Video Article

Glycopeptide Capture for Cell Surface Proteomics

M. C. Gilbert Lee[1], Bingyun Sun[1]

[1]Department of Chemistry, Simon Fraser University

Correspondence to: Bingyun Sun at bingyun_sun@sfu.ca

URL: https://www.jove.com/video/51349
DOI: doi:10.3791/51349

Keywords: Molecular Biology, Issue 87, membrane protein, N-linked glycoprotein, post-translational modification, mass spectrometry, HPLC, hydrazide chemistry, N-glycoproteomics, glycopeptide capture

Date Published: 5/9/2014

Citation: Lee, M.C., Sun, B. Glycopeptide Capture for Cell Surface Proteomics. *J. Vis. Exp.* (87), e51349, doi:10.3791/51349 (2014).

Abstract

Cell surface proteins, including extracellular matrix proteins, participate in all major cellular processes and functions, such as growth, differentiation, and proliferation. A comprehensive characterization of these proteins provides rich information for biomarker discovery, cell-type identification, and drug-target selection, as well as helping to advance our understanding of cellular biology and physiology. Surface proteins, however, pose significant analytical challenges, because of their inherently low abundance, high hydrophobicity, and heavy post-translational modifications. Taking advantage of the prevalent glycosylation on surface proteins, we introduce here a high-throughput glycopeptide-capture approach that integrates the advantages of several existing N-glycoproteomics means. Our method can enrich the glycopeptides derived from surface proteins and remove their glycans for facile proteomics using LC-MS. The resolved N-glycoproteome comprises the information of protein identity and quantity as well as their sites of glycosylation. This method has been applied to a series of studies in areas including cancer, stem cells, and drug toxicity. The limitation of the method lies in the low abundance of surface membrane proteins, such that a relatively large quantity of samples is required for this analysis compared to studies centered on cytosolic proteins.

Video Link

The video component of this article can be found at https://www.jove.com/video/51349/

Introduction

Cell surface proteins interact with the extracellular environment and relay signals from the outside to the inside of a cell. Thus, these proteins, including extracellular matrix proteins, play critical roles in all aspects of cellular biology and physiology ranging from proliferation, growth, migration, differentiation to aging and so forth. Surface proteins function by interacting with other cells, proteins and small molecules[1-3]. Molecular characterization of cell-surface proteins is of great interest not only for biologists but also for pharmaceutical companies, as more than 60% of drugs are targeted to cell-surface proteins[4].

Tandem mass spectrometry (MS), with its superior sensitivity, accuracy, and throughput for identification of proteins and peptides, has been a powerful tool for global proteomic studies[5,6]. Yet, surface proteins pose significant challenges to MS-based proteomics, as most surface proteins exist in low quantities and with heavy modifications. The membrane-spanning regions of the surface proteins render them hydrophobic; this is especially the case for multipass transmembrane proteins. It is thus difficult to dissolve membrane proteins in aqueous solutions without the help of a detergent; however the use of detergents generally suppresses the performance of HPLC and MS[1,7,8] in protein identification. Therefore, membrane proteins have been poorly characterized in direct LC-MS based proteomics.

Glycosylation is one of the most important and abundant post-translational modifications taking place in cell-surface proteins[9]. The enormous complexity and heterogeneity of glycans hamper peptides' MS signal[10]. Nevertheless, several proteomic methods have used this unique modification to enrich surface proteins and to remove the sugar moieties from proteins prior to LC-MS analysis. These methods include lectin-based affinity capture[11] and hydrazide-based or boric acid-based chemical capture[12] as well as hydrophilic chromatography separations[8,13]. The removal of glycans transforms membrane proteins to regular proteins and drastically simplifies the MS characterization. Because glycosylation also takes place in secreted proteins that have high solubility in contrast to membrane proteins, many glycoproteomic methods are optimized for soluble proteins, and tend to have lower glycopeptide selectivity and sensitivity when being deployed to membrane proteins[8,14]. Other methods also exist to enrich, in particular, cell-surface proteins, such as those using ultracentrifugation[15] and labeling strategies[16]. A detailed comparison between our method and other existing methods for characterizing membrane proteins was conducted recently[17], and the results indicated that our method can perform equally well, if not better, than all the compared membrane proteomics methods, but with higher simplicity.

