Analysis of 8000 proteins and reduced carry over significantly increase the throughput of single-shot proteomics

Karel Stejskal[2,§], Jeff Op de Beeck[4,§*], Manuel Matzinger[1], Gerhard Dürnberger[3], Oleksandr Boychenko[5], Paul Jacobs[4] and Karl Mechtler[1,2,3*]

[1] IMP - Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, A-1030 Vienna, Austria. [2] IMBA - Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr Gasse 3, A-1030 Vienna, Austria. [3] Gregor Mendel Institute of Molecular Plant Biology of the Austrian Academy of Sciences, Dr. Bohr Gasse 3, A-1030 Vienna, Austria. [4]Thermo Fisher Scientific, Technologiepark-Zwijnaarde 82, B-9052 Gent, Belgium. [5]Thermo Fisher Scientific, Dornierstrasse 4, 82110 Germering, Germany

[§] KS and JODB will contribute equal to the manuscript.

ABSTRACT

In the field of LC-MS based proteomics, increases in sampling depth and proteome coverage have mainly been accomplished by rapid advances in mass spectrometer technology. The comprehensiveness and quality of data that can be generated do however also depend on the performance provided by nano liquid chromatography (nanoLC) separations. Proper selection of reversed-phase separation columns can be of paramount importance to provide the MS instrument with peptides at the highest possible concentration and separated at the highest possible resolution. As an alternative to traditional packed bed LC column technology that uses beads packed into capillary tubing, we present a novel LC column format based on photolithographic definition and Deep Reactive Ion Etching (DRIE) into silicon wafers. With a next generation pillar array column (µPAC) designed for universal use in bottom-up proteomics, the critical dimensions of the stationary phase support structures have been reduced by a factor of 2 to provide further increases in separation power. To demonstrate the potential for single-shot proteomics workflows, we report on a series of optimization and benchmarking experiments where we combine LC separation on a new generation of µPAC columns using Vanquish Neo UHPLC with fast Orbitrap Tribrid MS data-dependent acquisition (DDA) and High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS). In addition to providing superior proteome coverage, robust operation over more than 1 month with a single nanoESI emitter and reduction of the column related sample carry over are additional figures of merit that can help improve proteome research sensitivity, productivity and standardization.

INTRODUCTION

Even though the practice of LC-MS based bottom-up proteomics has remained relatively unaltered over the past decade, researchers are progressively closing the gap between experimentally identified and theoretically expected proteoforms present in complex cell lysates[1–3]. Key aspects driving this progress are the continuous evolution of MS/MS instruments, the coming of age of additional ion mobility separation techniques and the combination with LC separation that delivers maximal resolving power and throughput[1,4,5]. Even though MS/MS instruments have evolved to a point where acquisition rates up to 133 Hz can be reached[6,7], these developments have struggled to materialize similar leaps in proteome coverage depth such as obtained by publications of Thakur et al., Hebert et al. and Scheltema et al in 2011, 2014[8–10]. As postulated several years ago by Shishkova et al[11], chromatographic separation performance is the key but perhaps underappreciated bottleneck limiting the speed and depth of single-shot proteomic analyses. Improvements in a chromatographic resolution or performance have historically been achieved by increasing column length or by decreasing the packing particle diameter[12–14]. However, reducing the particle diameter or extending the column length have a synergistic effect on operating pressures[15,16]. The pressure drop of a column is linearly related to its length and inversely proportional to the square of the particle diameter[17]. As a consequence, current state-of-the art nano LC columns often require ultra-high pressure liquid

chromatography (UHPLC) type of LC instruments which can accurately deliver nanoliter per minute flow rates at operating pressures up to 1,500 bar.

To cope with these pressure requirements, alternative formats such as monolithic columns have been introduced be it with limited adoption in the field of proteomics[18–20]. Alternatively, microfabricated pillar array columns (µPAC) have been proposed as a new promising technology that can redefine the boundaries of LC performance[21]. Using micromachining techniques rather than slurry packing, both chromatographic performance and column permeability can be controlled by design. Similar to the evolution that packed bed column technology has seen in the last 50 years, we are now on the verge of a new generation of pillar array columns where design specifications have been tightened in search of increased separation performance. Schematic drawings of the 'building' blocks or unit cells used to design different 'generations' of pillar array columns are shown in figure S1. The drawings have been complemented with transverse SEM images taken of respectively generation 1 (GEN1) and generation 2 (GEN2) pillar array column backbones. Whereas GEN1 µPAC had separation channels filled with 5 µm diameter pillars positioned in an equilateral triangular grid at a distance of 2.5 µm, pillar diameter and interpillar distance have been reduced to respectively 2.5 and 1.25 µm in the GEN2 µPAC columns[22]. Analogous to observations in packed bed columns, reduction of the pillar and inter pillar dimensions by a factor of 2 results in a net gain in separation resolution of 1.4 in half the analysis time (increase in productivity by a factor of 3)[23]. This increase in performance comes however at the cost of operating pressure: when scaling down particle (or pillar) dimensions by a factor of n and provided that column cross section and length are kept constant, the back pressure required to operate a column at its optimal flow rate will increase by a factor of n^3. The transition from generation 1 to generation 2 will therefore result in an increase in back pressure by a factor of 8.

To investigate potential benefits of this new µPAC design for nanoLC-MS applications, we report on an extensive benchmarking series where we coupled prototype second generation micro pillar array columns to the latest generation of tribrid MS systems, a Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) pro interface and a next-generation low-flow UHPLC system (Vanquish Neo UHPLC). As the goal of this study was to explore how deep we could dive into conventional bottom-up proteomics samples derived from complex cell lysates, we first executed a limited set of instrument optimization experiments before starting the actual benchmarking study. Such experiments are commonly performed with highly validated mammalian protein digest standards and provide unbiased data on instrument performance. Results achieved in benchmarking studies (which were performed under ideal sample loading and composition conditions) do however often differ from what can be achieved with biologically relevant samples and fail to provide information on day-to-day robustness and throughput. The current study aims to address these matters by providing additional data on performance over time, column related sample carry over and validation of results by implementing the workflow for the analysis of a synthetic library of cross-linked peptides[24].

RESULTS & DISCUSSION

MS settings optimization

Previous studies have already demonstrated the benefits of performing systematic optimization of MS settings when operating orbitrap based mass analyzers in either collision induced dissociation (CID) or higher energy collision dissociation (HCD)-DDA acquisition modes [25–28]. The settings resulting in the highest identification rates depend to a large extent on the availability of precursor ions and therefore vary with chromatographic performance, sample loading conditions and ionization efficiency. As this is an entirely new LC-MS/MS setup combining the recently introduced Vanquish Neo UHPLC, a second-generation prototype µPAC nanoLC column and an orbitrap Eclipse Tribrid MS equipped with a FAIMS Pro interface, we deemed it was mandatory to get a proper view on the optimal MS settings for different workflow demands before starting benchmarking experiments. To cover a broad operation range, three different gradient length and sample loading combinations were defined. The effect of maximum injection time in the (linear) ion trap (MIT) and dynamic exclusion time (DET) were evaluated while keeping all other MS and LC settings constant (Figure 1). Short gradients produce sharper peaks and generate higher relative detection responses, which reduces the amount of sample material needed to continuously trigger

MS/MS events. As no significant impact was anticipated by loading micrograms of sample material, a short method with relatively low sample loads (10 min gradient / 50 ng of HeLa cell digest) was used to optimize settings for high throughput analyses where high sensitivity is needed. Separation performance (peak capacity) can be increased by extending the LC solvent gradient, but this will result in a reduction of the relative concentration at which peptides elute[29–31]. As a result, the concentration of low abundant peptides will drop below the limit needed to trigger MS/MS and more material has to be loaded to convert increased LC separation performance into an increase in ID's. A routine method with standard sample loads (60 min gradient / 1 µg of HeLa cell digest) was used to cover typical nanoLC-MS bottom-up proteomics conditions and a long method with high sample loads (180 min gradient / 3 µg of HeLa cell digest) was used to explore deeper proteome coverage.

