bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ScanningSWATH enables ultra-fast proteomics using high-flow chromatography and minute-scale gradients

Christoph Messner*[1], Vadim Demichev*[1,2], Nic Bloomfield[3], Gordana Ivosev[3], Fras Wasim[3], Aleksej Zelezniak[1,4], Kathryn Lilley[2], Steven Tate[3] and Markus Ralser[1,5+]

1. The Francis Crick Institute, Molecular Biology of Metabolism laboratory, London, United Kingdom
2. Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom
3. SCIEX, Toronto, Canada
4. Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden
5. Department of Biochemistry, Charité Universitätsmedizin Berlin, Berlin, Germany

- These authors contributed equally

\+ To whom correspondence should be addressed. Markus.Ralser@crick.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Rapidly emerging applications in data-driven biology and personalised medicine call for the development of fast and reliable proteomic methods. However, the use of fast chromatographic gradients is limited by the mass spectrometers' sampling rate and signal interferences. Here we present scanningSWATH, a data-independent acquisition method, in which the DIA-typical stepwise windowed acquisition is replaced by continuous scanning with the first quadrupole. ScanningSWATH enables ultra-fast duty cycles as well as to assign precursor masses to the MS2 fragment traces. Furthermore, we have implemented the support for scanningSWATH in DIA-NN, a fully-automated software suite designed to deconvolute fast DIA experiments, which corrects for signal interferences. We show that the combination of scanningSWATH and DIA-NN enables the efficient application of high-flow liquid chromatography (800 µL/min flow rate) to proteomics. High-flow scanningSWATH increases proteomic sample throughput to the minute-scale, while the use of high-flow chromatography hardware improves reliability and robustness. Benchmarking on yeast and human cell lysates, we demonstrate that the proteomic depth achieved with five-minute high-flow scanningSWATH gradients is comparable to that obtained with several times slower nano- and microflow chromatographic gradients, even if compared to most recent studies that used microflow-SWATH or Evosep-DIA methods. The combination of scanningSWATH, advanced data processing, and industry-standard high-flow LC hardware paves a way for a new generation of cheaper and highly consistent proteomic methods. These allow the recording of hundreds of precise quantitative proteomes per day on a single instrument.

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Proteomes are inherently complex, generating huge analytical challenges. In mass spectrometry-based proteomics, a popular solution has been to decrease the complexity via sample pre-fractionation. Pre-fractionation strategies promote excellent proteome coverage, but at the cost of time and resources as well as the introduction of added variability between samples. Pre-fractionation methods are hence suboptimal for biological experiments and clinical studies that require the processing of large and very large sample series. Data-independent acquisition (DIA) approaches[1], such as SWATH-MS[21–3], have been developed as an alternative. In SWATH-MS, rather than selecting the most abundant precursor ions for fragmentation, the mass spectrometer is configured to cycle through a predefined set of wide precursor isolation windows, thus consistently fragmenting all the precursors within the mass range of interest[2]. This way SWATH-MS boosts the identification numbers and consistency in the analysis of complex samples in single injections[4,5].

However, SWATH-MS creates challenges when it comes to the analysis of raw data generated with fast gradients. Co-fragmentation of co-eluting precursors in SWATH results in highly complex spectra and signal interferences, an effect that is magnified with fast gradients. The development of algorithms that enable the efficient deconvolution of such spectra is still an ongoing process, but several major steps have been achieved recently, and have increased proteomic depth as well as quantification precision in DIA experiments[6–9]. On the hardware side however, SWATH methods still face similar bottlenecks as data-dependent methods, when it comes to acquisition speed of the mass spectrometer. When the gradient length is reduced below approximately 20-30 min, and peak widths become shorter than ~5 seconds, it has been so far challenging to both retain an acceptable number of data points sampled per chromatographic peak (at least 3-4 points at FWHM are essential for accurate peptide quantification[5]), while keeping the isolation window size small, to reduce interferences.

