Structural Analysis of *N*- and *O*-glycans Using ZIC-HILIC/DIALYSIS Coupled to NMR Detection

Yi Qu^{a1}, Ju Feng^{a1}, Shuang Deng^c, Li Cao^a, Qibin Zhang^a, Rui Zhao^b, Zhaorui Zhang^b, Yuxuan Jiang^b, Erika M. Zink^a, Scott E. Baker^b, Mary S. Lipton^a, Ljiljana Paša-Tolić^b, Jian Zhi Hu^a, Si Wu^{b*}

^aFundamental & Computational Sciences Directorate, Pacific Northwest National Laboratory,
P.O. Box 999, Richland, WA, 99352 USA

^bEnvironmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P.O.
Box 999, Richland, WA, 99352 USA

^cEnergy and Environment Directorate, P.O. Box 999, Pacific Northwest National Laboratory,
Richland, WA, 99352 USA

¹These authors contributed equally to this work.

*Corresponding author at: Environmental Molecular Sciences Laboratory, Pacific Northwest
National Laboratory, P.O. Box 999, MS K8-98, Richland, WA 99352, USA, E-mail:
Si.Wu@pnnl.gov

Abstract

Protein glycosylation, an important and complex post-translational modification (PTM), is involved in various biological processes, including the receptor-ligand and cell-cell interaction, and plays a crucial role in many biological functions. However, little is known about the glycan structures of important biological complex samples, and the conventional glycan enrichment strategy (i.e., size-exclusion column [SEC] separation) prior to nuclear magnetic resonance (NMR) detection is time-consuming and tedious. In this study, we developed a glycan enrichment strategy that couples Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) with dialysis strategies to enrich the glycans from the pronase E digests of RNase B, followed by NMR analysis of the glycoconjugate. Our results suggest that the ZIC-HILIC enrichment coupled with dialysis is a simple, fast, and efficient sample preparation approach. The approach was thus applied to the analysis of a biological complex sample, the pronase E digest of the secreted proteins from the fungus *Aspergillus niger*. The NMR spectra revealed that the secreted proteins from *A. niger* contain both *N*-linked glycans with a high-mannose core similar to the structure of the glycan from RNase B, and *O*-linked glycans bearing mannose and glucose with 1->3 and 1->6 linkages. In all, our study provides compelling evidence that ZIC-HILIC separation coupled with dialysis is very effective and accessible in preparing glycans for the downstream NMR analysis, which could greatly facilitate the future NMR-based glycoproteomics research.

Keywords

ZIC-HILIC, dialysis, glycan, NMR, secretome, *A. niger*

1. Introduction

Post-translational modifications play essential roles in controlling physiological cellular processes by affecting protein folding, conformation, interaction, and activities (Kamath et al., 2011). Protein glycosylation, an important and complex post-translational modification, is involved in various biological processes, including receptor-ligand and cell-cell interaction, and plays a crucial role in many biological functions (Mann and Jensen, 2003). In fungi, a significant number of secreted biomass-degrading enzymes are known to be glycosylated, and a variety of studies have assessed the functional effects of both *N*- and *O*-linked glycosylation (Beckham et al., 2012). Two types of glycosylation linkages exist: *N*-type and *O*-type. The *N*-glycans are linked to the amide nitrogen of asparagine (Asn) with the motif sequence of Asn-X-serine (Ser)/threonine (Thr) (X is any amino acid except proline [Pro] or aspartic acid [Asp]). The attached sugar residue is *N*-acetylglucosamine (GlcNAc). A common pentasaccharide core among *N*-linked oligosaccharides containing two *N*-acetylglucosamines and three mannoses exists (Kornfeld and Kornfeld, 1976). *O*-glycans can be significantly diverse, most of which are linked to the hydroxyl group of Ser/Thr (Kornfeld and Kornfeld, 1976). *N*-glycans can be removed from the proteins with an amidase such as PNGase F (Marino et al., 2010), while *O*-glycans are cleaved primarily using chemical approaches (Carlson, 1966; Merry et al., 2002).

Structural and functional characterization of glycoproteins presents significant challenges and is a growing research area, mainly because glycan analysis is involved in various disciplines (e.g., medical research and drug discovery) and is vital for many emerging areas including glycoproteomics. The investigation of glycan composition and structures cannot only explain the discrepancies in protein molecular weight and/or chromatographic retention time between the experimental data and the *in-silico* analysis of peptide sequence (Marino et al., 2010), but also

provides critical clues to the understanding of the physiological functions (Freeze and Aeby, 2005) and enzymatic activities (Wittwer et al., 1989) of glycoproteins. Therefore, the development of easy and fast methodologies for the analysis of glycan structures is important to advance glycoproteomics research.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for studying the molecular structure, molecular interaction, molecular kinetics—or dynamics—and composition of biological mixtures (Keeler, 2010). The analyte size within the capacity of NMR detection ranges from a small organic molecule or metabolite, to a mid-sized peptide or a natural product, up to proteins of several tens of kDa in molecular weight. In addition, two-dimensional NMR spectroscopy can be employed to provide the information complementary to one-dimensional NMR analysis, which is particularly useful in simplifying the determination of the molecular structures from a chemical mixture even without further chromatography separation. NMR-based methods have been extensively used in characterization of glycan structures for glycoproteins (Duus et al., 2000). For medium-sized polysaccharide molecules, NMR-based glycan analyses show significant advantage over chromatography and mass spectrometry (MS) analysis because NMR is highly quantitative and reproducible, and provides more detailed molecular structure information, such as the ring chain linkage and the conformation (Davis et al., 1994a; Keeler, 2010).

