A Universal Chemical Enrichment Method for Mapping the Yeast N-glycoproteome by Mass Spectrometry (MS)*§

Weixuan Chen‡, Johanna M. Smeekens‡, and Ronghu Wu‡§

Glycosylation is one of the most common and important protein modifications in biological systems. Many glycoproteins naturally occur at low abundances, which makes comprehensive analysis extremely difficult. Additionally, glycans are highly heterogeneous, which further complicates analysis in complex samples. Lectin enrichment has been commonly used, but each lectin is inherently specific to one or several carbohydrates, and thus no single or collection of lectin(s) can bind to all glycans. Here we have employed a boronic acid-based chemical method to universally enrich glycopeptides. The reaction between boronic acids and sugars has been extensively investigated, and it is well known that the interaction between boronic acid and diols is one of the strongest reversible covalent bond interactions in an aqueous environment. This strong covalent interaction provides a great opportunity to catch glycopeptides and glycoproteins by boronic acid, whereas the reversible property allows their release without side effects. More importantly, the boronic acid-diol recognition is universal, which provides great capability and potential for comprehensively mapping glycosylation sites in complex biological samples. By combining boronic acid enrichment with PNGase F treatment in heavy-oxygen water and MS, we have identified 816 N-glycosylation sites in 332 yeast proteins, among which 675 sites were well-localized with greater than 99% confidence. The results demonstrated that the boronic acid-based chemical method can effectively enrich glycopeptides for comprehensive analysis of protein glycosylation. A general trend seen within the large data set was that there were fewer glycosylation sites toward the C termini of proteins. Of the 332 glycoproteins identified in yeast, 194 were membrane proteins. Many proteins get glycosylated in the high-mannose N-glycan biosynthetic and GPI anchor biosynthetic pathways. Compared with lectin enrichment, the current method is more cost-efficient, generic, and effective. This method can be extensively applied to different complex samples for the comprehensive analysis of protein glycosylation. Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.036251, 1563–1572, 2014.

Glycosylation is an extremely important protein modification that frequently regulates protein folding, trafficking, and stability. It is also involved in a wide range of cellular events (1) such as immune response (2, 3), cell proliferation (4), cell-cell interactions (5), and signal transduction (6). Aberrant protein glycosylation is believed to have a direct correlation with the development of several diseases, including diabetes, infectious diseases, and cancer (7–11). Secretory proteins frequently get glycosylated, including those in body fluids such as blood, saliva, and urine (12, 13). Samples containing these proteins can be easily obtained and used for diagnostic and therapeutic purposes. Several glycoproteins have previously been identified as biomarkers, including Her2/Neu in breast cancer (14), prostate-specific antigen (PSA) in prostate cancer (15), and CA125 in ovarian cancer (16, 17), which highlights the clinical importance of identifying glycoproteins as indicators or biomarkers of diseases. Therefore, effective methods for systematic analysis of protein glycosylation are essential to understand the mechanisms of glyobiology, identify drug targets and discover biomarkers.

Approximately half of mammalian cell proteins are estimated to be glycosylated at any given time (18). There have been many reports regarding identification of protein glycosylation sites and elucidation of glycan structures (19–30). Glycan structure analysis can lead to potential therapeutic and diagnostic applications (31, 32), but it is also critical to identify which proteins are glycosylated as well as the sites at which the modification occurs. Despite progress in recent years, the large-scale analysis of protein glycosylation sites using MS-based proteomics methods is still a challenge. Without an effective enrichment method, the low abundance of glycoproteins prohibits the identification of the majority of sites using the popular intensity-dependent MS sequence method.

About a decade ago, a very beautiful and elegant method based on hydrazide chemistry was developed to enrich glycopeptides. Hydrazide conjugated beads reacted with aidel-
Universal Chemical Enrichment for N-glycoproteome

Hydrides formed from the oxidation of cis-diols in glycans (33). This method has been extensively applied to many different types of biological samples (34–41). Besides the hydrazide-based enrichment method, lectins have also been frequently used to enrich glycopeptides or glycoproteins before MS analysis (28, 29, 42–46). However, there are many different types of lectins, and each is specific to certain glycans (47, 48). Therefore, no combination of lectins can bind to all glycosylated peptides or proteins, which prevents comprehensive analysis of protein glycosylation. Because of the complexity of biological samples, effective enrichment methods are critical for the comprehensive analysis of protein glycosylation before MS analysis.