To help researchers use this method, we detail here a general protocol. This method integrates several advantages of existing glycoproteomics strategies and is devised specifically for membrane glycoproteins, yet the method works equally well for secreted proteins. The characteristics of this method include: 1) a complete solubilization of membrane proteins, 2) an enrichment of glycopeptides instead of glycoproteins to eliminate the potential steric hindrance when using a solid capturing substrate, 3) the use of hydrazide chemistry to form covalent bonds between glycopeptides and the capturing substrate, such that the bonded glycopeptides can tolerate stringent washes for high glycoselectivity, and 4) the

capability to conduct the entire capture procedure in one tube for reduced sample loss and shortened procedure duration. After implementing this method to studying a variety of biological samples including cells and tissues, we observed a high selectivity (> 90%) to glycoproteins[8,17,18].

Protocol

1. Harvest Membranes

1. Add hypotonic buffer (10 mM Tris-HCl, 10 mM KCl, 1.5 mM $MgCl_2$, pH 8.0) with protease inhibitor cocktail onto a cell pellet (~10^8 cells) and incubate for 15-30 min on ice. Lyse the cells by passing the sample through syringe needles (5-10 passes) or homogenizing the sample by a Dounce homogenizer (15-30 strokes). Use a hemocytometer and trypan blue staining to check the efficiency of lysis.
2. Obtain the microsomal fraction by a differential centrifugation of the lysate at 3,000 g x 15 min and then at 100,000 g x 2 hr. The first centrifugation obtains a pellet that is the nuclear fraction, and the subsequent centrifugation of the supernatant generates a second pellet, which is the microsomal fraction that contains the plasma membrane as well as membrane-bound organelles[8]. The final supernatant is the cytosolic fraction. Store the microsomal fractions in -80 °C freezer for further analysis as indicated below.

2. Dissolve, Denature, and Digest Membrane Proteins

1. Dissolve the microsomal fraction in the denaturation buffer (40 mM Tris, 10 mM EDTA, 10 mM TCEP, 0.5% Rapigest, pH 8) and then incubate the solution at 100 °C for 10 min.
2. Cool the heated solution to room temperature, add ultra-high purity urea powder into the solution to 8 M final concentration and incubate the sample at 37 °C for 30 min.
3. Add iodoacetamide stock solution to the sample to 15 mM final concentration, and incubate the solution in the dark for 30 min at room temperature to alkylate the free thiols in the sample. Add DTT stock solution to the sample to 10 mM final concentration and incubate the solution for another 10 min at room temperature to quench the excessive iodoacetamide.
4. Dilute the obtained solution 10x with 40 mM Tris buffer at pH 8, and subsequently add trypsin to the sample at a 1:20 ratio of trypsin to total protein. Maintain the digestion reaction in a 37 °C oven overnight to ensure the reaction is completed.
5. Terminate the digestion by acidifying the sample solution to pH 1 with HCl, a condition that also breaks down the detergent, *i.e.* Rapigest. Degrade the Rapigest at 37 °C for 1 hr, and remove the developed precipitation by centrifugation.
6. Clean the supernatant that contains sample peptides by a C-18 solid phase extraction (SPE) cartridge and dry the obtained sample by speedvac.
7. Perform a SDS-PAGE analysis of samples before and after trypsin digestion to confirm the digestion efficiency.

3. Glycopeptide Capture

1. Dissolve the cleaned peptides in the coupling buffer (100 mM sodium acetate, pH 5.5). Add sodium periodate into the peptide solution at a 10 mM final concentration for 30 min in the dark, at room temperature. This will oxidize the cis-diol groups in the glycans to aldehydes, which allow the glycans to couple to the resin through hydrazide chemistry. Quench the excessive periodate by sodium sulphite at a 20 mM final concentration and pH 5 for 10 min at room temperature.
2. Introduce hydrazide-derivatized resin into the peptide solution at a ratio of 1:4 (resin to solution) to couple glycopeptides to the resin. Incubate the reaction at 37 °C for 1-2 days with end-to-end rotation for complete coupling.
3. Remove the unbound peptides by washing the resin twice with 1 ml of each of the following: DI water, 1.5 M NaCl, methanol and 80% acetonitrile. Finally, wash the resin 3x with 1 ml of 100 mM NH_4HCO_3 at pH 8 to exchange the buffer of the system to 100 mM NH_4HCO_3.
    1. Collect the supernatant and the washes for the analysis of unbound peptides (optional).
4. Release the N-glycopeptides from the resin by PNGase F in an overnight incubation at 37 °C with an end-to-end rotation. Collect the released peptides by centrifugation and an 80% acetonitrile wash. Dry the obtained solution in the speedvac for LC-MS analysis.

