Digging deeper into the human proteome: A novel nanoflow LCMS setup using micro pillar array columns (µPAC™)

Jeff Op De Beeck[1], Jarne Pauwels[2,3], An Staes[2,3,4], Natalie Van Landuyt[1], Delphi Van Haver[2,3,4], Wim De Malsche[5], Gert Desmet[5], Andrea Argentini[2,3], Lennart Martens[2,3], Paul Jacobs[1], Francis Impens[2,3,4,$,*] and Kris Gevaert[2,3,$,*]

[1]*PharmaFluidics, Zwijnaarde, Belgium*

[2]*Department of Biomolecular Medicine, Ghent University, B-9000 Ghent, Belgium*

[3]*VIB Center for Medical Biotechnology, B-9000 Ghent, Belgium*

[4]*VIB Proteomics Core, UGent, Ghent, Belgium*

[5]*Department of Chemical Engineering, Vrije Universiteit Brussel, Brussels, Belgium*

[$]*These authors contributed equally*

[*]*To whom correspondence should be addressed: Kris Gevaert, VIB Center for Medical Biotechnology, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium, email: kris.gevaert@vib-ugent.be, and Francis Impens, VIB Proteomics Core, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium, email: francis.impens@vib-ugent.be.*

Abstract

In bottom-up proteomics, capillaries with internal diameters of 50 to 100 µm and packed with sub-2-µm C18-functionalized particles are routinely used in combination with a high resolution mass spectrometer. Typically, microgram amounts of samples are separated in 30 to 240 min long gradients on columns with lengths up to 75 cm. Unlike conventional LC columns, micro pillar array columns (µPAC™) are fabricated using micromachining technology initially developed for the semiconductor industry. Using this advanced technology, perfectly ordered chromatographic separation beds are fabricated by precisely etching micrometer scale channels out of a silicon wafer. As a result of the perfect order that is introduced into the separation bed, analyte dispersion in the column is reduced to an absolute minimum and column permeability is increased by one order of magnitude. This allows using very long columns (up to 200 cm) at a fraction of the pressure needed to operate packed bed alternatives. To validate the µPAC™ column performance for bottom-up proteomics, tryptic digests from HEK293T cell lysates were prepared over a wide concentration range (100 ng/µl to 3 µg/µl) and separated using a 200 cm µPAC™ column as well as a 40 cm long pulled tip packed bed capillary LC column. Using an Orbitrap Elite mass spectrometer, on average 25% more proteins could be identified when working with the µPAC™ column, and this over the entire concentration range. Peak width data showed that the rate at which the peak width increases with gradient time is much lower on the µPAC™ column. For a 10-hour long gradient, average peak widths below 0.5 min were observed, resulting in consistent identification of more than 5,000 proteins. Combining long solvent gradients and this new type of LC column, substantial improvements in plasma proteome coverage were obtained. Data are available via ProteomeXchange with identifiers PXD011547 and PXD011634.

Introduction

Mass spectrometry based proteomics has become an essential tool to provide biologists with accurate information about protein expression levels, posttranslational modifications and interaction partners (1–4). Whereas next generation sequencing offers the possibility to characterize genomes and transcriptomes at rates and costs that allow high throughput screening of a vast number of biological samples, poor correlation with actual protein expression levels that is often observed and the lack of data on posttranslational modifications still urges the use of proteomic profiling to obtain complementary data (5). With the ultimate goal to identify all proteoforms present in a biological sample (6), the practice of MS-based proteomic profiling has evolved enormously during the last decade (7, 8). Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) has been and still is the dominant technique for proteomic profiling and is continuously evolving to retrieve more data within shorter analysis times, while simultaneously decreasing operational complexity and increasing method robustness (9–11). Whereas two decades ago, state-of-the-art MS systems were able to acquire up to one MS/MS scan per second, technological advances in MS instrumentation are nowadays capable of acquiring over 40 MS/MS scans per second (12, 13). This evolution has had an impact on the practice of bottom-up proteomics, where complex protein samples are first extracted from biological tissues or cell populations, subsequently digested with (a) protease(s) and the resulting complex mixture of peptides is separated according to hydrophobicity by reversed phase liquid chromatography. Ionization using electrospray (ESI) ensures these peptides are transferred into the MS instrument, where tandem mass spectrometry analysis is performed. The MS/MS spectra that are generated are then searched against a database for the identification of peptides originating from a certain protein or protein group. It goes without saying that the number of peptides and proteins that can be identified in one LC-MS/MS run depends to a large extent on the scanning speed of the MS instrument, but also on the quality of the

chromatographic separation (14). Especially now MS instruments have become so fast, the main limiting factor appears to be chromatographic performance or resolving power (11, 13).

Chromatographic performance is commonly expressed as peak capacity (n_C) which takes both the average peak width and the elution window into account. Peak capacity is a theoretical measure for the number of peaks that can be separated within a given timeframe and previous studies have pointed out a linear relation to the amount of peptides that can be identified (14–16). With documented values ranging from several hundred up to 1,500, peak capacity mainly depends on the properties of the LC column, but is also affected by the conditions used to perform gradient elution chromatography such as column temperature, flow rate and gradient time (11, 17–19). When focusing on the column properties, an increase in peak capacity can be either achieved by extending the column length or by decreasing the plate height (h). The plate height (or efficiency) of a column depends on the diameter of the particles that are used to pack the column and the packing quality (20). Reducing the particle diameter will have a beneficial effect on chromatographic performance and will allow to obtain higher peak capacities. However, reducing the particle diameter and extending the column length comes at the cost of increased backpressure. The pressure drop of a column is linearly related to its length and inversely proportional to the square of the particle diameter (21). As a consequence, current state-of-the art nano LC columns often require UPLC type of LC instruments which can accurately deliver flow rates of 0.05 to 1 µl/min at operating pressures up to 1,200 bar. Such columns are typically quite long, with lengths up to 50 cm and more, and packed with sub-2-µm C18 functionalized porous silica particles, producing peak capacity values between 600 and 800 for very long gradient separations (> 240 min) (22). Peak capacities ranging from 1,000 to 1,500 have however been obtained when using longer columns or columns packed with even smaller particles, but custom built LC instrumentation that is capable of delivering operating pressures up to 1,360 bar was required (5, 17, 23). Today's commercially available nano LC systems do not yet allow working at pressures exceeding 1,200 bar and therefore the maximum column length used

when working with 2 µm particles is 75 cm, which delivers a peak capacity of about 827 at a gradient duration of 240 min (22). Operating columns at pressures exceeding 1,000 bar is not that straightforward as column fittings and LC instrument parts have to endure much more, hereby possibly compromising the system robustness. A strategy to circumvent the increased backpressure that is observed by extending column lengths beyond 100 cm is the use of monolithic stationary phase support structures as an alternative to silica particles. Both silica and polymer-based monolithic columns have been developed and several of them are commercially available (24). Due to the high permeability that is observed for monolithic column types, the use of columns with lengths exceeding 300 cm has been reported and such columns have been applied successfully to perform deep proteome profiling at operating pressures well below 400 bar (25, 26).

Following up on the pioneering work of the Regnier lab (27, 28), micro pillar array columns (µPAC™s) were introduced about two decades ago as a powerful alternative for classical packed bed columns and monoliths (29–33). By using micro fabrication technology initially developed for the microelectronics industry, the stationary phase backbone of pillar array columns can be precisely defined in high purity silicon wafers. The main advantage of this approach is that such columns can be manufactured with a high degree of uniformity and in a perfectly ordered manner, i.e. the stationary phase morphology (or bed of micro pillars) is perfectly uniform over the entire column width and length. SEM images shown in Figure 1 clearly illustrate the difference between packed bed, monolithic and µPAC™ column technology. Due to the high degree of uniformity, peak dispersion originating from heterogeneous flow paths in the separation bed is eliminated (no Eddy dispersion), and therefore components remain much more concentrated during separation (34, 35). Apart from an improved efficiency, pillar array columns can also be designed with substantially lower flow resistances compared to packed bed columns. When packing spherical particles into a cylindrical column or capillary, the relative volume that is taken by the particles to that of the inter particle void space (called external porosity ε) will always be restricted to a single value

of 0.4. This is the case for all packed bed columns, no matter what particle size or capillary diameter is used. The distance between the pillars can however be independently controlled from the pillar size, enabling the fabrication of columns over a range of external porosities. The higher the external porosity, the lower the flow resistance. Because of this low flow resistance, moderate LC pump pressures (below 350 bar) can be used to operate very long separation channels with lengths exceeding 100 cm. To fabricate very long columns on a silicon wafer with limited dimensions, the separation channel must be folded in a serpentine manner. By interconnecting a total of 40 5-cm long separation channels with low dispersion flow distributor and turn structures, a column with a total length of 200 cm could be fabricated (Figure 2) (36, 37).