One common feature of all glycoproteins and glycopeptides is that they contain multiple hydroxyl groups in their glycans. From a chemistry point of view, this can be exploited to effectively enrich them. Ideally, chemical enrichment probes must have both strong and specific interactions with multiple hydroxyl groups. The reaction between boronic acids and 1,2- or 1,3-cis-diols in sugars has been extensively studied (49–52) and applied for the small-scale analysis of glycoproteins (53–55). Furthermore, boronate affinity chromatography has been employed for the analysis of nonenzymatically glycated peptides (56, 57). Boronic acid-based chemical enrichment methods are expected to have great potential for global analysis of glycopeptides when combined with modern MS-based proteomics techniques. However, the method has not yet been used for the comprehensive analysis of protein N-glycosylation in complex biological samples (58).

Yeast is an excellent model biological system that has been extensively used in a wide range of experiments. Last year, two papers reported the large-scale analysis of protein N-glycosylation in yeast (59, 60). In one study, a new MS-based method was developed based on N-glycopeptide mass envelopes with a pattern via metabolic incorporation of a defined mixture of N-acetylglucosamine isotopologs into N-glycans. Peptides with the recoded envelopes were specifically targeted for fragmentation, facilitating high confidence site mapping (59). Using this method, 133 N-glycosylation sites were confidently identified in 58 yeast proteins. When combined with an effective enrichment method, this MS-based analysis will provide a more complete coverage of the N-glycoproteome. The other work combined lectin enrichment with digestion by two enzymes (Glu_c and trypsin) to increase the peptide coverage, and 516 well-localized N-glycosylation sites were identified in 214 yeast proteins by MS (60).

Here we have comprehensively identified protein N-glycosylation sites in yeast by combining a boronic acid-based chemical enrichment method with MS-based proteomics techniques. Magnetic beads conjugated with boronic acid were systematically optimized to selectively enrich glycosylated peptides from yeast whole cell lysates. The enriched peptides were subsequently treated with Peptide-N4-(N-acetyl-beta-glucosaminy1)asparagine amidase (PNGase F) in heavy-oxygen water. Finally, peptides were analyzed by an on-line LC-MS system. Over 800 protein N-glycosylation sites were identified in the yeast proteome, which clearly demonstrates that the boronic acid-based chemical method is an effective enrichment method for large-scale analysis of protein glycosylation by MS.

EXPERIMENTAL PROCEDURES

Cell Culture, Cell Lysis, and Protein Extraction and Digestion—BY4742 MAT alpha yeast were grown in YPD overnight. Cells were harvested by centrifugation in the exponential growth phase (O.D. was about 1.0 at 600 nm). Then they were resuspended at 4 °C in a buffer containing 50 mM Tris pH 8.2, 8 mM urea, 75 mM NaCl, 2% SDS and one protease inhibitor mixture tablet (complete mini, EDTA-free, Roche) per 10 ml. Cells were lysed using the MiniBeadbeater (Biospec) at maximum speed, four cycles of 40 s each, with 2 min pauses between cycles to minimize protein degradation. After centrifuging at 13,000 rpm, the supernatants were transferred to new tubes. The protein concentration was measured by BCA protein assay (Pierce, Rockford, IL) and proteins were subjected to disulfide reduction with 5 mM DTT (56 °C, 25 min) and alkylation with 15 mM iodoacetamide (RT, 30 min in the dark). Excess iodoacetamide was quenched with 5 mM DTT (room temperature, 15 min in the dark). Proteins were purified and SDS was removed by protein TCA precipitation. The purified proteins were digested with trypsin (Promega, Madison, WI) at the ratio of ~100:1 in 25 mM Tris pH 8.2, 1.5 mM urea, at 37 °C for 15 h.

The Boronic Acid Method for the Glycopeptide Enrichment—Digestion was quenched by the addition of 10% TFA to a final concentration of 0.4%, and the resulting peptides were purified using a tC18 Sep-Pak cartridge (Waters, Milford, MA). The peptide mixture was dissolved in 100 mM ammonium acetate buffer, and incubated for one hour with rotation with magnetic beads bound to boronic acid. After incubation, the beads were washed three times with the binding buffer. Enriched peptides were eluted first by incubation with a 5% formic acid solution at 37 °C for 30 min. The peptides were further eluted with a solution containing acetonitrile:H2O:trifluoroacetic acid at 50:49:1.

