Evaluation of an unbiased, deep, and scalable nanoparticle-based proteomics workflow for limited plasma sample volume from model organisms

Daniel Hornburg, Shao-Yung (Eric) Chen*, Lucy Williamson, Margaret Donovan, and Aaron S. Gajadhar

The Proteograph™ Product Suite enables rapid sample preparation for reproducible, deep plasma proteomic analysis

Model organisms, like mice, rats, pigs, and monkeys, are utilized to unveil biological insights to research in human diseases. However, blood volume in model organisms is limited and lower than the optimal 250 µL starting plasma volume required to perform deep plasma proteomics with the Proteograph Assay. This fact impacts the ability to conduct deep plasma proteomics in longitudinal sampling studies for model organisms, for example, collecting limited amounts of blood from the model organism while not sacrificing said model organism.

Herein, we explore the performance of the Proteograph Product Suite coupled with a state-of-the-art mass spectrometer with conventional capillary liquid chromatography workflows (nanoLC-MS/MS) when limited sample volumes are utilized. We evaluated Proteograph Product Suite SP100 sample processing with mouse serum starting volumes of 250, 125, 50, 25, and 10 µL. We investigated the LC-MS/MS performance using ~30-minute LC methods coupled to data-independent-acquisition strategies (DIA) and analyzed the data via our Proteograph Analysis Suite 2.1 platform, evaluating depth of proteome coverage, dynamic range, peptide yield, and reproducibility of the Proteograph platform.

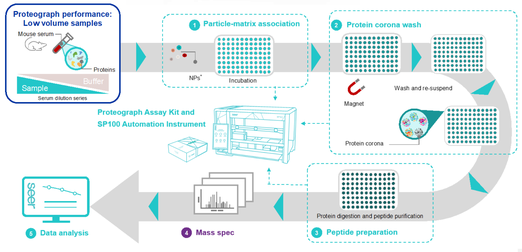

Figure 1. Proteograph workflow schematic for sample processing, LC-MS acquisition, and data analysis.

Triplicates of 250, 125, 50, 25, and 10 µL of mouse serum samples + triplicates of traditional neat digestion (10 µL) were processed in parallel. Data was acquired via a Thermo Orbitrap™ Exploris™ 480 MS with DIA method; data was then analyzed using Proteograph Analysis Suite 2.1.

Methods

Sample Preparation

The minimum volume required for each sample for Proteograph Assay is 250 µL. For processing samples with lower than the assay standard volume, sample volume was first mixed with an appropriate amount of Proteograph Assay Reconstitution Buffer A (Seer Inc.) to yield a final volume of 250 µL. After samples are loaded onto the SP100 Automation Instrument, samples are ready to undergo protein corona formation and processing to purified peptides. To form the protein corona, Seer's proprietary nanoparticles are mixed with diluted mouse serum samples and incubated at 37°C for 1 hr. Unbound proteins are removed, and downstream wash, reduction, alkylation, and protein digestion were all performed according to the Proteograph Assay protocol (Figure 1).

LC-MS Data Acquisition and Data Analysis

For LC-MS analysis in Data-Independent Acquisition (DIA) mode, 200 – 400 ng of peptides in 4 µL were reconstituted in a solution of 0.1% formic acid (FA) and 3% acetonitrile (ACN) spiked with 5 fmol/µL PepCalMix from SCIEX for constant mass MS injection between samples regardless of starting volume. Each samples were analyzed with a Ultimate 3000 HPLC system coupled with a Thermo Fisher Scientific Exploris 480 Mass Spectrometer. DIA data was processed using Proteograph Analysis Suite 2.1. Raw MS data was processed using the DIA-NN search engine (version 1.8.1) in library-free mode searching MS/MS spectra against an in silico generated spectral library of mouse protein entries (UP000000589_10090). Different starting volume conditions (250, 125, 50, 25, and 10 µl) were searched individually to avoid boosting of identifications for lower volume samples.