Here we present scanningSWATH, a new DIA acquisition method, which enables to overcome these limitations. ScanningSWATH achieves much faster cycle times than conventional DIA methods, as stepwise windowed acquisition is replaced by continuous scanning with the first quadrupole. Furthermore, a new spectral dimension provided by such scans is exploited to assign the intact precursor masses to the corresponding MS2 spectra. Further, we have expanded our all-in-one DIA data processing software based on deep neural networks[6], DIA-NN, so that scanningSWATH data can be processed in a fully automated fashion, and through a

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

simple-to-use graphical interface. We demonstrate that the combination of scanningSWATH and DIA-NN enables to conduct comprehensive ultra-fast analyses using industry-standard high-flow chromatography (800 µL/min) and minute-scale chromatographic gradients (e.g. 5 min gradients on a short 5 cm column). The resultant workflows hence fully profit from the advantages of ultra-high pressure standard flow chromatography when it comes to chromatographic quality and robustness, while the minute-scale short gradients boost sample throughput to hundreds of samples per day on a single instrument, substantially reducing the cost per proteome measured. Benchmarking our method on human (chronic myelogenous leukemia) lymphoblast cell line K562 and yeast (*Saccharmoyces cerevisae*) whole-cell tryptic digests, we demonstrate the precise and robust quantification of thousands of proteins in the scale of minutes, achieving proteomic depth comparable to what was previously only possible with much longer chromatographic methods. Thus, ScanningSWATH enables the measurement of hundreds of deep and precise proteomes per day on a single instrument platform.

Results And Discussion

Replacing windowed acquisition with continuous scans accelerates the duty cycle and enables to match precursor masses to fragment traces

In conventional SWATH-MS[2], precursor ions are selected for fragmentation via cycling through a predefined set of isolation windows. On a fast QTOF instrument, i.e. the TripleTOF(R) 6600 QqTOF, a maximum MS/MS sampling rate of 67Hz is achieved. As fast chromatographic gradients require short cycle times, in order to retain an acceptable number of data points per peak for precise quantification, SWATH-MS hence necessitates large isolation windows, often in the range of 15 - 30 m/z, resulting in typical duty cycles that are in the range of 2 - 4 seconds[3]. In conventional SWATH, a further reduction of the cycle time can only be achieved by broadening the isolation windows, which results in an increased degree of precursor co-fragmentation and thus more interferences at the MS2 level, affecting both the ID numbers and quantification accuracy.

ScanningSWATH addresses this problem from two angles. First, it draws upon the design of QqTOF instruments, which do not require the accumulation or trapping of ions. In scanningSWATH, precursor ions are fragmented using a “scanning” isolation window, which

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

moves across the m/z range of interest, with fragmentation spectra being continuously recorded by the TOF analyzer (Figure 1). This way the isolation window can be made arbitrary small, provided enough sample is injected for the acquisition not to be sensitivity-limited. These scans can be completed much faster than the canonical windowed SWATH acquisition. For example, our optimised method for a 5-minute chromatographic gradient uses a 5 m/z window with a duty cycle of just 500 milliseconds, to scan over a mass range from m/z 450 to m/z 850.

Second, scanningSWATH benefits from the ability to align precursor and MS^2 fragment masses. The matching is achieved via a chromatogram-type alignment enabled by an additional dimension recorded in scanningSWATH: the signal corresponding to each of the fragments of the precursor first appears and then disappears when the leading margin and the trailing margin of the "scanning" isolation window pass the precursor mass (Figure 1). ScanningSWATH data hence contain information on the mass of the precursor from which each MS/MS fragment trace observed has originated.

To be able to analyse scanningSWATH data, we have incorporated the support for scanningSWATH in our integrated software suite, DIA-NN. DIA-NN is an easy to use and fully-automated open source software for DIA data processing[6]. While DIA-NN improves protein quantification also in typical SWATH application, the tool has been specifically developed to allow effective handling of the highly complex data which arises in fast-gradient analyses, and is designed for the processing of large-scale proteomic data. DIA-NN is available as a command line tool, but is also equipped with an intuitive graphical user interface.

ScanningSWATH facilitates the efficient use of fast high-flow chromatography in proteomics

High-flow (also referred to as standard flow) chromatography is the method of choice in many analytical and industrial disciplines, since it is faster, more reliable and provides better analytical separation when compared to nanoflow or microflow chromatography. In DDA- and DIA-based proteomics, the application of high-flow chromatography has not, however, gained popularity, mainly, because a previously unachieved MS^2 sampling rate would have been required for comprehensive proteome coverage with fast gradients, that produce peak-widths in the seconds-range. This problem is now addressed by scanningSWATH.

