general deglycosylation by combination use of different Endo series, as complex tetraantennary structures can be effectively removed by several endoglycosidases with the assistance of exoglycosidases.

3.2. Assignment of N-glycosylation sites by the Endo H and PNGase F method

Four biological replicates for each method were performed in this study, resulting in an assignment of 1063 unique glycosylation sites from 560 glycoproteins by two deglycosylation methods. As shown in Fig. 2, a total of 622 sites from 390 glycoproteins

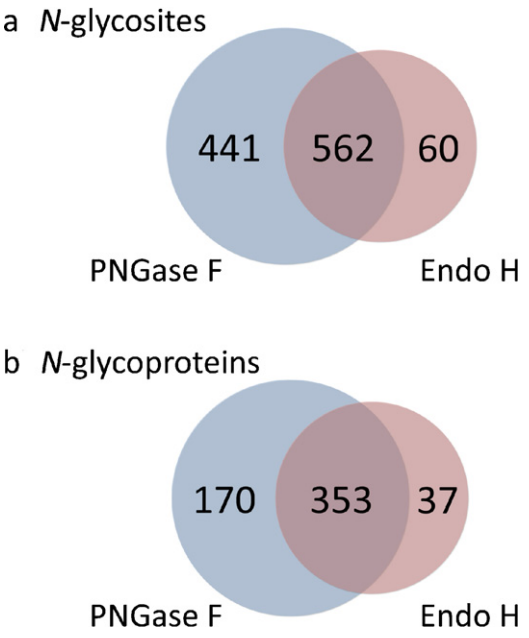

Fig. 2. Venn diagram of the overlap indicates the identified glycosylation sites (a) and the glycoproteins (b) between the two deglycosylation enzymes.

were identified by Endo H, whereas 1003 sites from 523 glycoproteins were identified by PNGase F. Among which, 562 sites from 353 glycoproteins were identified by both Endo H and PNGase F methods, showing a 52.9% overlapping for glycosites and a 63.0% overlapping for glycoproteins. The results evidently suggest that endoglycosidase should be also applicable and accessible for large-scale assigning glycosites compared with the traditional PNGase in glycoproteomics. Detailed information about the identified glycopeptides, glycosites, and corresponding glycoproteins is listed in Table S1 in the supporting information.

The ambiguity in the assignment of the 562 sites by both Endo H and PNGase F was avoided by the fact that the two enzymes provided absolute different diagnostic makers for glycosites (a GlcNAc or an H from deamidation on the motif). Fig. 3 shows the MS/MS spectra of an identified deglycosylated glycopeptide (LE**N**LSSTESGYTATLTR) from lysosomal alpha-glucosidase as an example. Fig. 3a depicts the localization of the glycosite determined by a mass increase of 0.98 Da after PNGase F treatment [$(M+2H)^{2+}$ at m/z 922.45]. Both the b- and y-series of the product ions clearly exhibit the mass shift from N to D in the N-X-S/T motif. Otherwise, as shown in Fig. 3b, a GlcNAc residue attached to the Asn residue was detected distinctly with a mass shift of 203.08 Da for the Endo H-treated sample [$(M+2H)^{2+}$ at m/z 1023.50], indicating the presence of the glycosite. False positive assignment is reduced in this case because of the large mass shift from the GlcNAc attached to the peptide. Furthermore, Fig. 3c illustrates a mass increase of 349.14 Da in the glycopeptide [$(M+2H)^{2+}$ at m/z 1096.53] detected also from the Endo H-treated sample, indicating the presence of fucosylated GlcNAc. The low intensity of the product ions was observed from the MS/MS spectrum, which could be due to two reasons: (1) the core-fucosylated glycans were relatively low in amount, accounting for only a low proportion of total glycans in the glycopeptides, and (2) the presence of carbohydrate residues on the glycoside could affect the signals. For example, within the theoretical 16 b-series ions, 10, 3, and 3 b-ions were found from Fig. 1a, b, and c, respectively, indicating the effect of carbohydrate residues. Our result indicates that core-fucosylation, which sometimes occurs in hybrid and complex glycans, can also be illuminated by the Endo H method.