The natural low abundance of glycoproteins in biological samples requires efficient strategy to enrich glycans for subsequent NMR measurement. Size-exclusion separation (SEC), based on the differences of the molecular sizes of analytes, was used almost extensively for the fractionation of oligosaccharides prior to NMR analysis (Gonzalez et al., 1998; Young et al.,

2002). However, the sample preparation is tedious and time-consuming (e.g., a single SEC running time is ~5 hours, and purification of glycans normally requires more than one type of SEC fractionation approach) (Harvey, 1997), and the peptides could be co-eluted with the glycans of the same size. In addition, the glycan analysis typically has special requirements, such as radiolabeling, fluorescent labeling, or mass spectrometry, which are not straightforward to couple online with SEC fractionation for localizing the glycan-containing fractions. Therefore, the lack of an easy approach to enrich glycans hinders the characterization of glycans by NMR.

On the other hand, ZIC-HILIC separation has recently been widely used in MS-based glycoproteomics research (Di Palma et al., 2011; Intoh et al., 2009; Takegawa et al., 2006). The stationary phase of the ZIC-HILIC column contains positively charged quaternary ammonium groups and negatively charged sulfonic acid groups on the surface (Di Palma et al., 2011; Intoh et al., 2009), forming a hydrophilic layer that served as the basis for the interaction between the stationary phase and analytes. Samples are dissolved in the high-concentration organic solvent and eluted by increasing the water gradient (Di Palma et al., 2011). Therefore, the ZIC-HILIC column is suitable for separating glycans from a peptide mixture because of the difference of hydrophobicity. To our knowledge, the ZIC-HILIC enrichment has not been used for NMR analysis of glycan structures.

Herein, we developed the strategy that coupled ZIC-HILIC separation with dialysis to enrich *N*- and/or *O*-linked glycans prepared by pronase E digestion of the pure glycoprotein or biological complex sample. For the proof-of-principle experiment, the known glycosylated protein RNase B (Fu et al., 1994a) was digested with pronase E followed by separation through the ZIC-HILIC column coupled with or without dialysis, and the enriched glycans were analyzed by NMR. The results demonstrate that the glycan enrichment using ZIC-HILIC coupled with

dialysis to remove non-specifically bound small molecules is ideal for the glycan preparation for the downstream NMR analysis. Therefore, this enrichment strategy was further applied to the enrichment of glycans from the *A. niger* secretome. *Aspergillus niger* is a well-known fungal species that has been used extensively for production of citric acid and polysaccharide hydrolases. The secreted glycoproteins contain the majority of the enzymes responsible for hydrolyzing biomass into fermentable sugars. *N*-glycosylation sites for *A. niger* secreted proteins have been determined (Wang et al., 2011a), and some genetic analysis of the pathway has been performed (Dai et al., 2013), but little is known about the glycan structures. Therefore, we took advantage of the ZIC-HILIC coupled with dialysis strategy to enrich the glycans derived from pronase E cleavage of the secretome, followed by NMR analysis of the glycan structure. We were able to simultaneously characterize both *N*- and *O*-linked glycan structures from the *A. niger* secretome: the *N*-linked glycans contain high-mannose core structure similar to RNase B, and the *O*-linked glycans contain mannose and glucose with 1->3 and 1->6 linkages. In all, our study demonstrates that ZIC-HILIC separation coupled with dialysis is ideal for the glycan enrichment because of its rapidity, ease, and effectiveness, and thus greatly facilitates the future glycoproteomics research.

2. Material and methods

2.1. Chemicals

RNase B, dithiothreitol, iodoacetamide, pronase E from *Streptomyces griseus*, methoxyamine hydrochloride, 3-(Trimethylsilyl)-1-propanesulfonic acid-d6 sodium salt (TSP), and *N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide were from Sigma-Aldrich (Milwaukee, Wisconsin, USA). Deuterium oxide was from VMR (Radnor, Pennsylvania, USA).

2.2. Growth of *A. niger* and preparation of secretome from *A. niger*

A. niger (ATCC11414) was maintained on complete agar medium, which is widely used for fungal growth (Bennette and Lasure, 1991). Conidia of spore inocula were grown on complete medium at 30°C and harvested after 4 days. 1×10^6 spores/ml were inoculated into 2×200 ml of modified minimal medium as described previously (Wang et al., 2011b). The cultivation was performed at 30°C in a 1-L baffled flask in the incubator shaker (New Brunswick Scientific) at 200 rpm. After 24 hours of growth, the supernatant was collected by filtering the culture through two layers of sterile miracloth and then centrifuging it at $15,000 \times g$ at 4°C for 10 minutes to remove cell debris. About 400 ml of supernatant was concentrated to 20 ml in a stirred cell (Millipore, Billerica, Massachusetts, USA) with an ultrafiltration membrane (NMWL 3 kDa, Millipore, Billerica, Massachusetts, USA) overnight at 4°C with continuous pressure of 40 psi N₂. Cold (-20°C) acetone was mixed with the protein solution at a volume ratio of 5:1, and incubated at -20°C for 1 hour. The precipitated proteins (~30 mg) were harvested by centrifugation at $18,000 \times g$ for 15 minutes at 4°C.

2.3. Protein digestion with pronase E

Twelve mg of RNase B or 30.6 mg of *A. niger* secretome dissolved in NH₄HCO₃ buffer (100 mM, pH 8) were reduced with 10 mM dithiothreitol at 60°C for 30 minutes and then alkylated with 40 mM iodoacetamide in the dark at 37°C for 60 minutes. With the addition of 1 mM CaCl₂, the chemically treated proteins were digested with pronase E from *S. griseus* (mass of protein substrate/mass of pronase E = 1/1) at 37°C for 6 hours and further digested with newly added fresh pronase E at 37°C for 18 hours. The concentration of the digested peptides was determined using bicinchoninic acid (BCA) assay. The digestion efficiency was examined by

running a gel with the loading amount of 10 µg. The completely or nearly completely digested products were used in the downstream experiments. Portion of the pronase E digest of *A. niger* secretome was dialyzed with Mini Dialysis Kit (1 kDa cutoff, GE Healthcare Life Sciences, Pennsylvania) 3 times with the total dialysis time of 17 hours.