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In order to illustrate whether chromatographic improvements are achieved by the use of high-flow chromatography in proteomics, we used an ultra-high pressure standard flow HPLC instrument (Agilent 1290 Infinity II) operated with a flow rate of 800 µL/min. We separated a K562 lymphoblast cell line tryptic digest with a 20 minute linear water to acetonitrile gradient. We compared median full width at half maximum (FWHM) of the elution peaks with a highly optimized microflow setup[6]. High flow-LC achieved much sharper peaks while using a shorter column (Figure 2).

In order to evaluate how fast we can go with the gradient-length while maintaining high peak capacity, we determined peak properties also for shorter gradients, ranging from 1 to 10 minutes. High-flow gradients as fast as 5 minutes resulted in peak separation at least as good as previously obtained with microflow-SWATH-MS (Figure 2B).

High-Flow scanningSWATH enables deep proteome coverage at previously unachieved throughput

The lead application for high-flow scanningSWATH MS is high-throughput proteomics. Although longer columns (e.g. 10cm) would provide narrower peaks, we decided to use a 5 cm column, allowing for shorter washing and equilibration times. We chose 5-minute chromatographic gradient as the one providing high throughput while still allowing a qualitatively excellent chromatographic separation of peptides (Figure 2B).

For a K562 cell line tryptic digest, the combination of high-flow-scanningSWATH and data processing with DIA-NN[6] allowed to quantify 2940 unique proteins (1% protein FDR; only proteotypic peptides used for identification and quantification) and 3800 protein groups (1% precursor FDR) on average in 5 injections, of which 2137 unique proteins had CV values less than 20% (Figure 3A). Analysis of the yeast lysate with 5-minute high-flow-scanningSWATH yielded 1688 unique proteins on average (Figure 3A), while the average number of protein groups identified reached 2509, which is about half of the expressed yeast proteome.

Remarkably, our workflow can measure around 180 samples/day, including all washing and equilibration steps. Furthermore, to visually illustrate the robustness and the advantage of high-flow LC proteomics, we injected 40 human (K562) samples and overlaid the TIC of the first and the last injection. No relevant changes in both chromatography and intensity can be spotted visually (Figure 2A).

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Recording a high number of data-points per peak despite the much faster chromatography, high-flow scanningSWATH retains protein identification characteristics competitive to other, much slower, high-throughput proteomic platforms

Finally, we conducted a simple illustrative benchmark, to compare high-flow-scanningSWATH to a first generation microLC-SWATH method on a TripleTOF(R) 5600 instrument[10] (50 min gradient, 'Microflow-SWATH 2015'), the latest generation microLC-SWATH method on a TripleTOF(R) 6600 instrument[6] (19 min gradient, 'Microflow-SWATH 2018'), and the data obtained with a recent EvoSep-DIA method as recorded on a QExactive HF-X instrument[11] (21 min gradient 'EvoSEP-DIA'). Despite enabling a much higher throughput, the 5-minute high-flow-scanningSWATH quantified more human proteins than the conventional (2015) microflow-SWATH[10], and only slightly less than the latest generation of microflow-SWATH or Evosep-DIA studies. Importantly, while the high-flow-scanningSWATH has the fastest duty cycle (Figure 3E) and demonstrates by far the highest throughput (Figure 3E), it is equal or even better in terms of recording a maximum number of data points per peak (Figure 3E). A high number of data points per peak is important for accurate quantification of samples that are heterogeneous in nature or that have not been analysed in consecutive injections within a short period of time, and thus are not characterised by near-perfect chromatographic overlap. Finally, high-flow-scanningSWATH demonstrates excellent protein quantification precision (Figure 3, C and D).

Concluding remarks

ScanningSWATH is a new acquisition method, that enables ultra-fast duty cycles, required to record proteomes with fast chromatographic gradients, as well as to align precursor and fragment masses in DIA-proteomic experiments. Integrating scanningSWATH support into a comprehensive proteomics software suite, DIA-NN, enabled us to effectively record deep and precise quantitative proteomes using robust high-flow (800µL/min) chromatography. Performing proteomic measurements with 5-minute chromatographic gradients on a 5 cm HPLC column using industry-standard HPLC hardware, we reached a so-far unachieved throughput in the processing of human and yeast quantitative proteomes. Effectively without major compromises in proteomic depth, high-flow scanningSWATH proteomes are precise, cost-effective, and robust, enabling a new generation of high-throughput proteomic workflows that profit from the

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

use of fast industry-standard chromatography to record hundreds of precise proteomes on a single day.