To investigate the benefits of the µPAC™ technology for RPLC-MS/MS based proteomics workflows, a series of benchmarking experiments was performed by injecting different amounts of tryptic protein digests of HEK293T cell lysates and separating these samples using an extensive range of solvent gradients. Employing an Orbitrap Elite mass spectrometer for MS/MS detection, we aimed at getting a clear assessment of both the chromatographic performance of this column type as well as its relation to the number of proteins and peptides that can be identified in a typical bottom-up proteomics experiment. In addition, the possible benefits for human plasma proteome profiling were evaluated. For comparison, all experiments were repeated on a 40 cm long nano LC column packed in-house with sub-2-µm C18 functionalized porous silica particles which is considered to be among the top performing nano LC columns for the separation of complex peptide mixtures (8).

Experimental procedures

Cell culture. Human HEK293T cells were purchased from the ATCC (TIB-152, American Type Culture Collection, Manassas, VA, USA) and grown in DMEM medium with Glutamax (Invitrogen, 31966-047) supplemented with 10% dialyzed foetal bovine serum (Gibco) and 1% penicillin-streptomycin (Gibco, 5,000 U/mL). Cell pellets containing 5 million cells were harvested, washed with PBS and frozen at -80°C until further use.

Sample preparation for proteome analysis. HEK293T cell pellets were lysed and digested with trypsin to prepare peptide samples. To this end, single cell pellets were re-suspended in 1 mL lysis buffer (8.0 M urea, 20 mM HEPES, pH 8.0) at room temperature (RT) and sonicated at 10°C in a water bath at high intensity for 10 cycles of 30 s (Diagenode). Cell debris was precipitated by centrifugation (20,000 *x g* for 15 min at RT) and the supernatant was transferred to a fresh Eppendorf tube. After measuring the protein concentration (Bradford assay), cysteine residues were reduced (5 mM dithiothreitol, 30 min at 50°C) and subsequently alkylated (10 mM iodoacetamide, 15 min at room temperature in the dark). Sample were then diluted two-fold using 20 mM HEPES pH 8.0 to a urea concentration of 4 M and from each pellet 500 µg of protein was digested by endo-LysC (Wako, 1:100 enzyme:protein ratio) for 4 h at 37°C. Samples were further diluted using 20 mM HEPES pH 8.0 to a urea concentration of 2 M and digested with trypsin (Promega, 1:100 enzyme:substrate ratio) overnight at 37°C. The next day, digested samples were acidified with TFA to a final concentration of 1% and incubated on ice for 15 min, followed by centrifugation at 1,780 *x g* for 15 min at RT to remove insoluble components. Peptides from each sample were purified using SampliQ C18 columns (Agilent) and eluted in 60% acetonitrile (ACN), 0.1% TFA in water. From each sample, aliquots containing approximately 100 µg of peptide material were vacuum-dried completely and stored dry at -20°C until further use. Prior to LC-MS/MS analysis, dried peptide aliquots were re-suspended in loading solvent (2% acetonitrile (ACN), 0.1% trifluoroacetic acid in water) and pooled to generate a single-vial master mix for each dilution series. The peptide concentration in every master-mix

was measured using a Lunatic instrument (Unchained Labs) with each mix containing around 500 µg of peptides in a volume of 1 mL. From this master-mix, aliquots with the required amount of peptide material in 1 µl were prepared for each dilution series (see experimental design).

Sample preparation for plasma analysis. From a commercially available pool of human plasma (Sigma-Aldrich, P9523), five replicate samples were prepared in parallel. For each replicate sample 1 µl (containing approximately 80 µg of plasma proteins) was suspended in 50 µL 8 M urea in 150 mM Tris-HCl buffer (pH 8) and cysteine residues were reduced (5 mM dithiothreitol, 30 min at 50°C) and subsequently alkylated (10 mM iodoacetamide, 15 min at room temperature in the dark). The samples were then diluted two-fold using 150 mM Tris-HCl buffer pH 8.0 to a urea concentration of 4 M and proteins were digested by endo-LysC (Wako, 1:100 enzyme:substrate ratio) for 4 h at 37°C. The samples were further diluted using 150 mM Tris-HCl buffer pH 8.0 to a urea concentration of 2 M and digested with trypsin (Promega, 1:100 enzyme:substrate ratio) overnight at 37°C. The next day, the digested samples were acidified with TFA to a final concentration of 1% and peptides from each sample were purified using Omix C18 tips (Agilent) and eluted in 60% ACN, 0.1% TFA in water. Purified peptides were vacuum-dried completely and at -20°C until further use.

LC-MS/MS and data analysis. LC-MS/MS analysis of either 100 ng, 500 ng, 1 µg, 2 µg or 3 µg HEK293T peptide material was performed using an Ultimate 3000 nanoRSLC system (Thermo) coupled to an LTQ Orbitrap Elite mass spectrometer (Thermo) equipped with either one of the two column types. Depending on the mounted column, the LC setup was slightly different. The first analytical column was a classical packed bed column (40 cm long, 75 µm inner diameter) packed in-house with ReproSil-Pur basic 1.9 µm silica particles (Dr. Maisch) mounted on a PneuNimbus dual-column nanoESI source (Phoenix S&T) combined with a butterfly portfolio heater (50°C, Phoenix S&T). Prior to packing the column, the fused silica capillary had been equipped with a laser pulled electrospray tip using a P-2000 Laser Based Micropipette Peller (Sutter Instruments). The set-up included a 20 µL loop and all microliter pickup

injections were performed using an in-house prepared trapping column (2 cm, 100 µm inner diameter) packed with ReproSil-Pur basic 5 µm silica particles (Dr. Maisch). Sample loading on the trapping column was performed at a flow rate of 10 µl/min during 4 min (2% ACN, 0.1% trifluoroacetic acid (TFA)) before the column was switched in line with the analytical column. The second column was a 200 cm long micro pillar array column (PharmaFluidics) with C18-endcapped functionality. This column consists of 300 µm wide channels that are filled with 5 µm porous-shell pillars at an inter pillar distance of 2.5 µm. With a depth of 20 µm, this column has a cross section equivalent to an 85 µm inner diameter capillary column. The µPAC column was mounted in the Ultimate 3000's column oven, set at 50°C. For proper ionization, a fused silica PicoTip emitter (10 µm inner diameter) (New Objective) was connected to the µPAC™ outlet union and a grounded connection was provided to this union. For this setup, we provided the system with a 1 µL loop and for all analyses 1 µl was directly injected onto the analytical column using microliter pickup.

Peptides were eluted from the analytical 40 cm packed bed column using 6 different non-linear gradients of 30, 60, 90, 120, 150 or 210 min. The gradients went from 2 to 30% solvent B (80% ACN, 0.1% formic acid (FA)) in respectively 21, 42, 63, 84, 105 and 147 min, to 56% solvent B in respectively, 8, 16, 24, 32, 40 and 56 min, ultimately reaching 99% solvent B in respectively, 1, 2, 3, 4, 5 and 7 min, all at a constant flow rate of 250 nl/min. The column was then washed for 10 min at 99% solvent B, followed by re-equilibration of the system with 98% solvent A (0.1% FA in water) for 20 min. In between samples, a blank was run to reduce sample carry-over. For the analyses on the µPAC™ column, the same gradients were applied using an increased flow rate of 300 nl/min with an additional 30 min at 98% solvent A at the end to ensure proper re-equilibration of the column due to the larger bed volume of the µPAC™ column (total volume of 10 µl). Blank runs were not performed on the µPAC column as previous results showed very little to no memory effect. Three extra-long non-linear gradients of 330, 450 and 570 min were performed additionally for the µPAC™ column going from 2 to 30% solvent B in respectively 231, 260 and 399 min, to 56% solvent B in respectively 88, 120 and 152 min, ultimately reaching 99% solvent B in respectively,

11, 15 and 19 min, all at a constant flow rate of 300 nl/min. As previously described, these runs were followed by a 10 min wash with 99% solvent B and re-equilibration of the column with 98% solvent A for 50 min.

LC separation of the protein digest of human plasma was performed using the same LC setups and the 210 min gradients as described above. In contrast to the previous analyses with the µPAC™ setup, a 5 µL injection loop was installed. Again microliter pickup was used as injection method, but the flow rate was adjusted for the first 8 min to 750 nl/min after which the injection valve was switched to bypass the sample loop and the flow rate was decreased to 300 nl/min.

