Rapid Proteome Analysis with Data-Independent Acquisition and Super-Resolution Orbitrap Mass Spectrometry

Florian Meier[1], Arne Kreutzmann[2], Daniel Mourad[2], Konstantin Aizikov[2], Dmitry Grinfeld[2], André C Michaelis[1], Oliver Lange[2], Alexander A Makarov[2], Matthias Mann[1, 3]

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

ABSTRACT

Purpose: Real-time ΦSDM (Phase-Constrained Spectrum Deconvolution Method) processing of full range mass spectra for rapid shotgun proteomics with data independent acquisition (DIA).

Methods: An auxiliary computer was developed to process MS^1 and MS^2 signals without introducing any overhead time and applied to rapid single run proteome analysis.

Results: At constant DIA cycle times with real-time ΦSDM- and eFT-processing, ΦSDM enabled us to distinguish otherwise unresolved precursor and fragment ions. This increased coverage of the yeast proteome to about 4,000 proteins in 23 min single runs.

INTRODUCTION

Data-independent acquisition (DIA) is gaining popularity in mass spectrometry based proteomics. Optimization of DIA methods often requires balancing spectral complexity with acquisition time. On the Orbitrap™ mass analyzer, narrower isolation windows and higher mass resolution facilitate spectral deconvolution, but the longer cycle times and lower quantitative accuracy.

To shorten the cycle times necessary, when using the standard signal processing enhanced Fourier transform (eFT) method [1], we present the Phase-Constrained Spectrum Deconvolution Method (ΦSDM) [2]**,** a signal processing approach that surpasses the Fourier uncertainty limit. ΦSDM translates into increased mass resolution without increasing acquisition time. Due to its high computational cost, ΦSDM was so far only applied for offline calculations or in narrow *m/z* regions of interest [3].

Here we apply ΦSDM to deconvolute full mass range spectra in real-time and demonstrate its applicability to rapid DIA experiments.

MATERIALS AND METHODS

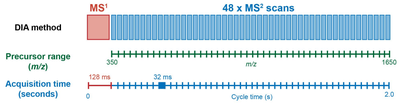
Sample Preparation

S. cerevisiae cells were grown to mid-log phase and harvested by centrifugation. Cell lysis, reduction and alkylation were performed in a single reaction vial as described earlier [4]. Proteins were digested overnight with equal amounts of LysC and trypsin. The digests were purified on styrenedivinylbenzene reversed-phase sulfonate (SDB-RPS) StageTips prior to LC-MS analysis.

Methods

Purified whole-proteome digests were separated in single runs within 23 min via nanoflow liquid chromatography on a Thermo Scientific™ Easy-nLC™ 1200 system coupled online to a research-grade Thermo Scientific™ Q Exactive™ HF-X Mass Spectrometer. Data acquisition was controlled by MaxQuant.Live [5]. The DIA acquisition cycle comprised one full scan (128 ms transient) and 48 DIA windows (32 ms transient) in the mass range *m/z* 350-1650. To cope with the higher computational demand of full range mass spectra processing and to ensure the real-time handling of up to 128 ms long transients, we configured an auxiliary computation unit to run a parallel CUDA C++ implementation of the ΦSDM algorithm. See **Figure 1** for a visualization of the workflow.

Figure 1. Sequencing scheme for the DIA workflow (not showing possible varying injection times).



Data Analysis

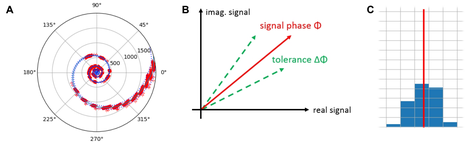
Targeted data extraction from DIA experiments was performed with the Biognosys Spectronaut™ Pulsar X software using a project-specific library and the default settings. The false discovery rate was controlled <1% at the precursor and protein group levels using a target-decoy approach.

SUPER-RESOLUTION FTMS WITH ΦSDM

The mass resolution of conventional Fourier Transform (FT)-based methods is inherently limited, in particular by signal uppression of interfering peaks closely spaced in mass. To overcome this limitation additional information is necessary. In Orbitrap mass analyzers, the phase information is readily available and can be used to establish more sophisticated signal processing schemes that achieve a higher frequency/mass resolution on a multiply refined grid.

ΦSDM builds upon a complex-valued FT spectrum and the calibrated initial phase of the ion signal in the time domain. A refined solution is obtained by an iterative algorithm (see the Supporting Information of Reference [4] for more details). After centroiding and conversion to the *m/z* domain, a refined mass distribution is obtained. See **Figure 2** for further visualization.

Figure 2. A) Calibration phase fit (mass on radial axis). B) Depiction of signal phase cone. C) Centroiding of the ΦSDM solution.



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 (isolated 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 1A). Using ΦSDM (Figure 1D), 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. 1B-C).