4. Further Fractionation (Optional)

To further simplify sample complexity, fractionate the obtained N-glycopeptides. For example, redissolve the dried peptides into 10 mM ammonia formate, pH 3 with 20% acetonitrile and use strong cation exchange (SCX) chromatography to fractionate the peptides. Dry the obtained eluent, and then analyze the obtained peptide fractions by reverse-phase LC-MS[8,17,18].

5. Cleaning of the Released N-glycopeptides (Optional)

If concerns rise for the potential contamination of the peptides, redissolve the dried peptides into 0.1% formic acid and use a MCX SPE column to further clean the peptides prior to reverse-phase LC-MS analysis.

Note: Database searching parameters.

During the selective cleavage of N-glycopeptides off the resin, PNGase F converts the N-glycan linked asparagine to an aspartic acid. Therefore, there is a 0.9840 Da mass shift of the liberated N-glycopeptides. To accurately identify these peptides, this modification needs to be added to the search parameters along with common modifications such as the carbamidomethylation of the cysteine and oxidation of the methionine.

Representative Results

A representative flow chart of the experimental procedure is summarized in **Figure 1**. The labeling and further fractionation steps are optional and details are described in a recent publication[18]. Another option is to analyze the unmodified peptides, which do not react with the resin. The advantages of analyzing the unmodified peptides include the potential identification of non-glycosylated peptides and proteins, such as claudins in tight junctions; an additional advantage is more accurate quantitation. Based on these advantages, we termed this method glycocapture-assisted-global quantitative proteomics (gagQP), and detailed the analysis in a recent publications[18].

A typical glycopeptide spectrum taken after the enrichment method is shown in **Figure 2**. In the obtained glycopeptide, the N-glycan was removed and the glycan-attached asparagine (N) was converted to an aspartic acid (D) by PNGase F; therefore, the spectrum can be readily searched using any proteomics search engine against common protein sequence databases.

The capture method can be evaluated by commercially available model glycoproteins prior to its application with complex and valuable biological samples. Some frequently used model glycoproteins include avidin (chicken), ovalbumin (chicken), invertase (yeast), α-1 antitrypsin (human), conalbumin (chicken), and ribonuclease B (bovine) (all can be obtained from Sigma). The frequently identified glycopeptides from these proteins can be found in a previous publication[8]. A customized protein-sequence FASTA database that is suitable for automatic searching of LC-MS results generated from these model proteins can be downloaded from the following link (http://www.sfu.ca/chemistry/groups/bingyun_sun/tools.html), which includes the above listed model proteins as well as a reversed yeast-sequence database, common contaminants and PNGase F.

A typical LC-MS result of the captured N-glycopeptides is shown in **Figure 3**, in which more than 100 glycoproteins can be identified from a single LC-MS run of a cell microsomal fraction. The enrichment selectivity to both glycoproteins and glycopeptides is generally more than 90%. A successful analysis can have an enrichment selectivity of 95%. Using an SCX column and step gradient to further fractionate the samples prior to LC-MS analyses will usually double the number of glycoprotein identifications[17]. Sometimes, when the quantity of the obtained glycopeptides is low, the impurity accumulated from the vials can be observed in the final sample. These contaminants can be removed by the method provided in step 5.

