Rapid Proteome Analysis Using DIA and Super-Resolution Orbitrap Mass Spectrometry

Sophia Steigerwald[1]; Lili Niu[2]; Kyle Fort[3]; Arne Kreutzmann[3]; Daniel Marc Mourad[3]; Konstantin Aizikov[3]; Dmitry Grinfeld[3]; Alexander A. Makarov[3]; Florian Meier[1,4]; Matthias Mann[1, 2]

[1]Max Planck Institute of Biochemistry, Planegg, Germany; [2]NNF Center for Protein Research University of Copenhagen, Copenhagen, Denmark; [3]Thermo Fisher Scientific (GmbH) Bremen, Germany; [4]Jena University Hospital, Jena, Germany

ABSTRACT

Purpose: Real-time ΦSDM (Phase-Constrained Spectrum Deconvolution Method) processing of full range mass spectra for rapid proteomics using the EvoSep One and data independent acquisition (DIA).

Methods: An auxilliary computer was used to enable real-time, full mass range ΦSDM processing of MS1 and MS2 spectra without introducing scan overhead.

Results: The two-fold increase of mass resolving power at a constant cycle time, enabled by ΦSDM signal processing, allowed for deeper analyses of full proteome samples of HeLa and blood plasma digests using short gradients. This approach increases protein identification by up to 14% while additionally improving overall protein quantification.

INTRODUCTION

With increasing focus on large-scale mass spectrometry studies, especially in clinical proteomics, there is a growing need for robust and sensitive high-throughput methods. Recent technological advances, such as preformed gradients in the Evosep One LC system, allow for the analysis of up to three hundred samples per day. Although appealing for high-throughput applications, these short gradients provide less coverage of the proteome and necessitate shorter cycle times when using standard signal processing enhanced Fourier transform (eFT) methods [1] in data-independent acquisition (DIA) strategies. Although this can be accommodated by increasing spectral complexity (larger DIA windows) or lowering mass resolution, data quality is reduced.

To counteract this, we implement the Phase-Constrained Spectrum Deconvolution Method (ΦSDM) [2] for Orbitrap™ signal processing. Previously limited to narrow m/z regions, e.g. TMT reporter ions, an auxiliary computer enables ΦSDM processing of full mass range spectra, thereby increasing acquisition speed for improved spectra quality and protein identification using short gradients.