2.4. Size-exclusion column separation of the pronase E digest of RNase B

The pronase E digest of RNase B was clarified using centrifugation at maximum speed for 5 minutes. The supernatant was run through the column (1 × 100 cm) containing Bio-Gel P4 gel (medium size of 90-180 µm, exclusion limit of 4 kDa, Bio-Rad Laboratories, Hercules, California, USA) using 50 mM NH₄HCO₃ buffer. The elute was monitored using an ultraviolet (UV) detector at 210 nm and collected into 80 fractions. Every 10 fractions were combined and analyzed using LTQ-Orbitrap-Velos (Thermo Scientific, San Jose, CA). The fractions exhibiting the sugar signature were freeze-dried for NMR analysis.

2.5. ZIC-HILIC enrichment of glycans

Pronase E digest was concentrated to dryness in a SpeedVac (GMI Inc., Ramsey, MN) and then diluted with buffer A (0.1% formic acid in acetonitrile) to reach the final concentration of 6.5 mg/ml. The glycans (500 µl for each injection) were enriched on the ZIC-HILIC column (4.6 × 150 mm id, 2.5 µm, PEEK, Merck KGaA, Darmstadt, Germany) at the flow rate of 0.6 ml/minute with the following gradient: the sample was loaded onto the column delivered by 99% buffer A and 1% buffer B (0.1% formic acid in nano-pure water) for 10 minutes; the nonspecifically bound peptides were washed by 80% buffer A and 20% buffer B for 30 minutes; and the glycans were eluted by 20% buffer A and 80% buffer B for 15 minutes (eluate 1), followed by 100% buffer B for 15 minutes (eluate 2). Each eluate was dialyzed with the filter (1k

MWCO, 2 ml capacity, GE Healthcare, Piscataway, NJ) in nanopure water for 4 hours and then overnight in the new nanopure water at 4°C. The dialyzed samples were completely dried in a SpeedVac for NMR analysis.

2.6. Glycan structure detection using NMR approaches

Before the NMR experiments, the water-soluble glycans were reconstituted in 250 µl of D₂O containing 0.5 mM TSP and then loaded into a 5-mm Shigemi advanced NMR microtube (Shigemi Inc., Japan) with magnetic susceptibility matched to D₂O at 5°C. Sodium azide (0.2%, w/v) was added into the solution to prevent the biodegradation. The one- and two-dimensional experiments were carried out on a Varian 600 MHz NMR spectrometer equipped with a Z-axis-gradient 5-mm HCN cold probe. All the NMR measurements were carried out at 25°C. The standard Varian PRESAT pulse sequence using a single pulse excitation and 2-s low-power pre-saturation at the H₂O peak position for H₂O suppression was used for the one-dimensional NMR measurement. To acquire each spectrum, an accumulation number of 2048 scans with an acquisition time of 2 s and recycle delay time of 1 s were used, resulting in a total time of ~3.5 s for each accumulation.

The two-dimensional NMR spectra (gradient correlation spectroscopy [gCOSY], total correlated spectroscopy [TOCSY], gradient heteronuclear single-quantum coherence [gHSQC], and gradient heteronuclear multiple-bond correlation [gHMBC]) were recorded at 25°C in D₂O on a Varian Inova 600 NMR spectrometer (¹H 600 MHz, ¹³C 150 MHz; Varian Medical Systems, Palo Alto, California, USA). The chemical shifts were referenced to the internal TSP. All two-dimensional NMR experiments were performed using the standard manufacturer's software (VNMRJ 3.2). Spectral widths of 7200 Hz (D₂O) were used for the one-dimensional ¹H NMR spectrum measurement (with 32k data points) and for all proton-detected experiments.

The gCOSY was acquired with 256 increments, which were used to determine the connectivity of a molecule by determining which protons are spin-spin coupled, such as anomeric proton H₁ with neighboring proton H₂ of the sugar ring (Aguilar et al., 2012; Inagaki et al., 1987). A sine-bell weighting was applied in both dimensions. The data were zero-filled to 2k × 2k data points prior to Fourier transform (FT), and absolute value presentation was chosen for the two-dimensional plot. TOCSY spectra were acquired with 256 increments and 64 scans, zero-filled to 2k × 2k data points, followed by sine-bell weighting in both dimensions prior to two-dimensional FT. An 80-ms MLEV17 mixing time was used in TOCSY, which is useful for dividing the proton signals into groups or coupling networks, especially when the multiplets overlap. A spectral width of 25,000 Hz was used for the ¹³C dimension in the multiplicity edited gHSQC experiment that correlates chemical shifts of directly bound nuclei (i.e., two types of chemical nuclei); for example, ¹H, ¹³C-gHSQC correlates chemical shifts within the CH group. Pulse sequences employing gradients and programmed adiabatic decoupling were used with these parameters for gHSQC, with 196 increments, 64 scans, 1024×512 data points, linear prediction in F1, and sine-bell processing in both dimensions.

3 Results and Discussion

3.1. ZIC-HILIC-based glycan enrichment strategy

Fig. 1 is the schematic representation of the overall experimental procedure. The glycoprotein sample was digested by pronase E, an enzyme mixture responsible for the cleavage of polypeptides into individual amino acids (Sweeney and Walker, 1993). The pronase E cleavage products contain the amino acids attached to the glycan moieties, which enable the determination of not only the glycan structure but also the glycan linkage type (*O*-linked or *N*-linked). This is the obvious advantage of the above-mentioned approach over the PNGase F-

based glycan analysis approach (Angel et al., 2007; Fu et al., 1994a), because only *N*-linked glycosylation sites can be identified. The pronase E digestion efficiency was examined by running an SDS-PAGE gel (i.e., no detected gel bands with M.W.>3kDa). The resulting pronase E digest was then loaded to a ZIC-HILIC column to enrich glycans with a step-wise gradient, and some portion of enrichment was further dialyzed to remove the small molecules such as the charged amino acids or small peptides non-specifically bound to the ZIC-HILIC column. Both the ZIC-HILIC enrichment and its dialyzed sample were subject to the NMR analysis.

