An Executive Summary

Expanding the Limits of Shotgun Proteomics Using Micro-Pillar Array Columns and Data-Independent Acquisition Methods

Jeff Op de Beeck, PhD
Application Development Manager
PharmaFluidics

Jan Muntel, PhD
Senior Scientist R&D
Biognosys AG

Perfectly ordered arrays of free-standing pillars obtained by lithographic etching of silicon wafers deliver unprecedented separation performance when used in chromatographic columns.

Introduction

The stationary phase of traditional liquid chromatography (LC) columns is a bed of microparticles packed in random fashion inside a cylinder or capillary tube. The particles have a certain size distribution, which combined with their random placement creates a level of heterogeneity throughout the bed. As a result, the paths available to the sample molecules for their migration through the column are not identical, resulting in slightly different migration times, which give rise to additional dispersion in the overall separation. To avoid this, PharmaFluidics has developed columns filled with accurately positioned, ordered arrays of free-standing pillars, instead of particles packed in a tube. With their perfect order, the pillars reduce dispersion to an absolute minimum, as all molecules follow identical paths through the column. This ordered configuration results in better peak resolution, sharper elution peaks, and increased sensitivity. In addition, the columns have much higher permeability, which allows for long columns to be operated at moderate pressures. This article explains specific features of the micro-pillar array-based columns (µPAC™) and their application to shotgun proteomics in combination with data-independent acquisition (DIA) methodology.

The Benefits of Ordered Media

The micro-pillar array columns (µPAC™) contain channels filled with pillars made by lithographic techniques and silicon etching using the same state-of-the-art technology involved in the manufacturing of electronic circuits and sensors (**Figure 1**). These pillars are the backbone of the stationary phase. They are subsequently anodized to make them porous and chemically modified to attach conventional reversed-phase C18 media onto them. Other phases will be available in the future. A clear benefit of the ordered array of pillars is the reduction of Eddy dispersion to an absolute minimum, as the bed is homogeneous, and all flow paths are identical. The columns have a transparent glass cover, which allowed the development team to use fluorescent imaging systems to follow what happens inside the column. They determined that the sample plug retains its narrow shape and its intensity stays high through the entire column.

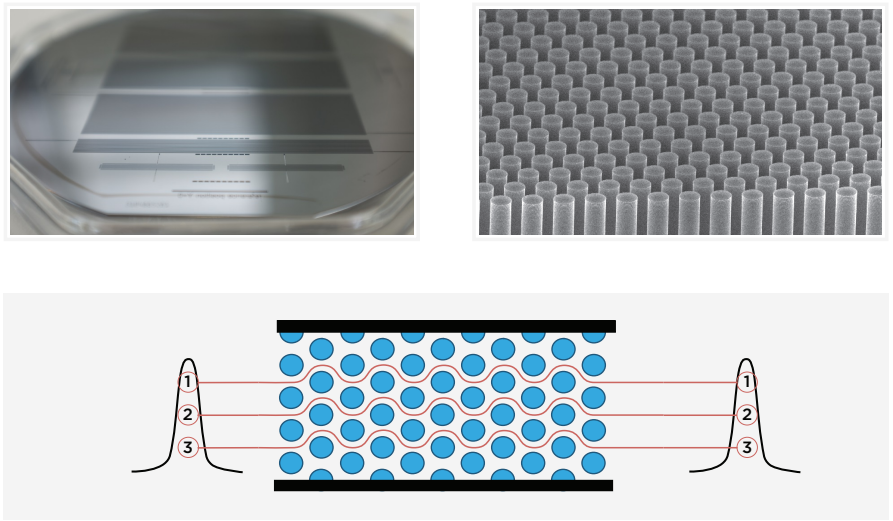
SPONSORED BY

PharmaFluidics BIOGNOSYS NEXT GENERATION PROTEOMICS

LC GC

An additional benefit comes from the ability to position the pillars at an accurately defined distance from each other: the column is very permeable, and the backpressure is low, even for long columns. There are currently two column lengths available: 50 cm and 200 cm. To make long columns on a compact footprint, the columns need to be folded. For example, the 200 cm column consists of 40 discrete lanes interconnected by 180-degree turns, resulting in a serpentine channel. To avoid dispersion effects through the turns (the classical racetrack effect in microfluidics), PharmaFluidics used proprietary flow distribution and collection structures at the beginning and at the end of each lane. In their columns, the pillar diameter is 5 µm and they are positioned at a distance of 2.5 µm from each other. Their height is 20 µm, and their porous layer is about 300 nm thick. The average pore size is estimated at approximately 10–30 nm.

