

Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints

Dorota F. Zielinska,^{1,3} Florian Gnad,^{1,2,3} Jacek R. Wiśniewski,^{1,*} and Matthias Mann^{1,*}

¹Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, Martinsried D-82152, Germany

²Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA

³These authors contributed equally to this work

*Correspondence: jwisniew@biochem.mpg.de (J.R.W.), mmann@biochem.mpg.de (M.M.)

DOI 10.1016/j.cell.2010.04.012

SUMMARY

N-linked glycosylation is a biologically important protein modification, but only a small fraction of modification sites have been mapped. We developed a “filter aided sample preparation” (FASP)-based method in which glycopeptides are enriched by binding to lectins on the top of a filter and mapped 6367 N-glycosylation sites on 2352 proteins in four mouse tissues and blood plasma using high-accuracy mass spectrometry. We found 74% of known mouse N-glycosites and discovered an additional 5753 sites on a diverse range of proteins. Sites almost always have the N-!P-[S|T]-!P (where !P is not proline) and rarely the N-X-C motif or nonconsensus sequences. Combining the FASP approach with analysis of subcellular glycosite localization reveals that the sites always orient toward the extracellular space or toward the lumen of ER, Golgi, lysosome, or peroxisome. The N-glycoproteome contains a plethora of modification sites on factors important in development, organ-specific functions, and disease.

INTRODUCTION

N-glycosylation is one of the most prominent posttranslational protein modifications and plays a major role in the assembly of complex multicellular organs and organisms (Varki et al., 2009). This modification is involved in many cellular functions including cell-cell and receptor-ligand interactions, immune response, apoptosis, and pathogenesis of many diseases (Varki et al., 2009; Woods et al., 1994). N-glycosyltransferases are predominantly located in the lumen of the ER and Golgi apparatus and attach this modification cotranslationally in a complex series of processing steps to a subset of the sites with the consensus sequence N-!P-[S|T] (where !P signifies any amino acid except proline). This motif has been extended to N-!P-[S|T]-!P in

C. elegans (Kaji et al., 2007). It is also possible that there are consensus motifs different from the canonical one.

Because of the topological location of the transferases, the modification is thought to be localized on secreted molecules, the extracellular part of plasma membrane proteins, and the luminal part of proteins in compartments of subcellular organelles such as the endoplasmic reticulum and the Golgi apparatus, endosomes, and lysosomes. A number of authors have raised the possibility that N-linked glycosylation may also be present in mitochondria (Chandra et al., 1998; Kung et al., 2009), in the nucleus (Reeves et al., 1981), and in the cytoplasm (Pedemonte et al., 1990). However, these studies do not map residue-specific N-linked glycosylation sites.

Despite great biological and clinical interest, our knowledge of in vivo N-glycosylation sites—a prerequisite for detailed functional understanding—is still very limited. Liquid chromatography coupled to high-resolution mass spectrometry (LC-MS) has emerged as the key technology for large-scale analysis of posttranslational modifications in general and N-glycosylated proteins in particular (Aebersold and Mann, 2003; Jensen, 2006; Medzihradzky, 2005; Witze et al., 2007). The large complexity of attached sugar molecules (North et al., 2010) and the low expression levels of many N-glycoproteins make the characterization of complete N-glycosylation structures very challenging. To detect low abundant N-glycosylated proteins or peptides in complex mixtures among the large excess of their nonglycosylated counterparts, specific enrichment methods have to be applied, most commonly based on lectin affinity (Bunkenborg et al., 2004) or chemical linkage of the sugar moiety to surfaces (Zhang et al., 2003). For determination of glycosylation sites—as opposed to the structure of the sugar—a universal deglycosylating enzyme (i.e., PNGase F) is used. This leads to deamidation of the asparagine residue to aspartic acid and a mass increase of 0.9848 Da of the modification site, which can be detected by tandem mass spectrometry (MS/MS) as a mass shift of the precursor peptide and of its fragments. If deglycosylation is performed in ¹⁸O-water, the mass shift is 2.9890 Da, adding confidence to the site assignment (Kuster and Mann, 1999).

Among large-scale N-glycoproteomic studies, the largest reported 1495 N-glycosylation sites from *C. elegans* (Kaji et al.,

2007). Others measured up to a few hundred N-glycosylation sites on cell surface proteins of the immune system (Wollscheid et al., 2009), of mouse C2C12 myoblasts (Gundry et al., 2009), in human blood plasma (Liu et al., 2005), in human serum (Bunkenborg et al., 2004), in human saliva (Ramachandran et al., 2006), and in rat liver (Lee et al., 2009).

The data of large-scale proteomics studies and some directed studies are combined in the Swiss-Prot database (Wu et al., 2006), which results in 830 mouse and 1998 human N-glycosylation sites. This is likely a drastic underestimate of the true extent of the mammalian N-glycoproteome. Notably, even though the Swiss-Prot database presents the most comprehensive resource of annotated N-glycosylation sites, it is not complete because of the difficulty in retrieving single sites from numerous literature studies.

Given its biomedical importance, we set out to map this modification in-depth and at very high accuracy using recent advances in proteomics technology. We have developed an N-glycopeptide enrichment method based on “filter aided sample preparation” (FASP) (Wisniewski et al., 2009b), which allows highly efficient capture of glycopeptides even from membrane proteins. We employ the ability of the recently introduced LTQ-Orbitrap Velos instrument to measure peptide fragments, and not only peptide precursor masses, with low ppm mass accuracy and at high sensitivity (Olsen et al., 2009). Our analysis of four different mouse tissues and blood plasma achieves very high confidence and covers a substantial part of the mouse N-glycoproteome—allowing in-depth characterization of this protein modification.

RESULTS

Development of a FASP-Based N-linked Glycopeptide Capture Method (N-Glyco-FASP)

Most N-linked glycosylations occur on membrane proteins, which have traditionally been difficult to analyze by proteomic methods. We have recently shown that the FASP method is especially well suited to analyze this class of proteins because it achieves complete protein solubilization in SDS while still allowing gel-free analysis (Wisniewski et al., 2009b). We reasoned that FASP could be combined with peptide affinity capture simply by adding the affinity reagent—in this case lectin—to the top of the filter after on-filter protein digestion. Glycosylated peptides are bound by lectin and thereby retained whereas nonglycosylated peptides can be washed through the filter. Next, glycopeptides are efficiently deglycosylated by PNGase F and released peptides are eluted, resulting in a peptide population of high purity (Figure 1A). We used two different endoproteinases, trypsin and Glu-C, to increase the number and localization confidence of glycosylation sites. In our experiments with the “N-glyco-FASP” method, sample amounts were typically 200 μg of total protein in 40 μl , but this can be scaled up or down as desired.