RNase B, catalyzing the reaction of RNA degradation, is a well-known *N*-linked glycoprotein that contains a single glycosylation site at Asn34 where the chitobiose core elongated with five to nine mannose residues is attached (i.e., Man₃₋₉GlcNAc₂) (Fu et al., 1994a). Because of the well-characterized glycan structure, RNase B was chosen as the model system for assessing the proposed glycan enrichment strategies. We first evaluated the glycan enrichment strategy using ZIC-HILIC enrichment followed by ¹H NMR detection (Fig. 2A). A group of well-resolved chemical shift peaks between 5.0 and 5.5 ppm were observed and identified as mannose and GlcNAc. However, we also observed many amino acid peaks with high intensities in the sample enriched with ZIC-HILIC alone (Fig. 2A): amino acids carrying methyl and methylene groups have peaks within 0 to 4 ppm. To improve the purity of the recovered glycans, following the ZIC-HILIC enrichment, dialysis (1 kDa MWCO) was employed to remove hydrophilic amino acids or oligopeptides non-specifically bound to the ZIC-HILIC columns. The results (Fig. 2B) indicated that dialysis can efficiently remove small molecules, particularly charged amino acids (e.g., lysine peak at 1.90 ppm).

The enrichment efficiency was further evaluated using 2D NMR (Fig. 3). Table 1 summarized the ¹H and ¹³C NMR chemical shifts obtained from ¹H one-dimensional NMR and

two-dimensional NMR analysis (i.e., TOCSY, gHSQC, and gHMBC). Our results are consistent with the previously reported glycan structure of RNase B (inset of Fig. 3), based on the following observations: 1) the chemical shifts (Table 1) were assigned to the corresponding sugars by comparison with previously published data on related asparagine- (Asn-) linked glycans (Davis et al., 1994a; Fu et al., 1994b; Marti et al., 1984; Paz-Parente et al., 1983; Pu et al., 2000b; van Leeuwen et al., 2008); 2) in the two-dimensional NMR spectra (e.g., TOCSY), anomeric proton peaks of 1- β -GlcNAc, 2- β -GlcNAc, and mannoses at different positions in the glycoconjugate (Fig. 3) were all clearly resolved; and 3) the sugars observed in TOCSY (Fig. 3) were also detected in the one-dimensional NMR spectra (Fig. 2): two major peaks were observed with chemical shifts around 2.03 ppm and 2.07 ppm, which were identified as the methyl peaks derived from the acetyl group of *N*-acetylglucosamine; mannose or GlcNAc was observed based on the characteristic sugar resonances belonging to the H₁ anomeric protons between 5.0 and 5.5 ppm.

We also include a one-column SEC enrichment approach for comparison purpose. NMR analysis of the RNase B glycans derived from the SEC enrichment indicated that the sugar yield was relatively lower, and a few major sugar peaks were not observed (data not shown), compared to the ZIC-HILIC enrichment. The results could be improved by combining the fractions containing glycans from the first SEC separation, which are subject to the separation through the second SEC column (Young et al., 2002), and/or using sugar-labeling techniques to better locate the sugar fractions. However, it can be even more laborious and time-consuming than the one-column SEC separation strategy.

Overall, the ZIC-HILIC coupled with dialysis enrichment method is reproducible, fast, easy, unbiased against particular type of glycans (Wohlgemuth et al., 2009), and can be applied

in small labs equipped with high-performance liquid chromatography (HPLC) systems. The NMR results demonstrate that the ZIC-HILIC enrichment coupled with dialysis is very efficient to enrich and characterize glycans, so it was used for the NMR analysis of glycan structures of fungal secretome as described in the following sections.

3.2 NMR analysis of the glycan structure of the secretome from *A. niger*

Unlike the well-studied *N*-linked glycan from RNase B, little is known about the structure of the glycans from *A. niger*, despite their importance in producing diverse products ranging from human therapeutics to biofuels (Wang et al., 2011a). Some studies have been performed on the individual glycoproteins from *A. niger*, such as α -galactosidase A (Wallis et al., 2001), endopolygalacturonase C (Woosley et al., 2006), and pectin methylesterase (Warren et al., 2002). The findings collectively reveal that the *N*-linked glycans are composed of *N*-acetyl glucosamine and different number of mannoses, *O*-glycosylation is diverse in sugar component and linkage mode (Goto, 2007b), and a slight difference was observed for the glycan structure of the same glycoprotein from different enzyme preparations or mutants (Goto, 2007b). However, to our knowledge, the high-throughput analysis of glycans from *A. niger* has not been conducted, and in most studies *N*- and *O*-linked glycans were not simultaneously characterized, and the glycan linkage information was not provided.

The glycan enrichment strategy using ZIC-HILIC coupled with dialysis was applied to analysis of the glycan structures of the secreted proteins from *A. niger*. Many of the nonexchangeable protons of the monosaccharide units in secretome glycans were identified based on the combination of gCOSY, TOCSY, and gHSQC. The key chemical shift data obtained from the NMR analysis are provided in Table 2. The sugar assignments were aided by

the comparison with the previous reports (see Table 2 for details) as well as the assignments of glycan structure of RNase B (Table 1).

The ^1H and ^{13}C NMR revealed that the glycoproteins from the *A. niger* secretome had the *N*-linked glycan structure similar to RNase B: 1) the chemical shifts of the glycans (e.g. 1- β -GlcNAc, 2- β -GlcNAc, and mannose) for *A. niger* secretome (Table 2) well match those for RNase B (Table 1); for example, anomeric proton peaks of Asn-linked glycoside “1- β -GlcNAc” were clearly resolved (Fig. 4A and B), and the 1- β -GlcNAc ring protons H_1 , H_2 , H_3 , H_4 , and H_5 have a distinguishing pattern similar to those of RNase B (Tables 1 and 2); 2) in TOCSY spectrum, the coupling constant between GlcNAc H_1 and H_2 ($^3J_{1-2}=8.0$ Hz) distinguished hexose protons from mannoses ($^3J_{1-2}=1-2$ Hz).