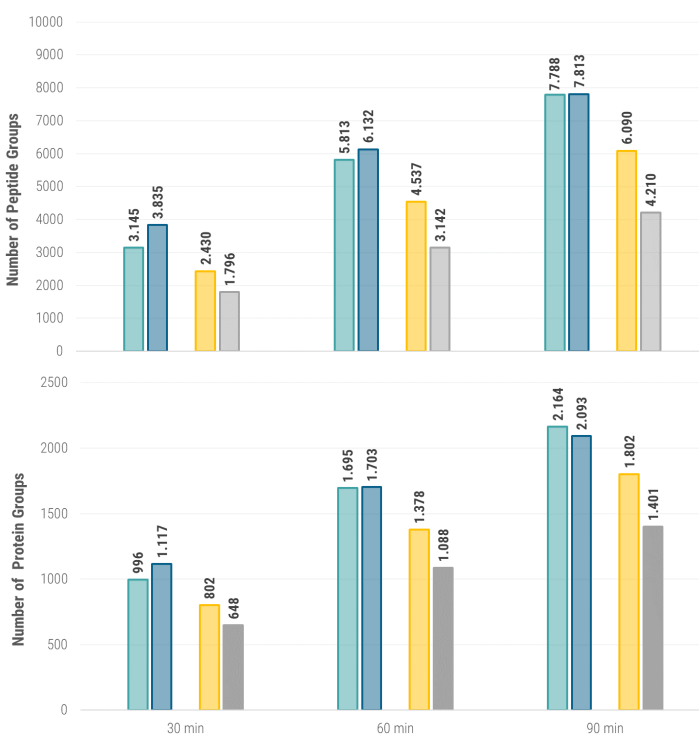
Figure 1: PharmaFluidics' game-changing technology.



Performance Benchmarking

The 50 cm column is ideally suited for short analysis and it does quite well in comparison to 15 cm packed-bed columns. In a study using HeLa cells digest, using short gradients from 30 minutes to 90 minutes and flow rates of either 300 nL/min or 1,000 nL/min, the micropillar array column shows clear reductions in peak widths, and 6–8 times lower backpressure than the conventional shorter columns. With these low backpressures, it is possible to use a

Figure 2: 50 cm µPAC™ - Benchmarking in bottom-up proteomics.



Peptides
300 nl/min → 28 - 75% increase
1000 nl/ min → 28 - 114% increase

50 cm µPAC™	300 nl/min
50 cm µPAC™	1000 nl/min
15 cm Packed Bed #1	300 nl/min
15 cm Packed Bed #2	300 nl/min

Potein Groups
300 nl/min → 20 - 56% increase
1000 nl/ min → 16 - 72% increase

wide range of flow rates, from 0.1 µL/min up to 2 µL/min. The superior chromatographic performance of the 50 cm column allows for increased peptide and protein identification, with up to more than double the number of peptides and 72% more proteins identified in a 60-minute analysis (see **Figure 2**). In addition, due to the solid backbone, the columns are highly durable, capable of running over 3,500 injections without having to be replaced.

The development team benchmarked the 200 cm micro-pillar array column against a 40 cm packed-bed column in a separation of human embryonic kidney (HEK) cell digest. They could identify 3,034 protein groups with the micro-pillar column, compared with only 2,334 protein groups with the packed bed one. The backpressure advantage was also significant: 75 bar versus 445 bar. Another advantage of the new technology is that peak widths do not increase as much when gradient times are extended, as opposed to what is observed with packed columns. Extending gradient times beyond four hours delivered over 5,000 protein and 40,000 peptide identifications in data-dependent acquisition (DDA) mode.

When using electrospray ionization in tandem with MS (ESI–MS) for detection, it is necessary to shunt the high voltage to the ground, as these columns are made of silicon and failure to do so will be detrimental to their chromatographic performance. The schematic drawings in **Figure 3** show how the columns must be connected for different types of ESI configurations.