MATERIAL AND METHODS

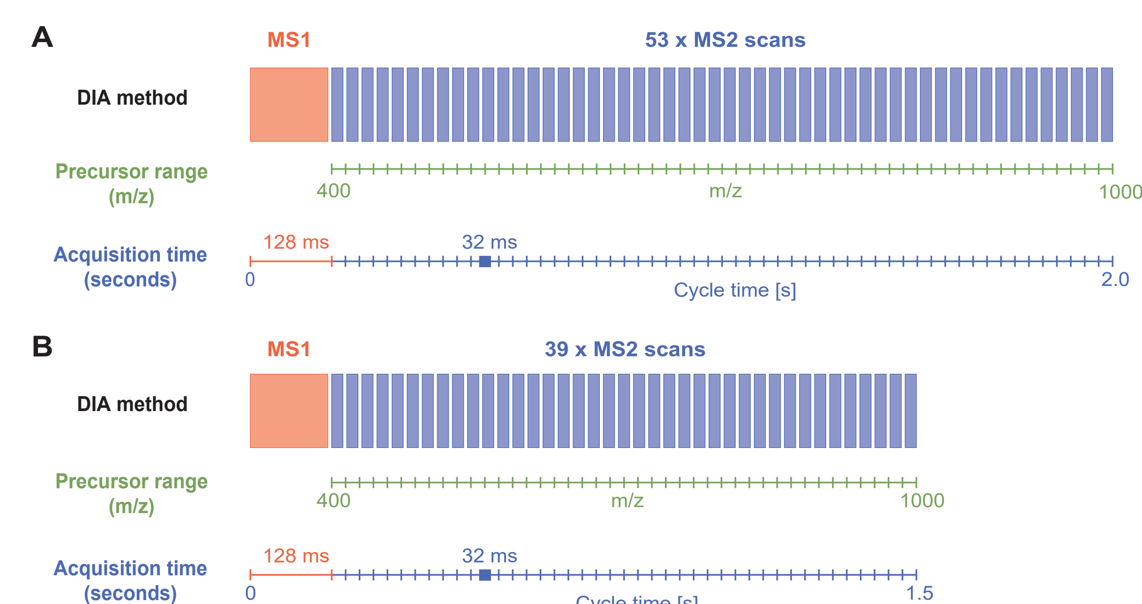
Sample Preparation

HeLa cells were subjected to cell lysis, reduction and alkylation followed by overnight digestion using equal amounts of LysC and trypsin. Digested samples were purified on Strata™-X-C cartridges prior to loading on C18 Evotips for online MS injection using the Evosep One liquid chromatography (LC) system. Pooled blood plasma samples were reduced and alkylated prior to overnight digestion using equal amounts of LysC and trypsin. Digests were directly loaded and cleaned up on C18 Evotips. All measurements were performed in triplicates.

Methods

Purified whole proteome HeLa and blood plasma digests were separated using the Evosep One LC system coupled to a Thermo Scientific™ Orbitrap Exploris™ 480 mass spectrometer. The DIA acquisition cycle comprised one full scan (128 ms transient) and 39 or 53 equidistant DIA windows (32 ms transient) in the mass range 400-1000 m/z for a 1.5 s or 2 s cycle time, respectively. Acquisition cycle times of the methods were adjusted to the set of standardized Evosep One gradients corresponding to 200 (5 min), 100 (12 min), 60 (21 min) and 30 (44 min) samples per day (SPD) to ensure ≥3 data points per chromatographic peak. To enable ΦSDM signal processing of up to 128 ms transients in real-time, we employed an auxiliary computer equipped with four Titan X Nvidia™ GPU. See for visualization of the acquisition schemes.

Figure 1. Optimized DIA acquisition schemes for different gradient lengths
A) DIA method with theoretical cycle time of 2 s. Optimized for Evosep One gradients corresponding to 30 and 60 SPD.
B) DIA method with theoretical cycle time of 1.5 s. Optimized for Evosep One gradients corresponding to 100 and 200 SPD.



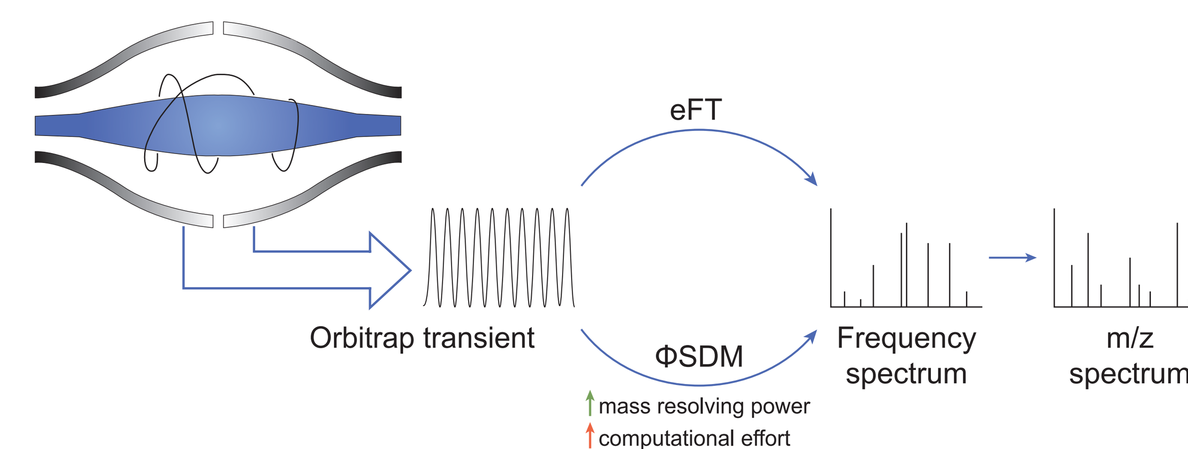
Data analysis

Data extraction from DIA experiments was performed using the Biognosys Spectronaut software (version 15) in the library-free (directDIA) mode or with a project- and gradient-specific library. DirectDIA analyses were performed using the default settings. For library-based DIA analys es default settings were used for all parameters except quantification. The quantification strategy was set to "MaxLFQ", row filtering was set to "Qvalue" and "Local Normalization" was selcted for data normalization.