Mass spectra were acquired on an LTQ Orbitrap Elite mass spectrometer (Thermo Fischer Scientific). The mass spectrometer was operated in data dependent mode, positive ionization mode, automatically switching between MS and MS/MS acquisition for the 20 most abundant peaks in a given MS spectrum. The source voltage was 3.5 kV for the classic analytical column, 2.5 kV for the µPAC column, and the capillary temperature was at 275°C. In the LTQ Orbitrap Elite, full scan MS spectra were acquired in the Orbitrap (m/z 300−2,000, AGC target 3×10^6 ions, maximum ion injection time 100 ms) with a resolution of 60,000 (at 400 m/z). In parallel, the 20 most intense ions fulfilling predefined selection criteria (signal threshold of 500, exclusion of unassigned and 1 positively charged precursors, dynamic exclusion time 20 s) were then isolated in the linear ion trap and fragmented in the high pressure cell of the ion trap (AGC target 5×10^3 ions, maximum ion injection time 20 ms, spectrum data type: centroid). The CID collision energy was set to 35 V and the polydimethylcyclosiloxane background ion at 445.120028 Da was used for internal calibration (lock mass).

Experimental design and statistical rationale. In order to compare the µPAC™ column with the conventional packed bed column, we separated each of the 5 different amounts of peptide material (100 ng, 500 ng, 1 µg, 2 µg or 3 µg) using 6 different gradients (30, 60, 90, 120, 150 and 210 min) in duplicate

on both columns, resulting in a total of 120 LC-MS/MS runs. For each injection on the packed bed column, we calculated the required volume based on the previously measured peptide concentration. The µPAC™ set-up, however, was equipped with a 1 µL injection loop for direct injection and required adjusted peptide concentrations in individual aliquots (i.e. 100 ng/µL, 500 ng/µL, 1 µg/µL, 2 µg/µL and 3 µg/µL). The run order was set so that per column all 6 gradients were run in one block for each of the different amounts of peptide material, starting with injecting the lowest amount (100 ng) and ending with the highest amount (3 µg). The second block of runs on both columns was done with fresh material. The different MS/MS-spectra acquired from the HEK293T proteome digest were divided in 20 batches for database searching with MaxQuant as described below. Each search contained the six different gradients on one type of column for one sample concentration. Three additional searches were performed for the extra-long gradient of 2 µg peptide material. For the comparison of the plasma samples between the different columns, the spectra from five replicates per column were searched in together.

Database search. All database searches were performed with MaxQuant (v1.6.1.0) (38) using the Andromeda search engine with default search settings including a false discovery rate set at 1% on the PSM, peptide and protein level. The mass tolerance for precursor and fragment ions was set to 20 and 4.5 ppm, respectively, during the main search. Enzyme specificity was set to C-terminal of arginine and lysine, also allowing cleavage at Arg/Lys-Pro bonds with a maximum of two missed cleavages. Recorded spectra were searched against the sequences of the human proteins in the Swiss-Prot/UniProtKB database (database release from January 2018 containing 20,231 human protein sequences for the HEK293T digests and database release from March 2018 containing 20,231 human protein sequences for the plasma samples, downloaded from www.uniprot.org). Variable modifications were set to oxidation of methionine residues and acetylation of protein N-termini. For the analysis of the plasma samples we excluded the contaminant list from the search. In all database searches the matching between runs option was turned off.

Data analysis. The comparison between the columns was primarily based on the overall number of peptide and protein identifications, the difference in column pressure, evolution of peak width and overall sensitivity. The results from the database search were further processed in Perseus (v1.6.1.3) (39). For the shotgun samples derived from HEK293T cells, we first removed the reversed hits, potential contaminants and proteins that were only identified by site. The intensities were then log2 transformed and a Venn diagram was created, returning the number of identifications per run. For the plasma samples, the same steps were taken, except for the removal of potential contaminants. Peak widths from several runs were extracted by the mOFF algorithm from each raw file and corresponding evidence.txt file. The parameters were set to 10 ppm peptide mass tolerance, 1 min retention time window for the XIC (default set to 3 min) and 0.1 min retention time window for the peak. Extracted peak widths at FWHM were converted to their corresponding value at 4σ using the relation $FWHM = 2\sqrt{2\sigma^2 ln2}$ and the average peak width of all identified peptides was calculated. Peak capacity values were calculated using the equation $n_c = (T_G/\overline{w}_{4\sigma}) + 1$ with n_c the peak capacity, T_G the solvent gradient time and $\overline{w}_{4\sigma}$ the average peak width at 4σ.

Note that all data are available via ProteomeXchange with identifiers PXD011547 (username: reviewer98026@ebi.ac.uk; password: 9ZGfEMs2) and PXD011634 (username: reviewer89029@ebi.ac.uk; password: ine13KvP).

Results & Discussion

In this study we aimed to investigate the possible benefits µPAC™ column technology can offer for bottom-up proteomics studies. A range of different amounts of HEK293T tryptic digest samples (0.1-0.5-1-2-3 µg) was injected in duplicate for LC-MS/MS analysis on a Thermo Orbitrap Elite mass spectrometer and separated using solvent gradient times ranging from 30 to 570 min. In this way, we intend to get a clear view of the chromatographic performance of µPAC™ columns and their effect on the number of peptide and protein identifications. When comparing the 200 cm µPAC™ column to a classical packed bed column with a length of 40 cm, the first observation we made was the low column pressure needed to operate this new type of column. Whereas classical nano LC columns packed with 1.9 µm porous particles easily generate pressure drops exceeding 400 bars, only 75 bar is needed to operate the µPAC™ column at a flow rate of 300 nl/min (Figure 3). This low back pressure can be attributed to the high permeability that is inherent to the pillar array column format and that is also observed for several types of monolithic columns (24–26) and columns packed with porous-shell beads (40).

Peak dispersion and peak capacity in perfectly ordered chromatographic media. In contrast to monolithic columns, the pillar array column format has the additional advantage that perfect order is introduced into the chromatographic bed, hereby reducing peak dispersion to an absolute minimum (34). The increased base peak intensity observed for the separation of 500 ng of HEK293T tryptic protein digest (Figure 3) already suggests that peptides have sharper elution peaks from the µPAC™ column as compared to the 40 cm packed bed nano LC column. The sharper peptides elute from the LC column, the more concentrated these peptides enter the MS and thus the higher the MS detector response and the better the signal to noise. In addition, sharper peptide peaks reduce peptide co-elution and result in decreased ionization suppression and hence enhanced MS sensitivity. This is expected to lead to a higher number and higher quality of recorded MS/MS spectra with a concomitant increase in peptide and protein identifications.

The average peak width can easily be monitored throughout an entire LC-MS/MS experiment and is often used to compare the chromatographic performance of different experimental set-ups or column types (14). We extracted the peak width for all identified peptides using the moFF algorithm (41). A comparison of the peptide peak widths for both column types is shown in Figure 4A, where peptide peak widths for all sample loads are plotted as a function of the gradient time that was applied. Even though the average peak width obtained for both column types is quite similar when short solvent gradients (30 to 60 min) were used, with average peak widths ranging from 0.22 to 0.27 min, the µPAC™ column showed more narrow peaks when extending gradient times beyond 120 min. Indeed, for a 210 min gradient separation, the average peak width obtained for the packed bed column was 0.40 min compared to 0.31 min obtained on the µPAC™ column. The rate at which the peptide peak width increased with gradient time is also much higher for the packed bed column. According to the linear relationships that were found between gradient time and average peak width for both column types ($y=0.0008x+0.2317$ with R^2 0.97, versus $y=0.0003x+0.2404$ with R^2 0.99), this rate was more than two times higher (Figure 4B). By extending gradient times beyond 210 min, the relative difference in peak width between the two column types was expected to increase. To highlight this, another 120, 240 and 360 min have been added to the gradient duration, resulting in total gradient durations of 330, 450 and 570 min. These very long gradients have only been applied for the separation of 2 µg tryptic digest on the µPAC™ column, resulting in an average peak width of 0.43 min for a 570 min gradient separation. Remarkably, a 450 min gradient separation performed on the µPAC™ column resulted in an average peak width that is comparable to what was achieved with a 210 min gradient on the 40 cm nano LC column. The improved separation performance or resolving power was even more apparent when peak capacities for both column types were compared (Figure 4C). Taking both the elution window and the average peak width for all identified peptides into account, a 28% increase in peak capacity for the µPAC™ column (668 versus 521) was obtained when

applying a 210 min solvent gradient. Extending the gradient time to 570 min resulted in an exceptional peak capacity value of 1,331 for the µPAC™ column.