To capture all three classes of N-glycosylated peptides, multi-lectin enrichment can be employed (Yang and Hancock, 2004). In N-glyco-FASP, lectins do not need to be coupled to a solid support because they are retained by the filter, and therefore any lectin or mixture of lectins can be employed. We selected concanavalin A (ConA), which binds to mannose, wheat germ agglutinin (WGA), which binds to sialic acid, as well as N-acetyl-

glucosamine and agglutinin RCA₁₂₀, which captures galactose modified at the 3-O position (e.g., with sialic acid or another galactose) as well as terminal galactose. Enrichment with this mixture of lectins was as efficient as separate experiments based on enrichment with all single lectins (Figure 1B). Overall, 63% of all N-glycosylation sites identified in a given tissue could be detected in a single LC-MS/MS experiment by multi-lectin enrichment. In comparison, 69% of a given tissue N-glycoproteome was covered when combining three LC-MS/MS experiments based on single lectin enrichment. WGA proved to have the highest binding efficiency among the applied lectins. The proportion of glycosylated peptides—measured as deamidated peptides after PNGase F digestion—to all identified peptides in single run analysis was 46%. In our experiments, this is comparable to or higher than the enrichment of phosphorylated peptides (Macek et al., 2009) and substantially higher than the enrichment of lysine acetylated peptides (Choudhary et al., 2009). Without lectin enrichment, glycopeptides were 0.5% of total peptides, indicating an enrichment factor of about 100-fold (Figure 1C). We also interrogated our datasets for other modifications but did not find large numbers of such peptides.

Precision Mapping of N-Glycosylation Sites

To identify deglycosylated peptides we used on-line liquid chromatography electrospray mass spectrometry (LC-MS/MS) on the recently introduced linear ion trap orbitrap instrument (LTQ-Orbitrap Velos). The LTQ-Orbitrap is capable of fragmenting peptides by “higher-energy dissociation” (HCD), in which the fragment mass spectrum is analyzed in the high-resolution part of the instrument without loss of low-mass ions (Olsen et al., 2007). The superior sequencing capabilities of HCD compared to ion trap fragmentation (CID) normally come at the cost of reduced sensitivity. However, the Velos instrument features 20-fold improved HCD performance (Olsen et al., 2009). We therefore tested if we could measure the N-glycosylation sites with HCD without loss of sensitivity. Comparison of orbitrap HCD and ion trap CID showed that HCD identified approximately the same number of glycosylated peptides and that it did not discriminate against low-abundance peptides (Figure S1 available online). We therefore performed all subsequent MS analyses in high-resolution precursor and high-accuracy fragment mode (“high-high” strategy).

We applied N-glyco-FASP combined with high-high MS measurement to four mouse organs (brain, liver, kidney, and heart) and blood plasma, which we group together with the other tissues for simplicity. Each tissue was independently prepared in triplicates and measured 11 times by single LC-MS/MS runs with 4 hr gradients after deglycosylation in ¹⁸O-water by both single and multi-lectin enrichment. Furthermore we measured N-glycosylation sites in four subcellular fractions of liver cells. Together, 59 LC-MS/MS runs were acquired (Table S1A). Additionally, we performed 64 experiments without ¹⁸O-water (Table S1B). Analysis of the data was performed with the MaxQuant software (Cox and Mann, 2008) specifying a false discovery rate of 1% at the peptide and site level. Average absolute mass deviation was 0.57 ppm for identified peptides and 3 ppm for all fragment ions contributing to peptide identification (Figure 2A). The median Mascot identification score for

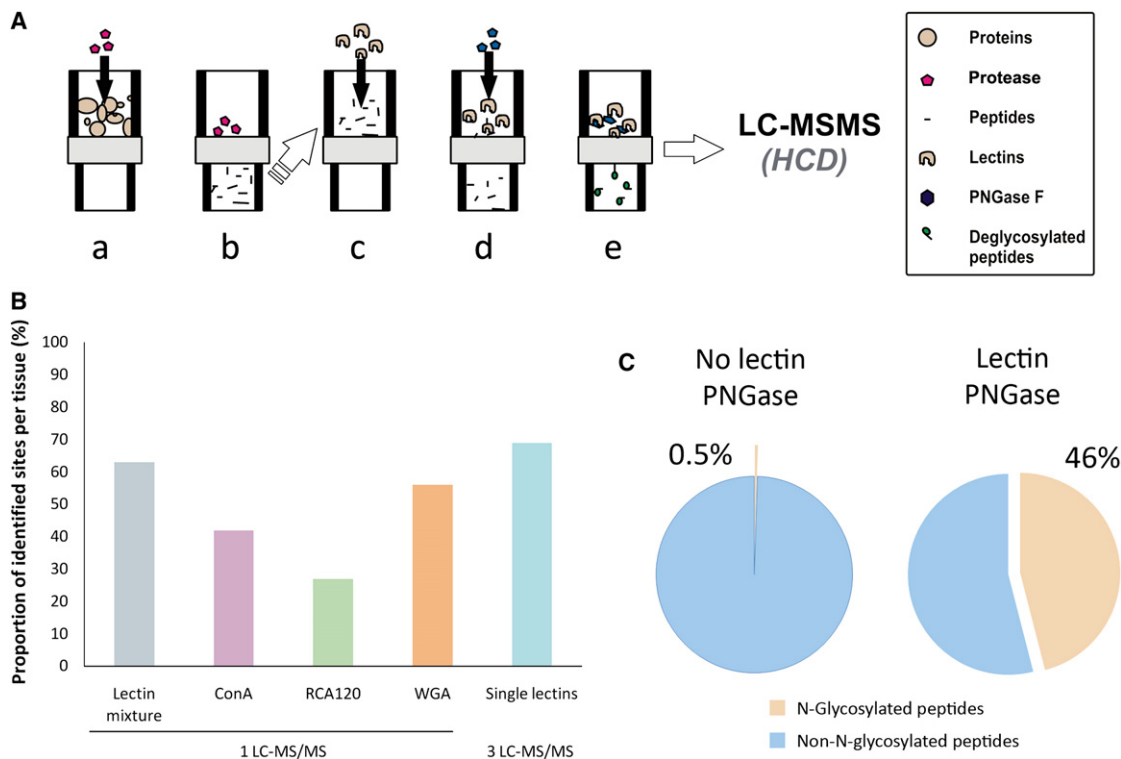


Figure 1. Sample Preparation and Enrichment of N-Glycosylated Peptides: N-Glyco-FASP

(A) A standard filtration device with 30k molecular cut-off is used as a proteomic reactor. Whole SDS-lysates of mouse tissues are processed according to the FASP protocol (a and b). To enrich for N-glycosylated peptides the digests are incubated with free lectins on the filter unit (c). Unbound peptides are removed by centrifugation. The bound glycopeptides are deglycosylated with PNGase F (d), eluted from the filter (e), and identified using high-accuracy LC-MS/MS. Peptide ions are fragmented via higher-energy dissociation (HCD).

(B) The proportion of N-glycosylated sites identified in a single LC-MS/MS experiment was lower in the case of single-lectin enrichment (ConA, RCA120, WGA) compared to multi-lectin enrichment. Using a mixture of lectins in one LC-MS/MS run was nearly as efficient as combining single lectin enrichment-based experiments in three LC-MS/MS runs.

(C) The proportion of detected N-glycosylated to unmodified peptides in an experiment without enrichment was 0.5%. With N-glycosylation enrichment the proportion increased by a factor of about 100 (to a median of 46%).

See also Table S1 and Table S2.

glycopeptides was 65 (Perkins et al., 1999) and the median posttranslational modification (PTM) score was 112 (Olsen et al., 2006) (Figures 2B and 2C). Because neither Mascot nor PTM scores incorporate the high fragment mass accuracy into database identification scores, confidence of glycopeptides identification is much higher than even those scores indicate. The average localization probability of all identified sites was 94.92%. This indicates that both peptide identification and localization of the modification in the peptide sequence with single amino acid resolution were unambiguous. Out of all identified N-glycosylated sites, we derived those that had a minimum residue localization probability of 95% and that were identified in two or more experiments with a minimum of one confirmation in the ^{18}O -water experiment to build a top confidence (class I) set (Table S2A). N-glycosites that do not satisfy these extremely strict criteria but show a residue localization probability higher than 90% are also high-confidence sites (class II, Table S2B), but they were not included in the global analyses described in this paper. Average localization probability of class I and II sites was 99.97%.