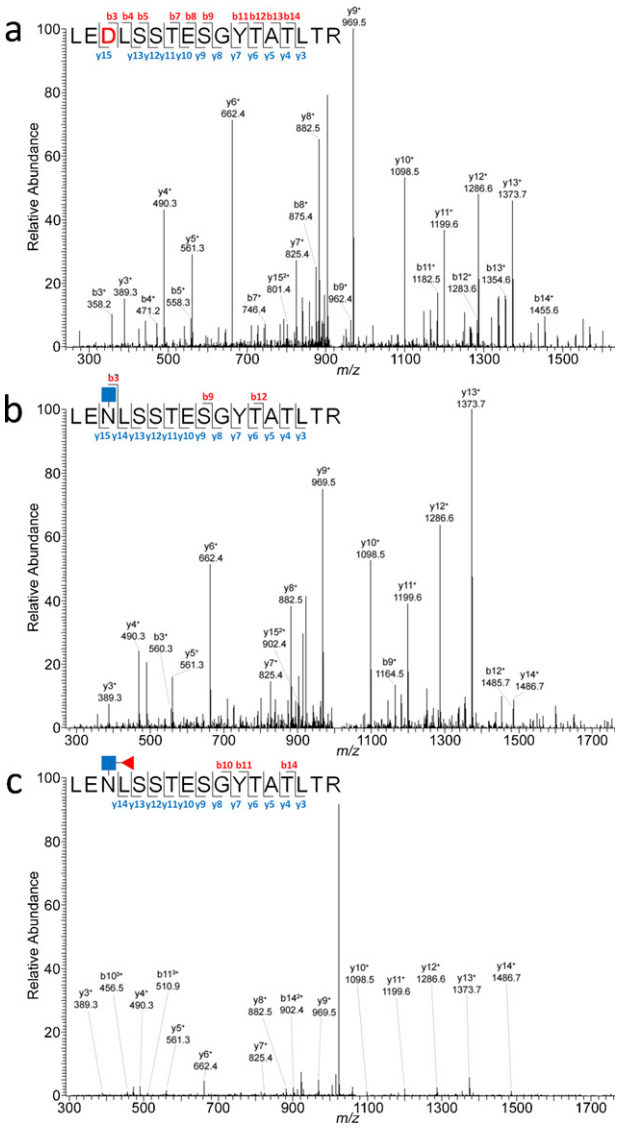

Fig. 3. Nano-LC–MS/MS mass spectra of doubly charged glycopeptide LE**N**LSSTESGYTATLTR from lysosomal alpha-glucosidase. (a) Deamidation of the glycopeptide [$(M+2H)^{2+}$ at m/z 922.45] from PNGase F treated sample; (b) *N*-GlcNAc modification of the glycopeptide [$(M+2H)^{2+}$ at m/z 1023.50] from Endo H treated sample; and (c) fucosylated *N*-GlcNAc modification of the glycopeptide [$(M+2H)^{2+}$ at m/z 1096.53] from Endo H treated sample. Blue square: GlcNAc; red triangle: Fuc. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

3.3. *Analysis of high-mannose and hybrid-type glycosylation sites in rat liver*

For the 1063 unique glycosylation sites detected, only 8 sites have been proven by direct experimental evidence and are currently documented in the Swiss-Prot database (November 2010 version, http://www.uniprot.org/). Moreover, 492 sites were annotated as potential (473 sites), probable (9 sites), or by similarity (10 sites), which are qualifiers indicating the computer prediction of the feature or the existence of indirect experimental evidence (Fig. 4). The other 563 sites (53.0%) were totally unknown in the Swiss-Prot database. These surprising results have two explanations. First, rat glycoproteome should be comprehensively investigated more. To the best of our knowledge, there are only a few reports on the study of rat glycoproteins. Second, the use of two different deglycosylation enzymes, especially Endo H, provides an efficient tool to map new glycosylation sites. Our dataset

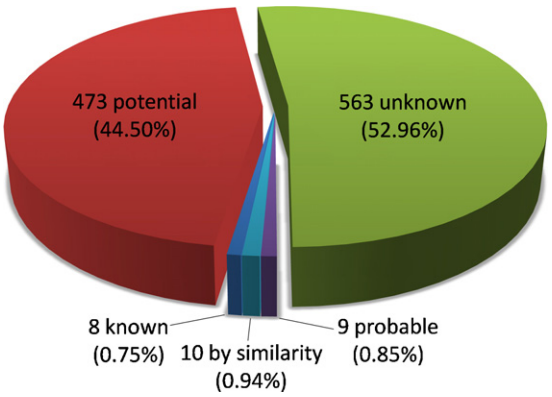

Fig. 4. Classification of glycosylation sites. The classification of known, unknown, potential, probable, and by similarity is based on the comparison of the experimental dataset with the annotated Swiss–Prot database.

