Method development for the selective enrichment and mass spectrometric characterization of posttranslational modifications of secreted and membrane proteins

We have aimed at method development for selective enrichment and mass spectrometric characterization of post-translational modifications including O-glycosylation, phosphorylation and sulfation using commercially available human serum as primary sample source.

We have tested multiple lectin-based affinity chromatography workflows for the selective enrichment of O-glycopeptides from the tryptic digest of human serum. Two approaches yielded satisfactory results, both applying Jacalin, a plant lectin (Figure 1). In sample preparation (SP) 1 we adapted our published workflow originally developed for fetal bovine serum, applying Jacalin affinity chromatography both on the protein and peptide level with an additional fractionation step, ERLIC (electrostatic repulsion hydrophilic interaction) chromatography in between in order to reduce nonspecific background from non-glycopeptides. We further optimized the sample preparation workflow by using MS-compatible volatile buffers during ERLIC chromatography in order to reduce sample clean-up steps. Although this approach did yield glycopeptide identifications, sample preparation was rather tedious and the number of identified glycosylation sites was far below our expectations. Therefore we tested another lectin, wheat germ agglutinin (WGA) which has recently been shown to bind a wide range of glycan structures. Unfortunately, the affinity of this lectin towards glycopeptides is weak, and glycopeptides (especially O-glycopeptides with relatively small glycan structures) partly co-elute with non-glycosylated peptides. We could identify only a few O-glycopeptides from PNGase F-treated human serum tryptic digest even after a two-round enrichment using WGA, presumably due to the high nonspecific background (unpublished results). Finally, we also tested a workflow which combined the two above lectins. First, we applied WGA for general glycopeptide enrichment (both N- and O-glycopeptides are isolated) followed by Jacalin to separate mucin core-1 type O-glycopeptides. (SP2, Figure 1). In order to reduce nonspecific background, we used WGA immobilized on POROS support (as opposed to agarose-bound lectin used previously). SP2 outperformed SP1 in terms of identified glycopeptides and glycosylation sites. SP2 is also much less labor-intensive compared to SP1. We have reported our results in [1].

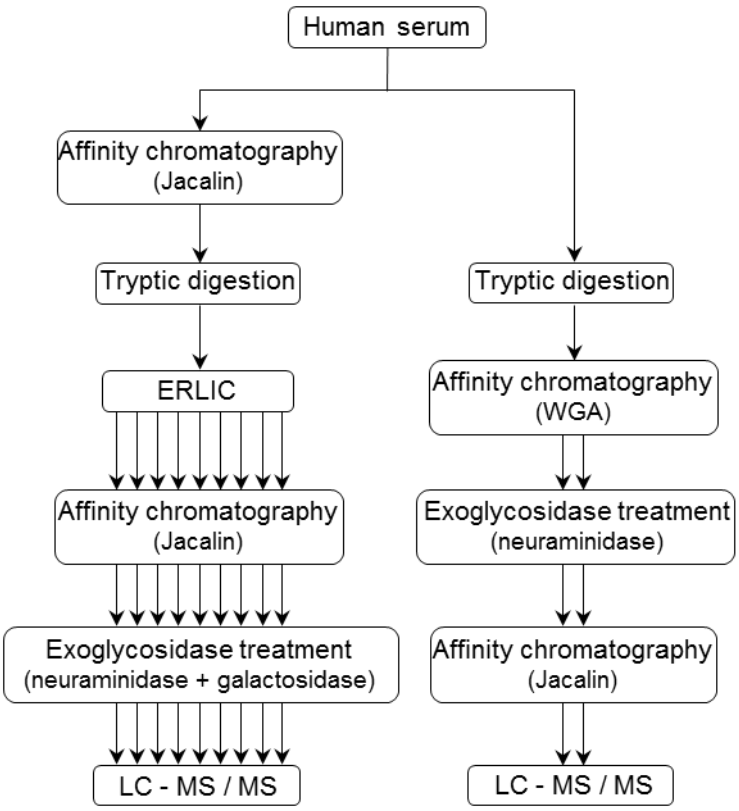

Figure 1. Outline of sample preparation workflows for O-glycopeptide enrichment from human serum tryptic digest