Proteograph assay provides high depth of coverage and precision while outperforming traditional direct digestion performance for low volume samples

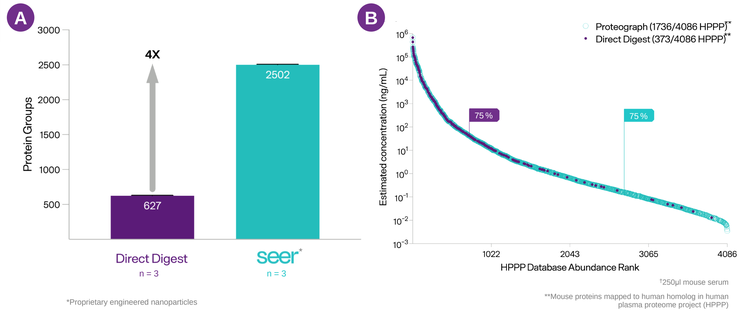

Figure 2. Performance of Proteograph vs. Traditional Direct Digest workflow using Mouse Serum. (A) Protein identifications for mouse serum samples using Proteograph Assay with standard input volumes and direct digest, with error bars denoting standard deviation. (n=3) **(B)** Waterfall plot mapped towards the human homolog in HPPP database showing the dynamic range and proteome coverage.

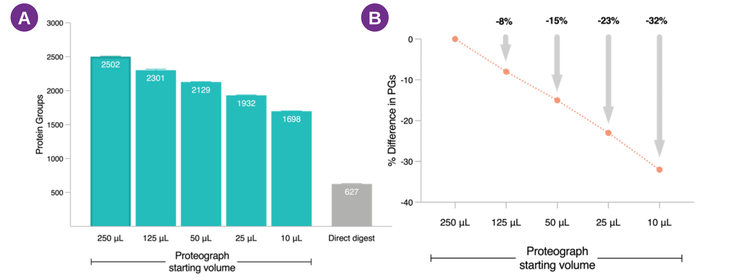

Figure 3. Evaluation of Assay Protein Identification Rates. (A) Protein identifications for triplicates of 250, 125, 50, 25, and 10 µL of mouse serum samples and triplicates of traditional neat digest (10 µL). **(B)** Percentage decrease of protein identification rates (relative to standard Proteograph Assay starting volumes) when lower starting volume is used for sample processing with Proteograph Assay,

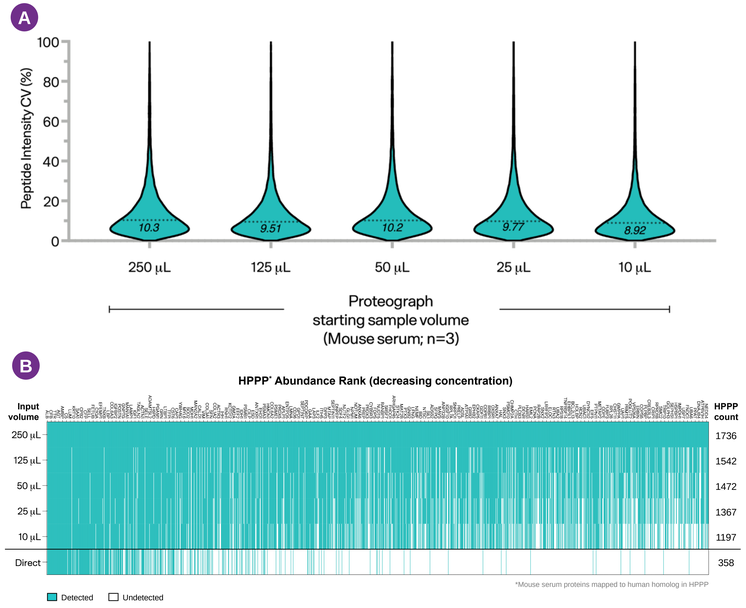

Figure 4. Evaluation of Measurement Precision and Protein Dynamic Range Coverage and Depth.

(A) Peptide intensity CV (%) plot across different Proteograph Assay with different starting sample volume (n=3), with dotted lines denoting median CV. **(B)** Protein identifications from 250, 125, 50, 25, 10, and traditional direct digests were mapped towards the human homolog in the HPPP database[1]. The protein concentrations are rank ordered in decreasing abundance from left to right. comparison was made towards the protein groups detected in the 250 µL starting condition.

[1]Schwenk, et al. *Journal of Proteome Research.* (2017

Conclusion

- Proteograph Assay performance across a range of low sample volumes outperforms direct digest workflow (2.5 – 4X increase)
- At low, non-standard Proteograph starting sample volumes, we observe a modest decrease in protein group identification (<15% drop at 50µl) relative to the standard input volume (250 µL) and higher coverage of low abundant proteins compared to direct digest workflows
- Proteograph workflow provides deep proteome coverage for a variety of small model organisms with low volume of plasma/serum samples
- The Proteograph Product Suite is compatible with low volume studies, including model organism longitudinal sampling studies. This application describes a simple, straightforward dilution strategy to extend the utility of the Proteograph Assay for cases where only limited sample volumes are available, but still provides deep proteome coverage with highly reproducible quantification.

Publications