Application to Shotgun Proteomics and Data-Independent Acquisition (DIA)

Protein molecules are difficult to analyze directly because of their large size and heterogeneous physicochemical properties. To facilitate their analysis, they are cut into smaller units (peptides) by a digestion process, and the peptides are then used as proxies for the quantification of proteins. Peptides are small (i.e., 7–25 amino acids in length) and have very homogeneous chemical properties. They are typically analyzed in a data-dependent acquisition (DDA) fashion, where after chromatographic separation, a mass spectrometry survey scan (MS1) is used to select a peptide for fragmentation and subsequent identification by the second mass spectrum (MS2). This approach often results in semi-stochastic sampling if the scan speed of the MS2 is not fast enough (**Figure 4**).

In contrast, DIA uses the whole MS mass range after the MS1 scan, fragmenting it in different windows to achieve comprehensive sampling. The resulting MS2 spectra are quite complex but can be analyzed using a library—a repository of previously acquired peptide sequences, fragment ions, intensities, and retention times. The available software tools detect the peaks in the MS2 spectra and generate extracted ion chromatograms at the MS1 and MS2 levels, match and

Figure 3: Connectivity - ESI-MS interfacing.

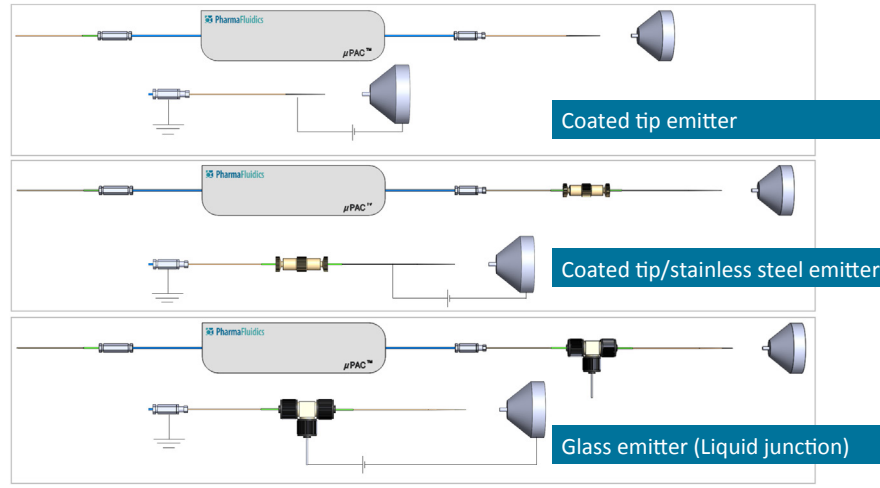
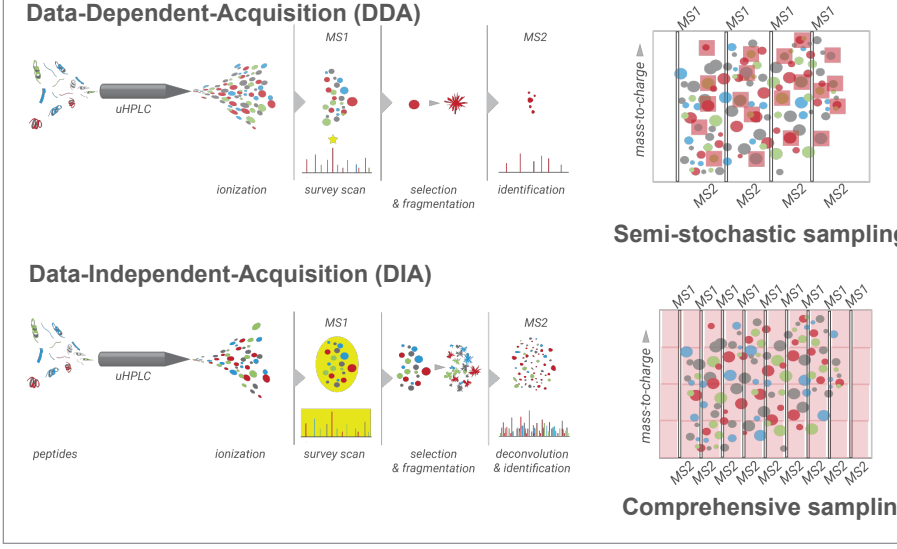


Figure 4: Introduction - Data-Independent-Acquisition (DIA).



score them, and estimate false discovery rates (FDR). A comparison of DIA and DDA methods using HeLa samples with results filtered down to a 1% FDR level shows that DIA delivers twice the number of identifications as DDA. It also shows that, for short gradients, the DIA approach identifies more peptides than what is theoretically possible with the DDA method. To quantitatively benchmark these methods, the development team used a mixed proteome sample of human, *C. elegans*, yeast and *E. coli* spiked at different ratios. Their results show that for all ratios, DIA quantification is closer to the theoretical levels and delivers much better precision and accuracy than DDA.