Improved proteome coverage. We wondered if the increase in peak capacity translated into higher numbers of identified peptides. The average numbers of unique peptides (n=2) that were identified for the different sample loads (0.1-0.5-1-2-3 µg) and gradient durations (30-60-90-120-150-210-330-450-570 min) are shown in Figure 5A. For all sample loads, consistently more peptides were identified when working with the µPAC™ column. The relative increase in peptide identifications gradually rised from 30% for short gradients to more than 50% for longer (210 min) gradients, obtaining an averaged total of 27,140 unique peptides with the µPAC™ column compared to 17,665 with the 40 cm packed bed nano LC column. The data shown in Figure 5A clearly indicate a relation between the absolute number of unique peptides that was identified and the amount of sample that was injected. For both column types, the optimal sample load in terms of peptide identifications was reached when injecting 1 µg or more. Increasing the sample load to 2 or 3 µg did not result in a noticeable increase in identifications, nor did this have an unfavorable effect on the chromatographic performance as can be observed in Figure 4A. Even though the absolute number of peptide identifications decreased when the sample load was reduced below 1 µg, decreasing the sample load definitely had a positive effect on the relative gain that could be achieved with the µPAC™ column. At the lowest injected amount of 0.1 µg, nearly double (198%) the number of peptides could be identified using a 210 min solvent gradient, implying that the improved chromatographic separation enhances the MS sensitivity as narrower peaks tend to be more intense. For this sample load, increasing the gradient time beyond 210 min was not believed to yield a substantial gain in identifications, as additional dispersion would cause the concentration of low abundant peptides to drop below the limit of detection of the MS. However, even when such low amounts were used, a considerable gain in identifications could still be achieved on the µPAC™ column when increasing the gradient time. This better performance with low amounts of sample makes it possible to analyze certain types of samples by LC-

MS/MS despite their low protein content, which is typically encountered when working with a limited number of cells or a single cell as starting material. When separating 2 µg of HEK293T tryptic digest sample using a 570 min solvent gradient, an average of 42,346 peptides were identified which correspond to 5,004 protein groups (Figure 5B). In line with the observations on the peptide level, considerable improvements of the proteome coverage were achieved with the µPAC™ column. On average 20% more protein groups were identified compared to the packed bed nano LC column (Figure 5B). Again, the difference was higher when injecting lower sample amounts. An increase in protein group identifications up to 30% could be achieved when injecting 0.1 µg of tryptic digest sample. To clarify this even further, the percentage of MS/MS spectra that were identified by MaxQuant was compared for both columns (reported in the MaxQuant summary file). An average increase of 8% in ID rate was observed when comparing the 0.1 µg HEK293T sample injections on the µPAC™ to those on the 40 cm packed bed column, resulting in more PSMs.

Single-shot versus replicate analyses. The percentage of identified peptides that was shared between duplicate runs was substantially higher on the µPAC™ column, with respective 77% and 68% overlapping peptides for the µPAC™ and the packed bed nano LC column (Figure 6). Combining duplicate runs, a total number of 33,501 versus 23,471 unique peptides were identified for a 210 min gradient separation. As the overlap was systematically lower on the packed bed nano LC column, relatively higher profits can be made with repetitive analysis of an identical sample and combining the results of a µPAC™ column. In absolute numbers however, the µPAC™ column still results in a net increase in unique peptide identifications of over 40% compared to the packed bed column. Moreover, 80% of the peptides that were identified by combining packed bed nano LC results were identified when working with the µPAC™ column. In terms of protein group identifications, an even higher overlap is observed between duplicate runs, likely as additional peptides often belong to protein groups that have already been identified. 87% overlapping peptides were observed for the packed bed column, compared to 90% for the µPAC™ column.

With a combined total of 4,048 protein groups on the µPAC™ column, this is a net increase of 16% when compared to the packed bed column. These values have been calculated for a head on comparison, where results of duplicate runs have been pooled for both columns. Comparing a single shot analysis on the µPAC™ column to the combined result of duplicates on the packed bed column, the µPAC™ column still outperforms the classical packed bed column with 15% more peptides and 6% more protein groups identified in half the analysis time. Combining the results of replicate runs is a conventional strategy to increase proteome coverage (11), and the benefits of running several shorter gradients versus a single long gradient separation have been a topic of discussion and research in the field of proteomics (14, 42). We already discussed the increased output that was generated by performing a single shot analysis on the µPAC™ column compared to duplicate runs on the packed bed nano LC column. But how about combining replicate runs when working with the µPAC™ column? What strategy is advised when very deep proteome coverage is desired and µPAC™ columns are used? To investigate this, we compared the output of several duplicate runs to the outcome of a single shot analysis (Figure 7). When combining two 210 min gradient separations, a total of 33,318 unique peptides and 4,292 protein groups were identified. By injecting the same sample and performing a single shot separation with a gradient duration of 330 min (which equals ¾ of the total analysis time) substantially more identifications were obtained. With an average of 34,834 unique peptides and 4,380 protein groups, extending the gradient time clearly is more efficient than running duplicates. These results become even more striking when the combined runs are compared to a single 450 min gradient separation. On average, 40,035 unique peptides and 4,814 protein groups could be identified in a single shot analysis. Increasing the gradient time is a valid strategy if deeper proteome coverage is desired. However, this will only make sense if considerable gain in peak capacity is obtained by doing so. When working with the µPAC™ column, extending gradient times beyond 300 min definitely has the potential to identify unique peptides that are not likely to be identified by combining

several shorter gradient runs, as a substantial gain in peak capacity can still be made in this gradient time regime (Figure 4C).

Human plasma proteome profiling. To further validate the results obtained with our benchmarking experiment, we performed another experiment were the goal was to achieve deep coverage of a non-depleted human plasma proteome sample. From a commercially available pool of human reference plasma, a set of five plasma proteomics samples was prepared. Each sample was separated using a 210 min solvent gradient, and this on both column types. Representative base peak chromatograms obtained for human plasma samples are shown in Figure 8. The number of unique peptides and protein groups that were identified in each run is given in Figure 9. Again, the µPAC™ column outperformed the classical packed bed nano LC column over the entire sample range. On average 45% more unique peptides and 35% more protein groups were identified when working with the µPAC™ column. By combining the results of all 5 samples, 252 protein groups were identified using the µPAC™ column. This equals a net gain of 38% when comparing the µPAC™ column to the packed bed nano LC column, which delivered a total of 182 protein groups for the same sample set.

Conclusions and outlook

Different approaches can be used to maximize the output of shotgun proteome profiling experiments (4, 5, 8, 10, 42). Compared to the labor intensive strategy of performing multidimensional separation prior to MS analysis, improving the chromatographic performance during standard LC-MS/MS analysis is substantially easier, more robust and can even be more sensitive when adequate LC-MS/MS conditions are used. As MS scanning speed is increasing progressively, the impact of the chromatographic separation quality on the depth of proteome profiling is becoming more significant. Nano LC columns are available in a wide range of formats and chemistries and are widely used among proteomics laboratories. The search for better chromatography has eventually lead to the use of relatively long columns (40-75 cm) that are packed with ≤ 2 µm silica particles. With peak capacity values reaching ~1000, very deep proteome coverage has already been documented (11). Further increase in chromatographic performance is however not that trivial as an increase in column length or a decrease of the packing particle diameter will have a significant impact on the LC pump pressure needed to operate these columns.

Micro pillar array columns (µPAC™) are a novel column type that show great potential in overcoming these limitations. Micro fabrication technology allows precise control over the column geometry, enabling the fabrication of very long columns with exceptional uniformity within the separation bed or stationary phase. On top of offering unrivaled column-to-column reproducibility, very high peak capacities can be achieved at LC pump pressures below 100 bar. We investigated how these new column types can be implemented for the analysis of complex human proteomes. When comparing a 200 cm long µPAC™ column to a packed bed nano LC column with a length of 40 cm, on average 30 to 50% more unique peptides and 20% more protein groups could be identified when injecting a 0.1-3 µg of tryptic digest samples obtained from HEK293T cell lysates. The µPAC™ column proved to yield an even bigger gain in identifications when limited sample amounts were injected. When reducing the sample load to 100 ng,

nearly double the amount of peptides and 30% more proteins could be identified in a single 210 min gradient separation. This increase in identifications is in line with the observed chromatographic performance for both columns. The rate at which peak width increases according to the gradient time was found to be more than two times lower. For a 210 min gradient, this resulted in an average peptide peak width of 0.31 min for the µPAC™ column compared to 0.40 min for the packed bed column. Extending the gradient time even further to 570 min, an average peptide peak width of 0.43 min was achieved, producing an exceptional peak capacity of 1,331 at operating pressures 6 times lower as observed on a 40 cm packed bed nano LC column. A significantly higher overlap between duplicate runs was also observed, which can most likely be attributed to an increase in run to run reproducibility. Combining several replicate runs to increase proteome coverage is a common strategy in bottom-up proteomics, however data generated in this study suggest that performing a single shot analysis on the µPAC™ column at gradient times exceeding 300 min offers deeper and more sensitive proteome coverage.