We performed PNGase digestion both with and without ^{18}O -water (see Extended Experimental Procedures). As mentioned above, ^{18}O -water deglycosylation events lead to a mass increment of 2.9890 Da that can be readily distinguished from spontaneous deamidation (mass increment of 0.9858 Da). Therefore the experimental confirmation with ^{18}O -water was considered as criteria for the definition of our top confidence set as described above. An example spectrum for N-glycopeptide identification is shown in Figure 2D.

The In Vivo Mouse N-Glycoproteome

Our large-scale analysis resulted in the high-confidence identification of 6367 N-glycosylated sites with a localization probability higher than 90% on 2352 proteins (class I and II; Table S2; all sites are available in the PHOSIDA database along with their corresponding spectra; Gnad et al., 2007). This dataset covers 74% of the 830 experimentally derived mouse N-glycosites recorded in Swiss-Prot (Version 57.12). Overall, 5753 sites of our set were not previously recorded as experimentally identified. Swiss-Prot also contains a large number of potential N-glycosites

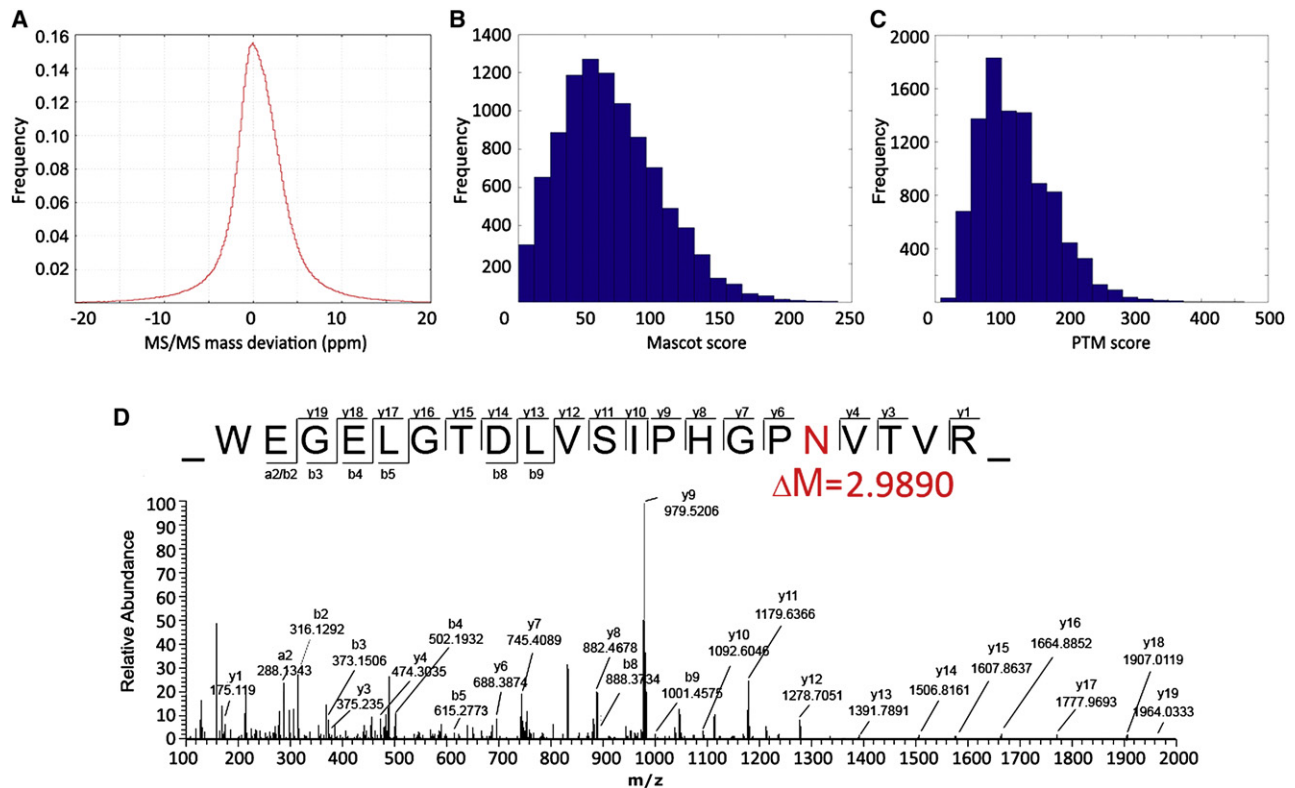


Figure 2. Identification of the Glycosylation Sites by High-Accuracy Tandem Mass Spectrometry using HCD Fragmentation

Distribution of MS/MS mass deviations (A), Mascot scores (B), and PTM scores (C) of the sequenced deglycosylated peptides.

(D) Representative MS/MS spectrum of the peptide WEGELGTDLVSIPIHGPNVTVR of β -secretase 1, identified in ^{18}O -water. The mass increment of 2.9890 Da is indicated.

See also Figure S1.

(11,846), partly by similarity to the 1998 known human sites but mainly derived from bioinformatic prediction. Of these sites, we cover 33%, which is excellent agreement given the fact that algorithms are generally adjusted toward overprediction.

Several lines of experimental evidence suggest that our dataset achieves very deep coverage of the mouse N-glycoproteome: Technical and biological repeats of N-glyco-FASP enriched tissue samples resulted in largely overlapping N-glycosites (on average 80% between any two single runs) and added only a small number of additional sites (Figure 3A). Additional fractionation either on the protein level by size-exclusion chromatography (Zielinska et al., 2009) or on the peptide level by anion exchange chromatography (Wisniewski et al., 2009a) resulted in only a few percent of additional glycosites. In contrast, performing these additional fractionation steps for the tissue proteome more than doubled the number of detected proteins compared to a single run (Figure 3B). Both observations are consistent with a glycoproteome that is thoroughly sampled by our analysis.

Sequence Recognition Motifs, Structure Preference, and Occupancy

The canonical N-linked glycosylation motif is N-IP-[S]T. We reasoned that our high precision and large-scale dataset might

provide a good basis to test the generality of this motif and to discover further consensus sequences. We compared the position-specific amino acid frequencies of the surrounding sequences (six amino acids to both termini) of glycosylated asparagines that have serine or threonine on the second position to the C terminus with those of all asparagines that occur in the mouse proteome. We found that proline is drastically underrepresented not only in the first position relative to the modification site (0% compared to 6.16% expected) but also in the third position (0.54% compared to 5.16% expected; $p = 0$; Table S3A). Notably, cysteine was highly enriched in the surrounding sequences of N-glycosites that match with the canonical N-IP-[S]T motif. Threonine occurs more frequently (1.4-fold) than serine at the second position—the reverse proportion as in the surrounding sequences of nonglycosylated asparagines that match with the consensus motif.