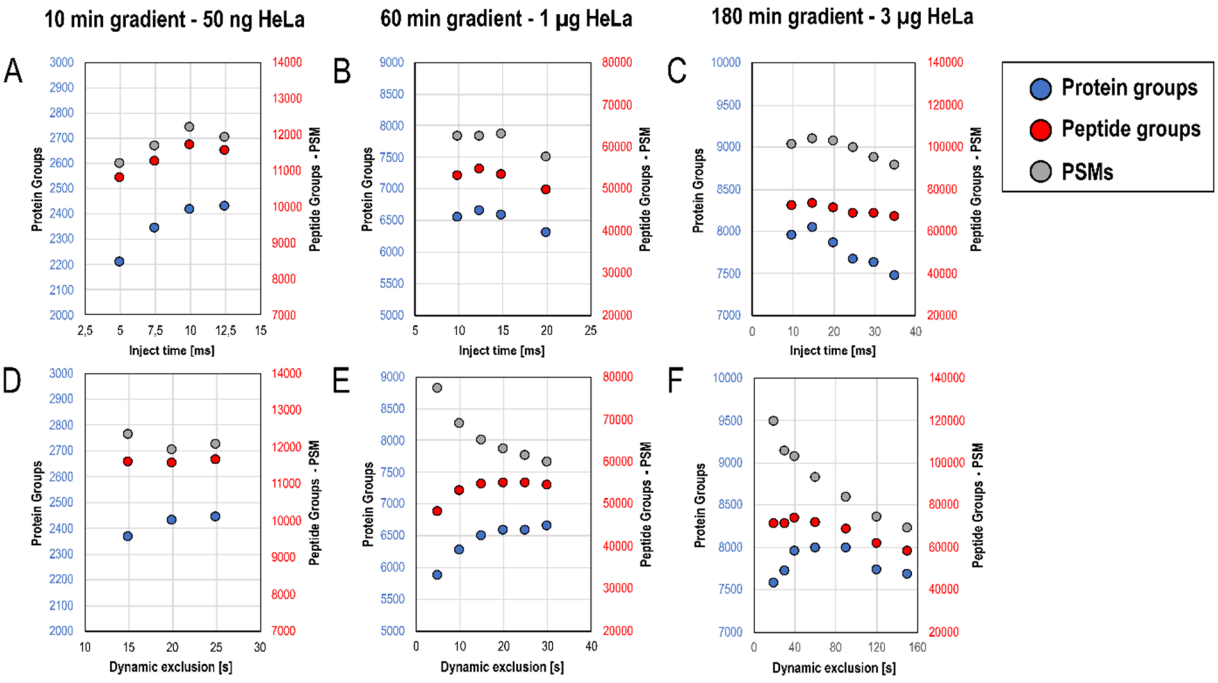

Figure 1: Impact of MS settings on identification rate – throughput (10 min gradient) versus routine (60 min gradient) and comprehensive (180 min gradient) (A-C): Impact of maximum injection time (MIT) in the ion trap, (D-F):): Impact of dynamic exclusion time (DET). Red = Protein Groups, Blue = Peptide Groups, Grey = PSMs.

In order to get the highest possible scan speed, the instrument was operated in high-low mode where the ion trap (IT) rather than the orbitrap (OT) is used for MS2 spectrum acquisition. Theoretically, operating the Orbitrap Eclipse in OT-IT mode with ion trap speed set at Turbo rate and covering a scan width of 1200 m/z allows an approximate MS scanning rate of 45 Hz [11,32,33]. Due to time spent for MS1 scans we could achieve MS2 scan rates of up to 35 Hz with our DDA methods, and this for all three conditions tested during LC-MS/MS optimization (Figure S2). The highest scan rates where consistently achieved at low MIT values, where ion accumulation times do not exceed MS/MS scan duration and associated overhead time. A stable MS2 scan speed of 35 Hz is observed up to MIT of 15 ms, which is in line with earlier reports where optimal MIT settings have been evaluated for a range of m/z ranges and ion trap speed settings[32]. MIT of approximately 15 ms also appeared to be the sweet spot in our analyses (Figure 1A-C). When MIT is increased above 15 ms, a linear decline in MS scanning speed with concomitant decrease in absolute identification numbers is observed. This highlights the importance of MS scan speed towards maximizing feature detection. Even though MS2 scanning speed was improved further, injection times below 15 ms did not result in higher MS2 identification rates nor absolute identification numbers. By collecting fewer ions for MS/MS, overall spectral quality deteriorates. Optimal MIT settings appear to be near the inflection point where the MS is still scanning as fast as possible, yet producing MS spectra with the highest possible quality.

DET is another indispensable setting in DDA acquisition mode. It enables the rejection of high abundant (and often 'broader' eluting) peptides from redundant sampling and fragmentation. Redundant sampling typically results in higher absolute PSM numbers but lower overall peptide and protein identifications. Time spent on the fragmentation of a peptide that has been already sampled reduces the time that can be spent to find new ones. This is observed

when low DET values were evaluated. The optimal DET setting typically depends on the elution width of high to medium abundant peptides impacted by the LC column performance and gradient length[34]. Even though we found that DET settings only affect MS scanning rates to a very limited extent, a significant impact on absolute identification numbers was observed (Figure 1 D-F). The optimum DET setting is directly related to the observed peak width and therefore can be gauged by plotting peak width distributions for a certain separation condition. Rather than FWHM, chromatographic peak widths obtained at 10 or 13.5% of the peak maximum (4σ) need to be used to deduct proper DET values. This is clearly illustrated in figure S3, where peak width distributions (4σ) have been visualized for all three conditions. Below the optimum DET value, identifications are lost due to redundant sampling. Above the optimum, identifications are lost because the MS instrument is running out of precursors and not using up all available speed.

Even though the absolute sample load is significantly different, the concentration distribution at which peptides are presented to the MS is not that divergent. With median PSM intensities between 0.5 and 1.5 x 10^6 (Figure S4), optimal MIT settings were found to be quite similar for all three LC methods. As expected, optimal DET settings do however shift with increased peak width, leading to higher optimal values for longer gradients and higher sample loads. Based on the comprehensive results obtained during this optimization, we defined a set of optimal MS settings for the subsequent benchmarking experiment (Table S1).