Except *N*-linked glycans, we also observed some *O*-linked glycans from *A. niger*, which is consistent with the literature report that *O*- and *N*-linked glycans often coexist in complex biological system such as the fungus *Aspergillus fumigatus* (Jin, 2012). The direct proof of *O*-linked glycan can be provided by gHBMC (Fig. 4C): the β -carbon signal (73.0 ppm) of Ser/Thr attached to hydroxyl group was correlated with anomeric H_1 of mannose ($\text{H}-1$, 5.06 ppm). This β -carbon signal (73.0 ppm) of Ser/Thr co-existed with the α -proton signal of Ser/Thr (4.60 ppm), which further correlated with the carbonyl group (~ 170 ppm), indicating the existence of *O*-linked glycopeptides on Ser/Thr with the structure of Ser-O-Man-X. The extended units X were complicated because of heterogeneous structures of *O*-linked glycans.

Extended units of *O*-linked glycan were further determined by two-dimensional NMR analysis and comparison with published data. 1) In gCOSY (Fig. 4B), the observed chemical shifts of H_1 from 4.96 to 5.42 ppm above water peak strongly indicated the presence of α configuration in the glycan chains (van Leeuwen et al., 2008), and the observed chemical shift of

H₂ around 3.60 ppm is often thought as typical peaks generated from α -glucose or glucose-like sugars (van Leeuwen et al., 2008). The characteristic anomeric H₁, suggested that there could be two types of alpha-glucose linkages: either 1->3 linkage because of the chemical shifts of H₁ (around 5.40 ppm) and H₂ (around 3.63 ppm), or 1->6 linkage because of the chemical shifts of H₁ (around 4.96 ppm) and H₂ (around 3.57 ppm) as described in literature (van Leeuwen et al., 2008). The glucose found in fungi was often linked either through 1->3 or 1->6 oxygen bridge bonds with mannose in glycopeptides, and further additional mannoses were often linked with Ser/Thr in *Aspergillus* species (Bennion et al., 2003; Gemmill and Trimble, 1999b; Goto, 2007a; Greenfield et al., 2012; Kim et al., 2003; Pazur et al., 1980); 2) We also observed α -galactofuranose present in the *O*-linked glycan chain, given that the significant signals from two-dimensional gCOSY and gHSQC (Table 2) gave distinguished peaks of α -galactofuranose (such as 5.19 ppm of H₁ correlated with H₂ of 4.14 ppm and C₁ of 110.2 ppm).

Taken together, we propose the possible structures of *O*-linked glycans from *A. niger* secretome in Fig. 5 as *O*-Man-X (X: either glucose or mannose units with or without α -galactofuranose). The *O*-linked glycans we observed could be initiated by mannose with *O*-linkage to Ser/Thr, followed by monosaccharide residues including glucose and mannose with alpha 1->2, 1->3, or 1->6 linkages. Because the structures of *O*-linked glycans are complicated, and the sample was a mixture containing *N*-linked and *O*-linked glycans, we could not completely assign the full structures of sugar sequences of individual glycans without additional separation approaches (e.g., graphite-activated carbon column). Overall, our approach demonstrated the feasibility of enriching *O*-linked and *N*-linked glycans from complex samples for NMR detection, which provides the global profiling of glycan structures that can be adapted for downstream analysis (e.g., MS-based glycolproteomics).

The carbohydrate is typically characterized using MS (Cao et al., 2014) and/or NMR. Both methods are complementary: NMR is the only powerful technique to characterize saccharide structure and sugar linkage based on the chemical shifts and coupling constants of protons without destruction of the analyte (Bubb, 2003; Fu et al., 1994a; Wohlgemuth et al., 2009), whereas the advantage of MS is its sensitivity and ability to provide the mass and glycosylation site information (Cao et al., 2014). Our ongoing research is to take advantage of the glycan information obtained using the NMR analysis to aid in bioinformatics analysis of the MS data to concurrently determine the *N*-/*O*-linked glycosylation sites and the corresponding glycan composition within the secreted proteins from *A. niger*.

According to the Fungal Secretome KnowledgeBase (Lum and Min, 2011), 886 proteins were predicted to be present in the secretome of *A. niger* (<http://proteomics.ysu.edu/secretomes/fungi.php>), most of which are hydrolases (e.g., glycosidase, peptidase/protease, and phosphatase), functionally uncharacterized proteins, and hypothetical proteins. Solid-phase glycopeptide enrichment followed by MS detection identified 156 *N*-linked glycoproteins from the *A. niger* secretome (Wang et al., 2011a). However, little is known about the glycan structure and *O*-linked glycosylation, so our work not only enriched the understanding of glycosylation of proteins from *A. niger*, but also provided useful information to facilitate future MS-based glycosylation study.

4.0. Conclusion and Perspective

Our work focused on the method for development of glycan enrichment for NMR detection. The results suggest that ZIC-HILIC separation coupled with dialysis is an efficient approach for both *N*-linked and *O*-linked glycan enrichment for NMR analysis. This

methodology also has the potential for the enrichment of *S*-linked glycans (through cysteine) which are involved in antimicrobial activities (Oman et al., 2011) but still poorly characterized (Wang and van der Donk, 2011), and could be applicable for the high-throughput analysis of glycan structures from any kind of biological system. Furthermore, the elucidation of the glycan structure using our method facilitates bioinformatics analysis of MS data to identify the sugar composition and the glycosylation sites within the glycoproteins by significantly minimizing the computational search against a large number of monosaccharides. MS analysis is very sensitive and could elucidate the glycosylation sites, so it can provide added insight into protein glycosylation. Last but not least, glycoproteins are involved in disease virulence (Szymanski et al., 2002) and biofuel production (Jeoh et al., 2008), so our sample preparation method could greatly facilitate the future glycoproteomics research involved in biomedicine and biofuel.