A similar outcome was observed for the analysis of human plasma proteomes, where on average 45% more unique peptides and 35% more protein groups were identified. Without any prior depletion, 252 protein groups could be identified when working with the µPAC™ column, whereas only 182 protein groups could be identified when separating the plasma samples using a classical packed bed nano LC column.

Future increases in mass spectrometric resolution and detection sensitivity will indisputably allow accomplishing even deeper proteome coverage with single run-analyses.

Acknowledgements

The authors would like to acknowledge the VIB Tech Watch Fund for supporting early access to innovative technology platforms, L.M. acknowledges support by the Research Foundation − Flanders (FWO) (Project No. G.0425.18N) and F.I. acknowledges support by FWO project S006617. The authors state that they have potential conflicts of interest regarding this work: JODB and NVL are employees of PharmaFluidics and PJ is one of the co-founders of PharmaFluidics.

References

1. Aebersold, R., and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355

2. Zhang, Y., Bryan R. Fonslow, Shan, B., Baek, M.-C., and Yates, J. R. (2014) Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* 113, 2343–2394

3. Walther, T. C., and Mann, M. (2010) Mass spectrometry-based proteomics in cell biology. *J. Cell Biol.* 190, 491–500

4. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2014) The One Hour Yeast Proteome. *Mol. Cell. Proteomics* 13, 339–347

5. Wang, H., Yang, Y., Li, Y., Bai, B., Wang, X., Tan, H., Liu, T., Beach, T. G., Peng, J., and Wu, Z. (2015) Systematic Optimization of Long Gradient Chromatography Mass Spectrometry for Deep Analysis of Brain Proteome. *J. Proteome Res.* 14, 829–838

6. Smith, L. M., and Kelleher, N. L. (2013) Proteoform: A single term describing protein complexity. *Nat. Methods* 10, 186–187

7. De Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Fröhlich, F., Walther, T. C., and Mann, M. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455, 1251–1254

8. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Cox, J., Mann, M., and Ms-based, I. (2011) Deep and Highly Sensitive Proteome Coverage by LC-MS / MS Without Prefractionation * □. *Mol. Cell. Proteomics*, 1–9

9. Riley, N. M., Hebert, A. S., and Coon, J. J. (2016) Previews Proteomics Moves into the Fast Lane.

Cell Syst. 2, 142–143

10. Anagnostopoulos, A. K., Stravopodis, D. J., and Tsangaris, G. T. (2017) Yield of 6,000 proteins by 1D nLC–MS/MS without pre-fractionation. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1047, 92–96

11. Shishkova, E., Hebert, A. S., and Coon, J. J. (2016) Commentary Now , More Than Ever , Proteomics Needs Better Chromatography. *Cell Syst.* 3, 321–324

12. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R. (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17, 676–682

13. Scheltema, R. A., Hauschild, J.-P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., and Mann, M. (2014) The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol. Cell. Proteomics* 13, 3698–3708

14. Köcher, T., Swart, R., and Mechtler, K. (2011) Reveals a Linear Relation between Peak Capacity and Number. *Anal. Chem.*, 2699–2704

15. Giddings, J. C. (1991) *Unified separation science*

16. Neue, U. D. (2008) Peak capacity in unidimensional chromatography. *J. Chromatogr. A* 1184, 107–130

17. Shen, Y., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., Zhao, R., Livesay, E. A., Udseth, H. R., and Smith, R. D. (2005) Automated 20 kpsi RPLC-MS and MS / MS with Chromatographic Peak Capacities of 1000 - 1500 and Capabilities in Proteomics and Metabolomics. *Anal. Chem.* 77, 3090–3100

18. Wang, X., Barber, W. E., and Carr, P. W. (2006) A practical approach to maximizing peak capacity by using long columns packed with pellicular stationary phases for proteomic research. *J. Chromatogr. A* 1107, 139–151

19. Hsieh, E. J., Bereman, M. S., Durand, S., Valaskovic, G. A., and Maccoss, M. J. (2013) Effects of Column and Gradient Lengths on Peak Capacity and Peptide Identification in Nanoflow LC-MS / MS of Complex Proteomic Samples. *J. Am. Soc. Mass Spectrom.* 24, 148–153

20. Macnair, J. E., Patel, K. D., Jorgenson, J. W., Hill, C., and Carolina, N. (1999) Ultrahigh-Pressure Reversed-Phase Capillary Liquid Chromatography : Isocratic and Gradient Elution Using Columns Packed with 1 . 0- µ m Particles. *Anal. Chem.* 71, 700–708

21. Majors, R. E. (2007) Column Pressure Considerations in Analytical HPLC. *LCGC North Am.* 25, 1074–1092

22. Lopez-Ferrer, D., Blank, M., Meding, S., Paulus, A., Huguet, R., Swart, R., and Huhmer, A. F. . (2016) Increased MS protein identification rates using 75 cm long nano LC C18 separation columns: Pushing the limits of Bottom-up Proteomics. *Thermo Fischer Sci. poster note* 64629,

23. Zhou, F., Lu, Y., Ficarro, S. B., Webber, J. T., and Marto, J. A. (2012) A Nano flow Low Pressure High Peak Capacity Single Dimension LC-MS/MS PLatform for High througput in Depth Analysis of Mammalian Proteomes. *Anal. Chem.* 84, 5133–5139

24. Josic, D., and Clifton, J. G. (2007) Use of monolithic supports in proteomics technology. *J. Chromatogr. A* 1144, 2–13

25. Iwasaki, M., Miwa, S., Ikegami, T., Tomita, M., Tanaka, N., and Ishihama, Y. (2010) One-Dimensional Capillary Liquid Chromatographic Separation Coupled with Tandem Mass Spectrometry Unveils the Escherichia coli Proteome on a Microarray Scale. *Anal. Chem.* 82, 2616–

2620

26. Iwasaki, M., Sugiyama, N., Tanaka, N., and Ishihama, Y. (2012) Human proteome analysis by using reversed phase monolithic silica capillary columns with enhanced sensitivity. *J. Chromatogr. A* 1228, 292–297

27. He, B., Tait, N., and Regnier, F. (1998) Fabrication of Nanocolumns for Liquid Chromatography. *Anal. Chem.* 70, 3790–3797

28. Slentz, B. E., and Regnier, F. (2002) Geometric effects of collocated monolithic support structures on separation performance in microfabricated systems. *J. Sep. Sci.*, 1011–1018

29. Malsche, W. De, Eghbali, H., Clicq, D., Vangelooven, J., Gardeniers, H., and Desmet, G. (2007) Pressure-Driven Reverse-Phase Liquid Chromatography Separations in Ordered Nonporous Pillar Array Columns. *Anal. Chem.* 79, 5915–5926

30. Mogensen, K. B., Gustafsson, O., Nikolajsen, R. P. H., and Kutter, J. P. (2004) Pure-silica optical waveguides , fiber couplers , and high-aspect ratio submicrometer channels for electrokinetic separation devices. *Electrophoresis* 25, 3788–3795

31. Illa, X., Malsche, W. De, Gardeniers, H., Desmet, G., and Romano-rodríguez, A. (2010) Experimental study of the depth influence on the band broadening effect in a cyclo-olefin polymer column containing an array of ordered pillars. *J. Chromatogr. A* 1217, 5817–5821

32. Aoyama, C., Saeki, A., Noguchi, M., Shirasaki, Y., Shoji, S., Funatsu, T., Mizuno, J., and Tsunoda, M. (2010) Use of Folded Micromachined Pillar Array Column with Low-Dispersion Turns for Pressure-Driven Liquid Chromatography. *Anal. Chem.* 82, 1420–1426

33. Taylor, L. C., Lavrik, N. V, and Sepaniak, M. J. (2010) High-Aspect-Ratio , Silicon Oxide-Enclosed Pillar Structures in Microfluidic Liquid Chromatography. *Anal. Chem.* 82, 9549–9556