Methods

Sample preparation

MS compatible yeast (*Saccharmoyces cerevisae*) and K562 tryptic digests were obtained from Promega. The digested peptides were dissolved in 3% ACN/0.1 FA and spiked with iRT peptides (Biognosys).

Liquid chromatography - mass spectrometry

Liquid chromatography was performed on an Agilent infinity II ultra-high-pressure system coupled to a Sciex TripleTOF(R) 6600. The peptides were separated in reversed phase mode using a C18 ZORBAX Rapid Resolution High Definition (RRHD) column 2.1mm x 50mm, 1.8 µm particles. A gradient was applied which ramps from 1% B to 35 % B in 5 min (Buffer A: 1%ACN/0.1%FA; Buffer B: ACN/0.1%FA) with a flow rate of 800 ul/min. Linear gradients were used for the FWHM peak estimations and non-linear gradients for all other measurements. For washing the column, the organic solvent was increased to 80 % B in 0.5 min and was kept for 0.2 min at this composition before going back to 1% B. The equilibration time between the runs was 1.2 min. The flow rates for washing and equilibration were 1.2 and 1.0 ml/min, respectively. A 2-pos/6-port post- column valve diverted the eluate to the waste for washing and equilibration. The scanningSWATH precursor isolation window was set to 5 m/z and a mass range from m/z 450 to m/z 850 was covered in 0.5s. Data was acquired in high sensitivity mode and the amount of total proteins injected was 2.5 ug (peak width estimations) and 5 ug (benchmarks). An IonDrive Turbo V Source was used with ion source gas 1 (nebulizer gas), ion source gas 2 (heater gas) and curtain gas set to 50, 40 and 25. The source temperature was set to 450 and the ionspray voltage to 5500V.

Microflow SWATH (K562)

Liquid chromatography was performed on a nanoAcquity UPLC (Waters). Peptides (2µg) were separated with a 19-minute non-linear gradient starting with 4% acetonitrile/0.1 % formic acid and increasing to 36% acetonitrile/0.1% formic acid. A Waters HSS T3 column (150mm x 300µm, 1.8µm particles) was used and the flow rate was set to 5µl/min. The DIA method consisted of an MS1 scan from m/z 400 to m/z 1250 (50ms accumulation time) and 40 MS2

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

scans (35ms accumulation time) with variable precursor isolation width covering the mass range from m/z 400 to m/z 1250.

Data analysis

Data analysis was performed with DIA-NN (version 30/05/2019, available at https://github.com/vdemichev/DiaNN/tree/ed84eddd248a88c786ebe9fd447e09c01ef70b2d) using the default settings, except the mass accuracy was fixed to 20ppm, as we have previously established this value to perform well for all the data types analysed.

Full width half maximum (FWHM) estimations

Median peak FWHM was estimated using Spectronaut (11.0.15038.23.25164 (Asimov); Biognosys), to allow for comparison of the values obtained with those generated with Spectornaut previously by other labs. Only precursors ubiquitously identified in all runs (1 min, 3 min, 5 min, 10 min and 20 min high-flow as well as with the 19 min microflow run) and with a Qvalue of < 0.000000001 were considered (in total 1755 precursors).

Library generation for Yeast

The library were generated by "gas-phase fractionation" using scanningSWATH and small precursor isolation windows. 5-10 ug of Saccharomyces cerevisiae extracts were injected and run on the microflow setup using a 55 min linear gradient ramping from 3% ACN/0.1FA to 40% ACN/0.1% FA. In total 11 injections were run with the following mass ranges: m/z 400-450, m/z 445-500, m/z 495 - 550, m/z 545-600, m/z 595-650, m/z 645 - 700, m/z 695 - 750, m/z 745 - 800, m/z 795- 850, m/z 845 - 900, m/z 895 - 1000 and m/z 995 - 1200. The precursor isolation window was set to m/z 1 except for the mass ranges m/z 895 - 1000 and m/z 995 - 1200, where the precursor windows were set to m/z 2 and m/z 3, respectively. The cycle time was 3sec consisted of high and low energy scan. A spectral library was generated using DIA-NN directly from these scanningSWATH acquisitions.

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank Roland Bruderer (Biognosys) for providing the human spectral library. This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001134), the UK Medical Research Council (FC001134), and the Wellcome Trust (FC001134), and received specific funding from the BBSRC (BB/N015215/1 and BB/N015282/1), as well as a Crick Idea to Innovation (i2i) initiative (Grant Ref 10658).