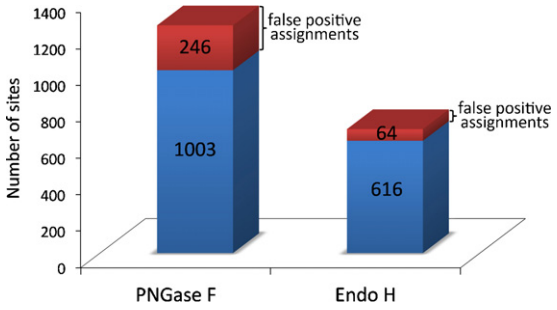

Fig. 5. Accuracy and confidence of the two methods. 19.7% (246 sites) of the total results (1249 sites) from the PNGase F method (N + 0.98 Da) are false assignments; 9.4% (64 sites) of the total results (680 sites) from the Endo H method (N + 203.08 Da) are false assignments. Blue square column: the number of sites assigned in the N-X-S/T motif (X ≠ P); red square column: the number of sites not assigned in N-X-S/T and assigned in N-P-S/T motif. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

contributes not only direct experimental evidence to a number of potential *N*-glycosylation sites but also a number of new *N*-glycosylation sites to the Swiss-Prot database.

An excellent work for glycoprotemics was recently conducted by Mathias Mann's group [15], where a dataset containing 6367 *N*-glycosylation sites were obtained from various mouse tissues by combination of lectin affinity enrichment, PNGase F treatment and MS analyses. In this dataset, 2768 sites were identified from liver tissue, by using different lectins, proteases or from subcellular fractions of liver cells. Within the 2768 assigned sites from mouse liver, 880 sites were obtained by only ConA affinity enrichment and trypsin digestion. By using the same Con A lectin, the obtained data in this study, 1003 sites by PNGase F and 562 sites by Endo H, were also considerable, confirmed the efficiency of the two methods and the ability of endoglycosidase for high throughput glycoproteomics.

Moreover, since Endo H displays enzymatic activity on core-fucosylated glycopeptides with a modification of 349.14 Da left on the parent peptide, 11 glycosites with core-fucosylation information from 11 glycoproteins were assigned by Endo H (Table 1). Core-fucose usually presents in hybrid and complex-type glycoproteins; thus, the core-fucosylated glycoproteins identified by Endo H are all hybrid types. For the 11 identified glycosites, 5 sites were annotated as potential, and the other 6 sites were unknown in the Swiss-Prot database. Most of these glycoproteins have vital roles in various biological processes. For example, lysosomal alpha-glucosidase is essential for the degradation of glycogen to glucose in lysosomes. Abnormal expression of this enzyme causes glycogen storage disease II (Pompe disease), which is an autosomal recessive disorder with a broad clinical spectrum [36].

Information concerning core-fucosylation attracts increasing attention from biological and clinical aspects. Core-fucosylated glycoproteins regulate the biological functions of adhesion molecules and growth factor receptors (e.g., TGF-β1 and EGF signaling pathways) [37]. Increased level of core-fucose has been found in a number of pathological processes, including inflammation and cancer [38]. Previous report also shows that core-fucosylation level, as a biomarker, is more sensitive and specific than the corresponding protein level [39]. Endoglycosidases, applied in high throughput glycoproteomics, provide an accessible way to determine new core-fucosylated glycoproteins, facilitating functional or clinical research as a therapeutic target for some diseases.