Next we asked if there were any motifs different from the canonical one. We applied the de novo method Motif-X (Schwartz and Gygi, 2005) to the surrounding sequences of all top confidence N-glycosylated asparagines (Extended Experimental Procedures). This resulted in the identification of three further significantly overrepresented consensus sequences (Figure 4A). Of 5052 sites, 177 did not match the N-IP-[S]T motif. These sites turned out to be heavily enriched

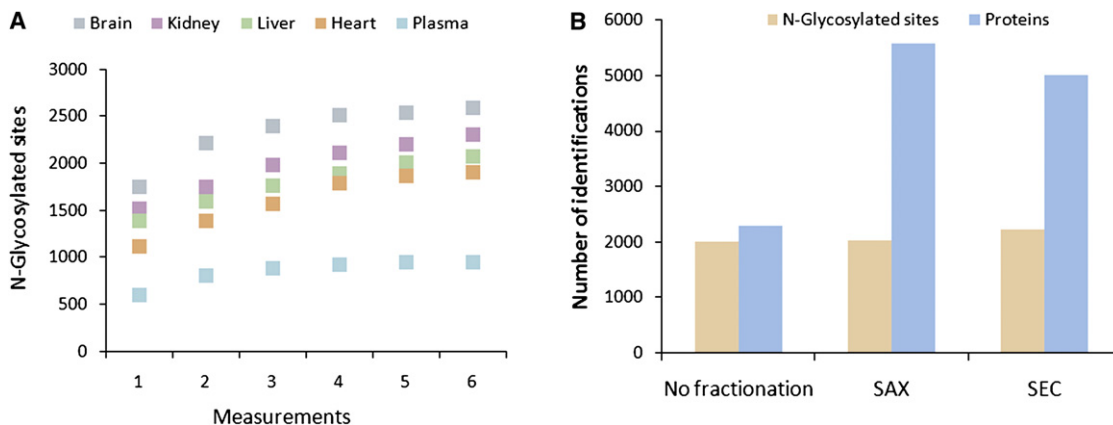


Figure 3. Depth of the Detected N-Glycoproteome

(A) Repeated measurements of each tissue yielded a minimal increase of identified N-glycosylation sites.

(B) Fractionation methods such as size-exclusion chromatography (SEC) and anion-exchange chromatography (SAX) did not result in a greatly increased number of identified N-glycosites, whereas the number of detected proteins in a proteome measurement increased more than 2-fold.

for cysteine or valine in place of S/T. In addition, glycine was enriched on the first position toward the C terminus. This result is interesting, as no other motifs except for N-IP-[S/T] and anecdotal evidence for the N-X-C motif are known (Zajonc et al., 2008). Overall, 112 N-glycosylated sites do not match

with any of the consensus sequences. To further verify the existence of N-glycosylation on sites that do not match with the known motifs, we performed western blotting on Apolipoprotein A1 (ApoA1) and Apolipoprotein E (ApoE) (Extended Experimental Procedures). In our large-scale dataset we found both

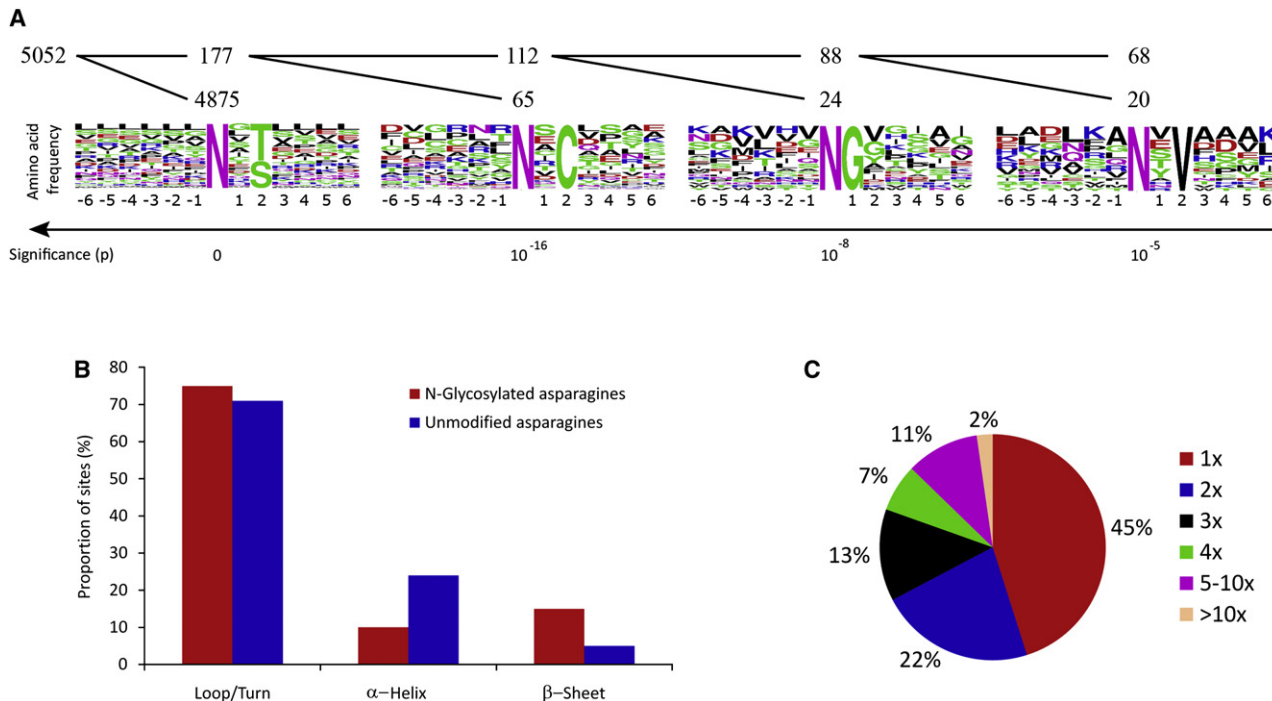


Figure 4. Sequence Recognition Motifs, Structure Preference, and Multiple Glycosylation

(A) N-glycosylation consensus sequence as derived using MotifX. WebLogo (Schneider and Stephens, 1990) was used to create relative frequency plots. The most significant sequence motif is the canonical one, with serine and threonine on position 2. In following iterative steps the consensus sequences N-X-C, N-G, and N-X-V were statistically identified.

(B) Proportion of N-glycosylated and non-N-glycosylated asparagines localized in loops, α helices, and β sheets.

(C) Distribution of singly and multiply glycosylated proteins.

See also Figure S2 and Table S3.

proteins to be N-glycosylated (ApoA1 on position 66, ApoE on position 130), even though they do not contain any asparagines that match with N-IP-[S/T] or N-X-C. In western blots, both proteins shifted their apparent molecular weights after PNGase treatment (Figure S2).

We predicted the secondary structure localizations and solvent accessibilities of N-glycosylated and non-N-glycosylated asparagines of N-linked mouse glycoproteins with SABLE 2.0 (Wagner et al., 2005) (Extended Experimental Procedures). As expected, N-glycosites are enriched on the protein surface. Like their unmodified counterparts N-glycosylated sites are mainly located in loop structures (71% versus 75%). Strikingly, we found evidence for a highly significant enrichment ($p < 10^{-10}$) in β sheets in comparison to non-N-glycosylated asparagines. Overall, 15% of glycosylated asparagines are predicted to be localized in β sheet structures in comparison to 5% of nonglycosylated asparagines (Figure 4B).

The percentage of proteins modified at a particular site (“site occupancy” or “stoichiometry”) is often very low for reversible PTMs. N-glycosylation is thought to be stable and frequently serves structural roles, which would favor high site occupancy. Studies in the literature frequently report high N-glycosylation site occupancy. For example, transferrin and α_1 -antitrypsin showed 98%–100% occupancy in human serum (Hulsmeier et al., 2007). However, experimental evidence for incomplete occupancies has also been reported. For example, in mouse brain the major prion protein has two different N-glycosylation sites with incomplete occupancy (66% of proteins glycosylated on both sites, 14% on one site, and 20% nonglycosylated) (Bradford et al., 2009).