LC conditions optimization

As from the first reports on the characteristics of nanoelectrospray, great potential has been anticipated for the combination of low liquid flow rates with ESI-MS[35]. The current supplied by electrospray is proportional to the square root of the flow rate[36]. Therefore, the liquid is ejected by electrospray at higher charge densities at low flow rates. This typically results in higher ionization efficiency of analytes and improves detection sensitivity. Increases in detection sensitivity have proven to be of key importance in the pursuit of comprehensive proteome characterization. Improvements in the field have typically been achieved by increasing the depth to which precursor molecules could be sampled. In search of improved detection sensitivity for limited sample amounts, the implementation of ultra-low flow (ULF) ESI-MS has seen quite a revival in the last few years. Major progress has been documented when ULF (flow rates below 100 nL/min) was combined with ESI-MS [20,37–41]. Robust and routine operation at these low flow rates is however not straightforward and often required highly specialized LC systems or customized pre-column flow splitting configurations. The importance of accurate flow rate control and gradient formation precision can not be underestimated for practical implementations as these parameters define quantitation accuracy to a large extent. Operating columns at low flow rates typically comes at the cost of increased analysis and overhead time. In order to restrict the impact on total analysis time and at the same time ensure good chromatographic performance, it is crucial to align LC column dimensions with the desired flow rate range. The prototype µPAC column has a reduced cross-section (equivalent to a packed bed column with ID of 60 µm) and a total column volume of approximately 1,5 µL. This holds great potential for operation at flow rates lower than 300 nL/min. To find the sweet spot for comprehensive proteome analysis, we evaluated the effect of decreased LC flow rate on chromatographic performance and proteome coverage. The recently introduced Vanquish Neo UHPLC system allows setting the flow from 1 nL/min up to 100 µL/min with 1 nL/min increments and running gradients at typical nano/cap/micro LC flow rates as well as ULF without flow splitting or hardware changes. This is enabled by active flow control and multipoint flow calibration that do not require re-adjustment or re-calibration during the LC usage. The wide flow range LC capabilities and low gradient delay volume together with system operation at constant maximum pressure specified for the column (400 bar) during the sample loading and column equilibration reduced the overhead time and made it possible to run gradients starting from 50 nL/min. Keeping all settings other than flow rate constant, we systematically compared the metrics obtained for a 180 min gradient separation (1 µg HeLa digest sample on column). Base peak chromatograms obtained at flow rates ranging from 50 to 300 nL/min demonstrate the impact of flow rate on the peptide elution start (Figure 2 C-H). Whereas the first peptides elute at approximately 6 min at 300 nL/min, operation at 50 nL/min postpones elution by a factor of 6. Even though the actual elution window in which digested protein material elutes was similar for all flow rates tested (180 min), a significant increase in absolute signal intensity was observed when reducing flow rate.

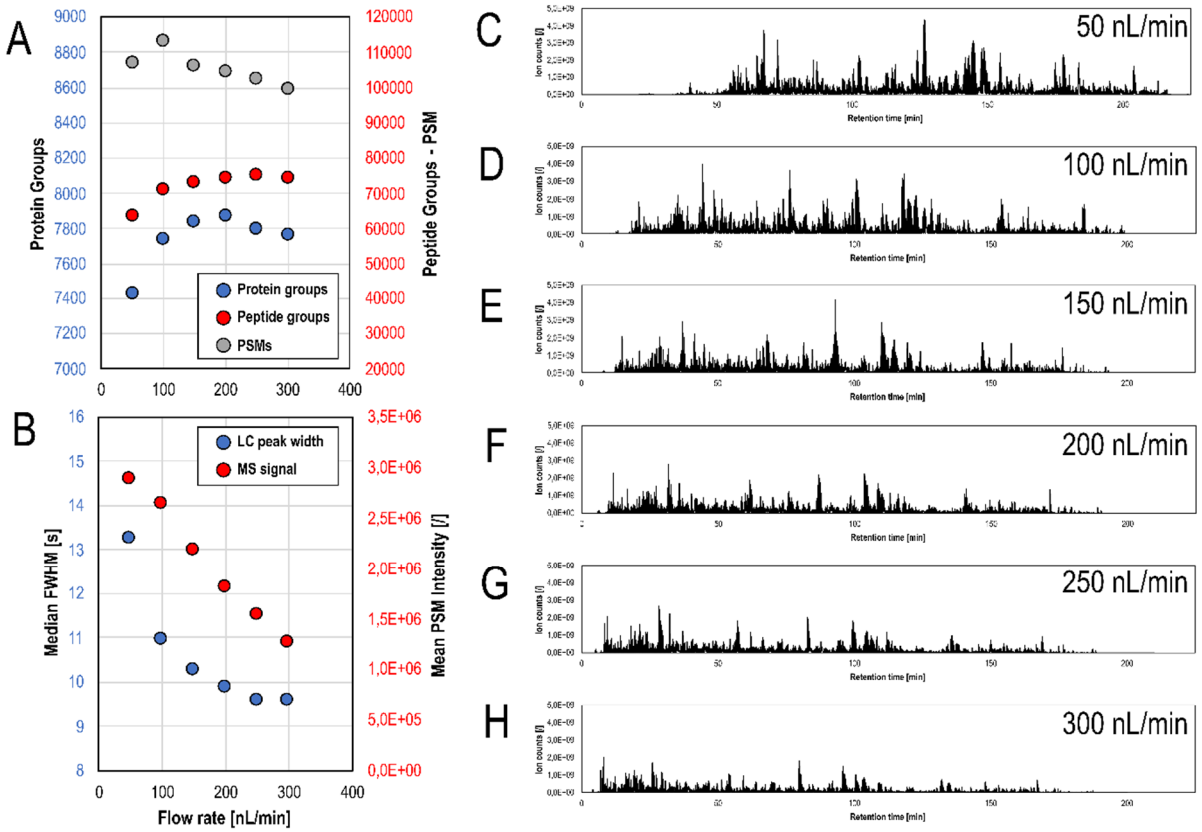

Figure 2: Impact of flow rate on identification rate, chromatographic performance, elution profile and signal intensity. (A):ID's as a function of flow rate, Red = Protein Groups, Blue = Peptide Groups, Grey = PSMs., (B): FWHM (Blue) and Mean PSM intensity (Red) as a function of flow rate, (C-H): Base Peak Chromatogram comparison / elution window for different flow rates.

This was confirmed when comparing mean PSM intensities obtained at different flow rates. Surprisingly, the increase in ionization efficiency did not result in improved proteome coverage. On the contrary, optimal proteome coverage (protein group IDs) was obtained at a flow rate of approximately 200 nL/min where a compromise between chromatographic performance (FWHM in figure 2B) and ionization efficiency (mean PSM intensity in figure 2B) was achieved. The increased ionization efficiency observed at low flow rates does however suggest that another cycle of MS settings optimization might have been needed to explore the full potential of ULF LC ESI MS on this column. We did not pursue low flow rate operation any further as the current evaluation was aimed at finding a balance between MS utilization time, sensitivity for typical bottom-up proteomics sample loads (250 - 2000 ng), and separation quality.

Column benchmarking

After optimization of a confined set of LC and MS parameters, we performed a comprehensive benchmarking experiment to evaluate the column's applicability for a range of LC gradient settings aimed at maximizing the output for single-shot DDA proteomics experiments. The prototype µPAC column was benchmarked against an integrated emitter packed bed nanoLC column with sub 2 µm fully porous particles. As peptide elution profiles and observed FWHM values were found to be comparable for both chromatographic set-ups (Figure S5), optimized settings for the µPAC column were also used for the packed bed column.

When applying a short 10 min gradient (Figure 3a), designed for increased throughput, low sample amounts or intermediate LC-MS quality control, over 2600 proteins could repeatedly be identified from 100 ng of HeLa digest. A significant increase in both peptide and protein group identifications was observed when comparing pillar array and packed bed column.