3.4. Evaluation of the two enzymatic deglycosylation methods

As a most widely used deglycosylation enzyme, PNGase F uniquely characterized 441 of the 1063 glycosites (41.5%), suggesting the high efficiency of this traditional method. However, only 60 of the total assigned glycosites (5.6%) were uniquely characterized by Endo H. The relatively low efficiency of Endo H method can be attributed to two factors: (1) the innermost GlcNAc (or fucosylated GlcNAc) specifically attached to the glycosites that influenced the ionization efficiency of peptides in the MS and MS/MS analyses, and (2) the cross-ring cleavages of the innermost GlcNAc (or fucosylated GlcNAc) that usually occurred in collision-induced dissociation (CID), causing the neutral loss of the GlcNAc fragments. Therefore, further improvements are still required for the endoglycosidase method, including Endo H and the other endoglycosidases used for previous glycosites assignment. At the same time, the improvement of a database searching algorithm is also worthwhile to realize the assignment of neutral loss during MS/MS analysis. Note that we also explored the use of novel electron-transfer dissociation (ETD) in the characterization of *N*-glycosites in this investigation. ETD fragments peptides by transferring an electron from a radical anion to a protonated peptide, preserving the PTMs that are labile in CID [40]. However, in the ETD dataset, only 11 sites assigned by PNGase F did not present in the corresponding CID dataset, and all the results obtained by Endo H were found to be covered by the results of CID (data not shown). This unexpected result may be due to the low speed of the ETD reaction and the limited charge states of the precursors, such that further optimization should be needed for a better ETD performance in large-scale glycosite analysis.

The accuracy and confidence of the two methods was also respectively assessed. The PeptideProphet evaluated data (before filtered by N-X-S/T motif) were employed to determine the false positive rates. The filtered data were not used here because most false positive assignments had been removed by N-X-S/T motif filtering, and thus results obtained from these data cannot represent the real accuracy of the two enzymes. The 1249 unique deamidated Asn (N + 0.98 Da) were characterized by MS with PNGase F method, among which 244 Asn were not assigned in the N-X-S/T motif, whereas 2 Asn were assigned in the N-P-S/T motif (Table S2 in the supporting information). These apparent false assignments accounted for 19.7% of the total results (Fig. 5, left column). Conversely, 680 unique GlcNAc modified Asn (N + 203.08 Da) were characterized by the Endo H method, among which 63 Asn were not assigned in the N-X-S/T motif, whereas 1 Asn was assigned in the N-P-S/T motif (Table S3 in the supporting information), accounting for 9.4% of the total assignments (Fig. 5, right column). The false rates (19.7% vs. 9.4%) show that false positive identifications were induced more by PNGase F than by Endo H. Although the modified Asn that not assigned in the N-X-S/T motif and that assigned in the N-P-S/T motif were both removed from

Table 1
List of identified glycosylation sites with core-fucosylation.

Peptide[a]	Protein name[b]	Protein IPI[b]	Site[c]	Annotation[d]
K.N#STKEEILAALEK.G	Sulfated glycoprotein 1	IPI00195160.1	456	Potential
P.QVSGLN#NSDDK.A	Interleukin-3 beta	IPI00206669.1	63	Potential
R.N#CSIFLADINQER.H	Palmitoyl-protein thioesterase 1	IPI00208382.1	197	Potential
P.DAGQQAGQVGFLNPN#GSSQGK.V	Nuclear factor 1	IPI00210163.1	420	Unknown
K.N#TTTALPLGPW.L	Similar to protein tyrosine phosphatase, receptor type, H precursor	IPI00214459.3	548	Unknown
-.MATELEKALSNVIEVYHN#YSGIKGNHHAL.Y	Protein S100-A8	IPI00231370.7	18	Unknown
R.LEN#LSSTESGYTATLTR.T	Lysosomal alpha-glucosidase	IPI00400579.1	140	Potential
K.VGPVDPSLN#TTYVFFDTFFK.E	Beta-hexosaminidase subunit beta	IPI00464518.2	306	Potential
F.ELVGN#TTTLCGENGQWLGGK.P	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	IPI00559054.3	2470	Unknown
T.SKHNN#DTWESDSNEFSVIADPR.G	54 kDa protein	IPI00777611.1	113	Unknown
L.VDN#NTWNNTHISR.V	Integral membrane protein 1	IPI00914170.1	592	Unknown

[a] #: identified glycosylation site (N).
[b] Protein name and protein accession number are from IPI database.
[c] Glycosylation sites are annotated according to Swiss–Prot database.
[d] Glycosylation sites are marked as unknown and potential according to the definition of Swiss–Prot database.

the final results, the falsity could not be absolutely avoided due to the mechanism of the method used, and the obtained false positive rates are helpful for evaluation of the confidence level of the datasets.