If glycosylation mainly occurred in high stoichiometry, then the nonmodified counterparts of the brain N-glycopeptides should rarely be detected. In a separately measured brain proteome, we found that 98.6% of 2714 N-linked glycosylated peptides were not present in their unmodified form among 33,198 identified peptides. The lack of such peptides suggests high occupancy of glycosylation sites.

Of the total of 1938 N-glycosylated proteins from the top confidence set, approximately half carried a single N-linked sugar chain (Figure 4C). For 22% we detected two N-glycosylation sites, 13% carried three identified sites, and the average degree of glycosylation was 2.6. Notably, there was a group of 247 proteins that contained 5 or more N-glycosylation sites and 44 with at least 10. Low-density lipoprotein receptor-related protein I and II were the most heavily glycosylated proteins with 41 and 32 asparagine-linked glycosylation sites, respectively.

We applied cluster analysis to test whether N-glycosites occur in specific sequence segments of the proteins or whether they are randomly distributed on the primary sequence (Extended Experimental Procedures). Bootstrapping-based comparison of primary sequence distances between N-glycosylated and non-N-glycosylated asparagines of multiply glycosylated proteins did not reveal significant differences (Table S3B). Instead, the sites in some proteins such as sortilin (858 residues; N160, N272, N404, N682) are evenly distributed over the sequence, whereas they tightly cluster in other proteins such as Latrophilin-1 (1466 residues; N526, N635, N736, N795, N800).

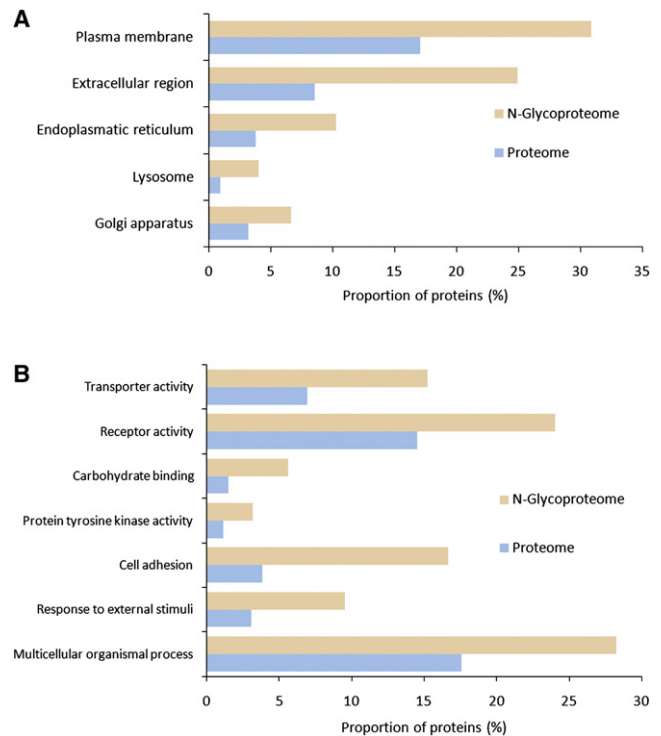


Figure 5. Gene Ontology Analysis

Cellular components (A) and molecular functions and biological processes (B) that are significantly overrepresented in the N-glycoproteome compared to the entire mouse proteome, according to Gene Ontology analysis. See also Table S4 and Table S5.

Cellular and Functional Classification of N-Glycosylated Proteins

We wished to obtain an overview of the subcellular compartments and the cellular functions that N-linked glycoproteins are preferentially associated with. We applied Cytoscape (Shannon et al., 2003) and BinGO (Maere et al., 2005) to determine Gene Ontology (GO) categories overrepresented in the glycoproteome compared to the entire mouse proteome.

A total of 31% of the N-linked glycoproteome was in the “plasma membrane” GO category and 25% in “extracellular region” (Table S4A). Taking into account nonexclusive localization in GO, 52% of the N-glycoproteome is located at the outside or beyond the plasma membrane (832 of 1594 N-glycoproteins with GO annotation). Furthermore, the ER, the Golgi apparatus, and the lysosome are overrepresented (Figure 5A). We also found N-glycosylated proteins associated with the peroxisome, endosome, and vacuole. Together, these intraluminal locations of cellular organelles accounted for 20% of the N-glycoproteome with GO annotation. Intriguingly, a number of proteins were assigned to compartments that are not topologically connected with the lumen of the ER or Golgi. However, in almost all cases, these annotations were nonexclusive or they were different in other databases such as Ensembl.

To directly address the long-standing question of N-glycosylation in unexpected cell compartments, we applied subcellular

fractionation to mouse liver using sucrose gradient separation (Extended Experimental Procedures). Consistent with the above results, we did not find any N-glycosites in proteins that were annotated on the basis of experimental evidence to be localized exclusively in the nucleus, in the mitochondria, or in the cytosol.

Within the set of plasma membrane proteins, 99% of 912 N-glycosylated sites were located in the extracellular region based on predicted Swiss-Prot topology assignments. Our dataset contains five examples of N-glycosylation, annotated to occur in cytoplasmic domains (Table S4B). However, each of these topology assignments was based on bioinformatic prediction rather than experimental evidence. Given the uncertainties of such predictions and the fact that only 5 out of 912 sites were mapped to an unexpected topological location, we conclude that our dataset contains no statistically significant evidence for N-glycosylation on the cytoplasmic face of proteins.

Many functions that are known to be characteristic for N-glycoproteins were enriched in our set, including transporter activity, receptor activity, and carbohydrate binding (Figure 5B, Table S4C). Cell adhesion, response to external stimuli, and multicellular organismal processes were the major overrepresented biological processes (Figure 5B, Table S4D). Most of the functional categories appear to be secondary to the location of the proteins at the membrane. For example, “transmembrane protein tyrosine kinase activity” is significantly overrepresented ($p < 10^{-22}$).

To test the robustness of our bioinformatic enrichment analysis, we repeated it by matching only the brain N-glycoproteome against the brain proteome derived as part of this study (5880 proteins). We obtained very similar results, indicating that the results are not tissue specific nor affected by using the total proteome instead of tissue proteome as background (Table S4E).

KEGG pathway enrichment analysis (Kanehisa and Goto, 2000) using DAVID (Dennis et al., 2003) led to similar results and additionally revealed that our N-glycoproteome is enriched for proteins that are involved in neurodegenerative diseases such as Alzheimer’s and prion disease (Table S5).

Tissue Distribution of the N-Glycoproteome

The function and extent of N-linked glycosylation are likely to be different between the tissues. To investigate this issue, we separately analyzed and overlapped the N-glycosites according to the tissue in which they had been identified.

The blood plasma N-glycoproteome comprised 1119 sites. According to GO annotation, they overwhelmingly mapped to “extracellular space” with only a few sites from luminal organellar localizations. As blood is present in all tissues—even after perfusion of mice, as done here—sites that were identified in both blood plasma and another tissue cannot be unambiguously assigned to one of the tissues (marked in orange in Figure 6).

The highest number of N-glycosylation sites (3162) was observed in brain. Of these, 1140 were not identified in any other tissue and this group includes many brain-specific proteins. Heart had the lowest number of identified sites (2213 total and 93 exclusive). All tissues had a large proportion of N-glycosylation sites that were found in at least one other tissue (purple in Figure 6).