References

1. Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).
2. Gillet, L. C. *et al.* Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (06/2012).
3. Ludwig, C. *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018).
4. Bruderer, R. *et al.* Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410 (05/2015).
5. Bruderer, R. *et al.* Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* **16**, 2296–2309 (12/2017).
6. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: Neural networks and interference correction enable deep coverage in high-throughput proteomics. *bioRxiv* 282699 (2018). doi:10.1101/282699
7. Ting, Y. S. *et al.* PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nat. Methods* **14**, 903–908 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8. Peckner, R. *et al.* Specter: linear deconvolution for targeted analysis of data-independent acquisition mass spectrometry proteomics. *Nat. Methods* **15**, 371–378 (2018).

9. Heaven, M. R. *et al.* microDIA (μDIA): data-independent acquisition for high-throughput proteomics and sensitive peptide mass spectrum identification. *Anal. Chem.* **90**, 8905–8911 (2018).

10. Vowinckel, J. *et al.* Cost-effective generation of precise label-free quantitative proteomes in high-throughput by microLC and data-independent acquisition. *Sci. Rep.* **8**, 4346 (2018).

11. Bache, N. *et al.* A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics. *Mol. Cell. Proteomics* **17**, 2284–2296 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

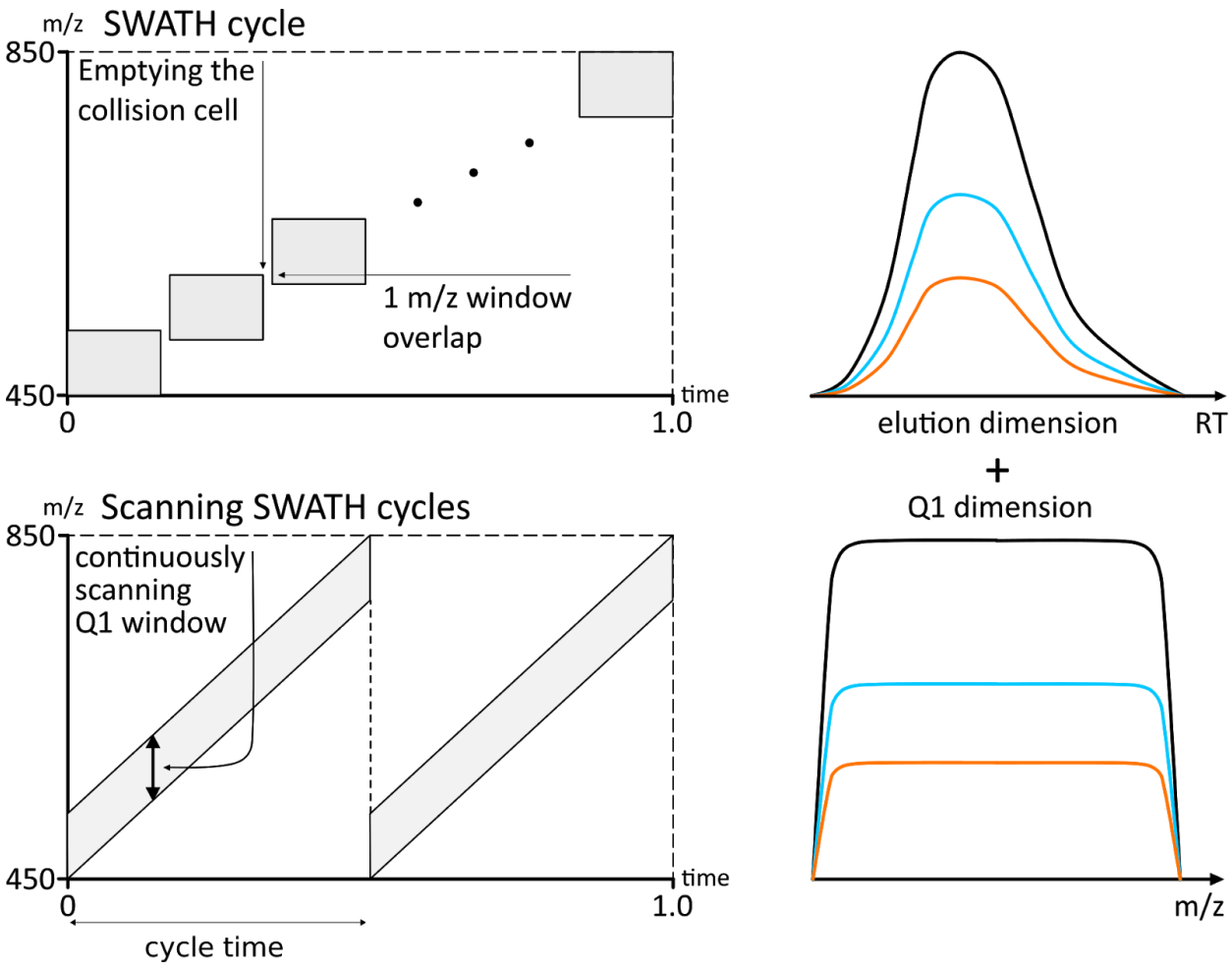
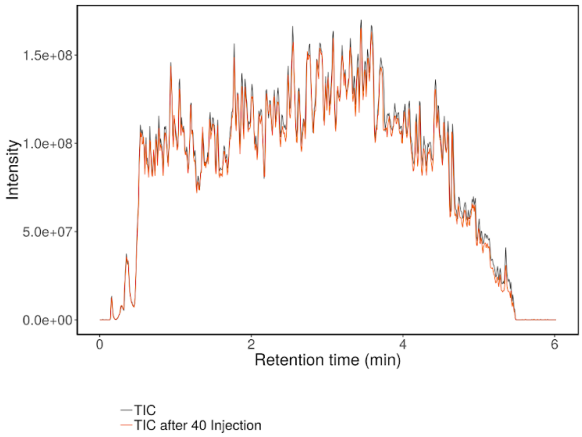