34. Gzil, P., Vervoort, N., Baron, G. V, and Desmet, G. (2003) Advantages of Perfectly Ordered 2-D Porous Pillar Arrays over Packed Bed Columns for LC Separations : A Theoretical Analysis. *Anal. Chem.* 75, 6244–6250

35. Malsche, W. De, Gardeniers, H., and Desmet, G. (2008) Experimental Study of Porous Silicon Shell Pillars under Retentive Conditions. *Anal. Chem.* 80, 5391–5400

36. De Malsche, W., Op de Beeck, J., De Bruyne, S., Gardeniers, H., and Desmet, G. (2012) Realization of 1 × 10 6 Theoretical Plates in Liquid Chromatography Using Very Long Pillar Array Columns. *Anal. Chem.*,

37. Callewaert, M., Op de Beeck, J., Maeno, K., Sukas, S., Thienpont, H., Ottevaere, H., Gardeniers, H., and De Malsche, W. (2014) Integration of uniform porous shell layers in very long pillar array columns using electrochemical anodization for liquid chromatography. *Analyst*, 618–625

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

39. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13, 731–740

40. Hayes, R., Ahmed, A., Edge, T., and Zhang, H. (2014) Core – shell particles : Preparation , fundamentals and applications in high performance liquid chromatography. *J. Chromatogr. A* 1357, 36–52

41. Argentini, A., Goeminne, L. J. E., Verheggen, K., Hulstaert, N., Staes, A., Clement, L., and Martens, L. (2016) MoFF: A robust and automated approach to extract peptide ion intensities. *Nat.*

Methods 13, 964–966

42. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R. A. (2013) Rapid and Deep Human Proteome Analysis by Single-dimension Shotgun Proteomics. *Mol. Cell. Proteomics* 12, 3330–3338

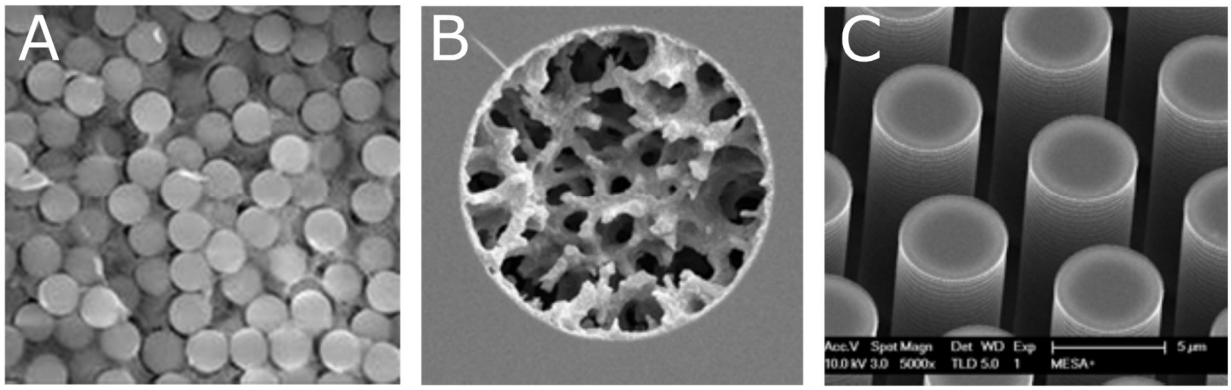

FIG. 1. **Scanning electron microscopy (SEM) images showing different stationary phase support types used for the analysis of complex peptide mixtures by reversed phase liquid chromatography**. A, 1.9 µm diameter silica particles, B, transverse section of a silica monolith column with an internal diameter of 25 µm, C, top view of a micro pillar array bed (5 µm diameter). Images A and B originating from G. Desmet, unpublished results.

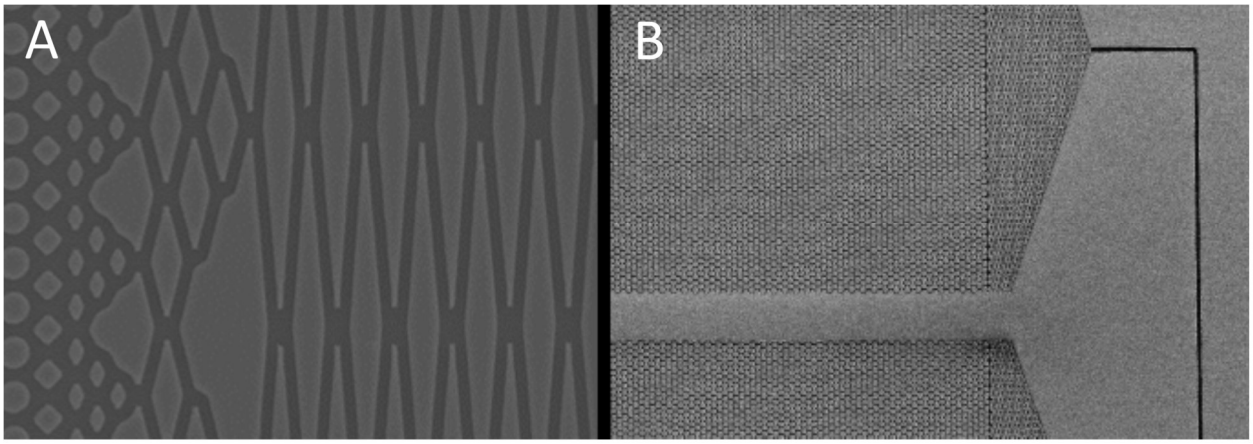

FIG. 2 **Scanning electron microscopy (SEM) images showing flow distributor structures that are used to interconnect parallel separation lanes of the micro pillar array column.** A, Top view of the proprietary flow distribution design. B, Two of these structures are used to interconnect parallel separation lanes with a narrow 180° turn channel.