SUPER-RESOLUTION FTMS WITH ΦSDM

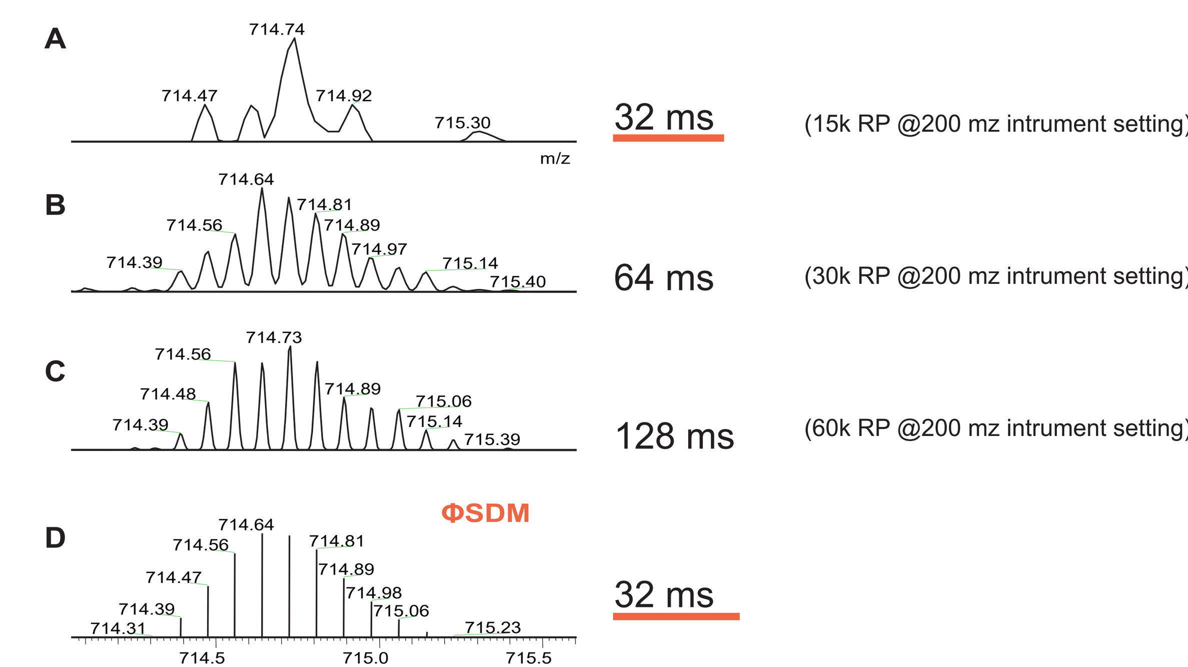
Fourier Transform (FT) signal processing is inherently limited in its resolution by the frequency uncertainty. While several interpolation techniques have been proposed to overcome the limitations, these cannot increase spectral information content, the main factor limiting the resolving power. ΦSDM, on the other hand, offers a computational approach to increase the spectral information content on the basis of information readily available in Orbitrap mass analyzers: oscillation phases and collisional decay constants [1]. The computational ΦSDM approach builds on complex-valued FT spectra and the calibrated initial phase of the ion signal in the time domain. Centroiding, followed by conversion to the m/z domain result in refined mass distribution and enables ΦSDM to achieve a superior resolving power in comparison to that of FT methods.

Figure 2 Schematic of the signal processing



The advantage of ΦSDM is that it requires substantially shorter transient times to achieve the same mass resolution relative to standard FT based approaches. As an example, **Figure 3** shows full scan spectra of a small protein, Ubiquitin, at charge state z = +12. Standard eFT processing of 32 ms transients (instrument setting 15,000 at m/z 200) does not provide sufficient resolution to separate individual isotopes (**Figure 3A**). Using ΦSDM (**Figure 3D**), all isotopes are baseline resolved at the very same 32 ms transient length. Conventional eFT processing requires ≥64 ms long transients to achieve comparable results (**Fig. 3 B-C**).

Figure 3. Real-time full mass spectra of charge state +12 of ubiquitin
A) Spectrum acquired with a 32 ms transient (resolution 15,000 at m/z 200 with eFT processing). Isotopes are not resolved.
B) Spectrum acquired with a 64 ms transient (resolution 30,000 at m/z 200 with eFT processing). Isotopes are not baseline resolved.
C) The isotope envelope of ubiquitin is baseline resolved using a 128 ms transient (resolution 60,000 at m/z 200 with eFT processing).
D) With ΦSDM processing, isotopes are resolved using a 32 ms transient.