Inspired by the performance of SP2 we further tested this workflow by using i) only volatile buffers during the whole sample preparation protocol and ii) lower amount of starting material. Furthermore, we assessed the reproducibility of the workflow. Using "spectral families" (glycopeptide groups representing the same glycosylation status and sequence stretch of the glycoprotein) in line with (at least low-confidence) MS/MS data and some retention time considerations we have shown that the technical reproducibility of the WGA-Jacalin approach is excellent, falling into the range of the repeatability of LC-MS/MS experiments [2]. This was the first report where the detailed assessment of a sample preparation aiming O-glycopeptide enrichment was performed. More importantly, based on our results, the WGA-Jacalin sample preparation workflow is highly reproducible and can be recommended for further studies such as glycoprofiling of body fluids from healthy and diseased individuals.
We originally planned to test other lectins for the enrichment of serum O-glycopeptides. However, while working on the optimization of the above workflows we came across a detailed study of the O-glycans of human serum [3] that established that the overwhelming majority represents core-1 structures. Thus, testing lectins with different preferences cannot yield good results.
Besides using soluble glycoproteins we attempted to isolate glycopeptides from membrane proteins, namely from exosomes (their phosphorylation analysis is described below) as well as from a mouse liver membrane preparation. Glycopeptides from the exosomes were clearly below our detection limit, and we could not increase the quantity of starting material. With the 'pure' membrane fraction there were basic sample preparation issues that have to be optimized prior to the glycopeptide enrichment. Most likely we have to find a better model system, perhaps a laboratory cell line. Nonetheless, we identified O-glycopeptides in the same range as published in [4], although only partial overlap of the identified glycosylation sites was observed (unpublished results), probably due to differences between the two sample preparation workflows.

All glycopeptide studies on complex samples heavily rely on ETD (electron transfer dissociation) MS/MS analysis of the samples. Unlike conventional collisional activation (CID and HCD), labile side-chain modifications including glycans remain mostly intact during ETD allowing site assignment. This activation method works best for higher charge states, more precisely, for high charge density ($m/z<850$) molecules - a criteria that tryptic glycopeptides might not fulfill (tryptic peptides are typically doubly charged and of low charge density). Therefore, we tested whether using an endoproteinase other than trypsin would allow the identification of more/additional glycopeptides. We have chosen LysArgiNase, an endoproteinase that cleaves N-terminally to Lys and Arg residues yielding peptides with preferential charge retention at the N-terminus. Thus, MS/MS signals are not split between N- and C-terminal fragment series leading to higher sensitivity. (In our original workplan endoprotease LysN was listed for such an experiment, but the recently discovered LysArgiNase cleaves at both basic amino acids, well characterized and is readily available). Parallel to tryptic samples, we also prepared technical replicates for SP2 using LysArgiNase and showed that the reproducibility of the method was similar for both digests [5]. On the other hand, LysArgiNase was not found to be superior to trypsin – although it yielded some novel O-glycosylation site identifications, trypsin afforded four times more of confidently identified O-glycosylation sites compared to LysArgiNase.

Data interpretation is the bottleneck for many PTM studies. Due to the lack of truly reliable softwares for automated glycopeptide data interpretation, we carefully inspect all novel glycopeptide identifications. Although manual validation of the spectra is a tedious task it might yield unexpected gems. We have discovered a side reaction of the commonly used buffering agent, Tris (tris(hydroxymethyl)aminomethane) with sialic acid which is a common capping unit both in N- and O-glycans. We have shown that this side reaction is not negligible (~5-10% of the acidic glycopeptides were found modified), occurs both on N- and O-glycopeptides, and may compromise data interpretation [6].
Furthermore, we have also shown that another side reaction, overalkylation may also lead to glycopeptide data misinterpretation [7]. Although carbamidomethylation of amino acid side chains other than cysteine has been described for unmodified peptides, this was the first report that discussed glycopeptide data interpretation issues related to this undesired side reaction.
Our experience with glycopeptide analysis is valued by the international glycoscience community. K.F. Medzihradszky was invited as a keynote speaker [8] and session chair (Glycoproteomics - Technical Limitations & Prospects) to the 2016 HUPO World Congress in Taipei. In addition, we were invited to contribute a minireview on the status of O-glycosylation analysis (in preparation) to *Mol. Cell. Proteomics*. We also actively participate in development of search engines (ProteinProspector, Byonic and Pinnacle) for improving automated interpretation of glycopeptide MS/MS data as illustrated by conference presentations [9,10].

In parallel to the O-glycosylation studies we also aimed at the characterization of extracellular phosphorylation and sulfation. The regulatory role of phosphorylation within the cell is well established, however, its importance outside