As mentioned above, the relatively high percentage false rate of the PNGase F method may be due to two reasons. First, the conversion between N and D can occur both *in vivo* and *in vitro*, leading to false assignment. Second, large-scale glycoproteomic analysis requires high scan speed and sensitivity, which come at the trade-off between mass accuracy and resolution. This also decreases the accuracy of assignment due to the low mass shift obtained by PNGase F. Moreover, the Endo H method also introduces some false results, which may be induced by *O*-GlcNAc glycosylation. *O*-GlcNAc glycosylation is a type of *O*-linked glycosylation where the monosaccharide GlcNAc is α-glycosidically linked to Ser or Thr residues [41]. In the present study, 61 unique *O*-GlcNAc glycosites on Ser/Thr from 48 unique peptides were identified from PNGase F treated sample (Table S4), whereas 20 of the 48 peptides contain the N-X-S/T motif. Thus, in Endo H treated sample, incomplete b- and y-series obtained from an *O*-GlcNAc glycosylated peptide inevitably lead to the false assigning of the *O*-GlcNAc of 203.08 Da to Asn if the peptide contains Asn residue. We are now performing enzymatic removal of this *O*-GlcNAc modification from Endo H treated sample before MS analysis to try to avoid the false assignment.

All the comparison conditions indicate the complementary nature of the two *N*-glycosite assignment methods. Approximately 47.1% of the total 1063 glycosites can only be assigned by either of the two methods. Furthermore, the two methods show different features and merits. More glycosites can be identified by the PNGase F method, with a 0.98 Da mass shift in the Asn residue as a diagnostic maker. In contrast, the Endo H method results in the more confident identification of glycosites because of a larger modification in Asn as a diagnostic maker. Combining the use of the two methods, a larger and a more confident dataset can be achieved for *N*-glycosylation site analysis.

4. Conclusions

In this study, we have reported a large-scale assignment of *N*-glycosylation sites for glycoproteome using two complementary deglycosylation enzymes, Endo H and conventional PNGase F, coupled with lectin affinity glycopeptide enrichment and nano-LC–MS/MS analysis. A total of 1063 *N*-glycosites were identified from rat liver proteins, most of which were unknown in the Swiss-Prot database. This investigation has the following contributions. (1) The possibility of large-scale glycosite analysis by endoglycosidase was confirmed, which can be a complementary method to PNGase with a similar simple procedure. (2) Combining the results of PNGase F and Endo H, a larger dataset of *N*-glycosites, as well as many new glycosites from rat liver glycoproteins, were obtained. (3) Information concerning glycan subtypes and core-fucosylation on each glycosite was provided by Endo H, and this is valuable for further functional study. (4) The characters and merits of each method were discussed in-depth in this study. Briefly, the PNGase F method results in a larger dataset, whereas the Endo H method provides more confident assignment.

Acknowledgments

This work was supported in part by National Science and Technology Key Project of China (2007CB914100, 2010CB912700 and 2011CB910600), and National Natural Science Foundation of China (20735005 and 20975024).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.talanta.2011.04.019.

References

[1] J.E. Turnbull, R.A. Field, Nat. Chem. Biol. 3 (2007) 74.
[2] A. Varki, Glycobiology 3 (1993) 97.
[3] P.M. Rudd, T. Elliott, P. Cresswell, I.A. Wilson, R.A. Dwek, Science 291 (2001) 2370.
[4] J.B. Lowe, Cell 104 (2001) 809.
[5] S.M. Hanash, S.J. Pitteri, V.M. Faca, Nature 452 (2008) 571.
[6] J.N. Arnold, R. Saldova, U.M.A. Hamid, P.M. Rudd, Proteomics 8 (2008) 3284.
[7] A.S.V. Edouart, M.C. Slomianny, O.D. Beseme, J.F. Haeuw, J.C. Michalski, Proteomics 8 (2008) 3236.
[8] J. Hirabayashi, Glycoconjugate J 21 (2004) 35.
[9] E. Bause, H. Hettkamp, FEBS Lett. 108 (1979) 341.
[10] E. Bause, G. Legler, Biochem. J. 195 (1981) 639.
[11] R. Kornfield, S. Kornfield, Annu. Rev. Biochem. 54 (1985) 631.
[12] H. Lis, N. Sharon, Chem. Rev. 98 (1998) 637.
[13] M. Geng, X. Zhang, M. Bina, F. Regnier, J. Chromatogr. B 752 (2001) 293.
[14] H. Kaji, H. Saito, Y. Yamauchi, T. Shinkawa, M. Taoka, J. Hirabayashi, K. Kasai, N. Takahashi, T. Isobe, Nat. Biotechnol. 21 (2003) 667.
[15] D.F. Zielinska, F. Gnad, J.R. Wisniewski, M. Mann, Cell 141 (2010) 897.
[16] R. Apweiler, H. Hermjakob, N. Sharon, Biochim. Biophys. Acta 1473 (1999) 4.
[17] H. Geyer, R. Geyer, Biochim. Biophys. Acta 1764 (2006) 1853.
[18] W. Morelle, K. Canis, F. Chirat, V. Faid, J.C. Michalski, Proteomics 6 (2006) 3993.
[19] J. Cao, C. Shen, H. Wang, H. Shen, Y. Chen, A. Nie, G. Yan, H. Lu, Y. Liu, P. Yang, J. Proteome Res. 8 (2009) 662.
[20] N.E. Robinson, A.B. Robinson, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 12409.
[21] Z.M. Segu, A. Hussein, M.V. Novotny, Y. Mechref, J. Proteome Res. 9 (2010) 3598.
[22] M.S. Kim, K. Kandasamy, R. Chaerkady, A. Pandeya, J. Am. Soc. Mass Spectrom. 21 (2010) 1606.
[23] B. Kuster, M. Mann, Anal. Chem. 71 (1999) 1431.