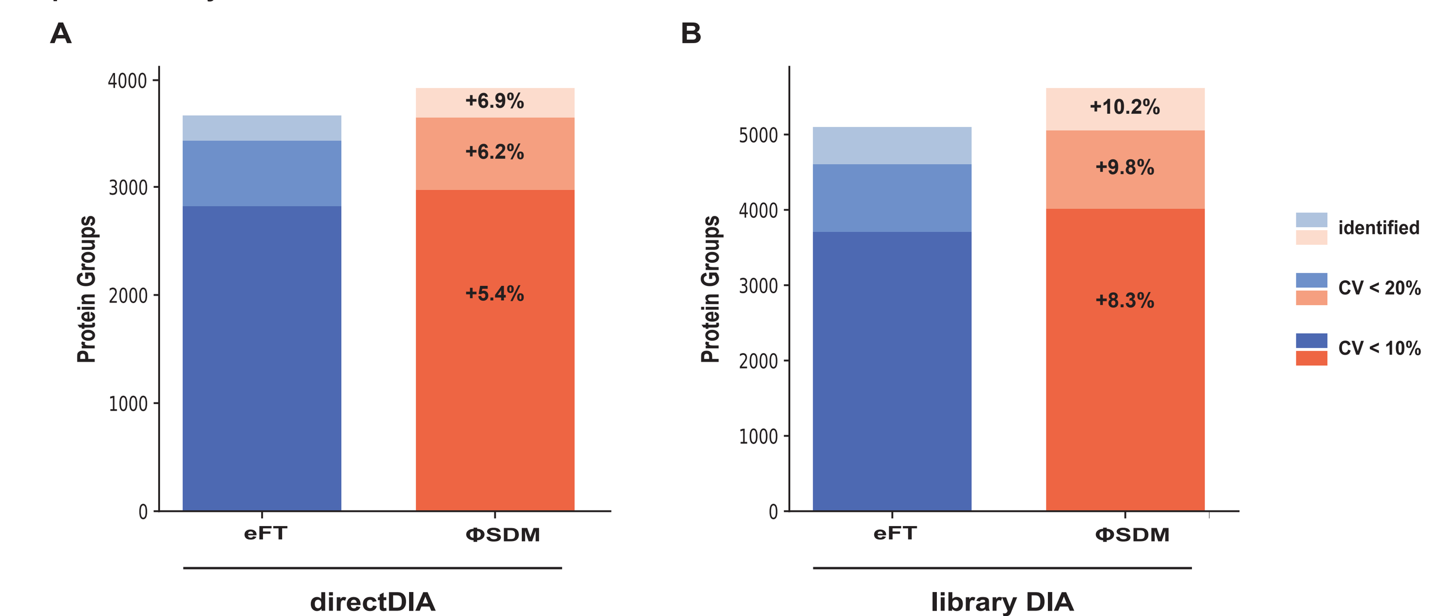
While ΦSDM processing of narrow m/z ranges can be computed on the standard instrument PC, applying the algorithm to the full mass range comes with a significantly higher computational effort. To avoid additional scan overhead time, which is almost 4 times longer than the corresponding transient, an auxiliary computer equipped with four Titan Xp Nvidia™ GPU cards is used for efficient data transfer and parallelizing these calculations. In proteomics practice, we found that the mass resolving capability can be increased two-fold without increasing acquisition cycle times.

DIA WITH SHORT GRADIENTS

Technical advances in LC technology, such as preformed gradients in the Evosep One LC system, enable rapid proteome analysis with minimal measurement overhead time, allowing the acquisition of up to 300 samples per day [4]. While greatly increasing throughput, these methods often necessitate shorter cycle times in DIA acquisition strategies, increasing spectral complexity as well as decreasing proteome depth. By increasing mass resolving power, while maintaining transient times, ΦSDM enables faster acquisition cycles and helps to overcome these limitations without compromising spectral quality.

In combination with the Evosep 21 min gradient (60 SPD) and a DIA cycle time of 2 s (Figure 1A), ΦSDM increased precursor identifications by ~15% and ~19% and protein IDs in triplicate HeLa measurements by ~7% and 10% for directDIA and library-based DIA, respectively (**Figure 4**). Collectively, we identified almost 4000 protein groups (PG) in directDIA (**Figure 4A**) and 5600 PG using a project-specific deep library (**Figure 4B**).

