Supplementary Information

Robust micro-flow LC-MS/MS for proteome analysis – 38,000 runs and counting

Yangyang Bian[1,2], Florian P. Bayer[1], Yun-Chien Chang[1], Chen Meng[3], Stefanie Hoefer[1], Nan Deng[4], Runsheng Zheng[5], Oleksandr Boychenko[5], Bernhard Kuster[1,3*,]

Affiliations

1, Chair of Proteomics and Bioanalytics, Technical University of Munich, Emil Erlenmeyer Forum 5, 85354, Freising, Germany

2, College of Life Science, Northwest University, Taibai North Road 229, Xi'an, Shaanxi, 710069, P.R. China

3, Bavarian Biomolecular Mass Spectrometry Center (BayBioMS), Technical University of Munich, Gregor-Mendel-Straße 4, 85354,Freising, Germany

4, Instrumental analysis center, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an, Shaanxi, 710049 P.R. China

5, Thermo Fisher Scientific, Dornierstraße 4, 82110, Germering, Germany

* Corresponding authors:

Bernhard Kuster (kuster@tum.de)

Tel: +49 8161 715696, Fax: +49 8161 715931

Experimental Section. Detailed description of sample preparation

Figure S1. Column to column variation. (A) Chromatographic peak width (FWHM) distributions of peptides separated on five PepMap C18 columns (2 µm or 3 µm particle size) that were manufactured in different years. (B) Number of unique peptides identified for each column (5 µg protein digest injected; 60 min gradient time).

Figure S2. Pressure curves of two raw mass spectrometry files. The red line represents a sample run immediately before the column partially blocked (red) and the black line shows the separation during which partial blockage was observed (Orbitrap HF-X data).

Figure S3. Number of peptide identifications per minute for Hela peptides separated on the same column in 2018 (red) and 2020 (blue) using gradients of 10 min (A), 15 min (B), 30 min (C), 60 min (D), 90 min (E), and 120 min (F).

Figure S4. Venn diagrams comparing the number of proteins identified from plasma (A) and urine (B) protein digests. Data from this study is shown in light blue and data from the published literature is shown in dark blue. The number of samples and LC gradient times required to generate these results are also given.

Table S1. Summary of raw files generated by two micro-flow LC-MS/MS systems, the corresponding month and column information is annotated.

Table S2. Summary of proteins and peptides identified from plasma and urine samples in the published literatures and this study.

Experimental Sections

Sample preparation

PROCAL peptides were obtained from JPT Peptide Technologies GmbH (Berlin, Germany), and dissolved following the described protocol[1]. HeLa cells were cultured in DMEM (Gibco, Invitrogen), supplemented with 10% fetal bovine serum, 100 U/mL penicillin (Invitrogen), 100 µg/mL streptomycin (Invitrogen), at 37 °C, in a humidified incubator with 5% CO2. Cells were harvested at ~80% confluence by washing twice with PBS buffer and lysed in 8 M urea, 40 mM Tris/HCl (pH 7.6), 1 × EDTA-free protease inhibitor mixture (Complete Mini, Roche), and 1 × phosphatase inhibitor mixture (Sigma-Aldrich). The cell lysate was clarified by centrifuged at 20,000 × g for 20 min. The supernatants were used for in-solution trypsin digestion. The urine and plasma samples were collected and processed as described before[2]. Briefly, plasma was diluted by 5 volumes of 8 M urea buffer containing 80 mM Tris–HCl, pH = 7.6, and stored at −80 °C until further use. Urine was collected and centrifuged at 4000 × g at 4 °C for 30 min to remove cell debris, the supernatant was vacuum concentrated 5-fold using a SpeedVac and proteins were precipitated using five volumes of ice-cold ethanol, and followed by centrifugation at 20,000 × g, 4 °C for 30 min. The protein pellet was dissolved in 8 M urea buffer containing 80 mM Tris-HCl, pH = 7.6, and stored at −80 °C until further use. Protein concentration was determined by the BCA method. Proteins were reduced by 10 mM DTT at 37 °C for 1 h, and alkylated using 55 mM chloroacetamide (CAA) at room temperature for 30 min in the dark. Proteins were digested with sequencing grade trypsin (Roche) at a protease-to-protein ratio of 1:50 (w/w) and incubating overnight at 37 °C. Digestion was quenched by addition of formic acid (FA) to a final concentration of ~1%, and the peptides were desalted on Sep-Pak C18 Cartridges (Waters) and dried in a SpeedVac. Samples were stored at −80 °C until further use.