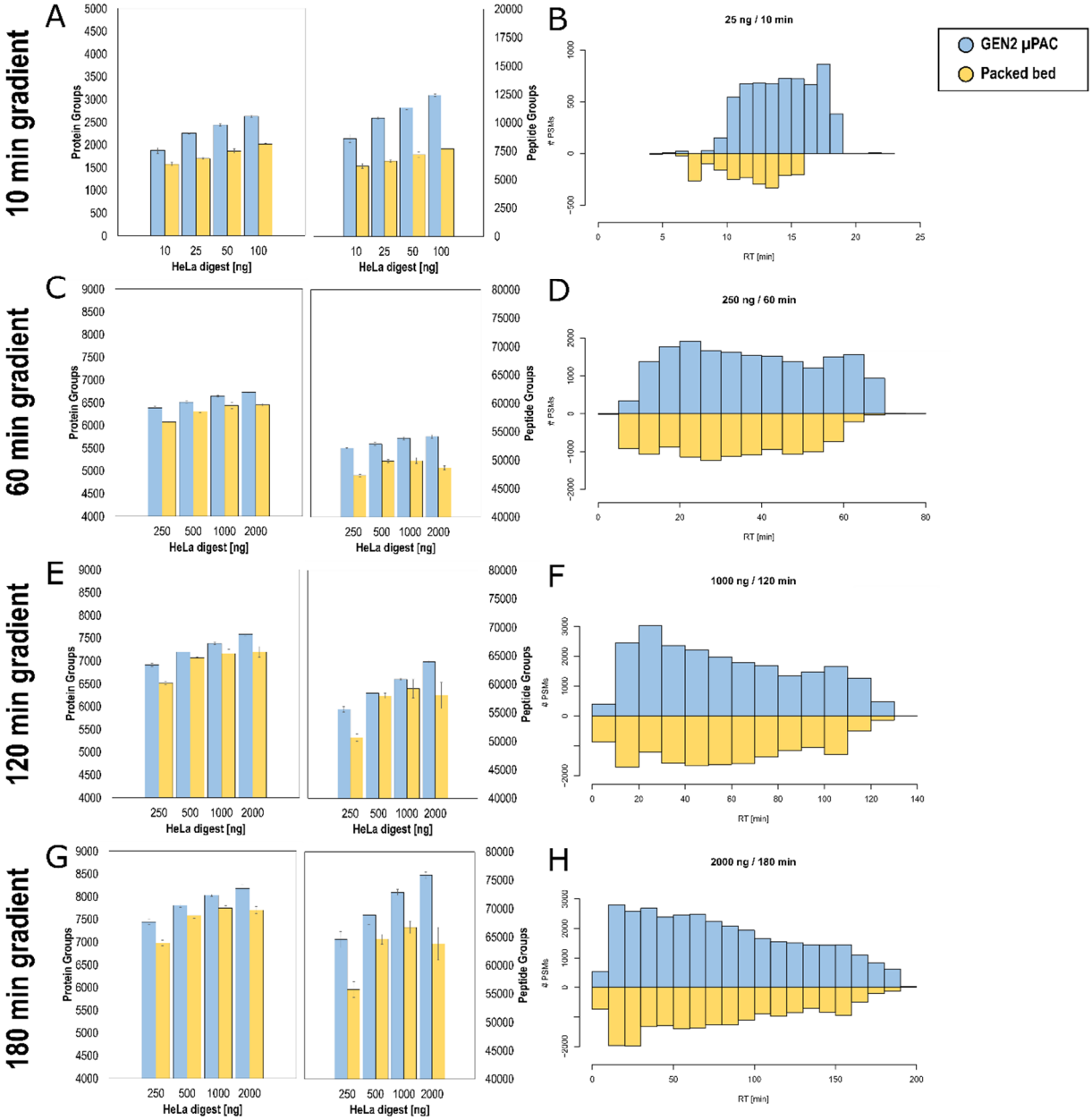

Figure 3: Proteome coverage (protein and peptide Group ID's) obtained for different gradient lengths and sample loads during the extensive benchmarking experiment. Four different methods are tested, the GEN2 pillar array column (blue) is compared to a packed bed column (Yellow). Unique PSMs are plotted as a function of elution time to the right. A-B: 10 min gradient separation, 3 CV FAIMS method, 10-100 ng HeLa digest sample, C-D: 60 min gradient separation, 3 CV FAIMS method, 250-2000 ng HeLa digest sample, E-F: 120 min gradient separation, 3 CV FAIMS method, 250-2000 ng HeLa digest sample, G-H: 180 min gradient separation, 4 CV FAIMS method, 250-2000 ng HeLa digest sample

Even though processed results did not reveal a significant impact of raw chromatographic performance (median FWHM of peptides – Figure S5), 20-30% more protein groups and 40-60% more peptide groups could be identified when using the pillar array format. When plotting the amount of unique PSMs versus retention time (Figure 3B), a clear trend is revealed with additional unique ID's towards the end of the gradient. These data suggest that the morphology of this column technology has an impact on the elution behavior of hydrophobic peptide species. We hypothesize that the use of superficially porous rather than fully porous chromatographic media promotes elution and prevents persistent adsorption of analytes from and to the chromatographic support material. Additional data that confirm this statement are provided later on when evaluating sample carry over and analyzing cross-linked peptides on both LC column formats. It should be noted that we have focused here on the evaluation of fast gradient performance for low sample amounts without optimizing MS utilization. These methods are not optimal when efficient instrument occupation is of primary interest.

As opposed to the high throughput method described above, more efficient MS utilization can be achieved when using longer solvent gradients (75% for 60 , 86% for 120 and 90% for 180 min). Protein identification rates observed for short gradients (approx. 250 protein group ID's per min gradient) do not endure and attenuate according to gradient length. This can be attributed to the fact that the first proteins to be identified from a complex mixture are high abundant ones that can be picked up relatively easily. Further increase in proteomic depth progressively becomes more challenging as undiscovered proteins are of ever decreasing abundance. This is clearly illustrated in Figure S6, where the abundancies of proteins uniquely discovered by extending gradient length or sample load have been compared to those shared with shorter analyses. The relative increase in protein ID's attenuates with increasing gradient length, reaching an averaged maximum of close to 8100 protein groups identified out of 2 µg of HeLa digest sample (Figure 3C, E, G). Again, consistently more features were identified when using the pillar array as compared to the packed bed column. Even though the relative increase in ID's was smaller as compared to the high throughput method (3-6% on the protein group level, 6-19% on the peptide group level), unique hits were again predominantly originating from later eluting peptide species (Figure 3D, F, G), confirming earlier observations.

Carry over study

In many cases, very few or no intermediate washes are performed between runs. So researchers struggle to understand whether sample material is transferred from one analysis to another. Those who do run blank or intermediate wash runs often assume that a single blank injection is sufficient to clear persistent sample material, without actually acquiring or analyzing MS/MS data. In practice, these assumptions can have a serious impact on results and affect the outcome of a study. The newly introduced Vanquish Neo UHPLC enabled an unbiased investigation of LC column related sample carry over, as after each injection and in parallel to the peptide separation step, the autosampler executes stringent system washing cycles with a high volume of organic solvent to wash the needle outside and the complete injection fluidics path including needle inside. To assess LC column related sample carry over, blank injections (pure sample solvent -1% ACN in 0.1% FA – analyzed using a 15 min gradient QC method) were included in the benchmarking series. By running these blank methods immediately after each sample concentration, we intend to quantify how sample carry over relates to the amount of protein digest loaded onto the column during the analytical run. Figure 4A shows the number of protein groups identified from blank runs for both column set-ups. Up to a sample load of 1 µg, no protein groups were identified from the blank injections on the µPAC prototype, as there were too few spectra present for FDR assessment in Percolator (200 peptides required)[42]. This however does not mean there is absolutely no sample carry over, which is illustrated in Figure S7, where basepeak chromatograms obtained for blank runs after each sample concentration are compared. Even after loading 250 ng of HeLa digest, well-defined peaks are present in the basepeak chromatograms for both column types. Traces obtained after similar sample loads are however much lower in intensity when comparing column types More data is provided by analyzing results obtained for consecutive washes (n=2) that have been performed after 100 ng HeLa QC runs. Using a fixed value validator for FDR assessment, apQuant areas obtained for the top 50 most abundant peptides have been compared (Figure 4 C-D). Wash runs immediately after the analytical run (1st wash) still show up quite some quantifiable signals for both columns, respectively 43 and 46 out of 50 peptides were quantified on pillar array and packed bed respectively. There is however a significant difference when analyzing data from the 2nd wash run. Respectively 5 and 19 peptides were quantified in the 2nd wash. This indicates that additional column washes are needed to remove persistent peptide material when using LC columns packed with fully porous silica materials. As mentioned before when discussing the impact of stationary phase support morphology on peptide elution, we believe this is a result of the intrinsic difference in interaction surface between both column types. This consistently results in decreased carry over related identifications on the µPAC column, 3-4 times less on the peptide group and 2-3 times less on the protein group level. In addition to providing comprehensive data on LC column type related sample carry over, these data also clearly identify the LC column rather than the LC autosampler as a major source of cross-contamination in the current LC-MS set-up. Implementation of blank runs in between biologically relevant samples should be considered best practice, even though it affects sample throughput and decreases instrument productivity. As an example, incorporation of a single column wash after each analytical run reduces instrument productivity from 75 to 55, from 86 to 71 and from 90 to 78% for 60, 120 and 180 min gradient analyses respectively. In the case an additional wash is needed to get rid of adsorbed sample material, productivity decreases even further to 43, 60 and 69%. The impact of carry over on

global analysis time should not be neglected, proper selection of LC and column hardware can save time and increase throughput.