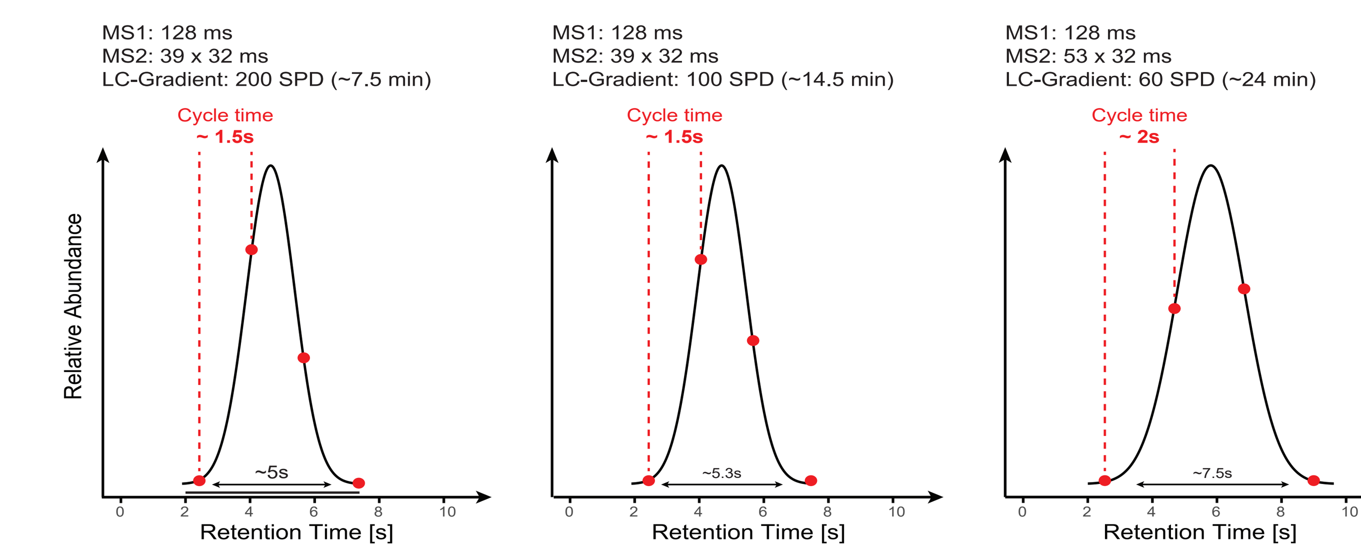
Figure 4. Improved protein identification and quantification using full range ΦSDM signal processing for triplicate HeLa measurements using a 21 min gradient and A) a directDIA (library-free) approach or B) a project-specific spectral library.



Further extending our study to even shorter gradients, we optimized the DIA cycle time to balance mass resolution and cycle time for each gradient. As short LC gradients often lead to peak widths of <10 s, this can be quite challenging. On the one hand, longer cycle times and MS transients result in a decrease of spectral complexity, which can increase identification rates. However, when working with such short as the 200 SPD method, longer cycle times can also significantly decrease protein quantification accuracy by reducing the number of data points per chromatographic peak.