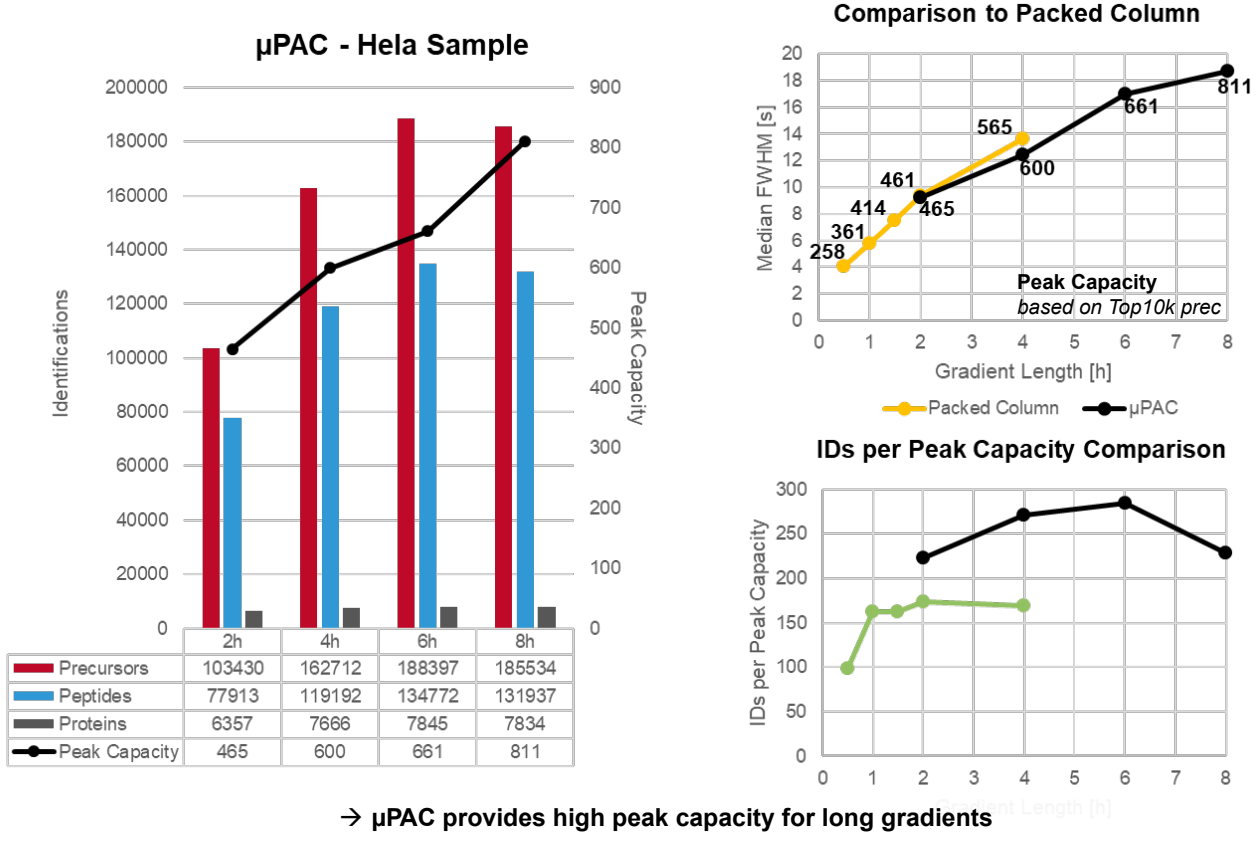
Overcoming the Limitations of Packed-Bed Columns

A team of scientists from Biognosys AG ran a typical single-shot proteomics analysis using a 50 cm column packed with sub-2µm spherical silica particles and quantified nearly 6,000 proteins using a 2-hour gradient. This falls short of the 10,000-mark considered acceptable for comprehensive coverage. Longer columns, which in principle could increase the number of identifications, were not an option because they quickly reached the upper-pressure limit of their LC system. Increasing the gradient length did help, but the efficiency of the analysis, measured in identifications per minute, dropped sharply after a one-hour gradient. In addition, the ratio of identifications to peak capacity remained fairly constant at gradient times longer than an hour, which suggests that to increase the proteomic coverage, the peak capacity for this LC setup must be increased.