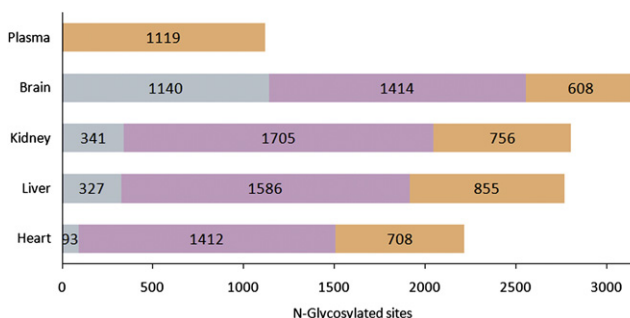


Figure 6. N-Glycoproteomes of Different Mouse Tissues

Number of identified N-glycosylated sites in blood plasma, brain, kidney, liver, and heart. Sites detected in blood plasma are in orange. Gray: sites only detected in one organ; purple: sites detected in at least two organs. See also Figure S3 and Table S6.

Accurate Relative Quantification of the In Vivo N-Glycoproteome

To demonstrate the capability of our approach to quantify N-glycoproteome changes under different conditions in vivo, we applied stable isotope labeling of amino acids in cell culture (SILAC) (Ong et al., 2002). Using brains from non-SILAC and SILAC mice (Kruger et al., 2008) we compared the N-glycosylation site pattern in old versus young mice (Extended Experimental Procedures). We performed two independent experiments, using different proteolytic enzymes, and each experiment was repeated after swapping the SILAC labels between old and young mice. We quantified 763 N-glycosylation sites using LysC and 1118 N-glycosylation sites using trypsin. In both sets we found the same N-glycosylation sites to be under- or overrepresented in forward and reverse experiments (Figure S3A, Table S6A). To determine if the apparent glycosylation changes were instead due to changes in protein abundance, we also quantified the brain proteins between the mice. Indeed, we found that the detected regulations occur not on the site but at the protein level (Table S6B and Table S6C). Although these experiments do not exclude subtle changes between the N-glycoproteome as a function of age, they suggest that there are no drastic changes. They also demonstrate that the N-glyco-FASP method is fully compatible with accurate SILAC-based quantification.

Evolution of N-Glycosylation

To derive orthologous proteins in 36 eukaryotic species, we assigned N-glycosylated proteins to their corresponding gene entries in Ensembl (Flicek et al., 2008; Gnad et al., 2009) and retrieved evolutionary annotation data from the Ensembl Compara Database (Vilella et al., 2009). We defined only “one-to-one” orthologs as interspecies homologs in our analysis. This strict definition excludes between-species paralogs and “one-to-many” or “many-to-many” orthologs in the homology set. Mouse N-glycoproteins have more orthologs in vertebrates compared to all mouse proteins (Figure S3B). However, in contrast to acetylated and phosphorylated proteins (Choudhary et al., 2009; Gnad et al., 2007), mouse N-glycosylated proteins have less orthologs in invertebrates and yeast. For example,

8% of mouse N-glycoproteins have orthologs in fly in comparison to 14% non-N-glycosylated mouse proteins. This finding also held true for each tissue-specific N-glycoproteome and may reflect the specific functions of many N-glycosylated proteins in multicellular organisms. Consistent with the drop in conservation to lower organisms, only 32 of the 829 N-glycosylated proteins identified in a *C. elegans* N-glycoproteome (Kaji et al., 2007) had one-to-one orthologs in our mouse N-glycosites. Finally, as expected, N-glycoproteins were not significantly conserved to prokaryotes compared to the entire proteome, as assessed by BLAST-based conservation analysis (Figure S3C).

DISCUSSION

A Proteomic Approach for In-Depth Mapping of the In Vivo N-Glycoproteome

We have developed and applied a strategy based on FASP, multi-lectin affinity, and high-accuracy mass spectrometric characterization. The N-glyco-FASP method allowed the determination of over 6000 N-glycosylation sites from five mouse tissues. In single LC-MS/MS runs it is possible to map over 2000 sites from 200 μ g starting material, which opens up interesting areas of application. Enrichment efficiency is reflected by the detection of more than 1000 N-glycosites on almost 500 proteins in blood plasma—thus N-glyco-FASP may be an efficient method for plasma proteome characterization in a biomarker context (Zhang et al., 2005). Importantly, the difficult class of membrane proteins, on which most N-glycosylation occurs, was excellently covered in our dataset due to the use of SDS solubilization in the FASP protocol. Further highlighting the deep coverage of the brain N-glycoproteome, we found that 58% of 1296 identified N-glycoproteins were not contained in our brain proteome consisting of 5880 proteins. Although demonstrated for multi-lectin affinity here, the method is equally applicable to the enrichment of other modified peptides. For example, phosphotyrosine peptides could be enriched with a mixture of anti-phosphotyrosine antibodies, which would not have to be coupled to any support.

The data in our study were acquired with the LTQ-Orbitrap Velos mass spectrometer with which peptide ions can be efficiently and sensitively fragmented via HCD (Olsen et al., 2009). Mass accuracy for fragments was in the low ppm range, about a factor 100 better than what we achieved in recent large-scale studies of the phosphoproteome and the acetylome (Choudhary et al., 2009; Olsen et al., 2006). Thus a “high high” strategy with HCD is a powerful technology for in-depth posttranslational modification identification, as it allows very high-confidence identification of individual peptides and ensures precise localization of modifications within the peptide sequence.

N-Glycoproteome Characterization

Our dataset covers 74% of all known mouse N-glycosylation sites and increases the number of all experimentally identified N-glycosites by a factor of about seven compared to the Swiss-Prot reference database. Interestingly, the number of identified N-glycosylation sites does not increase significantly with repeated measurements and additional fractionation, sug-

gesting that our dataset covers a substantial part of the mouse N-glycoproteome. This is in contrast to phosphoproteome analysis, in which repeated runs and fractionation add a large percentage of additional sites. Detection of 2352 N-linked glycoproteins demonstrates that more than 10% of the mouse proteome is N-glycosylated. The fact that glycosylated peptides are rarely found in their unmodified form points to high site occupancy (stoichiometry). In contrast, phosphorylation and acetylation are generally attached to the proteins in a substoichiometric manner, and nonmodified counterparts can be found for a large proportion of the phosphopeptides in proteomics studies (Olsen et al., 2010).

Biology and Evolution of the N-Glycoproteome

We find N-glycosites to be predominantly located in loops and turns on the protein surface, similarly to what we previously found for phosphorylation or acetylation sites (Choudhary et al., 2009; Gnad et al., 2007). However, unlike either acetylation sites or phosphorylation sites, N-glycosylation is also preferentially located in β sheets. This points to a more stable and rigid binding of cotranslationally attached sugar molecules in contrast to reversible phosphorylation for which sites have to be accessible posttranslationally to kinases and phosphatases.