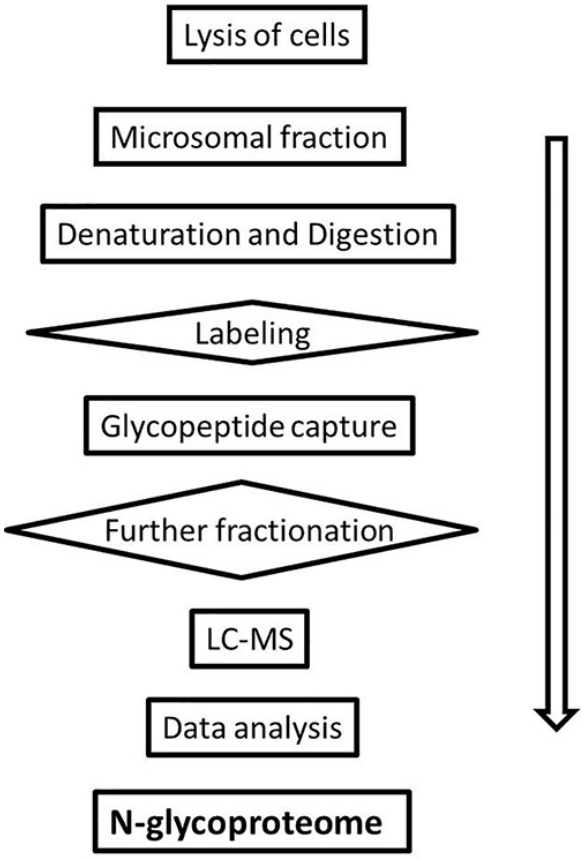

Figure 1. Flow chart of the experimental procedure. Rectangles are required steps and diamonds are optional steps.

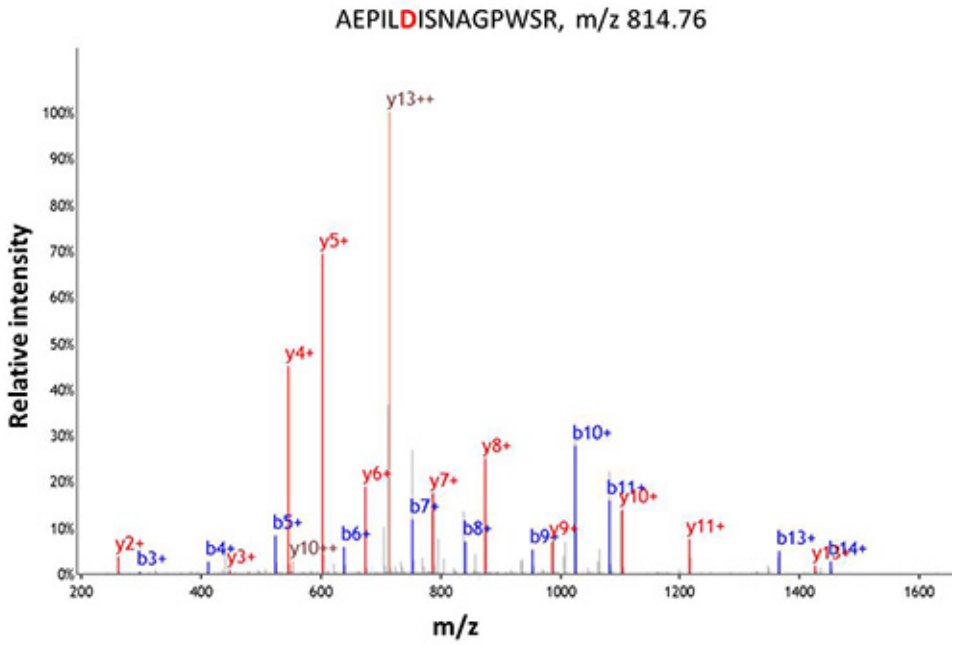

Figure 2. Collision induced dissociation (CID) spectrum of AEPILNISNAGPWSR from Baker's yeast after the glycopeptide-capture method. The underscored asparagine (N) is converted to the aspartic acid (D) as highlighted in red in the peptide sequence in the figure.