Off-line high pH reversed phase peptide fractionation

A Dionex Ultra 3000 HPLC system operating a Waters XBridge BEH130 C18 3.5 µm 4.6 × 250 mm column was used to fractionate peptides at a flow rate of 1000 µl/min. Buffer A was 25 mM ammonium bicarbonate (pH = 8.0), buffer C was 100% ultrapure water (ELGA), buffer D was 100% ACN, buffer B was not used in this system. The proportion of buffer A was kept at 10% throughout the separation. 2.0 mg of urine or plasma protein digest was separated by a linear gradient from 4% D to 32% D in 47 min, followed by a linear gradient from 32% D to 80% D in 8 min. 96 fractions were collected with 0.5 min time interval, and pooled into 48 fractions by adding fraction 49 to fraction 1, fraction 50 to fraction 2, and so forth. Finally, all samples were vacuum-dried in a SpeedVac (Thermo Fisher Scientific), and stored at -80 °C until further use.

LC-MS analysis

Samples were analyzed on a micro-flow LC-MS/MS system using a modified Vanquish pump (Thermo Fisher Scientific) coupled to a Q Exactive Orbitrap HF-X or an Orbitrap Funsion Lumos mass spectrometer (both Thermo Fisher Scientific). Chromatographic separation was performed via direct sample injection onto the head of a 15 cm Acclaim PepMap 100 C18 column (2 µm particle size, 1 mm ID, Thermo Fisher Scientific) at a flow rate of 50 µL/min. Solvent A was 0.1% FA, 3% DMSO in water, and solvent B was 0.1% FA, 3% DMSO in ACN. Samples were separated with a linear gradient of 3% to 28% B using different gradient lengths. The HF-X was operated in positive ion mode, using a electrospray spray voltage at 4.0 kV, a funnel RF lens value of 40, capillary temperature of 320 °C, auxillary gas heater temperature of 200 °C. The flow rates for sheath gas, aux gas and sweep gas were set to 35, 5, and 0, respectively. The Orbitrap Fusion Lumos was operated as follows: positive ion mode; electrospray voltage of 3.5 kV, capillary temperature of 325 °C; vaporizer temperature of 125 °C. The flow rates of sheath gas, aux gas and sweep gas were set to 32, 5, and 0, respectively. The detailed mass spectrometer parameters of MS1 and MS2 spectra were adjusted according to the type of sample analyzed. For the deep-proteome analysis of plasma and urine, MS1 resolution

was set to 60,000 at m/z 200 and a AGC target value of 3E6 with a maximum injection time (IT) of 50 ms. The MS1 mass range was set to 360–1300. The AGC target value for fragment spectra was set to 1E5. For MS2 spectra, the minimum AGC target was kept at 2E3. The isolation width was set to 1.3 m/z, and the first mass was fixed at 100 m/z. The normalized collision energy was set to 28%. Peptide match was set to 'preferred', and isotope exclusion was enabled. Up to 10 precursors per cycle were picked for MS2 using a maximum IT of 86 ms and fragments were recorded at 15,000 resolution. MS1 and MS2 spectra were acquired in profile and centroid mode, respectively. The dynamic exclusion value was set to 15 s.

Data processing and analysis

Raw data files were processed with MaxQuant v1.6.2.3[3] using the integrated Andromeda Search engine and searched against the Uniprot human reference database containing 20,230 (downloaded 06.07.2017) canonical entries. Default MaxQuant parameters were used. Trypsin was specified as the enzyme, cleaving after all lysine and arginine residues and allowing up to two missed cleavages. Carbamidomethylation of cysteine was specified as fixed modification and protein N-terminal acetylation and oxidation of methionine were considered as variable modifications. The false discovery rate (FDR) was set to 1% on the peptide-spectrum match (PSM) and protein levels. Data analysis downstream of MaxQuant output results was performed in R[4].

References

1 Zolg, D. P. *et al.* PROCAL: A Set of 40 Peptide Standards for Retention Time Indexing, Column Performance Monitoring, and Collision Energy Calibration. *Proteomics* **17** (2017).