Figure 1. **ScanningSWATH 'decouples' duty cycle and isolation window size, and enables to assign precursor masses to the MS2 space**. In SWATH [2] (**top panel**), the mass spectrometer repeatedly fragments all the precursor ions (i.e. peptides bearing a specific charge) within a specified mass range via cycling through a predefined set of fairly wide Q1 isolation windows. The resulting fragmentation spectra are then recorded by the mass analyser. (**bottom panel**) ScanningSWATH uses instead a 'scanning' Q1 isolation window, which continuously moves across the precursor mass range of interest, with the mass analyzer continuously recording the fragmentation spectra of the precursors that fall into the window at a given moment in time. This effectively 'decouples' the window size from the cycle time, enabling high selectivity of precursor isolation with ultra-fast gradients. Furthermore, an extra dimension of information is provided, and enables to assign precursor masses to the MS2 space. When the window scans across the mass range, each precursor only reaches the collision cell when its mass has been passed by the leading margin of the window but has not yet been passed by the trailing edge, its fragments having approximately 'square' detection profiles in the Q1 dimension (bottom right; x-axis corresponds to the window position and y-axis to the signal intensity for each of the fragments, represented with different colours). Since the position (in Q1 dimension) of the isolation window is known at each moment in time, each fragment trace observed in the data contains information about the respective precursor mass, allowing to improve the confidence of precursor identification.

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

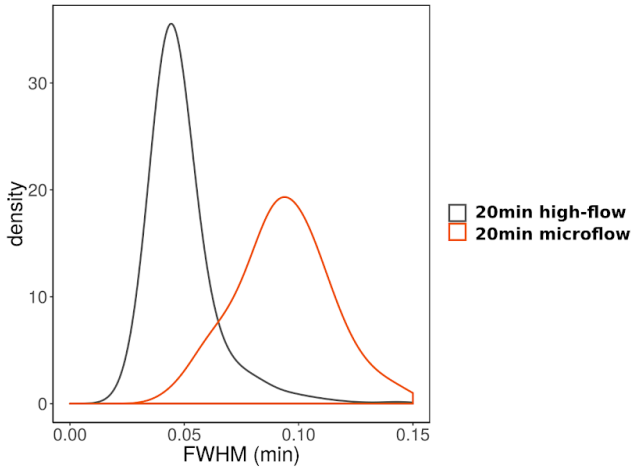
(A)



(B)

Gradient Length (min)	Flow rate (ul/min)	Column length (mm)	Median FWHM (s)	Gradientlength/ FWHM
1	800	50	1	61
3	800	50	1.1	163
5	800	50	1.3	225
10	800	50	1.9	323
20	800	50	2.7	440
19	5	150	5.7	201