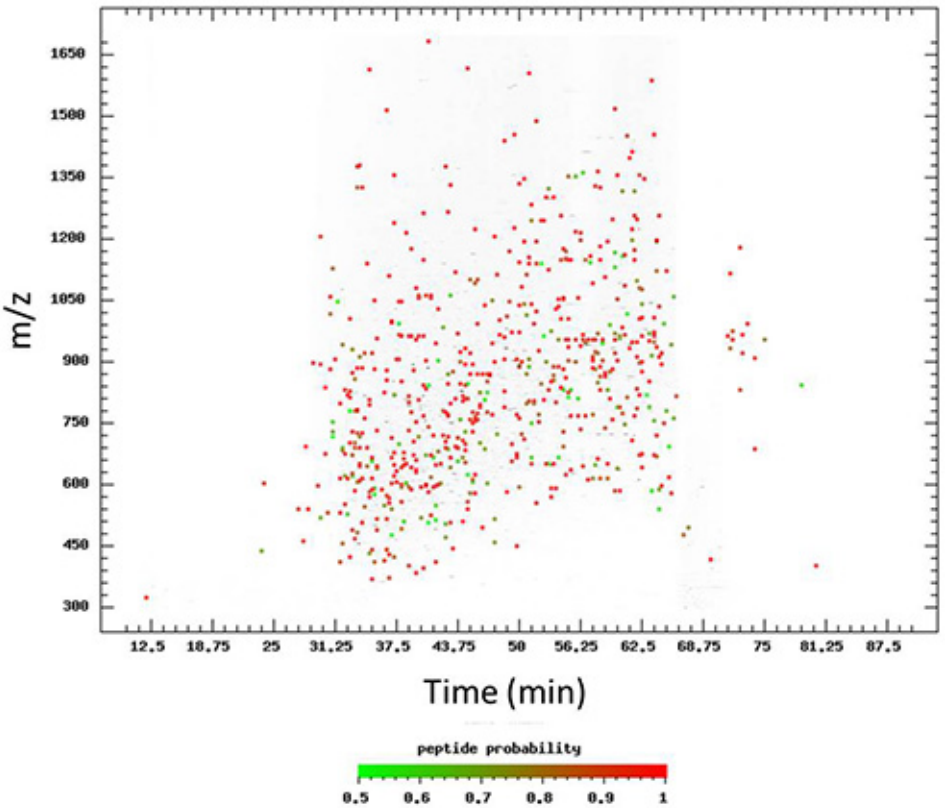

Figure 3. 2D plot of the LC-MS result of the N-glycopeptides obtained from mouse embryonic stem cells (E14.Tg2a). The dots are the detected peptides, and the color of the dot represents the identification confidence obtained statistically (*i.e.* peptide probability).

Discussion

Here we introduce a glycopeptide-capture strategy for profiling cell-surface proteins. The method can be applied to study secreted proteins, such as those in blood, as well as in other body fluids or in cell culture media.

The success of the method relies on the complete digestion of samples; therefore, a SDS-PAGE characterization of the digestion efficiency is necessary, especially for the first-time analysis of a sample. A complete digestion can be challenging for membrane proteins, and can only be possible after thorough solubilization of the membrane fraction. The solubilization process begins when the detergent is introduced and ends after the incubation of urea. Therefore, if cloudiness presents in the sample before the addition of urea but disappears after urea incubation, the solubilization is sufficient. If the membrane fraction is difficult to dissolve, increase the amount of detergent used. For membrane-rich tissue samples such as brain and adipose tissues, 5% Rapigest can be utilized. Sometimes, precipitation can appear during the dilution of samples prior to the trypsin digestion, which is more frequently observed for human serum samples. The cause of this precipitation is mainly the decreased concentrations of urea and detergent. This precipitation will generally be removed by trypsin during the digestion and is not a concern. However, when the precipitation forms, it is important to rotate the sample vial during the digestion step to ensure a good mixing. The pH of the glycocapture step is important because the primary amines in peptides, such as those from N-termini and lysines, will react with the newly formed aldehydes after the periodate oxidation. Thus, it is important to protonate these amines by adjusting the pH of the capturing solution to below 6.0 using acetate buffer, to prevent them from interfering with the capture.

Using a ratio of resin to capture solution around 1:3 to 1:4 ensures sufficient mixing during the capture step. A minimum of 50 µl of resin is necessary based on our experience; lower quantities will render the subsequent series of washes difficult to perform and will introduce severe sample loss due to the loss of resin. We have obtained good results using 100-300 µl of resin for 0.5-2 mg of total protein. However, the ratio between the amount of protein and resin can be sample dependent, we recommend you optimize this condition for your specific samples and applications.

The hydrazide chemistry captures both O- and N-linked glycopeptides on the substrate; due to the lack of an effective enzyme to release all the O-linked glycopeptides, we only used PNGase F for N-glycopeptide studies[8,12]. The possibility of studying the O-type of glycopeptides may require the discovery or bioengineering of appropriate hydrolases.

This method can be paused at several places including: 1) after obtaining the membrane fraction, 2) after trypsin digestion and cleaning of the peptides, and 3) after release of the N-glycopeptides. Additional procedures can be introduced during these intervals. For example, the digested peptides can be differentially labeled by N-isotags as indicated in **Figure 1**, for quantitative analysis. Using a method called gagQP, in which the unmodified peptides are analyzed in parallel with the glycopeptides, the accuracy of quantitation can be significantly improved as we demonstrated in a recent publication[18].