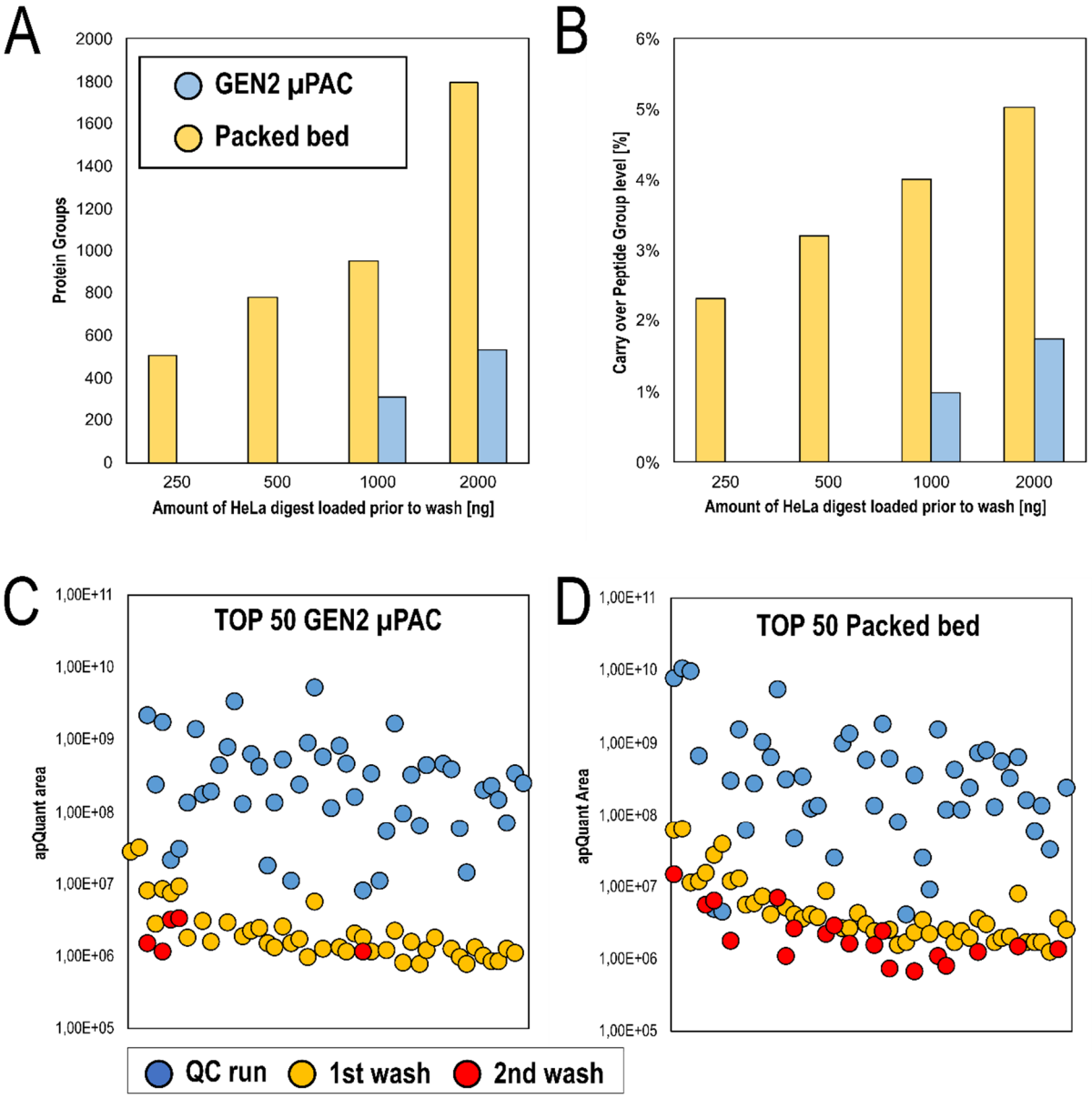

Figure 4: Comparison of sample carry over obtained after increasing sample loads. Blank wash runs immediately after each concentration have been analyzed. Comparison of GEN2 pillar array (blue) to packed bed column (yellow). A: Number of protein group ID's, B: Relative percentage of carry over on the peptide group level, C-D: Comparison of apQuant area obtained for top 50 most abundant HeLa peptides, 100 ng QC run is compared to results from a first and a second wash, (C = GEN2 pillar array column, D = packed bed column)

Cross-linking experiments

In addition to providing a transparent LC column performance comparison for standardized HeLa digest samples, we performed a limited set of experiments with cross-linked peptide samples. During the last decade cross-linking mass spectrometry was established as a potent technique to investigate protein-protein interaction networks as well as in the field of structural proteomics. This technique, including a wide variety of applications was already described in several excellent reviews (e.g. [43–45]). Briefly, two amino acid residues are covalently connected by application of the cross-linker reagent, followed by proteomic sample preparation, and yielding two interconnected peptides for detection by mass spectrometry. Depending on the used linker type, cross-linkers can target amines (lysines), sulfhydryl groups (cysteines), carboxylic acids (glutamic- or aspartic-acid) or they can form radical species reactive to any amino acid. The broad variety of linker-types, acquisition techniques as well as data analysis

algorithms makes it difficult to find an optimal workflow for a specific protein system. To alleviate this issue, we previously developed a synthetic peptide library based on sequences of the Cas9 protein[24]. The peptides contain exactly one targeted (i.e. lysine) amino acid for cross-linking. They were mixed into groups that were separately cross-linked, followed by quenching and pooling to a single peptide-library. This system allows an exact FDR calculation as only interpeptide connections within a group are possible. Furthermore, the maximal theoretical cross-link number is known (426 unique combinations), which allows estimating the efficiency of a detection workflow based on the reached identification numbers. Such a synthetic library, in combination with the linker reagent ureido-4,4-dibutyricacidbis(hydroxysuccinimide) ester (DSBU), therefore represents an ideal benchmarking tool for the comparison of two different chromatographic setups, as done in this study.