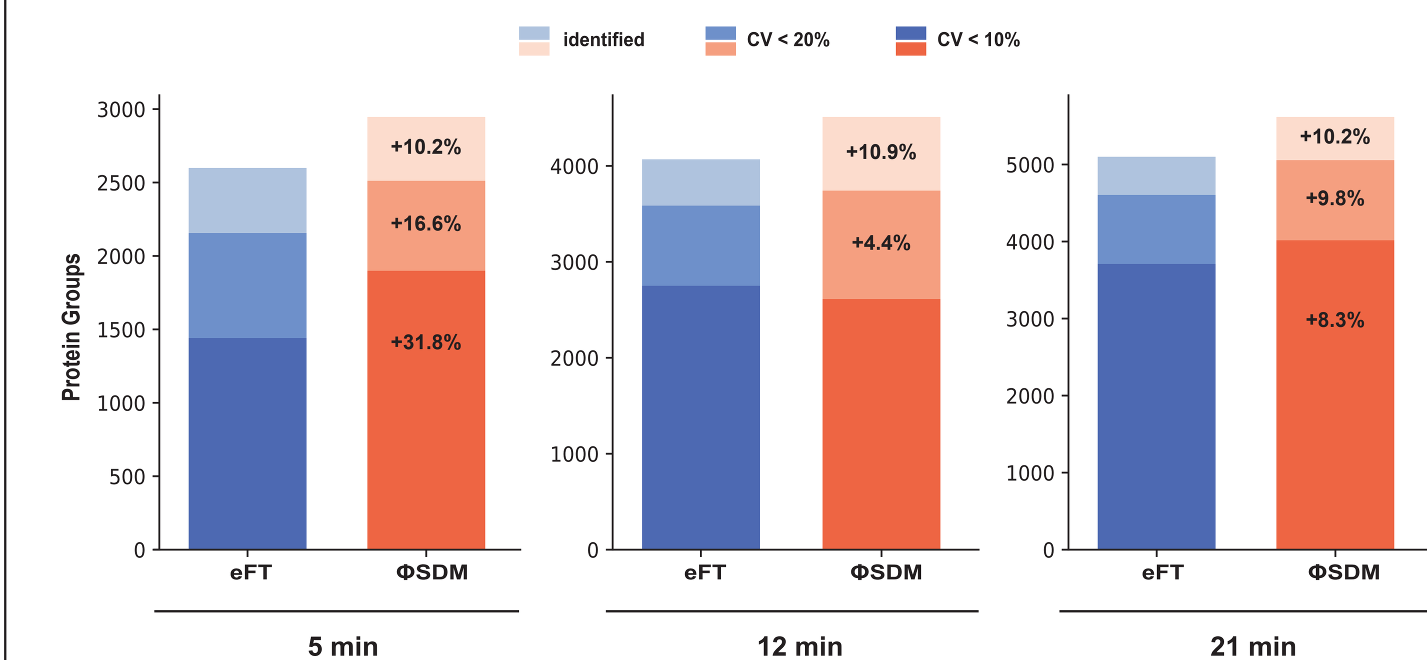
We found that the 200 and 100 SPD methods (corresponding to 5 min and 12 min gradient lengths respectively) showed similar average peak widths around 5 - 5.3 s (at baseline), while for the 60 SPD (21 min gradient length) the average peak width was around 7.5 s. Consequently, we used DIA met hods using cycle times of 1.5 s and 2 s respectively with transients of 128 ms (MS1) and 32 ms (MS2), as shown in **Figure 5**.

Figure 5. Balancing mass resolving power and cycle time in DIA acqusition strategies using short gradients



Using the aforementioned short gradients in triplicate HeLa measurements, we show that applying ΦSDM signal processing to the full mass range robustly improves protein identification and quantification (**Figure 6**), increasing proteome depth by at least 10% for all tested gradients. Particularly the 5 min gradient (200 SPD) benefits from the more sophisticated processing algorithm. This enables the identification of almost 3000 PG from 100 ng of HeLa digest in such a short gradient, as well as increasing the number of proteins quantified with a coefficient of variation (CV) below 20% and 10%.

Figure 6. Average number of identified and quantified proteins in triplicate HeLa measurements using short Evosep gradients.



CONCLUSION

ΦSDM achieves comparable mass resolution as standard eFT signal processing in half or less than the transient time. An additional auxiliary computing unit now enables full mass range ΦSDM signal processing in real-time without introducing any scan time overhead. Application to rapid DIA analysis of HeLa using short Evosep gradients consistently showed improvements in protein identification and quantification. Especially very short gradients, such as the 5 min Evosep method, benefitted greatly from the increased acquisition speed. In conclusion, we have demonstrated that super-resolution FTMS signal processing using the ΦSDM algorithm has a great potential for DIA combined with short LC gradients.

REFERENCES

[1] O. Lange, et al., Int. J. Mass Spectrom. 377, 338–344, 2015.
[2] D. Grinfeld et al., Anal. Chem., 89 (2), 1202–1211, 2017.
[3] C. D. Kelstrup et al., J. Proteome Res., 17 (11), 4008–4016, 2018.
[4] N. Bache et al., Mol Cell Proteomics., 17(11), 2284-2296, 2018
[5] V. Ignjatovic et al., J Proteome Res., 18(12):4085-4097, 2019

ACKNOWLEDGEMENTS

This research is partially supported by the European Union's Horizon 2020 research and innovation program under grant agreement 686547 (MSmed project).

TRADEMARKS/LICENSING