Acknowledgments

We thank Dr. Ziyu Dai for providing relevant protocols. We also thank anonymous reviewers for their helpful comments to improve the manuscript. Portions of this work were supported by funds from EMSL intramural research projects and EMSL capability development projects, the U.S. Department of Energy Office of Biological and Environmental Research (DOE-BER) Genome Sciences Program under the Pan-omics project, and the National Institute of Environmental Health Sciences of the National Institutes of Health (NIH) under Award Number R01ES022176. The work was performed at EMSL, a national scientific user facility sponsored by DOE-BER and located at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated by Battelle for DOE under Contract DE-AC05-76RL01830.

References

- Aguilar, J. A., et al., 2012. Decoupling Two-Dimensional NMR Spectroscopy in Both Dimensions: Pure Shift NOESY and COSY. *Angewandte Chemie-International Edition*. 51, 6460-6463.
- Angel, P. M., et al., 2007. A potential pitfall in 18O-based N-linked glycosylation site mapping. *Rapid Commun Mass Spectrom*. 21, 674-82.
- Beckham, G. T., et al., 2012. Harnessing glycosylation to improve cellulase activity. *Curr Opin Biotechnol*. 23, 338-45.
- Bennette, J. W., Lasure, L. L., 1991. *More Gene Manipulations in Fungi*. Academic Press. 441-458.
- Bennion, B., et al., 2003. Glycosphingolipids of the model fungus *Aspergillus nidulans*: characterization of GIPCs with oligo- α -mannose-type glycans. *Journal of Lipid Research*. 44, 2073-2088.
- Bubb, W. A., 2003. NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. *Concepts in Magnetic Resonance Part A*. 19A, 1-19.
- Cao, L., et al., 2014. Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. *Anal Biochem*. 452, 96-102.
- Carlson, D. M., 1966. Oligosaccharides isolated from pig submaxillary mucin. *J Biol Chem*. 241, 2984-6.
- Dai, Z., et al., 2013. Impact of *alg3* gene deletion on growth, development, pigment production, protein secretion, and functions of recombinant *Trichoderma reesei* cellobiohydrolases in *Aspergillus niger*. *Fungal Genet Biol*. 61, 120-32.
- Davis, J. T., et al., 1994a. ^1H -NMR Studies on an Asn-Linked Glycopeptide: GlcNAc-1 C2-N2 Bond Is Rigid in H_2O . *Journal of Biological Chemistry*. 269, 3331-3338.
- Davis, J. T., et al., 1994b. ^1H NMR studies on an Asn-linked glycopeptide. GlcNAc-1 C2-N2 bond is rigid in H_2O . *J Biol Chem*. 269, 3331-8.

- Di Palma, S., et al., 2011. Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC) provide high resolution separation and increase sensitivity in proteome analysis. *Anal Chem.* 83, 3440-7.
- Duus, J., et al., 2000. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem Rev.* 100, 4589-614.
- Freeze, H. H., Aebi, M., 2005. Altered glycan structures: the molecular basis of congenital disorders of glycosylation. *Curr Opin Struct Biol.* 15, 490-8.
- Fu, D., et al., 1994a. A detailed structural characterization of ribonuclease B oligosaccharides by ¹H NMR spectroscopy and mass spectrometry. *Carbohydr Res.* 261, 173-86.
- Fu, D. T., et al., 1994b. A Detailed Structural Characterization of Ribonuclease-B Oligosaccharides by H-1-NMR Spectroscopy and Mass-Spectrometry. *Carbohydrate Research.* 261, 173-186.
- Gemmill, T. R., Trimble, R. B., 1999a. Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochimica Et Biophysica Acta-General Subjects.* 1426, 227-237.
- Gemmill, T. R., Trimble, R. B., 1999b. *Schizosaccharomyces pombe* produces novel Gal(0-2)Man(1-3) O-linked oligosaccharides. *Glycobiology.* 9, 507-515.
- Gonzalez, J. E., et al., 1998. Biosynthetic control of molecular weight in the polymerization of the octasaccharide subunits of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Proc Natl Acad Sci U S A.* 95, 13477-82.
- Goto, M., 2007a. Protein O-glycosylation in fungi: Diverse structures and multiple functions. *Bioscience Biotechnology and Biochemistry.* 71, 1415-1427.
- Goto, M., 2007b. Protein O-glycosylation in fungi: diverse structures and multiple functions. *Biosci Biotechnol Biochem.* 71, 1415-27.
- Greenfield, L. K., et al., 2012. Biosynthesis of the Polymannose Lipopolysaccharide O-antigens from *Escherichia coli* Serotypes O8 and O9a Requires a Unique Combination of Single- and Multiple-active Site Mannosyltransferases. *Journal of Biological Chemistry.* 287, 35078-35091.
- Harvey, D. J., Structural Determination of ProteinBound Oligosaccharides. In: D. G. Large, C. D. Warren, Eds.), *Glycopeptides and Related Compounds: Synthesis, Analysis, and Applications.* Taylor & Francis, 1997, pp. 607.
- Inagaki, F., et al., 1987. Analysis of Nmr-Spectra of Sugar Chains of Glycolipids by Multiple Relayed Cosy and 2d Homonuclear Hartman-Hahn Spectroscopy. *Febs Letters.* 212, 91-97.
- Intoh, A., et al., 2009. Separation with zwitterionic hydrophilic interaction liquid chromatography improves protein identification by matrix-assisted laser desorption/ionization-based proteomic analysis. *Biomed Chromatogr.* 23, 607-14.
- Iwashkiw, J. A., et al., 2012. Identification of a general O-linked protein glycosylation system in *Acinetobacter baumannii* and its role in virulence and biofilm formation. *PLoS Pathog.* 8, e1002758.
- Jeoh, T., et al., 2008. Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Biotechnology for Biofuels.* 1, 10.
- Jin, C., 2012. Protein Glycosylation in *Aspergillus fumigatus* Is Essential for Cell Wall Synthesis and Serves as a Promising Model of Multicellular Eukaryotic Development. *Int J Microbiol.* 2012, 654251.
- Kamath, K. S., et al., 2011. Proteomic databases and tools to decipher post-translational modifications. *J Proteomics.* 75, 127-44.
- Keeler, J., 2010. *Understanding NMR Spectroscopy*, Second Edition. Wiley.
- Kim, S., et al., 2003. Structural determination of the N-glycans of a lepidopteran arylphorin reveals the presence of a monoglucosylated oligosaccharide in the storage protein. *Glycobiology.* 13, 147-157.
- Kornfeld, R., Kornfeld, S., 1976. Comparative aspects of glycoprotein structure. *Annu Rev Biochem.* 45, 217-37.