The team then switched to a 200 cm micro-pillar array column operating at 300 nL/min heated to 50 °C, using a nanospray ion source and stainless-steel emitters for the MS instrument to perform the same analysis. Using the available dedicated software, they developed optimized non-linear gradients based on calculations from linear gradients running from two to eight hours, with similar FWHM across the whole gradient. Under these conditions, they were able to identify about 100,000 precursors in their analysis of analogue HeLa samples. More impressively, the peak capacity increased in an almost linear way with gradient time (**Figure 5**). In terms of identifications per peak capacity, they found an almost stable line with respect to gradient time, except for the eight-hour gradient, which was derived from the six-hour data and was not fully optimized.

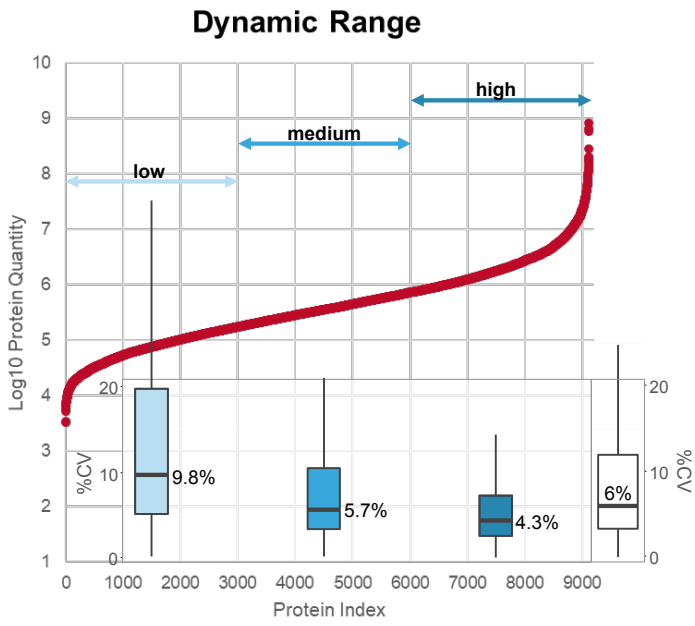
To optimize the DIA workflow the team first studied the effect of increasing the peptide sample loading amount and

Figure 5: Introduction of µPAC at Biognosys – Initial data.



→ **µPAC provides high peak capacity for long gradients**

Figure 6: Overview of results – Quantified proteins.



- 9,123 protein IDs in AVG per run
  - 1% protein FDR
  - 9,173 protein IDs in triplicates
  - 9,121 to 9,124 protein IDs per run
- > 5 orders of magnitude
- median technical CV = 6%

→ among the highest ever reported protein IDs within a single-shot analysis (1% FDR)

found that 4 µg was the optimal load for getting maximum identifications by using longer (>120 min) gradient times. They then determined that the best scan resolution for the MS2 was 30k, based on a target of five data points per peak. Finally, based on overall performance, they chose the six-hour gradient for further benchmarking.

To build the required library for the targeted analysis of the DIA data, the team used high pH reversed-phase fractionation of HeLa samples, using the same LC setup to generate 15 fractions. This yielded more than 300,000 precursors and slightly more than 10,000 protein groups. With this library, they identified 9,123 proteins per run on average, keeping the FDR at 1% or less. These results rank among the highest ever reported protein identifications in a single-shot analysis mode and with controlled protein FDR (**Figure 6**). The DIA approach is also applicable to a variety of analyses, such as DNA replication, because it provides good coverage of major metabolic and cell growth pathways. It can be used for systems involving highly abundant, medium, and lowly abundant proteins.

Conclusion

Micro-pillar array columns (µPAC™) deliver high peak capacity even at long gradient times, with high reproducibility, sensitivity, and precision. They produce lower backpressure than conventionally packed columns. They have excellent scalability for proteomic experiments and they deliver robust separations. When used in combination with DIA methods, they enable single-shot comprehensive proteome analysis delivering more than 9,000 protein identifications in about six hours.