2 Bian, Y. *et al.* Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nat Commun* **11**, 157 (2020).

3 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).

4 Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J Comput Graph Stat* **5**, 299-314 (1996).

Figure S1

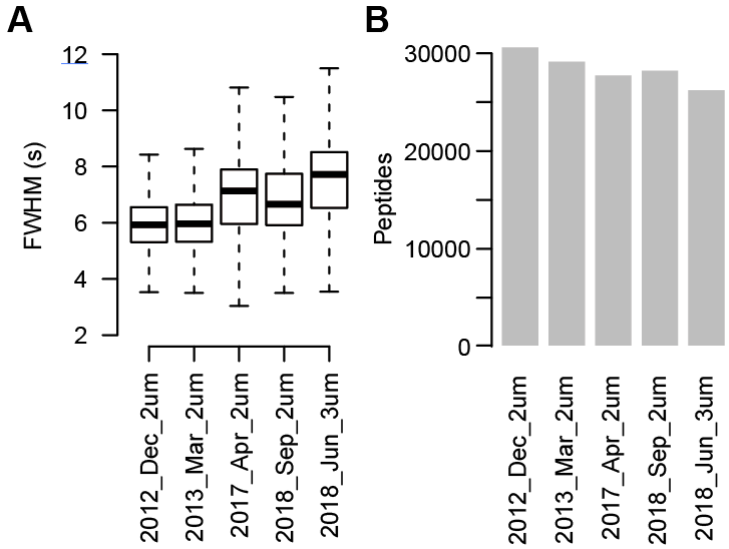

Figure S1: Column to column variation. (A) Chromatographic peak width (FWHM) distributions of peptides separated on five PepMap C18 columns (2 µm or 3 µm particle size) that were manufactured in different years. (B) Number of unique peptides identified for each column (5 µg of protein digest injected; 60 min gradient time).

Figure S2

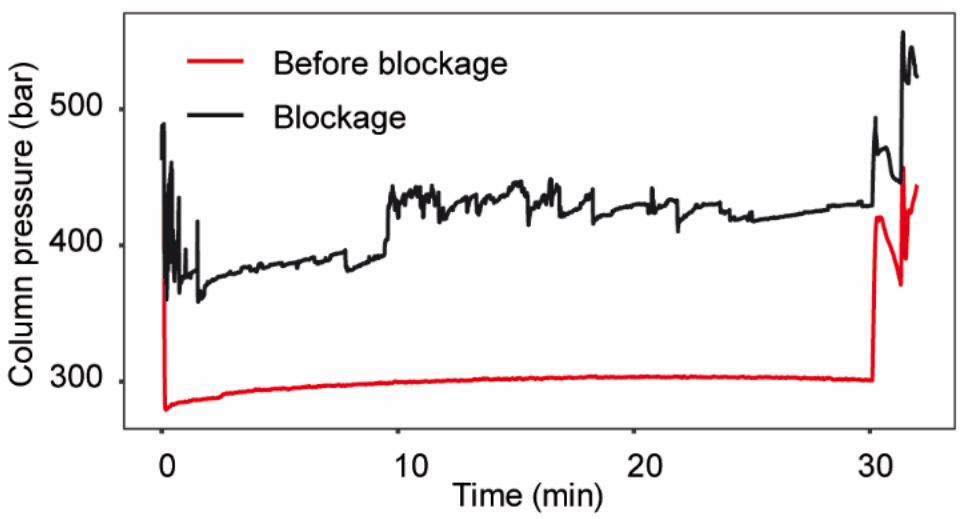

Figure S2. Pressure curves of two raw mass spectrometry files. The red line represents a sample run immediately before the column partially blocked (red) and the black line shows the separation during which partial blockage was observed (Orbitrap HF-X data).