- Lum, G., Min, X. J., 2011. FunSecKB: the Fungal Secretome KnowledgeBase. Database (Oxford). 2011, bar001.
- Mann, M., Jensen, O. N., 2003. Proteomic analysis of post-translational modifications. *Nat Biotechnol.* 21, 255-61.
- Marino, K., et al., 2010. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol.* 6, 713-23.
- Marti, T., et al., 1984. The structure of the carbohydrate units of the carboxyl-terminal peptide of procollagen as elucidated by 500 MHz ¹H-NMR spectroscopy. *Biochimica et Biophysica Acta (BBA) - General Subjects.* 799, 305-312.
- Merry, A. H., et al., 2002. Recovery of intact 2-aminobenzamide-labeled O-glycans released from glycoproteins by hydrazinolysis. *Anal Biochem.* 304, 91-9.
- Oman, T. J., et al., 2011. Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat Chem Biol.* 7, 78-80.
- Paz-Parente, J., et al., 1983. Primary structure of a novel N-glycosidic carbohydrate unit, derived from hen ovomucoid: A 500-MHz ¹H-NMR study. *FEBS Letters.* 152, 145-152.
- Pazur, J. H., et al., 1980. Glycoenzymes: an unusual type of glycoprotein structure for a glucoamylase. *Carbohydrate Research.* 84, 103-114.
- Pu, Z., et al., 2000a. Structural studies of the glycopeptides of B-chain of cinnamomin--a type II ribosome-inactivating protein by nuclear magnetic resonance. *Glycoconj J.* 17, 749-59.
- Pu, Z., et al., 2000b. Structural studies of the glycopeptides of B-chain of cinnamomin - a type II ribosome-inactivating protein by nuclear magnetic resonance. *Glycoconjugate Journal.* 17, 749-759.
- Sweeney, P. J., Walker, J. M., *Enzymes of Molecular Biology.* In: M. M. Burrell, (Ed.), *Methods in Molecular Biology*, 1993, pp. 271-276.
- Szymanski, C. M., et al., 2002. Campylobacter protein glycosylation affects host cell interactions. *Infect Immun.* 70, 2242-4.
- Takegawa, Y., et al., 2006. Separation of isomeric 2-aminopyridine derivatized N-glycans and N-glycopeptides of human serum immunoglobulin G by using a zwitterionic type of hydrophilic-interaction chromatography. *J Chromatogr A.* 1113, 177-81.
- van Leeuwen, S. S., et al., 2008. Development of a ¹H NMR structural-reporter-group concept for the primary structural characterisation of α -D-glucans. *Carbohydrate Research.* 343, 1114-1119.
- Wallis, G. L. F., et al., 2001. Galactofuranoic-oligomannose N-linked glycans of α -galactosidase A from *Aspergillus niger*. *European Journal of Biochemistry.* 268, 4134-4143.
- Wang, H., van der Donk, W. A., 2011. Substrate selectivity of the sublancin S-glycosyltransferase. *J Am Chem Soc.* 133, 16394-7.
- Wang, L., et al., 2011a. Mapping N-Linked Glycosylation Sites in the Secretome and Whole Cells of *Aspergillus niger* Using Hydrazide Chemistry and Mass Spectrometry. *J Proteome Res.*
- Wang, L., et al., 2011b. Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. *Theor Appl Genet.* 122, 327-40.
- Warren, M. E., et al., 2002. Studies on the glycosylation of wild-type and mutant forms of *Aspergillus niger* pectin methylesterase. *Carbohydrate Research.* 337, 803-812.
- Wittwer, A. J., et al., 1989. Effects of N-glycosylation on in vitro activity of Bowes melanoma and human colon fibroblast derived tissue plasminogen activator. *Biochemistry.* 28, 7662-9.
- Wohlgemuth, J., et al., 2009. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Analytical Biochemistry.* 395, 178-188.
- Woosley, B., et al., 2006. Comprehensive glycan analysis of recombinant *Aspergillus niger* endo-polygalacturonase C. *Analytical Biochemistry.* 354, 43-53.

Young, N. M., et al., 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. J Biol Chem. 277, 42530-9.

Figure Captions

Fig. 1. Schematic representation of the enrichment of glycans for NMR analysis. The protein sample containing glycoproteins was digested with pronase E, and the digestion efficiency was examined using SDS-PAGE. The completely or nearly completely digested products were loaded onto the ZIC-HILIC column to enrich *N*- and *O*-linked glycans, and then the enrichment was subject to dialysis with a filter (molecular weight cutoff: 1 kDa) to remove the small molecules non-specifically bound to the ZIC-HILIC column. Finally, the collected samples were analyzed by NMR.

Fig. 2. One-dimensional ^1H -NMR spectra for the RNase B glycan enriched using A) ZIC-HILIC separation only or B) ZIC-HILIC separation followed by dialysis. To compare the sugar part, all spectra was raised to the same height around chemical shifts 4.95 to 5.45 ppm. The actual peptides contents in ZIC-HILIC without dialysis were enormous and make the typical 1-GlcNAc and 2-GlcNAc acetyl peaks (2.01 and 2.07 ppm) in glycan invisible. The chemical shifts of amino acids, peptides, and their linear side chains fall into the region of 0.5 to 4.4 ppm, and the chemical shifts of polysaccharide chains into the region of 3.0 to 5.5 ppm.

Fig. 3. Two-dimensional NMR analysis (TOCSY) of glycans from RNase B. Man = mannose; GlcNAc = *N*-acetylglucosamine.

