The impact of sample processing on data quality in plasma proteomics

Introduction

Medical research, especially molecular diagnostics depends on generation of reliable data. Proteins proven to be associated with a specific disease need to be reproducibly detected in human biopsy samples.

Sample preparation procedures can have a major impact on reproducibility, therefore we tested three methods for processing plasma samples.

In this whitepaper we compare *in solution* digest, filter aided sample preparation (FASP) and *in gel* digest with respect to the reproducible measurement of quality values like overall signal intensity, number of identified peptides/proteins, missed cleavages etc.

Method evaluation

Signal intensity

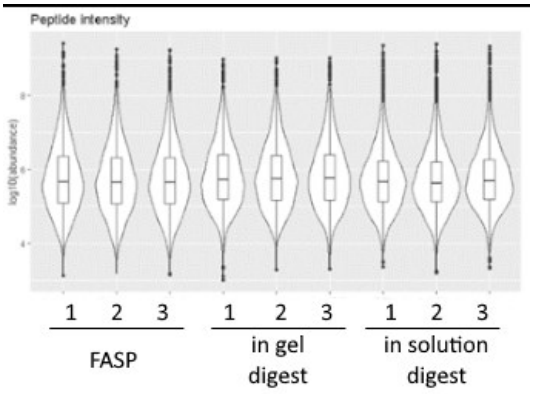
The median of the signal intensity was similar for each sample preparation method and in between samples (Fig 1A). Clustering analysis (Fig 1B) showed variations in signal intensity between sample preparation methods but the pattern of peptide signal intensities was repeatable for each method. The highest number of peptides could be identified from the *in solution* digested samples while *in gel* digested samples returned the least amount of peptide IDs.

Quality of digestion

Quality of digestion is analysed through detection and evaluation of missed protease cleavage sites (Fig 2). Numbers of missed cleavages were comparable in FASP and *in gel* digest, with 21 % and 23 % missed cleavages, respectively. *In solution* digest resulted in a very high number of missed cleavages, with 52 % of all detected peptides.

The high number of missed cleavages may be caused by insufficient removal of disturbing substances. While *in gel* and FASP samples were washed several

A)



B)

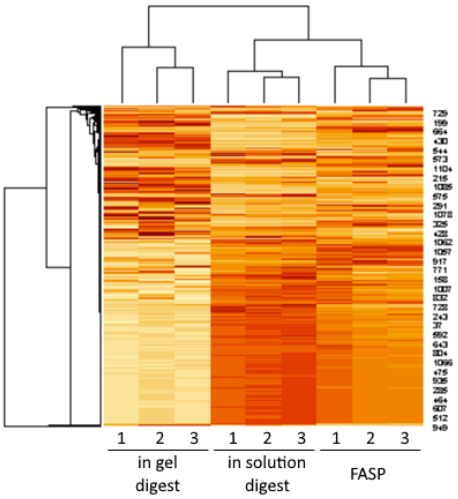


Fig1.
A) Violin/Box plot of overall distribution of measured peptide intensities. B) Clustering analysis of peptide signal intensities.

times before the digest, during the *in solution* digest the sample was precipitated and solubilized and may still contain substances negatively affecting protease efficiency.

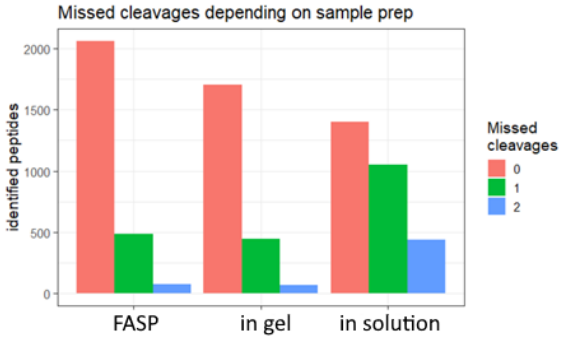


Fig. 2 Illustration of detected missed cleavages for each sample prep. No missed cleavages (red bars), one missed cleavage (green bars), two missed cleavages (blue bars).