Glycopeptide capture itself can effectively decrease sample complexity, and it is generally not necessary to further fractionate the enriched N-glycopeptides. Exceptions apply to situations where the samples are abundant but with substantial concentration dynamics, and the proteins of interest are in low abundance, such as in the discovery of blood protein biomarkers or for a complete survey of cell surface proteins. Under those circumstances further fractionation can be implemented for the captured N-glycopeptides to provide a better penetration of the glycoproteome. As the bottom of the glycoproteome is being approached through fractionation, the identification of non-glycopeptides introduced by nonspecific binding will increase. Thus a decrease of glycopeptide and glycoprotein selectivity (to ~85%) will typically be observed after a further fractionation of N-glycopeptides[17]. Therefore, researchers need to weigh the pros and cons when designing the most suitable procedure.

For researchers who have never used an N-glycopeptide-capture method, it is best to practice the method with a few pure N-glycoprotein(s) having known glycosylation sites as listed previously[8]. This method is robust and the selectivity to N-glycopeptides is high. A drawback of this method lies in the inherent low abundance of surface N-glycoproteins; for a comprehensive characterization of the sample, higher quantity is generally required than that of cytosolic proteomics. However, if cultured cells can be expanded *in vitro* or the tissues of interest are abundant, this drawback is negligible.

Disclosures

The authors declare that they have no competing financial interests.

Acknowledgements

This research has been supported by the startup fund of Simon Fraser University.

References

1. Cho, W. & Stahelin, R. V. Membrane-protein interactions in cell signaling and membrane trafficking. *Annual review of biophysics and biomolecular structure* **34**, 119-151, doi:10.1146/annurev.biophys.33.110502.133337 (2005).
2. Tadevosyan, A., Vaniotis, G., Allen, B. G., Hebert, T. E. & Nattel, S. G protein-coupled receptor signalling in the cardiac nuclear membrane: evidence and possible roles in physiological and pathophysiological function. *The Journal of physiology* **590**, 1313-1330, doi:10.1113/jphysiol.2011.222794 (2012).
3. White, S. H. Biophysical dissection of membrane proteins. *Nature* **459**, 344-346, doi:10.1038/nature08142 (2009).
4. Heijne, G. Membrane-protein topology. *Nature reviews. Molecular cell biology* **7**, 909-918, doi:10.1038/nrm2063 (2006).

5. Cox, J. & Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry* **80**, 273-299, doi:10.1146/annurev-biochem-061308-093216 (2011).
6. Lamond, A. I. *et al.* Advancing cell biology through proteomics in space and time (PROSPECTS). *Mol Cell Proteomics* **11**, O112 017731, doi:10.1074/mcp.O112.017731 (2012).
7. Savas, J. N., Stein, B. D., Wu, C. C. & Yates, J. R., 3rd. Mass spectrometry accelerates membrane protein analysis. *Trends in biochemical sciences* **36**, 388-396, doi:10.1016/j.tibs.2011.04.005 (2011).
8. Sun, B. *et al.* Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol Cell Proteomics* **6**, 141-149 (2007).
9. Lowe, J. B. Glycosylation, immunity, and autoimmunity. *Cell* **104**, 809-812 (2001).
10. Dodds, E. D. Gas-phase dissociation of glycosylated peptide ions. *Mass spectrometry reviews* **31**, 666-682, doi:10.1002/mas.21344 (2012).
11. Kaji, H. *et al.* Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* **21**, 667-672 (2003).
12. Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**, 660-666 (2003).
13. Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N. & Roepstorff, P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* **3**, 556-566 (2004).
14. Bond, M. R. & Kohler, J. J. Chemical methods for glycoprotein discovery. *Current opinion in chemical biology* **11**, 52-58, doi:10.1016/j.cbpa.2006.11.032 (2007).
15. Pasini, E. M. *et al.* In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood* **108**, 791-801, doi:10.1182/blood-2005-11-007799 (2006).
16. Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**, 378-386 (2009).
17. Sun, B. *et al.* N-Glycoproteome of E14.Tg2a mouse embryonic stem cells. *PLoS ONE* **8**, e55722, doi:10.1371/journal.pone.0055722 (2013).
18. Sun, B. *et al.* Glycocapture-assisted global quantitative proteomics (gagQP) reveals multiorgan responses in serum toxicoproteome. *J Proteome Res* **12**, 2034-2044, doi:10.1021/pr301178a (2013).