Glycopeptide Enrichment by Lectin—The enrichment of glycopeptides by lectin has been extensively reported in the literature (12, 22, 61–64). Purified peptides from a yeast whole cell lysate were incubated with lectin (Con A and WGA) bound to agarose beads in a binding buffer (pH 7.4) containing 20 mM Tris, 0.5 mM NaCl, 1 mM CaCl2 and 1 mM MnCl2 for 10 min. After incubation, the beads were washed with the binding buffer four times. Finally, glycosylated peptides were eluted with a buffer containing 10 mM EDTA, 20 mM ethylenediamine, 200 mM methyl α-D-mannoside, and 200 mM N-acetyl-α-D-glucosamine.

PNGase F Treatment—The enriched samples were dried in a lyophilizer for at least 24 h. The completely dried samples were dissolved in heavy-oxygen water and treated with PNGase F (lyophilized powder from Sigma Aldrich) overnight. PNGase F removed sugar groups in Asn and as a result, Asn was converted to Asp with a tag of O18.

Glycopeptide Fractionation—The enriched glycopeptide samples were desalted again using a tC18 Sep-Pak cartridge. Then we per-
formed the separation of glycopeptides using high-pH reversed phase HPLC (pH = 10). The sample was separated into 10 fractions using a 4.6 × 250 mm 5 µm particle reversed phase column (Waters) with a 40-min gradient of 5%-30% acetonitrile (ACN) with 25 mM ammonium acetate. Every fraction was further purified with stage tips.

**LC-MS/MS Analyses**—Purified and dried peptide samples were resuspended in a solvent of 5% ACN and 4% formic acid (FA), and 2 µl was loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 5 µm,200 Å, 100 µm x 16 cm) using a WPS-3000TPLRS autosampler (Ultimate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler, Dionex). Peptides were separated by reversed-phase chromatography using an Ultimate 3000 binary pump with a 90-min gradient of 4%-30% ACN (in 0.125% FA) and detected in a hybrid dual-cell quadrupole linear ion trap - orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher, with a software of Xcalibur 2.0.7 SPI) using a data-dependent Top20 method (65). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10th AGC target was followed by up to 20 MS/MS in the LTQ for the most intense ions. Selected ions were excluded from further analysis for 90 s. Ions with a single or unassigned charge were not sequenced. Maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

**Database Searches and Data Filtering**—The raw files recorded by MS were converted into mzXML format. Precursors for MS/MS fragmentation were checked for incorrect monoisotopic peak assignments while refining precursor ion mass measurements (66). All fragmentation were checked for incorrect monoisotopic peak assignments while refining precursor ion mass measurements (66). All MS/MS spectra were then searched using the SEQUEST algorithm (version 28) (67). Spectra were matched against a database encompassing sequences of all proteins (6607 protein entries) in the yeast ORFs database (S288C 2010) downloaded from SGD. Each protein sequence was listed in both forward and reversed orientations to provide positive and negative training data. This approach was employed to evaluate and further control FDRs of glycopeptide identification (68). Linear discriminant analysis (LDA) was used to distinguish correct and incorrect peptide identifications using numerous parameters such as Xcorr, ΔCn, and precursor mass error (66). Separate linear discriminant models were trained for each raw file using forward and reversed peptide sequence to estimate false positive rate (FDR) of peptide identification. The following parameters were used for the database search: 20 ppm precursor mass tolerance; 1.0 Da product ion mass tolerance; fully tryptic sequencing; up to two missed cleavages; variable modifications: oxidation of methionine (+15.9949) and O18 tag of Asn (+2.9883); fixed modifications: carbamidomethylation of cysteine (+57.0214).