Besides such structural constraints, the vast majority (96.5%) of N-glycosylated asparagines match the stringent glycosylation consensus sequence N-IP-[S]T. Sequence motif analysis reveals additional constraints on the known sequence recognition pattern. We find that proline is also underrepresented on the third position relative to the asparagines. This has already been shown in *C. elegans* (Kaji et al., 2007) but not in vertebrates. Threonine occurs more often than serine on the second position, in contrast to nonglycosylated sites that match the motif by chance and to phosphorylation sites, which occur much more often on serine than threonine. Furthermore, there are scattered reports of N-X-C motif on individual proteins. Our results show that this motif is widespread but that it occurs with a low frequency of about 1.3%. Furthermore, 2.2% of the N-glycosites did not match with either of the two motifs. We found these sites to be enriched for either valine on the second position or glycine on the first position relative to the N-glycosylated asparagine. The N-G and N-X-V sequence overrepresentation is statistically significant (Figure 4), and it will be interesting to investigate if it represents further minor motifs for the N-glycosylation machinery. Additionally, western blotting validated the occurrence of N-glycosylation on sites that do not match with the known motifs (Figure S2). Thus our study provides in silico as well as experimental evidence for N-glycosylation on consensus sequences different from N-IP-[S]T and N-X-C.

Evolutionary analysis reveals that N-glycosylated mouse proteins are more highly conserved throughout vertebrates but not invertebrates compared to non-N-glycosylated mouse proteins. This finding underlines the essential role of N-glycosylation on proteins that evolved with the rise of vertebrates, in concordance with the role of N-glycosylation in complex multicellular organisms. In particular, the adaptive immune system evolved in vertebrates and is thought to be intimately connected with recognition glycostructures by host and pathogen,

providing a further reason why N-glycoproteins are more highly conserved than their nonglycosylated counterparts.

Cellular Organization of N-Glycosylation

We almost exclusively found N-glycosylation to occur on secreted molecules, on the extracellular face of membrane proteins, and on the luminal side of ER, Golgi apparatus, or lysosomes. This is in agreement with the topology of glycosyltransferases that attach the sugar chains and usually have their active sites within the lumens of the ER and Golgi.

There have been suggestions regarding the presence of N-glycosylated proteins in the nuclear, cytoplasmic, and mitochondrial compartments for many years (Pedemonte et al., 1990; Reeves et al., 1981; Chandra et al., 1998; Kung et al., 2009). Several experimental and conceptual attempts have been undertaken to explain and prove this hypothesis. Nonconventional soluble glycosyltransferases may exist in the cytoplasm or nucleus and directly modify the proteins in these compartment, or soluble N-glycosylated proteins may be flipped across membranes or originate from secretory pathways (Varki et al., 2009). Experimental evidence for such mechanisms was mostly based on protein binding to lectins or radioactive labeling but did not include mapping of the sugar attachment sites. For example, sodium/potassium-transporting ATPase subunit alpha from dog kidney was reported to contain N-glycan in its cytoplasmic domain but without defining the sites (Pedemonte et al., 1990). Here, we detected a few N-glycosylation sites on parts of the proteins sequence that were annotated as cytoplasmic domains of plasma membrane proteins; however the topology of these proteins is based on prediction methods. Therefore this observation does not supply experimental evidence for the presence of N-glycosylation in the cytoplasm. Instead, despite in-depth mapping of these cellular compartments, we did not detect any proteins that were annotated to occur exclusively in the nucleus, in the mitochondria, or in the cytosol. In summary, our data point to a universal requirement of N-glycosylation sites to be in topological continuity with the lumen of the ER and Golgi.

Tissue-Specific and Disease-Related N-linked Glycoproteins

Many novel glycosylation sites have been detected for tissue-specific proteins, such as neurotransmitter receptors and contactins for brain or cubilin and megalin for kidney. For example, we found 30 kidney-specific sites on megalin (low-density lipoprotein receptor-related protein 2, 520 kDa), out of which only one was reported in a recent N-glycoproteomic study (Wollscheid et al., 2009).

Besides tissue-specific N-glycosylation sites, we also detected a number of N-glycoproteins that are associated with various diseases. For example, we found many N-glycosites on proteins that are involved in Alzheimer's disease (AD), the most common type of dementia (Price et al., 1998). Because N-glycosylation is involved in many processes impaired in AD, such as protein folding, protein anchoring to cell membranes, and protein delivery to organelles, it is possible that N-glycosylation is directly involved in cause or progress of AD (Selkoe, 2004; Suzuki et al., 2006). Surprisingly, using both high-accuracy mass spectrometry and western blotting, we found N-glycosyla-

tion on Apolipoprotein E, an important AD protein (Kim et al., 2009), which does not contain any asparagines that match the known motifs N-IP-[S|T] and N-X-C.

Summary and Outlook

Here we provide a practical and highly efficient method for mapping the attachment sites of N-glycostructures. Modification sites occur on many proteins of pivotal importance in development, multicellular communication, and many other basic biological functions. These high-confidence N-glycosylation sites can now be used by the community for detailed functional studies. We have also shown that N-glyco-FASP is fully compatible with quantitative proteomics methods such as SILAC, which could be used to study the many diseases that directly or indirectly involve N-glycosylation.

EXPERIMENTAL PROCEDURES

Tissue Preparation, Protein Extraction, and Digestion

Tissues used in our study were dissected from inbred mouse strain C57BL/6, males, aged 3 months and proceeded according to the FASP protocol (Wisniewski et al., 2009b). The resulting peptides were eluted in lectin binding buffer (1 mM CaCl₂, 1 mM MnCl₂, 0.5 M NaCl in 20 mM TrisHCl, pH 7.3). For more details, see the [Extended Experimental Procedures](#).

Lectin Enrichment and Deglycosylation

Approximately 100 µg of digested peptides were mixed with lectin solution containing either a combination of ConA, WGA, and RCA₁₂₀ (Sigma) or just one lectin at a time, resulting in mixtures of peptides and lectins with a mass proportion of 1:2. The mixtures were transferred to new YM-30 filter units (Microcon, Millipore). After 1 hr incubation at room temperature the unbound peptides were eluted by centrifugation. The captured peptides were washed followed by deglycosylation with PNGase F (Roche). After incubation for 3 hr at 37°C deglycosylated peptides were eluted. A more detailed protocol is available in the [Extended Experimental Procedures](#) section.

Mass Spectrometric Analysis

The eluted peptides were purified on StageTips (Rappsilber et al., 2003) and analyzed on the LTQ-Orbitrap Velos mass spectrometer; (Thermo Fisher Scientific, Germany) coupled to HPLC via a nanoelectrospray ion source. Full MS scans were acquired with resolution of 30,000 at m/z 400 in the orbitrap analyzer. The ten most intense ions were fragmented by higher-energy collisional dissociation (HCD) in the octopole collision cell. The HCD fragment ion spectra were acquired in the orbitrap analyzer with resolution of 7500 at m/z 400. The MS data were analyzed using MaxQuant (Cox and Mann, 2008). Proteins were identified by searching MS and MS/MS data using the MASCOT search engine (Matrix Science, UK) (Perkins et al., 1999). See the [Extended Experimental Procedures](#) for further information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and six tables and can be found with this article online at [doi:10.1016/j.cell.2010.04.012](https://doi.org/10.1016/j.cell.2010.04.012).

ACKNOWLEDGMENTS

We thank Alexandre Zougman, Nagarjuna Nagaraj, Korbinian Mayr, Mara Monetti, Ina Rohwedder, Markus Moser, and Pawel Ostasiewicz for help and fruitful discussions. This work was supported by PROSPECT, a 7th framework program of the European Union (grant agreement HEALTH-F4-2008-201648/PROSPECTS).