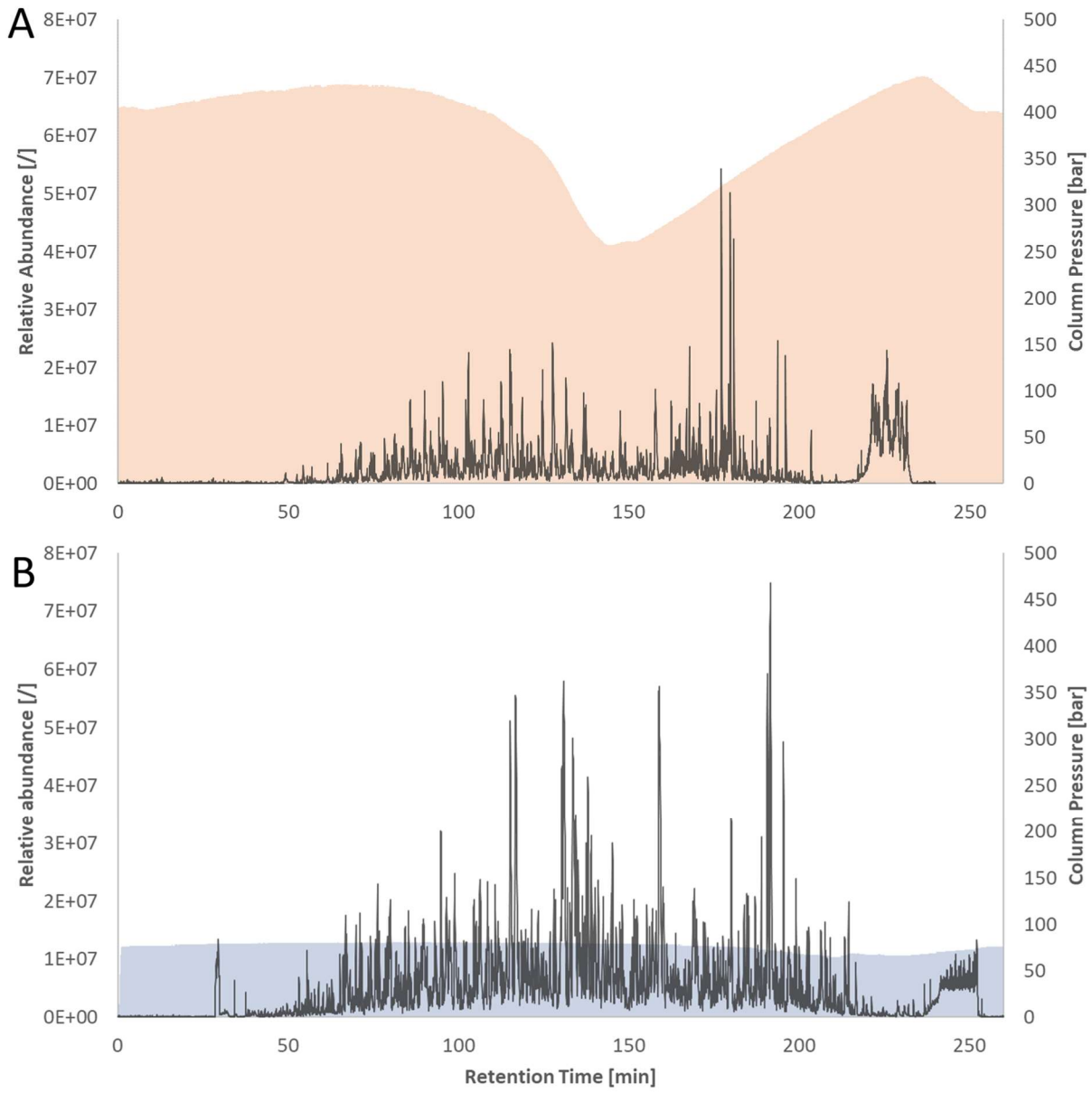

FIG. 3. **Base peak chromatograms obtained by analysis of 500 ng HEK293T cell lysate digest with a 4 hour LC-MS/MS gradient.** A, (orange) obtained on a 40 cm packed bed nano LC column that was prepared in house. B, (blue) obtained on a 200 cm PharmaFluidics µPAC™ column). Corresponding column pressures are plotted on the secondary X-axis.

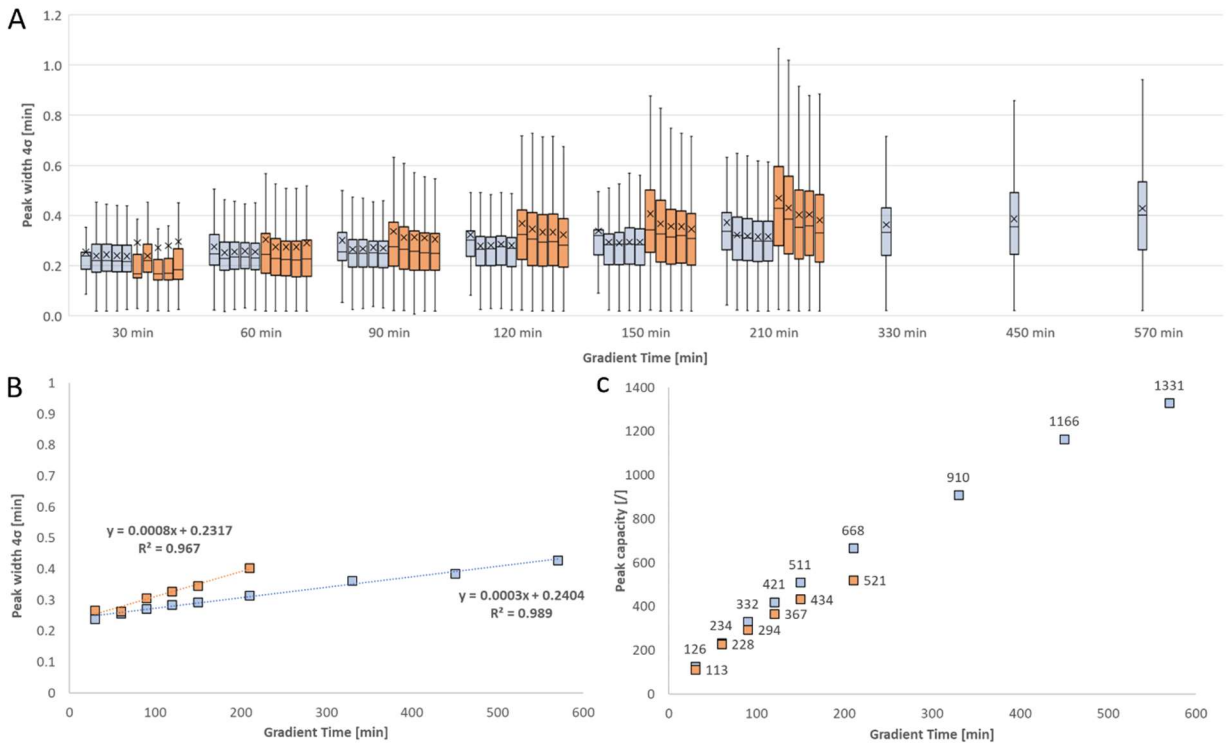

FIG. 4 **Chromatographic metrics calculated for the separation of HEK293T cell lysate digests on a µPAC™ (blue) and a packed bed nano LC column (orange).** A, Box plots showing the peak width (4σ, calculated from FWHM values) for identified peptides. Results are shown for the different gradient lengths and for all injected amounts. From left to right: 0.1-0.5-1-2-3 µg of HEK293T digest. Average values have been marked with X. B, Average peak widths (4σ, calculated from FWHM values) plotted as a function of gradient time. C, Peak capacity values plotted as a function of gradient time for the separation of 2 µg HEK293T digest.

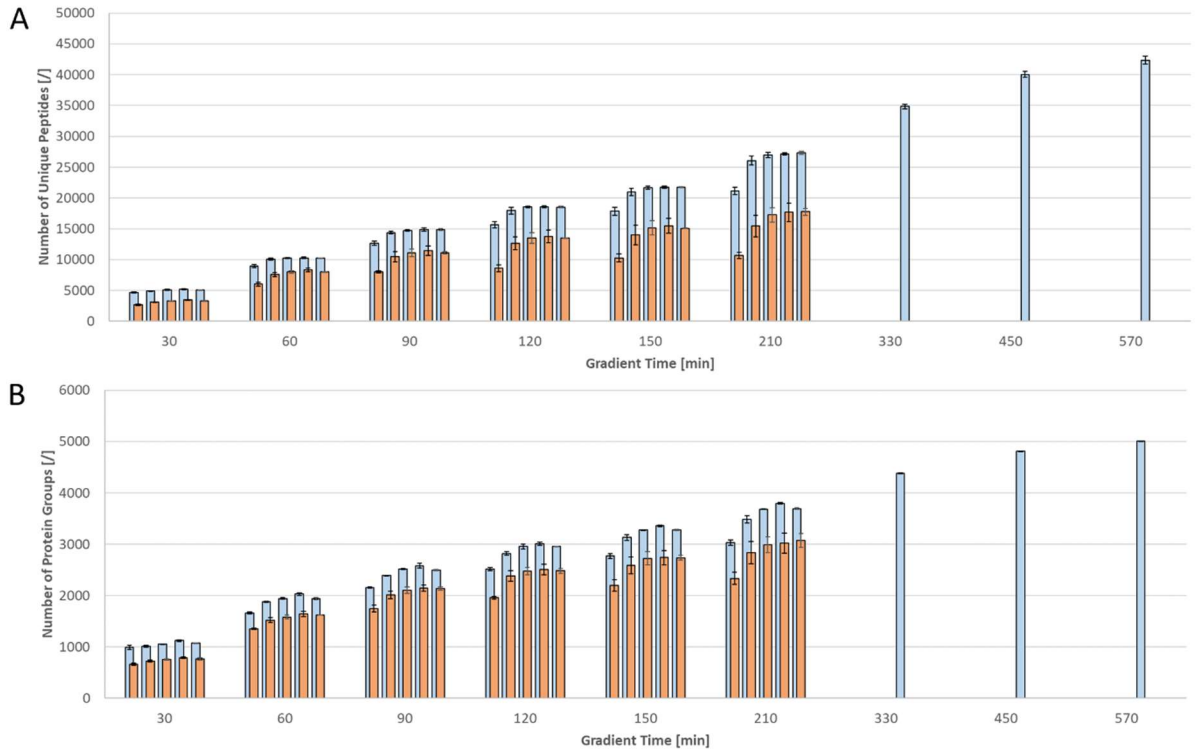

FIG. 5. **Number of peptide and protein identifications achieved for the separation of HEK293T digest samples on a µPAC™ (blue) and a packed bed nano LC column (orange).** A, Number of unique peptides plotted against gradient time, from left to right: 0.1-0.5-1-2-3 µg HEK293T digest. B, Number of protein groups plotted against gradient time, from left to right: 0.1-0.5-1-2-3 µg HEK293T digest.