Repeatable protein quantification with FASP

The distribution of protein intensities varied between sample preparation methods but the distribution was still highly similar for the *in gel* digested samples and the FASP samples. *In solution* digested samples showed a higher degree of variation (data not shown). Most proteins were identified in samples prepared with the FASP method while the least number of proteins was identified using the *in gel* digest.

Variation of quantification

To determine the data quality, we examined the variability of the protein signal intensities between the triplicate samples (Fig. 3). The least variation was observed when using the FASP sample preparation method. The highest variation was observed for the *in solution* digested samples.

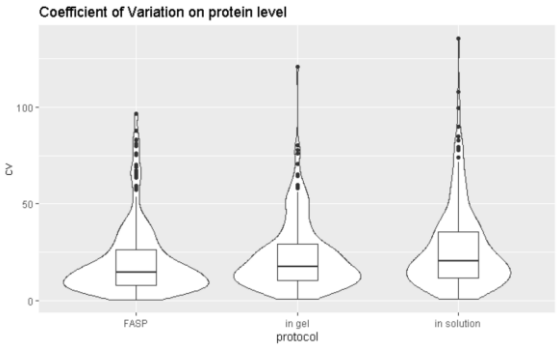


Fig. 3 Violin/Box plot, displaying the coefficient of variation on the protein level.

Conclusion

In solution digest enabled detecting the most peptides but the sample was not completely digested and still contained a significant amount of missed cleavages. *In gel* digested samples contained less missed cleavages but also less peptides and proteins could be identified. We identified the highest number of proteins in the samples processed with the FASP protocol. The performance of this protocol was also demonstrated by low variation between samples and an efficient digest. Thus, we recommend the FASP protocol for generating reproducible and reliable data for clinical proteomics.

Sample processing

Each method was performed in triplicates.

In solution digest

10 µl of human plasma were denatured in 6 M Guanidine Hydrochloride (GdnHCl) and heat-inactivated. The diluted plasma was precipitated with ice-cold Methanol. The air-dried pellet was solubilized in GdnHCl.

The solubilized plasma was diluted in Ammoniumbicarbonate (ABC) buffer followed by reduction and alkylation with Dithiothreitol (DTT) and Iodoacetamide (IAA). Proteins were digested overnight at 37°C with Trypsin and LysC.

In gel digest

The plasma was diluted in ABC and run briefly into an SDS-PAGE. Proteins were stained with Coomassie Brilliant Blue. Each band containing all proteins from one sample was excised and sliced into smaller pieces, followed by washing and destaining. After drying the gel pieces, the proteins were reduced and alkylated. Gel pieces were dehydrated and dried. The proteins were digested with Trypsin overnight at 37°C.

FASP

The heat-inactivated plasma was diluted in GdnHCl/ABC. After reduction and alkylation the sample was loaded on an appropriate FASP filter and washed with GdnHCl and ABC. Proteins were digested with Trypsin and LysC overnight at 37°C in a wet chamber. Peptides were eluted by centrifugation.

Peptide clean-up and MS measurement

Peptides were desalted with C18 stage tips following the manufacturer's instructions and separated by liquid chromatography (LC) using a nano-LC column. Discovery MS analyses were performed on a quadrupole orbitrap mass spectrometer equipped with a nano-electrospray ion source (Orbitrap Exploris 480). The mass spectrometer was operated in positive ion mode and data-independent acquisition mode (DIA).

Data analysis

Data was analysed with DIA-NN using a complete human database. Statistical analyses were computed using R programming language.

Summary

	In solution digest	In gel digest	FASP
Peptide IDs	+++	+	++
Protein IDs	++	+	+++
Missed cleavages	-	++	++
Repeatability	+	+	++
Variation	+	++	+++

POLYQUANT

PolyQuant GmbH
Industriestr. 1
93077 Bad Abbach
Germany

Tel. +49 (0) 9405 96999 10
Fax. +49 (0) 9405 96999 28
www.polyquant.com