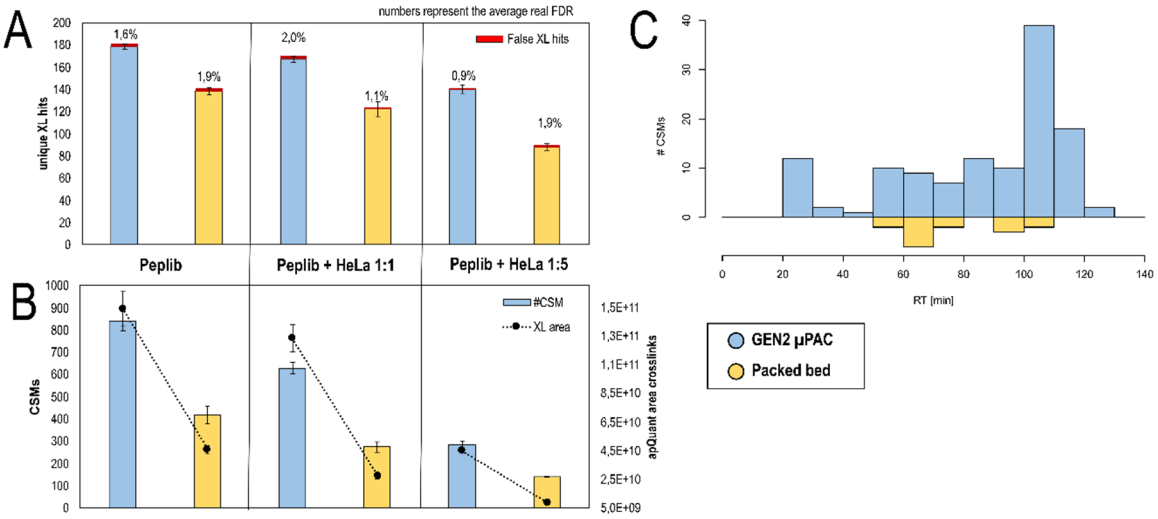

Figure 5: Benchmarking of GEN2 pillar array column vs packed bed column using a DSBU cross-linked synthetic peptide library. A: Number of unique cross-links identified on 1% estimated FDR level and real FDR printed above. B: Number of cross-linked peptides exclusively identified with either pillar array or packed bed chromatographic setup vs retention time in on representative replicate, summed to 10 min windows. C: Number of identified cross-linked peptides (CSMs) and its relative abundance based on LFQ. A and C: All values represent average values (n=3) with error bars depicting standard deviations.

The number of identified unique cross-links, as well as the number of cross-link spectrum matches, is reproducibly boosted when using the pillar array column setup compared to the packed bed setup (Figure 5 A, C). The number of identifications decreases upon increasing background of linear peptides present in the sample, which is likely not only a result of increased sample complexity but also of decreased amounts of cross-linked peptides present in spiked samples (i.e. 1µg XL-peptides without spiking vs. 200 ng XL peptides + 800 ng tryptic HeLa peptides in 1:5 spiked sample). Of note, the advantage of the pillar array over packed bed increases in complex sample mixtures. On average, we observed a boost in cross-link IDs of ~29 % without spiking but of 37% and 59% upon 1:1 and 1:5 spiking respectively. In line with ID numbers also the relative abundance of cross-linked peptides is increased in the pillar array setup for all three test samples. The average real FDR rate is close to the expected 1 % in all sample types and is independent of the used column, highlighting the quality of the obtained data and a properly working target decoy-based FDR approach using MS Annika. As shown in Figure 5B and in line with the results obtained using different sample loads and gradient lengths (Figure 3B), we observed most of the extra CSM identifications with the µPAC column at high retention times. This could indicate fewer losses of larger species which are expected predominant among cross-linked peptides as two peptides are connected.

CONCLUSION

Data compiled in this manuscript provide a transparent perspective on the benefits that next-generation µPAC technology can bring to nanoLC-MS proteomics workflows. By combining this technology with the latest innovations in LC-MS/MS instrumentation, significant improvements in proteome coverage can be obtained with high

reproducibility, robust operation and minimal sample carry over. Improvements in chromatographic performance were achieved by reducing pillar diameter and interpillar distance by a factor of 2, resulting in separation channels filled with 2.5 µm diameter at an interpillar distance of 1.25 µm. As opposed to packed bed columns with integrated emitter tips, LC column and ESI emitter lifetime can be detached, providing a potentially more sustainable LC-MS solution without compromising separation performance. After optimization of a confined set of MS and LC parameters, systematically higher proteome coverage could be obtained as compared to pulled-tip packed bed nanoLC columns. Compared to earlier nanoLC µPAC generations, separation channel cross-sections have also been further reduced, making both low flow nanoLC and fast gradient operation more attractive. Low delay volumes help to reduce the impact of loading and equilibration on total analysis times, hereby increasing instrument productivity.

For short gradients (10 min) and limited sample amounts (10-100 ng of cell lysate digest), the impact on proteome coverage was found to be most pronounced, with gains in proteome coverage between 20 and 30% on the protein group level. When extending gradient lengths to (60, 120 and 180 min) and injecting sample amounts typically encountered in the analysis of whole cell lysates (250-2000 ng) increases in coverage were less distinct, producing an increase in proteome coverage between 3 and 6% on the protein group level. The highest proteome coverage was obtained with an optimized 180 min gradient separation where 2 µg of HeLa cell digest was injected, resulting in an average number of identified protein groups of 8280. To verify that these findings were not biased by instrument performance, a limited set of experiments was conducted on a second µPAC column immediately after disconnecting the packed bed column. Similar performance as observed on the first column was confirmed and an immediate boost in identifications obtained from QC runs was observed.

Comparison of peptide elution behavior revealed that a larger portion of uniquely identified peptides was acquired at later eluting times, suggesting that the intrinsic difference in surface morphology (superficially porous versus fully porous) produces better distribution of peptides across the solvent gradient. These differences in surface morphology are also thought to be the main contributor to the reduced sample carry over that was observed in the current experiments. Column-related carry over could be reduced by a factor of 2-3 by switching from fully porous packed bed to superficially porous microfabricated column types. In-depth investigation of carry over revealed that at least 2 wash cycles were needed to wash away highly abundant peptides on traditional fully porous silica based LC columns. Conversely, only a single wash cycle was sufficient when operating µPAC columns, hereby providing better quality data at increased instrument productivity.

Even though benchmarking studies and LC-MS/MS instrument optimization are typically performed with highly validated mammalian protein digest standards, results are often interpreted as deceiving by application scientists as experiments are performed under ideal sample loading and composition conditions. To confirm results obtained in the benchmarking study, both column types were subsequently used in the analysis of a DSBU crosslinked synthetic library. When using the µPAC column, 29 to 59 more unique crosslinked peptides could be identified at a true FDR of 1-2%, providing additional proof for the general applicability of the next-generation µPAC technology in a range of nanoLC-MS proteomics workflows.

SUPPORTING INFORMATION

The Supporting Information is available free of charge and includes:

- Experimental section covering (1) sample preparation, (2) LC-MS set-up, (3) FAIMS settings selection, (4) MS acquisition settings,(5) Data analysis
- Results section on performance consistency evaluation
- Design, scanning electron micrographs and performance specifications of GEN1 and GEN2 pillar array columns
- MS settings optimization, effect of MIT and DET on MS2 scan and identification rate
- Comparsion of the effect of 3 versus 4 FAIMS compensation voltages on proteome coverage
- Table with LC, MS and FAIMS settings used for final benchmark study

- Systematic comparison of PSM intensity, FWHM, datapoints per peak and quantitation accuracy for different sample loads and gradient lengths
- FWHM comparison between packed bed and µPAC GEN 2 column
- apQuant peak area distribution comparison for unique versus shared proteins, 180 min gradient separation is compared to combined 60 and 120 min gradient separations
- Basepeak chromatograms obtained for blank injections immediately after increasing sample loads.
- Protein group identifcation numbers obtained for QC runs throughout the entire experiment

AUTHOR INFORMATION

Corresponding Authors

Karl Mechtler - IMP - Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, A-1030 Vienna, IMBA - Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr Gasse 3, A-1030 Vienna, Austria, Austria, Gregor Mendel Institute of Molecular Plant Biology of the Austrian Academy of Sciences, Dr. Bohr Gasse 3, A-1030 Vienna, Austria; orcid.org/ 0000-0002-3392-9946; E-mail: Karl.Mechtler@imp.ac.at

Jeff Op de Beeck –Thermo Fisher Scientific,Technologiepark Zwijnaarde 82, 9052 Gent, Belgium, E-mail: Jeff.Opdebeeck@thermofisher.com

Competing interests

JODB, OB and PJ are employees of Thermo Fisher Scientific.