(C)



(D)

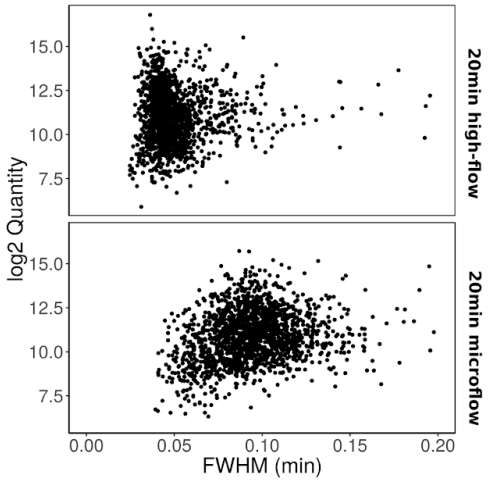


Figure 2. **High-flow chromatography is highly robust and yields significantly sharper peaks, allowing for better peptide separation in proteomics**. (**A**) Human (K562) cell lysate tryptic digest was injected 40 times in row and analysed on TripleTOF(R) 6600 with a 5 min high-flow (800 µl/min) gradient. Total ion chromatograms for the first and the last injections (plotted) show almost perfect overlap, reflecting the high robustness of this LC-MS setup. (**B**) We also compared the median peak widths at FWHM (full width at half maximum) for high flow (800 µl/min; 1, 3, 5, 10, and 20 min gradients) and microflow (5 µl/min; 20 min gradient) setups (see Methods) We observed that despite a three-times shorter column, the use of high-flow LC resulted in a significantly sharper peaks, with the peak capacity (right column, FWHM) being higher for high-flow at 5 minutes than for microflow at 20 minutes. We also illustrate the advantage of high-flow at the same 20 minute gradient, evidenced as a marked difference in the peptide peak width distributions (**C**, **D**).