Received: October 27, 2009

Revised: February 8, 2010

Accepted: April 7, 2010

Published: May 27, 2010

REFERENCES

- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Bradford, B.M., Tuzi, N.L., Feltri, M.L., McCorquodale, C., Cancellotti, E., and Manson, J.C. (2009). Dramatic reduction of PrP C level and glycosylation in peripheral nerves following PrP knock-out from Schwann cells does not prevent transmissible spongiform encephalopathy neuroinvasion. *J. Neurosci.* 29, 15445–15454.
- Bunkenborg, J., Pilch, B.J., Podtelejnikov, A.V., and Wisniewski, J.R. (2004). Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* 4, 454–465.
- Chandra, N.C., Spiro, M.J., and Spiro, R.G. (1998). Identification of a glycoprotein from rat liver mitochondrial inner membrane and demonstration of its origin in the endoplasmic reticulum. *J. Biol. Chem.* 273, 19715–19721.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4, 3.
- Flicek, P., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., et al. (2008). Ensembl 2008. *Nucleic Acids Res.* 36, D707–D714.
- Gnad, F., Orosi, M., Birney, E., and Mann, M. (2009). MAPU 2.0: high-accuracy proteomes mapped to genomes. *Nucleic Acids Res.* 37, D902–D906.
- Gnad, F., Ren, S., Cox, J., Olsen, J.V., Macek, B., Orosi, M., and Mann, M. (2007). PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* 8, R250.
- Gundry, R.L., Raginski, K., Tarasova, Y., Tchernyshyov, I., Bausch-Fluck, D., Elliott, S.T., Boheler, K.R., Van Eyk, J.E., and Wollscheid, B. (2009). The mouse C2C12 myoblast cell surface N-linked glycoproteome: Identification, glycosite occupancy, and membrane orientation. *Mol. Cell. Proteomics* 8, 2555–2569.
- Hulsmeier, A.J., Paesold-Burda, P., and Hennet, T. (2007). N-glycosylation site occupancy in serum glycoproteins using multiple reaction monitoring liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics* 6, 2132–2138.
- Jensen, O.N. (2006). Interpreting the protein language using proteomics. *Nat. Rev. Mol. Cell Biol.* 7, 391–403.
- Kaji, H., Kamiie, J., Kawakami, H., Kido, K., Yamauchi, Y., Shinkawa, T., Taoka, M., Takahashi, N., and Isobe, T. (2007). Proteomics reveals N-linked glycoprotein diversity in *Caenorhabditis elegans* and suggests an atypical translocation mechanism for integral membrane proteins. *Mol. Cell. Proteomics* 6, 2100–2109.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Kim, J., Basak, J.M., and Holtzman, D.M. (2009). The role of apolipoprotein E in Alzheimer's disease. *Neuron* 63, 287–303.
- Kruger, M., Moser, M., Ussar, S., Thievensen, I., Lubner, C.A., Forner, F., Schmidt, S., Zanivan, S., Fassler, R., and Mann, M. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* 134, 353–364.
- Kung, L.A., Tao, S.C., Qian, J., Smith, M.G., Snyder, M., and Zhu, H. (2009). Global analysis of the glycoproteome in *Saccharomyces cerevisiae* reveals new roles for protein glycosylation in eukaryotes. *Mol. Syst. Biol.* 5, 308.
- Kuster, B., and Mann, M. (1999). 18O-labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal. Chem.* 71, 1431–1440.
- Lee, A., Kolarich, D., Haynes, P.A., Jensen, P.H., Baker, M.S., and Packer, N.H. (2009). Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *J. Proteome Res.* 8, 770–781.
- Liu, T., Qian, W.J., Gritsenko, M.A., Camp, D.G., 2nd, Monroe, M.E., Moore, R.J., and Smith, R.D. (2005). Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* 4, 2070–2080.
- Macek, B., Mann, M., and Olsen, J.V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu. Rev. Pharmacol. Toxicol.* 49, 199–221.
- Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449.
- Medzihradsky, K.F. (2005). Characterization of protein N-glycosylation. *Methods Enzymol.* 405, 116–138.
- North, S.J., Huang, H.H., Sundaram, S., Jang-Lee, J., Etienne, A.T., Trollope, A., Chalabi, S., Dell, A., Stanley, P., and Haslam, S.M. (2010). Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J. Biol. Chem.* 285, 5759–5775.
- Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635–648.
- Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 4, 709–712.
- Olsen, J.V., Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., et al. (2009). A dual pressure linear ion trap - Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 8, 2759–2769.
- Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* 3, ra3.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386.
- Pedemonte, C.H., Sachs, G., and Kaplan, J.H. (1990). An intrinsic membrane glycoprotein with cytosolically oriented n-linked sugars. *Proc. Natl. Acad. Sci. USA* 87, 9789–9793.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.
- Price, D.L., Tanzi, R.E., Borchelt, D.R., and Sisodia, S.S. (1998). Alzheimer's disease: genetic studies and transgenic models. *Annu. Rev. Genet.* 32, 461–493.
- Ramachandran, P., Boonthung, P., Xie, Y., Sondej, M., Wong, D.T., and Loo, J.A. (2006). Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res.* 5, 1493–1503.
- Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* 75, 663–670.
- Reeves, R., Chang, D., and Chung, S.C. (1981). Carbohydrate modifications of the high mobility group proteins. *Proc. Natl. Acad. Sci. USA* 78, 6704–6708.
- Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100.
- Schwartz, D., and Gygi, S.P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* 23, 1391–1398.

- Selkoe, D.J. (2004). Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases. *Nat. Cell Biol.* 6, 1054–1061.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Suzuki, T., Araki, Y., Yamamoto, T., and Nakaya, T. (2006). Trafficking of Alzheimer's disease-related membrane proteins and its participation in disease pathogenesis. *J. Biochem.* 139, 949–955.
- Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., and Etzler, M.E. (2009). *Essentials of Glycobiology* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19, 327–335.
- Wagner, M., Adamczak, R., Porollo, A., and Meller, J. (2005). Linear regression models for solvent accessibility prediction in proteins. *J. Comput. Biol.* 12, 355–369.
- Wisniewski, J.R., Zougman, A., and Mann, M. (2009a). Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* 8, 5674–5678.
- Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009b). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362.
- Witze, E.S., Old, W.M., Resing, K.A., and Ahn, N.G. (2007). Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* 4, 798–806.
- Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R., and Watts, J.D. (2009). Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotechnol.* 27, 378–386.
- Woods, R.J., Edge, C.J., and Dwek, R.A. (1994). Protein surface oligosaccharides and protein function. *Nat. Struct. Biol.* 1, 499–501.
- Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., et al. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* 34, D187–D191.
- Yang, Z., and Hancock, W.S. (2004). Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J. Chromatogr. A* 1053, 79–88.
- Zajonc, D.M., Striegl, H., Dascher, C.C., and Wilson, I.A. (2008). The crystal structure of avian CD1 reveals a smaller, more primordial antigen-binding pocket compared to mammalian CD1. *Proc. Natl. Acad. Sci. USA* 105, 17925–17930.
- Zhang, H., Li, X.J., Martin, D.B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* 21, 660–666.
- Zhang, H., Yi, E.C., Li, X.J., Mallick, P., Kelly-Spratt, K.S., Masselon, C.D., Camp, D.G., 2nd, Smith, R.D., Kemp, C.J., and Aebersold, R. (2005). High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* 4, 144–155.
- Zielinska, D.F., Gnad, F., Jedrusik-Bode, M., Wisniewski, J.R., and Mann, M. (2009). *Caenorhabditis elegans* has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. *J. Proteome Res.* 8, 4039–4049.