Resource

Cell

# 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,*]
[1]Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, Martinsried D-82152, Germany
[2]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA
[3]These 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 lumenal part of proteins in compartments of subcellular organelles such as the endoplasmatic 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; Medzihradszky, 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 $^{18}$O-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-acetylglucosamine and agglutinin $RCA_{120}$, which captures galactose modified at the 3-0 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 $^{18}O$-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 $^{18}O$-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