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

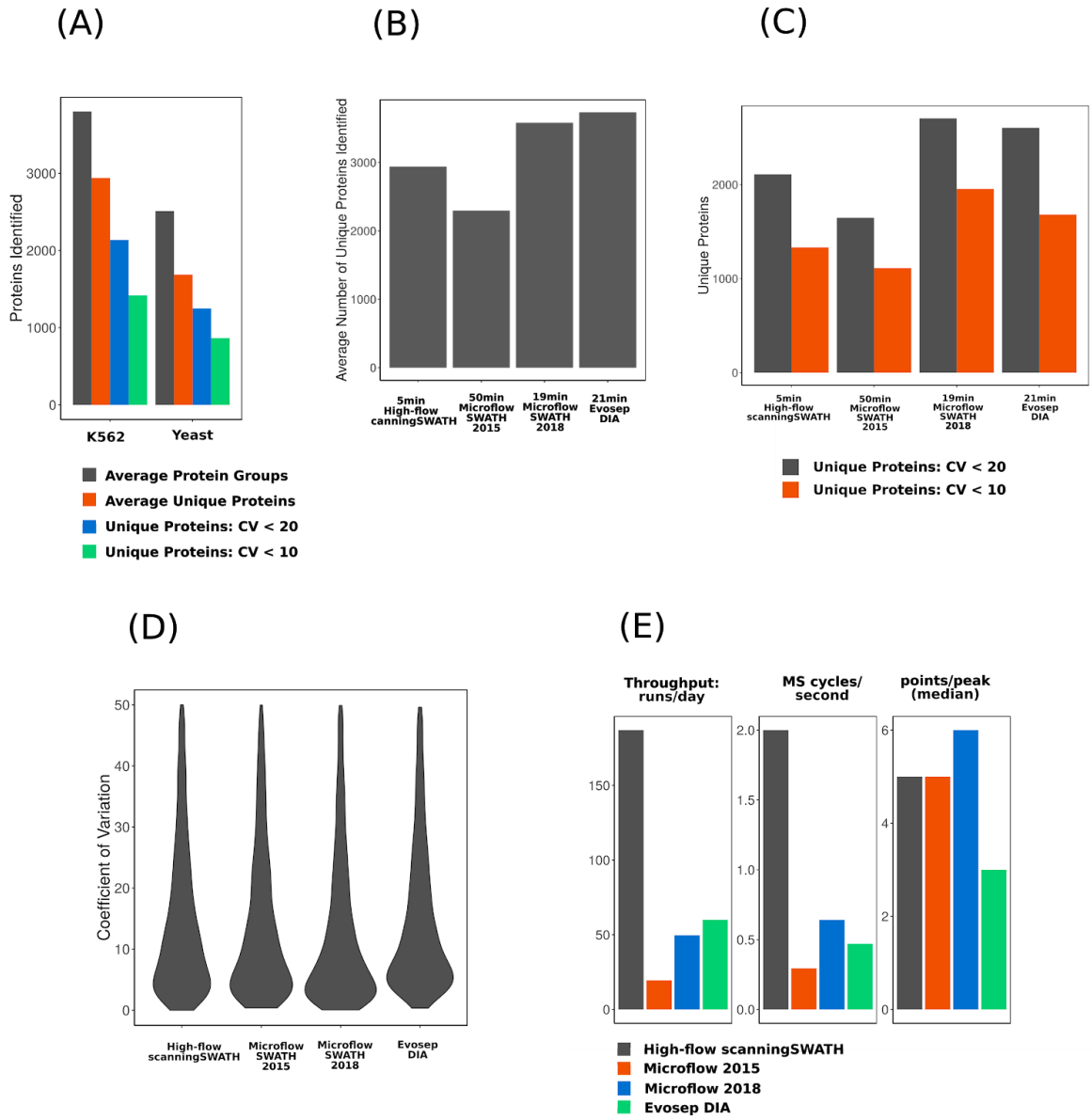


Figure 3. **5-minute gradient ScanningSWATH coupled with high-flow HPLC allows for comprehensive protein quantification and is competitive in comparison with methods that have several-times lower throughput**. **(A)** Protein identification with high-flow ScanningSWATH. We benchmarked the 5-minute gradient high-flow-scanningSWATH workflow on TripleTOF(R) 6600 using human (K562) and yeast (*S. cerevisiae*) cell lysate tryptic digests. The data were processed with DIA-NN[6]. A public HeLa-HEK293T spectral library (204098 precursors) previously generated via pre-fractionated sample analysis using data-dependent acquisition on an Orbitrap instrument[5] was used for the analysis of the human samples. The *S. cerevisiae* sample was analysed with a spectral library obtained from the analysis of 11 scanningSWATH gas-phase fractionation acquisitions with DIA-NN (Methods). Protein group numbers identified were obtained at 1% precursor q-value and the numbers of unique proteins (i.e. proteins identified and quantified using proteotypic peptides only) were calculated using

bioRxiv preprint doi: https://doi.org/10.1101/656793; this version posted May 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

simultaneous 1% precursor q-value and 1% protein q-value filtering (**B**). Identification numbers and quantification performance of the 5-minute gradient high-flow scanningSWATH on TripleTOF(R) 6600 method, in comparison with a first-generation (2015)[10] and a latest-generation (2018; see Methods)[6] microflow SWATH (method run on TripleTOF(R) 5600 and 6600, respectively) as well as a recently published Evosep-DIA study[11], with the data recorded on a Q Exactive HF-X Orbitrap. Human cell lysate tryptic digests were used for comparison. All SWATH methods were benchmarked with a K562 digest (first generation microflow SWATH data were generated previously for the respective publication[10]). Please note that the Evosep-DIA data were acquired using a HeLa digest and not K562 digest, and hence the expected ID numbers are similar but not identical. The comparison is however designed to give an advantage for EvoSep-DIA [11], as an Orbitrap-based HeLa-HEK293T spectral library was used for the analysis with DIA-NN[6]. Five injection replicates were analysed for each of the SWATH methods, while as inherent to the Evosep method, the Evosep-DIA data were obtained with each of the five replicates undergoing separate solid-phase extraction in an Evotip. The numbers of unique proteins (i.e. proteins identified and quantified using proteotypic peptides only) were calculated, using simultaneous 1% precursor q-value and 1% protein q-value filtering. Averages (**B**), as well as numbers of unique proteins quantified with CV values below 20% or 10% (**C**) were calculated. We also present CV value distributions (**D**) as well as the throughput estimates and the numbers of data points per peak for the methods benchmarked (**E**). In conclusion, despite boosting the sample throughput to hundreds of proteomes per day, high-flow-scanningSWATH maintains an identification and quantification performance that is comparable to that of slower high-end proteomic methods.