The target-decoy method was employed to evaluate and further control FDRs of glycopeptide identification (68). Linear discriminant analysis (LDA) was used to distinguish correct and incorrect peptide identifications using numerous parameters such as Xcorr, ΔCn, and precursor mass error (66). Separate linear discriminant models were trained for each raw file using forward and reversed peptide sequences to provide positive and negative training data. This approach is similar to other methods in the literature (69–71). After scoring, peptides less than six amino acids in length were discarded and peptide spectral matches were filtered to a less than 1% FDR based on the number of decoy sequences in the final data set. The data set was restricted to glycopeptides when determining FDRs.

**Glycosylation Site Localization**—To assign glycosylation site localizations and measure the assignment confidence, we applied a probabilistic algorithm (72) that considers all glycoforms of a peptide and uses the presence or absence of experimental fragment ions unique to each to create a Modscore. The Modscore, which is similar to Ascore, indicates the likelihood that the best site match is correct when compared with the next best match. We considered sites with a score ≥19 to be confidently localized.

**Data Dissemination**—Four supplementary tables as Excel files are available (supplemental Tables S1-S4): N-glycosylation sites identified in the first experiment (supplemental Tables S1), N-glycosylation sites identified in the second experiment (supplemental Tables S2), total N-glycosylation sites identified in this work (supplemental Tables S3), and N-glycosylation sites identified in proteins in the Glycosylphosphatidylinositol (GPI) anchor biosynthesis pathway (supplemental Tables S4). Annotated, mass labeled spectra for all modified peptides from the parallel experiments can be found online with the supplemental data. All raw files and annotated spectra are accessible in the following public accessible website (http://www.peptideatlas.org/PASS/PASS00443, Data set Identifier: PASS00443, Password: TA3755yv).

**RESULTS AND DISCUSSION**

**Protein N-glycosylation Site Identification**—The basic principle of the boronic acid enrichment method is displayed in Fig. 1. It is well known that boronic acids react with 1,2- or 1,3-cis-diols, which are present in glycans of glycopeptides, saccharides, and many other molecules, including nucleic acids. To minimize the interference from other molecules containing 1,2- or 1,3-cis-diol groups, proteins were further purified after extraction by protein TCA precipitation.

Many membrane proteins are known to be glycosylated. During protein extraction, 2% SDS was added into the lysis buffer to dissolve membrane proteins. However, detergents like SDS should be avoided for MS analysis because it will dramatically affect the stability of electrospray ionization and suppress ion intensity. Protein TCA precipitation can also be used to remove SDS from solutions.

The reaction between boronic acid and sugars has been extensively studied (49, 51) and it has been established that boronic acid can form reversible covalent bonds with sugars under basic conditions which can be broken in acidic solutions. To obtain the most effective and comprehensive enrichment of glycopeptides from yeast whole cell lysates, several experimental parameters were optimized. First, the pH of the binding buffer was varied. As shown in Fig. 2A, the largest number of unique glycopeptides was identified when the pH of the binding buffer was 10. When the pH was below 10, the binding between boronic acid and glycans was not very effective. At pH values above 10, fewer glycopeptides and total peptides were identified, which may be caused by not only peptide hydrolysis but also the hydrolysis of the amide bonds connected to magnetic beads that release 3-amino boronic acid. Next, the optimal number of washes after binding was investigated. After the one hour binding incubation, the magnetic beads were washed with the binding buffer to remove non-specifically bound peptides. The boronic acid forms covalent bonds with sugars in glycopeptides under basic conditions, which should be stable enough for the stringent wash. However, extensive washing steps could cause the loss of magnetic beads and consequently the loss of bonded peptides. Finally, four washes resulted in the largest number of unique glycopeptides identified, as shown in Fig. 2B.

After enrichment, samples were thoroughly dried and treated with PNGase F in heavy-oxygen water. This method has been extensively applied for protein N-glycosylation site identification. Theoretically, PNGase F treatment can be carried out in normal water and Asn would get converted into Asp, which would result in a mass shift of 1 Dalton. That shift could be used to search against the yeast database to identify...
N-glycosylation sites. However, the deamination of Asn may occur in vivo and in vitro, which introduces the possibility of false positives (73). The PNGase F treatment in heavy-oxygen water can allow us more confidently identify N-glycosylation sites and minimize potential false positive identification.