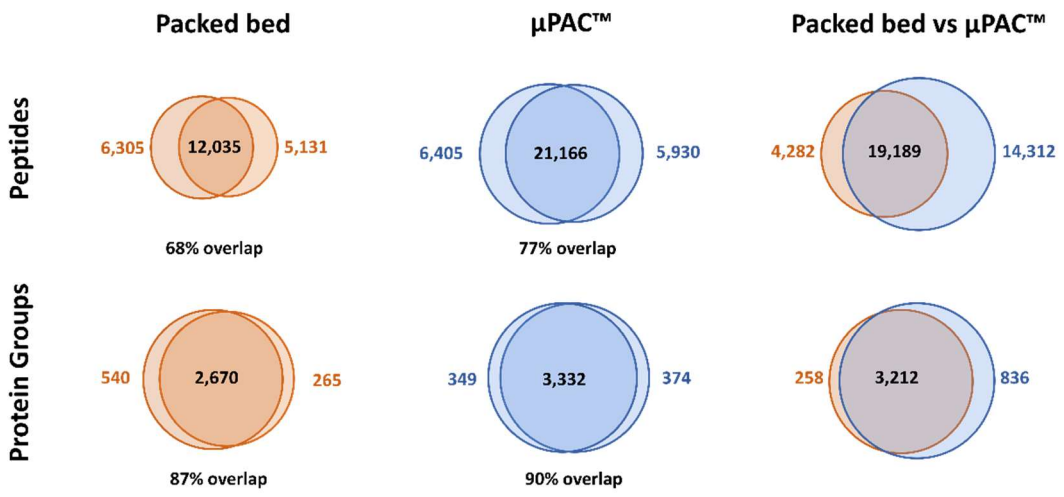

FIG. 6. **Venn diagrams showing the overlap in peptide and protein identifications between duplicate runs and between both column types when duplicate runs were combined.** 3 µg of HEK293T digest was injected on both columns and separated using a 210 min solvent gradient. µPAC™ (blue), packed bed nano LC column (orange).

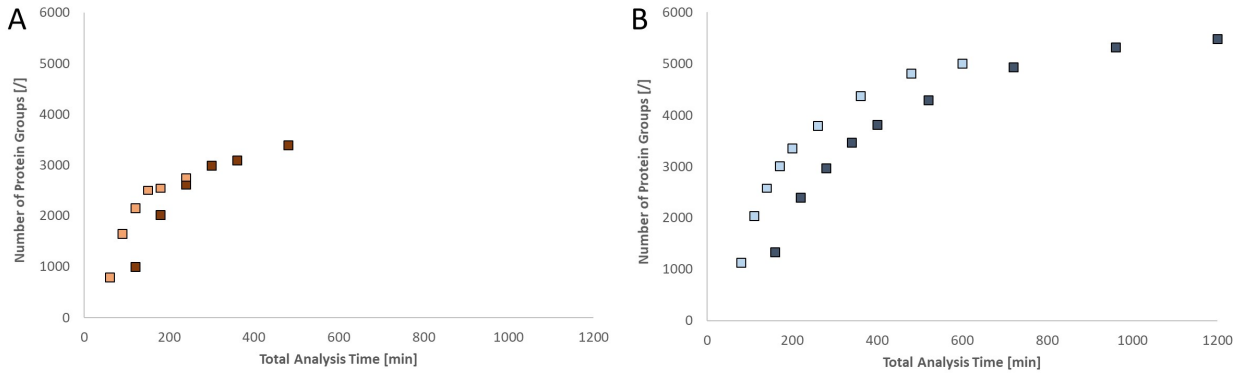

FIG. 7. **Comparison of the number of protein groups that were identified within a given time frame by running single shot analyses (light markers) and by combining the results of duplicate runs (dark markers).** A, results obtained for the analysis of 2 µg HEK293T digest on a 40 cm packed bed column. B, results obtained for the analysis of 2 µg HEK293T digest on a µPAC™ column.

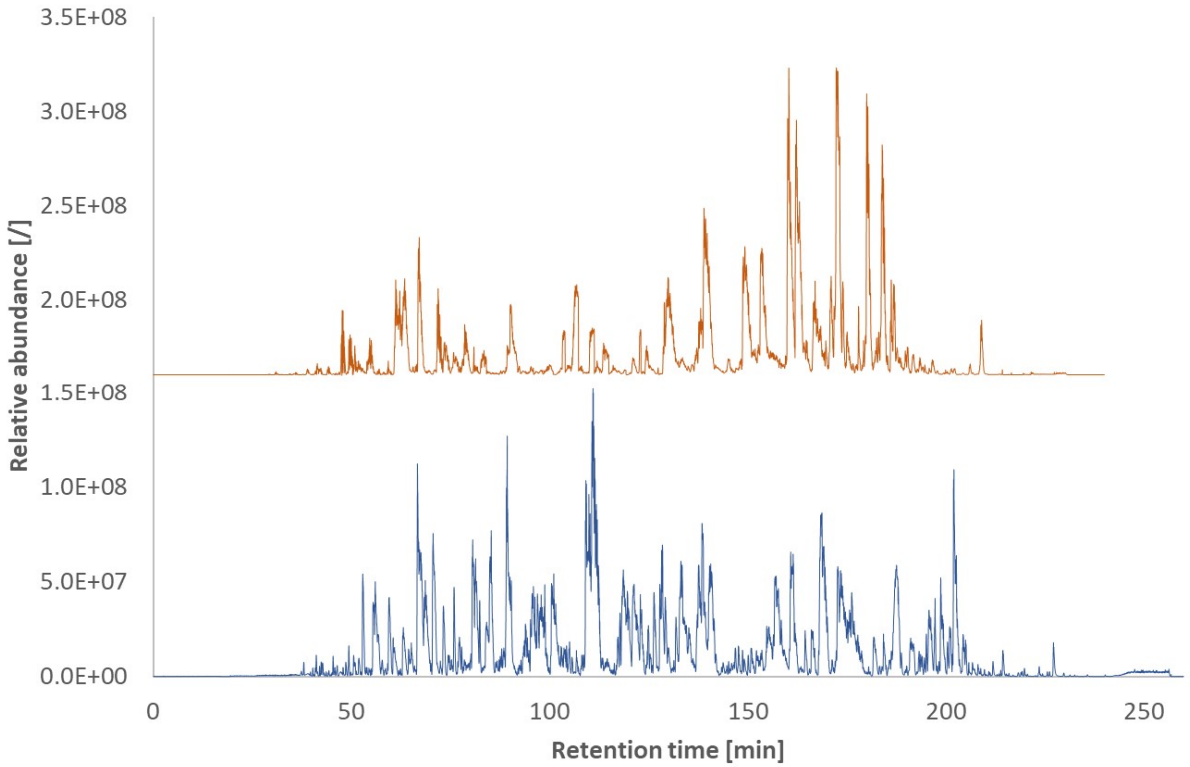

FIG. 8. **Base peak chromatograms obtained for the separation of a human plasma tryptic digest sample.** µPAC™ column (blue), packed bed nano LC column (orange).

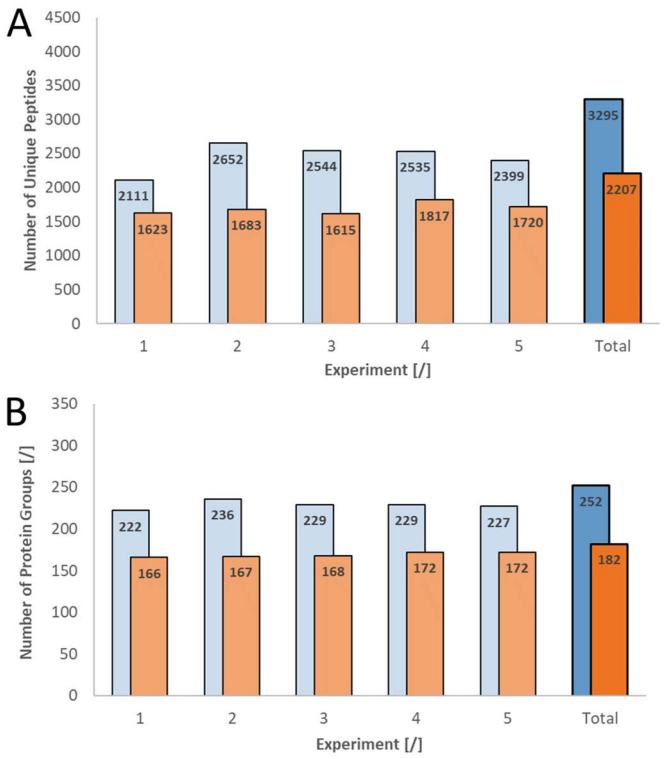

FIG. 9. **Number of peptide and protein identifications achieved for the separation of human plasma digest samples on a µPAC™ (blue) and a packed bed nano LC column (orange).** A, Number of unique peptides identified. B, Number of protein groups identified.