ACKNOWLEDGEMENTS

This work supported by the EPIC-XS, Project Number 823839, funded by the Horizon 2020 Program of the European Union, by the project LS20-079 of the Vienna Science and Technology Fund and the by the ERA-CAPS I 3686 and P35045-B project of the Austrian Science Fund. We thank the IMP for general funding and access to infrastructure and especially the technicians of the protein chemistry facility for continuous laboratory support.

The authors are thankful to Christopher Pynn, Roman Huguet, and Guenther Muellauer for providing technical support with LC and MS measurements.

REFERENCES

(1) Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10 000 Identified and Quantified Proteins in a Single Run by Optimizing Current LC-MS Instrumentation and Data Analysis Strategy. *Mol. Omi.* **2019**, *15* (5), 348–360. https://doi.org/10.1039/c9mo00082h.

(2) Bian, Y.; The, M.; Giansanti, P.; Mergner, J.; Zheng, R.; Wilhelm, M.; Boychenko, A.; Kuster, B. Identification of 7 000-9 000 Proteins from Cell Lines and Tissues by Single-Shot Microflow LC-MS/MS. *Anal. Chem.* **2021**, *93* (25), 8687–8692. https://doi.org/10.1021/acs.analchem.1c00738.

(3) Bekker-Jensen, D. B.; Martínez-Val, A.; Steigerwald, S.; Rüther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol. Cell. Proteomics* **2020**, *19* (4), 716–729. https://doi.org/10.1074/mcp.TIR119.001906.

(4) Bache, N.; Geyer, P. E.; Bekker-Jensen, D. B.; Hoerning, O.; Falkenby, L.; Treit, P. V.; Doll, S.; Paron, I.; Müller, J. B.; Meier, F.; Olsen, J. V.; Vorm, O.; Mann, M. A Novel LC System Embeds Analytes in Pre-Formed Gradients for Rapid, Ultra-Robust Proteomics. *Mol. Cell. Proteomics* **2018**, *17* (11), 2284–2296. https://doi.org/10.1074/mcp.TIR118.000853.

(5) Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J. J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal. Chem.* **2018**, *90* (15), 9529–9537. https://doi.org/10.1021/acs.analchem.8b02233.

(6) Meier, F.; Park, M. A.; Mann, M. Trapped Ion Mobility Spectrometry and Parallel Accumulation–Serial Fragmentation in Proteomics. *Mol. Cell. Proteomics* **2021**, *20*, 100138. https://doi.org/10.1016/j.mcpro.2021.100138.

(7) Yu, Q.; Paulo, J. A.; Naverrete-Perea, J.; McAlister, G. C.; Canterbury, J. D.; Bailey, D. J.; Robitaille, A. M.; Huguet, R.; Zabrouskov, V.; Gygi, S. P.; Schweppe, D. K. Benchmarking the Orbitrap Tribrid Eclipse for Next Generation Multiplexed Proteomics. *Anal. Chem.* **2020**, *92* (9), 6478–6485. https://doi.org/10.1021/acs.analchem.9b05685.

(8) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Cox, J.; Mann, M.; Ms-based, I. Deep and Highly Sensitive Proteome Coverage by LC-MS / MS Without Prefractionation. *Mol. Cell. Proteomics* **2011**, 1–9. https://doi.org/10.1074/mcp.M110.003699.

(9) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The One Hour Yeast Proteome. *Mol. Cell. Proteomics* **2014**, *13* (1), 339–347. https://doi.org/10.1074/mcp.M113.034769.

(10) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. 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* **2014**, *13* (12), 3698–3708. https://doi.org/10.1074/mcp.M114.043489.

(11) Shishkova, E.; Hebert, A. S.; Coon, J. J. Commentary Now , More Than Ever , Proteomics Needs Better Chromatography. *Cell Syst.* **2016**, *3* (4), 321–324. https://doi.org/10.1016/j.cels.2016.10.007.

(12) Gritti, F.; Guiochon, G. Mass Transfer Kinetics, Band Broadening and Column Efficiency. *J. Chromatogr. A* **2012**, *1221*, 2–40. https://doi.org/10.1016/j.chroma.2011.04.058.

(13) Hsieh, E. J.; Bereman, M. S.; Durand, S.; Valaskovic, G. A.; Maccoss, M. J. 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.* **2013**, *24*, 148–153. https://doi.org/10.1007/s13361-012-0508-6.

(14) Sorensen, M. J.; Anderson, B. G.; Kennedy, R. T. Liquid Chromatography above 20,000 PSI. *TrAC - Trends Anal. Chem.* **2020**, *124*, 115810. https://doi.org/10.1016/j.trac.2020.115810.

(15) Giesche, H.; Unger, K. K.; Esser, U.; Eray, B.; Trüdinger, U.; Kinkel, J. N. Packing Technology, Column Bed Structure and Chromatographic Performance of 1-2-Mm Non-Porous Silicas in High-Performance Liquid Chromatography. *J. Chromatogr. A* **1989**, *465* (2), 39–57. https://doi.org/10.1016/S0021-9673(01)83571-8.

(16) Unger, K. K.; Kumar, D.; Grün, M.; Büchel, G.; Lüdtke, S.; Adam, T.; Schumacher, K.; Renker, S. Synthesis of Spherical Porous Silicas in the Micron and Submicron Size Range: Challenges and Opportunities for Miniaturized High-Resolution Chromatographic and Electrokinetic Separations. *J. Chromatogr. A* **2000**, *892* (1–2), 47–55. https://doi.org/10.1016/S0021-9673(00)00177-1.

(17) Broeckhoven, K.; Desmet, G.; Vos, J. De. Pressure, Particles, and System Contribution: The Holy Trinity of Ultrahigh-Performance Liquid Chromatography. *LCGC Eur.* **2017**, *30* (11), 618–625.

(18) Iwasaki, M.; Sugiyama, N.; Tanaka, N.; Ishihama, Y. Human Proteome Analysis by Using Reversed Phase Monolithic Silica Capillary Columns with Enhanced Sensitivity. *J. Chromatogr. A* **2012**, *1228*, 292–297. https://doi.org/10.1016/j.chroma.2011.10.059.

(19) Josic, D.; Clifton, J. G. Use of Monolithic Supports in Proteomics Technology. *J. Chromatogr. A* **2007**, *1144*, 2–13. https://doi.org/10.1016/j.chroma.2006.11.082.

(20) Greguš, M.; Kostas, J. C.; Ray, S.; Abbatiello, S. E.; Ivanov, A. R. Improved Sensitivity of Ultralow Flow LC-MS-Based Proteomic Profiling of Limited Samples Using Monolithic Capillary Columns and FAIMS Technology. *Anal. Chem.* **2020**, *92* (21), 14702–14712. https://doi.org/10.1021/acs.analchem.0c03262.

(21) Gzil, P.; Vervoort, N.; Baron, G. V; Desmet, G. Advantages of Perfectly Ordered 2-D Porous Pillar Arrays over Packed Bed Columns for LC Separations : A Theoretical Analysis. *Anal. Chem.* **2003**, *75* (22), 6244–6250.