Enriched and deglycosylated peptides were fractionated by high-pH reversed HPLC, and subsequently analyzed with an on-line LC-MS system. In the biological duplicate samples, 8427 and 10,153 total glycopeptides corresponding to 1274 and 1289 unique glycopeptides, respectively, were identified. Glycosylated peptides are expected to be more hydrophilic because each glycan contains several hydroxyl groups. In the current experiments, however, sugars were removed by PNGase F before fractionation. According to our results, more glycopeptides were found in the fractions collected at late elution times, which suggests that glycopeptides are more hydrophobic after glycans are removed.

An example of a tandem mass spectrum of a glycopeptide (LFN#SSSALN#ITELYNVAR, # denotes glycosylation sites) is displayed in Fig. 1B. Based on the high mass accuracy of the precursor m/z (0.7 ppm) and many fragments present in the spectrum, this glycopeptide was confidently identified to be from the protein Ygp1, with an Xcorr of 6.3. An example of a tandem mass spectrum of a glycopeptide (LFN#SSSALN#ITELYNVAR, # denotes glycosylation sites) is displayed in Fig. 1B. Based on the high mass accuracy of the precursor m/z (0.7 ppm) and many fragments present in the spectrum, this glycopeptide was confidently identified to be from the protein Ygp1, with an Xcorr of 6.3, containing two glycosylation sites with the consensus motif of NXS/T (X is any amino acid residue except proline). Ygp1 is a cell wall-related secretory glycoprotein that is induced by nutrient de-
privation-associated growth arrest and upon entry into stationary phase.

Although a 20 ppm window was used to match the spectra, the vast majority of identified peptides had a mass shift of less than 3 ppm. The mass accuracy distribution of the 8427 total glycopeptides identified in the first experiment is displayed in Fig. 2C. Very few glycopeptides had a mass shift greater than 5 ppm, which shows the high confidence associated with identification. Among 1274 unique peptides, approximately three-quarters of glycopeptides contain one modification site, whereas one-fifth contain two sites, as shown in Fig. 2D.

Comparison of the Current Boronic Acid Method with the Lectin Enrichment Method—Considering the lectin method has been most commonly used for enriching glycopeptides or glycoproteins, we ran parallel experiments to compare the current boronic acid method with the lectin enrichment method. The identical peptide samples from a yeast whole cell lysate digest were used for the parallel experiments. Besides the enrichment methods, all other steps were exactly the same. Finally, we identified 348 and 288 N-glycosylation sites by using the boronic acid and lectin enrichment methods, respectively, as shown in the supplemental Fig. S1. Based on the current results, the enrichment specificity of the boronic acid method is slightly higher than that of the lectin-based method. The specificity can be improved by more stringent washes, but some glycopeptides could be lost. Currently state-of-the-art MS instruments with rapid speed and high sensitivity can compensate low specificities. Overall, in these parallel experiments more N-glycosylation sites were identified by the boronic acid enrichment method, which is consistent with the universal sugar recognition of boronic acid, in contrast to the glycan structural restrictions of lectin.

Comparison of Biological Duplicate Experiments—To evaluate the effectiveness of the boronic acid-based chemical enrichment method, we performed two parallel experiments using biological duplicates. Yeast cells were grown separately, but all other experimental conditions were consistent. From the two experiments, 665 and 687 N-glycosylation sites were identified in 1274 and 1289 unique glycopeptides, respectively (supplemental Table S1 and S2). A total of 816
glycosylation sites were identified (supplemental Table S3), among which 536 common sites were found in both experiments (Fig. 2E), i.e. about 80% of total sites from each experiment. The overlap between the two experiments was quite high considering protein glycosylation is a reversible and dynamic process. In previous biological duplicate experiments for large-scale analysis of protein phosphorylation, the number of common sites identified in the two experiments were about 60% of the total sites in each experiment, despite the fact that over 8000 total phosphorylation sites were identified (74). The overlap is much lower than that found here, which may be caused by differences in the nature of these two modifications. Both phosphorylation and glycosylation are reversible in living organisms. However, phosphorylation may be more dynamic to rapidly execute signal transductions, so more enzymes are responsible for phosphorylation than N-glycosylation in yeast.

We also investigated the overlap at the glycoprotein level. In the two experiments, 266 and 282 glycoproteins were identified, respectively, among which 216 were common to both, as shown in Fig. 2F. A total of 332 glycoproteins containing 816 sites were identified (supplemental Table S3); 675 sites were well-localized with a ModScore>19, which represents over 99% confidence.