of the cell is just being explored. Similarly, our knowledge about serum sulfation is scarce. We have recently reviewed the current status of protein sulfation analysis in an invited book chapter [11].
In the analysis of these PTMs we tested different enrichment techniques that target the acidic character of these PTMs: TiO_2 enrichment, Fe(III)-immobilized metal affinity chromatography (IMAC) and ERLIC. The generally applied conditions of TiO_2 enrichment resulted in the co-isolation of phosphorylated and glycosylated peptides. As glycosylation represents a more abundant modification of extracellular proteins this interference of the glycopeptides compromised the phosphorylation analysis of serum samples. For Fe(III)-IMAC various conditions in terms of the acidic component and the organic modifier have been reported in previous large scale phosphorylation studies. We tested different enrichment conditions to selectively isolate phosphopeptides from the tryptic digest of human serum and found that similarly to the TiO_2 enrichment, co-isolation of the glycopeptides by Fe(III)-IMAC occured under conditions promoting hydrophilic interaction as a result of a mixed mode of metal affinity and hydrophilic interaction. Thus, changing the enrichment conditions can yield the selective isolation of phosphopeptides or the joint enrichment of phospho- and glycopeptides as we presented in [12]. Further optimization of the enrichment conditions resulted in an improved and reproducible isolation of phosphopeptides from human serum, as reported in [13]. We also discussed additional aspects of extracellular phosphorylation regarding enrichment, mass spectrometry analysis, data interpretation and database mining in a minireview [14]. In parallel to the metal affinity enrichment we also tested ERLIC for the fractionation of phosphopeptides from human serum. Unfortunately, ERLIC could not handle the large complexity and wide dynamic range of serum and co-elution of highly acidic unmodified peptides compromised phosphopeptide identifications.
Using sulfopeptides and artificially sulfated fetuin tryptic digest as standard we found that the sulfopeptides modified on Tyr residues are not retained at all by metal affinity and sulfopeptides modified on Ser or Thr residues show only partial retention on TiO_2. Although ERLIC showed some separation of the unmodified and sulfated peptides using our standard, we failed to identify sulfopeptides from human serum samples.
As a pilot study for membrane phosphoproteomics we performed phosphorylation analysis of exosomes released from mouse macrophages. Approximately half of the phosphoproteins identified were membrane proteins according to GO annotation indicating that our Fe(III)-IMAC phosphopeptide enrichment protocol is applicable to membrane preparations as well. The mouse liver membrane sample that was used for O-glycosylation analysis was also subjected to phosphorylation analysis and resulted in high number of phosphopeptides representing over 100 membrane associated proteins (unpublished results).

3 MSc and 1 BSc students were involved in the above detailed studies and prepared their theses from the results.

References

1. Darula Z., Sarnyai F., Medzihradszky KF.: O-glycosylation sites identified from mucin core-1 type glycopeptides from human serum, *Glycoconj. J.* 33(3):435-45 (2016)
2. Pap A, Medzihradszky KF, Darula Z.: Using "spectral families" to assess the reproducibility of glycopeptide enrichment: human serum O-glycosylation revisited, *Anal. Bioanal. Chem.* 409(2):539–550 (2017)
3. Yabu M, Korekane H, Miyamoto Y. Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology* 24(6):542-53 (2014)
4. Medzihradszky KF, Kaasik K, Chalkley RJ. Tissue-Specific Glycosylation at the Glycopeptide Level. *Mol. Cell. Proteomics* 14(8):2103-10 (2015)
5. Pap A., Medzihradszky K.F., Darula Z.: Exploring the human serum O-glycoproteome using different proteases, 10th Central and Eastern European Proteomic Conference (2016)
6. Darula Z, Medzihradszky KF: Glycan Side Reaction May Compromise ETD-Based Glycopeptide Identification, *J. Am. Soc. Mass Spectr.* 25(6):977-987 (2014)
7. Darula Z, Medzihradszky KF.: Carbamidomethylation Side Reactions May Lead to Glycan Misassignments in Glycopeptide Analysis. *Anal. Chem.* 87(12):6297-302, 2015
8. Medzihradszky KF.: Extracellular Glycosylation: How to See the Forest Despite All the Trees, 15th Human Proteome Organization World Congress (2016)
9. Hunyadi-Gulyas E; Darula Z; Medzihradszky K; Bern M.: Byonic: a highly efficient new bioinformatic tool for intact glycopeptide analysis, 32nd Informal Meeting on Mass Spectrometry (2014)
10. Darula Z, Hunyadi-Gulyas E, Bern M, Medzihradszky KF: N- and O-glycopetide analysis from HCD data, Proteomic Forum (2015)

11. Klement E, Hunyadi-Gulyas E, Medzihradszky KF. Biological Significance and Analysis of Tyrosine Sulfation in "Analysis of Protein Post-Translational Modifications by Mass Spectrometry" edited by J. Griffiths and R. Unwin; Wiley Production; DOI: 10.1002/9781119250906.ch9.
12. Klement E., Medzihradszky K.F.: IMAC enrichment with two-step elution for the simultaneous characterization of phosphorylation and glycosylation on extracellular proteins, 10th Central and Eastern European Proteomic Conference (2016)
13. Klement E, Raffai T, Medzihradszky KF.: Immobilized metal affinity chromatography optimized for the analysis of extracellular phosphorylation, Proteomics 16(13):1858-62. doi: 10.1002/pmic.201500520. (2016)
14. Klement E, Medzihradszky KF.: Extracellular Protein Phosphorylation, the Neglected Side of the Modification, Mol. Cell. Proteomics 16(1):1-7 (2017)