Fig. 4. Two-dimensional NMR analysis of glycans from *A. niger*. A) TOCSY analysis. B) gCOSY analysis. C) gHMBC analysis. Man = mannose; Glc = glucose; Ser = serine; Thr = threonine; Gal = galactose; GlcNAc = *N*-acetylglucosamine.

Fig. 5. Proposed structures of *O*-linked glycans from *A. niger*. Ser = serine; Glc = glucose; Man = mannose; Galf = galactofuranose

Figure 1.

Figure 2.

Fig. 3.

Fig. 4A.

Fig 4C.

Fig. 5.

Tables and captions

Table 1

^1H and ^{13}C NMR chemical shifts (ppm) of the glycan moiety of RNase B. NA: Not available (not all chemical shifts were identified). The sugar numberings (1, 2, 3, 4, 5, 6), based on the position of the carbon in the sugar ring, are listed in the table header (see Fig. S1 for details).

Linkage	RNase B	Chemical shifts	1	2	3	4	5	6	References
N-linked glycan	1- β -GlcNAc	H	5.03	3.86	3.75	3.65	3.61	NA	(Davis et al., 1994b; Fu et al., 1994a; Marti et al., 1984; Paz-Parente et al., 1983; Pu et al., 2000a; van Leeuwen et al., 2008)
		C	105.3	73.4	NA	NA	NA	NA	
	2- β -GlcNAc	H	4.61	3.77	3.77	3.77	3.68	NA	
		C	103.6	76.2	NA	NA	NA	NA	
	Man-3	H	4.84	4.14	3.81	3.77	NA	NA	
		C	102.9	71.3	NA	NA	NA	NA	
	Man-4	H	5.11	4.02	3.91	NA	NA	NA	
		C	104.9	71.9	NA	NA	NA	NA	
	Man-4'	H	4.88	3.99	3.91	NA	NA	NA	
		C	102.6	75.5	NA	NA	NA	NA	
	Man-B	H	5.13	4.01	NA	NA	NA	NA	
		C	110.3	NA	NA	NA	NA	NA	
	Man-C	H	5.28	4.2	NA	NA	NA	NA	
		C	110.3	NA	NA	NA	NA	NA	
	Man-A	H	5.42	4.09	NA	NA	NA	NA	
		C	117.3	NA	NA	NA	NA	NA	

Table 2

^1H and ^{13}C NMR chemical shifts (ppm) of the glycan moiety of the secretome from *A. niger*. The sugar numberings (1, 2, 3, 4, 5, 6), based on the position of the carbon in the sugar ring, are listed in the table header (see Fig. S1 for details). The majority of glycans from the secretome were *O*-linked polysaccharides. *: mannoses present in either *N*-linked or *O*-linked glycans; ***: the mannose was *O*-linked to Ser/Thr. NA: Not available (not all chemical shifts were identified). Galf: galactose furanose. Different chemical shifts were observed for the same sugar unit, which indicated impurities (Iwashkiw et al., 2012).

Linkage	Secretome	Chemical shifts	1	2	3	4	5	6	References
N-linked glycan	1- β -GlcNAc	H	5.07	3.86	3.75	3.66	3.58	NA	(Davis et al., 1994b)
		C	103.6	73.2	NA	NA	NA	NA	
	2- β -GlcNAc	H	4.66	3.75	3.79	3.71	3.61	NA	
		C	104.5	74.2	NA	NA	NA	NA	
	2- β -GlcNAc	H	4.62	3.84	NA	NA	NA	NA	
		C	104.9	76.2	NA	NA	NA	NA	
	Man-3	H	4.87	4.15	NA	NA	NA	NA	
		C	102.6	79.7	NA	NA	NA	NA	
	Man-4	H	5.10	4.1	NA	NA	NA	NA	
		C	105.5	79.4	NA	NA	NA	NA	
O-linked glycan	*Man	H	5.27	4.19	NA	NA	NA	NA	(Bennion et al., 2003; Greenfield et al., 2012)
		C	110.0	84.8	NA	NA	NA	NA	
	*Man	H	5.22	4.16	NA	NA	NA	NA	
		C	103.9	84.6	NA	NA	NA	NA	
	*Man	H	5.18	4.05	NA	NA	NA	NA	
		C	105.6	76.2	NA	NA	NA	NA	
	***Man	H	5.05	4.07	NA	NA	NA	NA	
		C	105.5	73.3	NA	NA	NA	NA	
	*Man	H	4.91	4.01	NA	NA	NA	NA	
		C	105.5	76.2	NA	NA	NA	NA	
	1->3- α -glucose	H	5.41	3.64	3.95	3.67	4.03	3.85/3.73	(van Leeuwen et al., 2008)
		C	102.8	74.1	82.2	70.1	72.6	63.8	
	1->3- α -glucose	H	5.37	3.68	3.99	3.67	4.06	3.82	
		C	102.8	74.1	82.2	70.1	73.3	63.8	
	1->3- α -glucose	H	5.32	3.59	4.01	3.61	4.03	3.80/3.74	
		C	103.7	74.1	82.2	70.1	72.6	63.8	
	1->3- α -glucose	H	5.24	3.58	3.97	3.62	NA	NA	
		C	104.9	74.4	82.2	70.1	NA	NA	
	1->3- α -glucose	H	5.12	3.65	3.91	3.62	4.03	3.85/3.76	
		C	101.2	NA	NA	NA	NA	NA	
	1->6- α -glucose	H	5.01	3.57	3.75	3.45	3.78	4.02	
		C	101.2	74.2	75.9	NA	NA	64.4	
	1->6- α -glucose	H	4.97	3.63	3.75	3.45	3.78	4.02	
		C	101.2	74.4	NA	NA	NA	NA	
	1->5- β -Galf	H	5.19	4.15	4.08	4.05	3.95	3.81	(Gemmill and Trimble, 1999a; Goto, 2007b)
		C	110.2	84.6	79.4	82.1	NA	NA	
	1->6- β -Galf	H	5.04	4.13	4.05	4.05	NA	NA	
		C	109.5	84.6	81.8	NA	NA	NA	