**N-glycosylation Site Distribution**—To determine whether protein N-glycosylation is biased for the N terminus, C terminus or other fragments of proteins, we performed position analysis for the N-glycosylation sites along the protein sequences. In this analysis, the protein length was divided into 1% bins and the number of glycosylation sites within each bin was quantified. The results show a general trend of fewer glycosylation sites near the C terminus (Fig. 3A). This is possibly caused by steric hindrance in the secondary structures formed in the near-complete nascent polypeptide, which has been reported in previous studies (75–77).

Another interesting phenomenon is that the first three bins have significantly fewer glycosylation sites compared with the remaining bins. To further investigate this observation, we compared the number of glycosylation sites found in each set of 10 amino acid residues starting from the N terminus (Fig. 3B). Only the first 180 amino acid residues were compared. Within the first 20 amino acid residues, only about half of the average number of sites (per ten residues) were found. These findings are consistent with the computational results that have been previously reported (78). One potential explanation for the fewer glycosylation sites found in the first 20 residues is that signal sequences at the N terminus could have been cleaved and thus could not be identified in this experiment.

**Highly Glycosylated Proteins**—Considering 816 sites in 332 proteins, every glycoprotein contained an average of 2.6 sites. As shown in Fig. S1A, more than half (54%) of glycoproteins were found to have a single site, and 17% contained two sites. In our data set, 40 proteins (12%) had at least five sites. Among these heavily glycosylated proteins, 30 were membrane proteins, and 11 were involved in the external encapsulating structure based on Gene Oncology analysis from the Database for Annotation, Visualization and Integrated Discovery (DAVID) (79). Several highly glycosylated proteins are listed in Fig. 4B.

The protein containing the greatest number of glycosylation sites in our data set is Rax2, with 20 sites. Rax2 is a large protein containing 1220 amino acid residues that is required for the maintenance of the bipolar budding pattern. It is involved in selecting bud sites at both the distal and proximal poles of daughter cells. Based on predictions from the Uniprot website, this protein is highly glycosylated with 52 potential sites.

Several acid phosphatases were also found to be highly glycosylated. Here 13 sites were identified in Pho11, which is one of three repressible acid phosphatases (72). This glycoprotein is transported to the cell surface by the secretory pathway. Pho5 is another phosphatase that facilitates extracellular nucleotide-derived phosphate hydrolysis. We found 11 sites, whereas 12 sites were predicted and listed in the Uniprot website. In addition, we identified 10 sites in Pho3, which is brought to the cell surface by transport vesicles. Heavy glycosylation in these proteins could play an important role in their transportation.

**Glycoprotein Clustering**—To understand which groups of glycoproteins were enriched, we used DAVID (79) to cluster all 332 of the glycoproteins identified here. Fig. 5 shows the results based on cellular compartment. The analysis assigned over half (194) as membrane proteins, which were highly...
enriched with a \( p \) value of 7.4E-19. The endoplasmic reticulum (ER) glycoproteins were the next most enriched group, with a total of 44 proteins and a \( p \) value of 5.3E-16. We also found 31 glycoproteins involved in the cell wall, which is the same number found in the external encapsulating structure.

Clustering based on biological processes demonstrates that the carbohydrate metabolic process was the most enriched, with 68 glycoproteins corresponding to the lowest \( p \) value of 2E-20. Almost half of them (148) were relevant to the protein metabolic process. Additionally, based on molecular function analysis, 36 glycoproteins were assigned to be responsible for the transfer of glycosyl groups and 19 were found to have hydrolase activity on glycosyl bonds.

Glycoproteins Involved in Sugar-related Biosynthesis Pathways—High-mannose \( N \)-glycans are very common in yeast, and 12 proteins are involved in their biosynthesis. Pathway analysis shows that 7 proteins get glycosylated; sites identified in these proteins are listed in Table I. Furthermore, 10 glycoproteins were involved in the \( O \)-mannosyl glycan biosynthesis: Kre2, Ktr1, Pmt1, Pmt2, Pmt3, Pmt5, Pmt6, Mnn1, Mnt3, and Mnt3. These suggest that the proteins responsible for mannosyl glycan biosynthesis are commonly modified by \( N \)-glycans.