[24] P. Hagglund, J. Bunkenborg, F. Elortza, O.N. Jensen, P. Roepstorff, J. Proteome Res. 3 (2004) 556.
[25] P. Hagglund, R. Matthiesen, F. Elortza, P. Hojrup, P. Roepstorff, O.N. Jensen, J. Bunkenborg, J. Proteome Res. 6 (2007) 3021.
[26] R.B. Trimble, A.L. Tarentino, J. Biol. Chem. 266 (1991) 1646.
[27] R.A. O'Neill, J. Chromatogr. A 720 (1996) 201.
[28] W. Jia, Z. Lu, Y. Fu, H. Wang, L. Wang, H. Chi, Z. Yuan, Z. Zheng, L. Song, H. Han, Y. Liang, J. Wang, Y. Cai, Y. Zhang, Y. Deng, W. Ying, S. He, X. Qian, Mol. Cell. Proteomics 8 (2009) 913.
[29] R.B. Trimble, F. Maley, Anal. Biochem. 141 (1984) 515.
[30] K. Mills, A.W. Johnson, O. Diettrich, P.T. Clayton, B.G. Winchester, Tetrahedron: Asymmetry 11 (2000) 75.
[31] J. Colangelo, R. Orlando, Rapid Commun. Mass Spectrom. 15 (2001) 2284.
[32] L. Izquierdo, B.L. Schulz, J.A. Rodrigues, M.L.S. Guther, J.B. Procter, G.J. Barton, M. Aebi, M.A.J. Ferguson, EMBO J. 28 (2009) 2650.
[33] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Anal. Chem. 74 (2002) 5383.
[34] H.J. An, S.R. Kronewitter, M.L.A. de Leoz, C.B. Lebrilla, Curr. Opin. Chem. Biol. 13 (2009) 601.
[35] B. Blomme, C. Van Steenkiste, N. Callewaert, H. Van Vlierberghe, J. Hepatol. 50 (2009) 592.
[36] M. Arad, B.J. Maron, J.M. Gorham, W.H. Johnson, J.P. Saul, A.R. Perez-Atayde, P. Spirito, G.B. Wright, R.J. Kanter, C.E. Seidman, J.G. Seidman, New Engl. J. Med. 352 (2005) 362.
[37] B. Ma, J.L. Simala-Grant, D.E. Taylor, Glycobiology 16 (2006) 158.
[38] E. Miyoshi, K. Moriwaki, T. Nakagawa, J. Biochem. 143 (2008) 725.
[39] R.R. Drake, E.E. Schwegler, G. Malik, J. Diaz, T. Block, A. Mehta, O.J. Semmes, Mol. Cell. Proteomics 5 (2006) 1957.
[40] J. Wiesner, T. Premsler, A. Sickmann, Proteomics 8 (2008) 4466.
[41] N. Khidekel, S.B. Ficarro, P.M. Clark, M.C. Bryan, D.L. Swaney, J.E. Rexach, Y.E. Sun, J.J. Coon, E.C. Peters, L.C.H. Wilson, Nat. Chem. Biol. 3 (2007) 339.