(22) Stejskal, K.; De Beeck, J. O.; Durnberger, G.; Jacobs, P.; Mechtler, K. Ultrasensitive NanoLC-MS of Subnanogram Protein Samples Using Second Generation Micropillar Array LC Technology with Orbitrap Exploris 480 and FAIMS PRO. *Anal. Chem.* **2021**. https://doi.org/10.1021/acs.analchem.1c00990.

(23) Mazzeo, J. R.; Neue, U. D.; Kele, M.; Plumb, R. S.; Corp, W. A New Separation Technique Takes Advantage of Sub-2-Mm HPLC. *J. Am. Chem. Soc.* **2005**, 460–467.

(24) Beveridge, R.; Stadlmann, J.; Penninger, J. M.; Mechtler, K. A Synthetic Peptide Library for Benchmarking Crosslinking-Mass Spectrometry Search Engines for Proteins and Protein Complexes. *Nat. Commun.* **2020**, *11* (1). https://doi.org/10.1038/s41467-020-14608-2.

(25) Nie, L.; Zhu, M.; Sun, S.; Zhai, L.; Wu, Z.; Qian, L.; Tan, M. An Optimization of the LC-MS/MS Workflow for Deep Proteome Profiling on an Orbitrap Fusion. *Anal. Methods* **2016**, *8* (2), 425–434. https://doi.org/10.1039/c5ay01900a.

(26) Kalli, A.; Smith, G. T.; Sweredoski, M. J.; Hess, S. Evaluation and Optimization of Mass Spectrometric Settings during Data-Dependent Acquisition Mode: Focus on LTQ-Orbitrap Mass Analyzers. *J. Proteome Res.* **2013**, *12* (7), 3071–3086. https://doi.org/10.1021/pr3011588.

(27) Espadas, G.; Borràs, E.; Chiva, C.; Sabidó, E. Evaluation of Different Peptide Fragmentation Types and Mass Analyzers in Data-Dependent Methods Using an Orbitrap Fusion Lumos Tribrid Mass Spectrometer. *Proteomics* **2017**, *17* (9), 1–21. https://doi.org/10.1002/pmic.201600416.

(28) Huang, P.; Liu, C.; Gao, W.; Chu, B.; Cai, Z.; Tian, R. Synergistic Optimization of Liquid Chromatography and Mass Spectrometry Parameters on Orbitrap Tribrid Mass Spectrometer for High Efficient Data-Dependent Proteomics. *J. Mass Spectrom.* **2021**, *56* (4), 1–10. https://doi.org/10.1002/jms.4653.

(29) Neue, U. D. Theory of Peak Capacity in Gradient Elution. *J. Chromatogr. A* **2005**, *1079* (1-2 SPEC. ISS.), 153–161. https://doi.org/10.1016/j.chroma.2005.03.008.

(30) Gilar, M.; Neue, U. D. Peak Capacity in Gradient Reversed-Phase Liquid Chromatography of Biopolymers. Theoretical and Practical Implications for the Separation of Oligonucleotides. *J. Chromatogr. A* **2007**, *1169* (1–2), 139–150. https://doi.org/10.1016/j.chroma.2007.09.005.

(31) Petersson, P.; Frank, A.; Heaton, J.; Euerby, M. R. Maximizing Peak Capacity and Separation Speed in Liquid Chromatography. *J. Sep. Sci.* **2008**, *31* (13), 2346–2357. https://doi.org/10.1002/jssc.200800064.

(32) Trujillo, E. A.; Hebert, A. S.; Brademan, D. R.; Coon, J. J. Maximizing Tandem Mass Spectrometry Acquisition Rates for Shotgun Proteomics. *Anal. Chem.* **2019**, *91* (20), 12625–12629. https://doi.org/10.1021/acs.analchem.9b02979.

(33) Riley, N. M.; Hebert, A. S.; Coon, J. J. Previews Proteomics Moves into the Fast Lane. *Cell Syst.* **2016**, *2* (3), 142–143. https://doi.org/10.1016/j.cels.2016.03.002.

(34) Johnson, D.; Boyes, B.; Fields, T.; Kopkin, R.; Orlando, R. Optimization of Data-Dependent Acquisition Parameters for Coupling High- Speed Separations with LC-MS/MS for Protein Identifications. *J. Biomol. Tech.* **2013**, *24* (2), 62–72. https://doi.org/10.7171/jbt.13-2402-003.

(35) Wilm, M.; Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996**, *68* (1), 1–8. https://doi.org/10.1021/ac9509519.

(36) De La Mora, J. F.; Loscertales, I. G. The Current Emitted by Highly Conducting Taylor Cones. *J. Fluid Mech.* **1994**, *260*, 155–184.

(37) Brunner, A.; Thielert, M.; Vasilopoulou, C.; Ammar, C.; Coscia, F.; Mund, A.; Horning, O. B.; Bache, N.; Apalategui, A.; Lubeck, M.; Raether, O.; Park, M. A.; Richter, S.; Fischer, D. S.; Theis, F. J.; Meier, F.; Mann, M. Ultra-High Sensitivity Mass Spectrometry Quantifies Single- Cell Proteome Changes upon Perturbation. *bioRxiv* **2020**. https://doi.org/10.1101/2020.12.22.423933.

(38) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: Mass Spectrometry of Single Mammalian Cells Quantifies Proteome Heterogeneity during Cell Differentiation 06 Biological Sciences 0601 Biochemistry and Cell Biology 06 Biological Sciences 0604 Genetics. *Genome Biol.* **2018**, *19* (1), 1–12. https://doi.org/10.1186/s13059-018-1547-5.

(39) Specht, H.; Emmott, E.; Petelski, A. A.; Huffman, R. G.; Perlman, D. H.; Serra, M.; Kharchenko, P.; Koller, A.; Slavov, N. Single-Cell Proteomic and Transcriptomic Analysis of Macrophage Heterogeneity Using SCoPE2. *Genome Biol.* **2021**, *22* (1), 1–27. https://doi.org/10.1186/s13059-021-02267-5.

(40) Liang, Y.; Acor, H.; McCown, M. A.; Nwosu, A. J.; Boekweg, H.; Axtell, N. B.; Truong, T.; Cong, Y.; Payne, S. H.; Kelly, R. T. Fully Automated Sample Processing and Analysis Workflow for Low-Input Proteome Profiling. *Anal. Chem.* **2021**, *93* (3), 1658–1666. https://doi.org/10.1021/acs.analchem.0c04240.

(41) Xiang, P.; Zhu, Y.; Yang, Y.; Zhao, Z.; Williams, S. M.; Moore, R. J.; Kelly, R. T.; Smith, R. D.; Liu, S. Picoflow Liquid Chromatography-Mass Spectrometry for Ultrasensitive Bottom-Up Proteomics Using 2-Mm-i.d. Open Tubular Columns. *Anal. Chem.* **2020**, *92* (7), 4711–4715. https://doi.org/10.1021/acs.analchem.9b05639.

(42) MacCoss M, M. J.; Noble, W. S.; Käll, L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (11), 1719–1727. https://doi.org/10.1007/s13361-016-1460-7.

(43) Belsom, A.; Rappsilber, J. Anatomy of a Crosslinker. *Curr. Opin. Chem. Biol.* **2021**, *60*, 39–46. https://doi.org/10.1016/j.cbpa.2020.07.008.

(44) Matzinger, M.; Mechtler, K. Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein-Protein Interaction Networks in Vivo. *J. Proteome Res.* **2021**, *20* (1), 78–93. https://doi.org/10.1021/acs.jproteome.0c00583.

(45) Piersimoni, L.; Sinz, A. Cross-Linking/Mass Spectrometry at the Crossroads. *Anal. Bioanal. Chem.* **2020**, *412* (24), 5981–5987. https://doi.org/10.1007/s00216-020-02700-x.

(46) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450. https://doi.org/10.1093/nar/gky1106.