GPI anchor biosynthesis produces the same oligosaccharide core in all organisms analyzed so far: protein-CO-NH-(CH\(_{2}\)\(_{2}\))\(_{2}\)-PO\(_{4}\)-6Man\(_{\alpha}1\sim2Man\(_{\alpha}1\sim6Man\(_{\alpha}1\sim4GlcNH_{2}\)-myo-inositol-PO\(_{4}\)-lipid (80). Among 23 annotated proteins in this pathway, our results show that 9 of them were glycosylated;

| Protein | Sites found in this work | Sites predicted in Uniprot | Annotation |
|---------|--------------------------|-----------------------------|------------|
| Rax2    | 20                       | 52                          | N-glycosylated protein involved in the maintenance of bud site selection |
| Ath1    | 15                       | 28                          | Acid trehalase required for utilization of extracellular trehalose |
| Pat1    | 14                       | 15                          | Cell wall mannoprotein |
| Pho11   | 13                       | 10                          | Acid phosphatase |
| YLR413W | 13                       | 18                          | Cell membrane protein |
| PLB1    | 13                       | 19                          | Lysoospholipase 1 |
| Pho5    | 11                       | 12                          | Repressible acid phosphatase |
| Pho3    | 10                       | 12                          | Constitutively expressed acid phosphatase |

\[ \text{Fig. 4. A, Distribution of the numbers of glycosylation sites in glycoproteins; B, List of the numbers of glycosylation sites identified in this work compared with those predicted by Uniprot.} \]

\[ \text{Fig. 5. Glycoprotein clustering based on (A) cellular compartments, and (B) biological processes. The } y \text{ axis is } -\log(p) \text{ where the } p \text{ value reflects the enrichments of glycoprotein groups.} \]
all sites identified in these proteins are listed in supplemental Table S4. For example, Pbn1, a homolog of mammalian PIG-X, is an essential component of glycosylphosphatidylinositol-mannosyltransferase I, and required for the autocatalytic post-translational processing of the protease B precursor Prb1. Two sites (N212 and N365) were confidently identified in Pbn1. Because there are four sugars contained in the core structure, it is expected that protein glycosylation plays an important role in the GPI anchor biosynthesis.

In conclusion, protein N-glycosylation is a very important co/post-translational modification. The low abundance of glycoproteins and high heterogeneity of glycans make it extremely challenging to comprehensively identify protein glycosylation in complex biological samples. Lectin-based methods are the most common enrichment techniques used before MS analysis. However, because of their inherent specificity, no combination of lectins can bind to all glycoproteins or glycopeptides. In this work, we have mapped the yeast N-glycoproteome by combining MS with a boronic acid-based chemical enrichment method that takes advantage of the common feature of multiple hydroxyl groups in all glycans. The strong and reversible covalent bond interactions between boronic acid and diols provide a great opportunity to catch glycopeptides and glycoproteins by boronic acid under basic conditions, and release them in an acidic solution. More importantly, the universal sugar recognition by boronic acid provides great capability and potential for comprehensively mapping glycosylation sites in complex biological samples. In our experiment, over 800 N-glycosylation sites were identified in 332 yeast proteins by LC-MS, among which 675 sites were well-localized with over 99% confidence. The results demonstrated that the current method is cost-efficient, universal, and effective. This boronic acid-based chemical enrichment method can be extensively applied to many complex biological samples for the comprehensive analysis of protein glycosylation.

* This work was supported by a start-up fund from Georgia Institute of Technology.

□ This article contains supplemental Fig. S1 and Tables S1 to S4.

§ To whom correspondence should be addressed: School of Chemistry and Biochemistry and the Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332. Tel.: 404-385-1515; Fax: 404-894-7452; E-mail: Ronghu.Wu@chemistry.gatech.edu.

REFERENCES

1. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and E., E., M. (2008) Essentials of glycobiology (2nd ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

2. Erbacher, A., Gieseke, F., Handregtering, R., and Muller, I. (2009) Dendritic cells: Functional aspects of glycosylation and lectins. Hum. Immunol. 70, 308–312.

3. van Kooyk, Y., and Rabinovich, G. A. (2008) Protein-glycan interactions in the control of innate and adaptive immune responses. Nat. Immunol. 9,