Figure S3

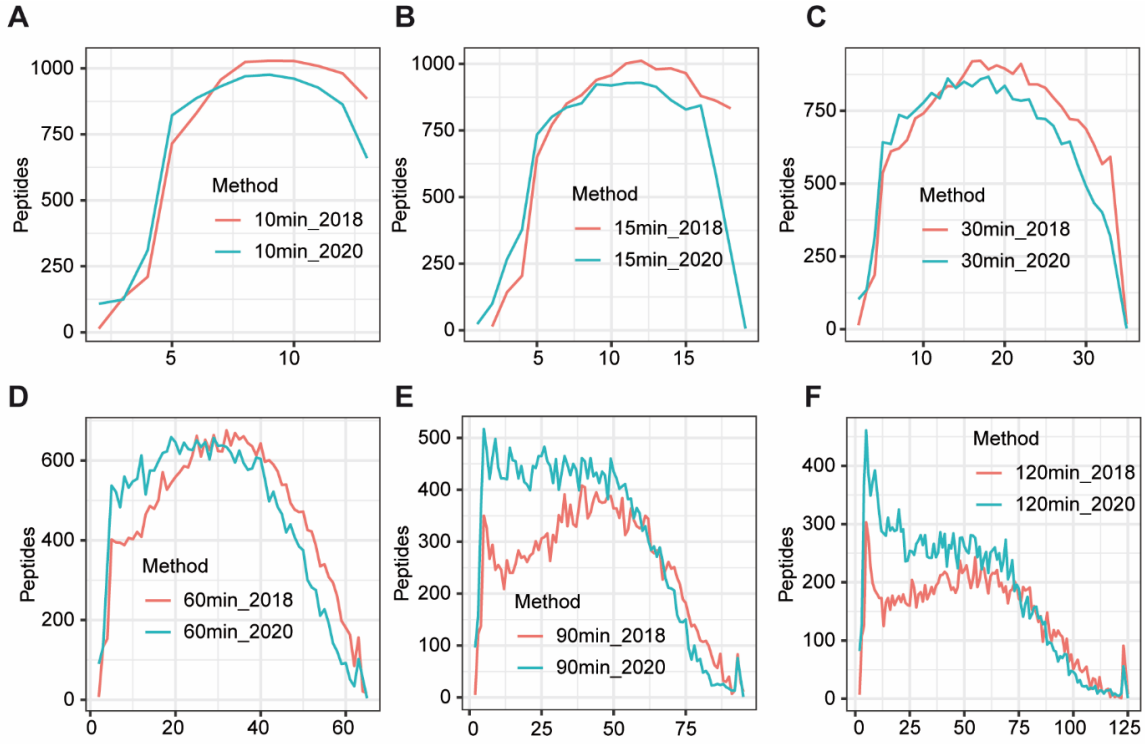

Figure S3. Number of peptide identifications per minute for Hela peptides separated on the same column in 2018 (red) and 2020 (blue) using gradients of 10 min (A), 15 min (B), 30 min (C), 60 min (D), 90 min (E), and 120 min (F).

Figure S4

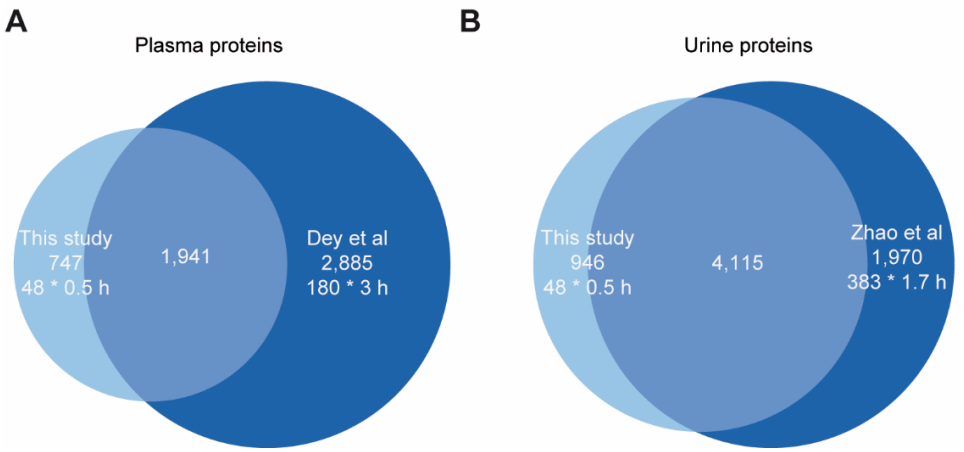

Figure S4. Venn diagrams comparing the number of proteins identified from plasma (A) and urine (B) protein digests. Data from this study is shown in light blue, and data from the published literatures is shown in dark blue. The number of samples and LC gradient times required to generate these results are also given.