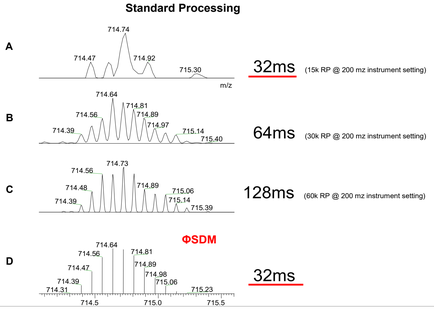
Figure 3. Real-time Full MS spectra of charge state 12+ of ubiquitin.

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 baseline resolved using a 32 ms transient.

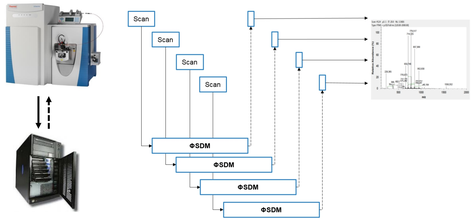


REAL-TIME ΦSDM PROCESSING

Full mass range ΦSDM processing involves a significantly higher computational effort than processing selected spectral windows (for example in TMT™ label applications, where the processing can be performed on the standard instrument PC). To cope with the increased numerical demands, we perform the calculations on an auxiliary computer equipped with four Titan Xp Nvidia™ GPU cards. The flow of signal information is shown in **Figure 4**. The instrument and the auxiliary computer are connected through the same gigabit network connection as used by the instrument and data PC. To reduce I/O traffic, lossless compressions are required when transferring the raw acquisition data.

As the ΦSDM processing time is almost 4 times longer than the corresponding transient time, all calculations are performed on 4 graphic cards in parallel. Distributing a single spectrum acquisition across multiple cards would increase the processing even further by introducing I/O overhead. The DIA workflow used does not involve data-dependent decisions. Consequently, we chose a parallel processing pipeline approach, where each transient is processed on a single card.

Figure 4. Parallel ΦSDM spectrum processing pipeline using the auxiliary computer.

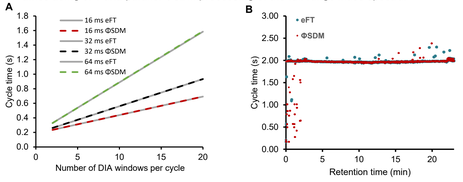


This setup should allow on-the-fly spectra processing without introducing any additional scan time overhead, regardless of the mass range. To test this, we performed direct infusion and online LC-MS experiments using either the standard eFT processing algorithm on the internal acquisition computer or our implementation of parallel ΦSDM on the auxiliary computer to process full mass range MS^1 (*m/z* 350-1650) and MS^2 (*m/z* 100-2000) spectra (**Figure 5**).

Figure 5. Real-time full mass range signal processing with ΦSDM.

A) Experimentally measured cycle time for one full scan versus the number of DIA windows at nominal MS^2 mass resolutions of 7,500 to 30,000 (corresponding to Orbitrap transient times 16 to 64 ms) with eFT and ΦSDM signal deconvolution. No measurable overhead is observed.

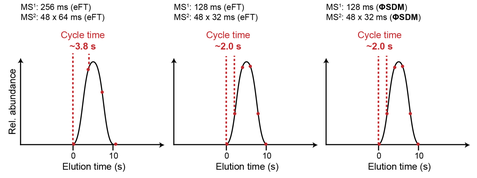
B) Cycle time in LC-MS DIA experiments with a fixed number of DIA windows and transient times using eFT and parallel ΦSDM processing of the full range mass spectra.



DIA WITH SHORT GRADIENTS

For robust quantification, 6-10 data points per peak are recommended to reconstruct the chromatographic elution profile of a peptide accurately. In contrast, longer cycle times with more and narrower precursor selection windows and longer Orbitrap transients decrease spectral complexity and thereby increase identification rates. In short LC gradients, peak widths can be 10 s or less, which makes DIA method optimization very challenging and typically requires wider isolation windows and lower mass resolution. ΦSDM achieves faster acquisition cycles without compromising precursor and fragment mass resolution (**Figure 6**).

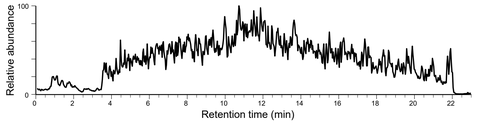
Figure 6. Balancing mass resolving power and cycle time in DIA.



THE 20 MIN YEAST PROTEOME

To test the performance of ΦSDM in proteomics practice we analyzed a full proteome digest of yeast in short 23 min single runs (**Figure 7**) with a DIA cycle time of 2 seconds using either standard eFT or ΦSDM**.**

Figure 7. Total ion chromatogram .



With real-time ΦSDM processing, we identified on average 16% more precursors per run as compared with standard eFT (**Figure 8**). This translated into an average of 3,951 identified protein groups per run, covering essentially the core yeast proteome in 23 min with very high quantitative accuracy (**Figure 9**).

Figure 8. Average number of identifications in triplicate analyses of yeast.

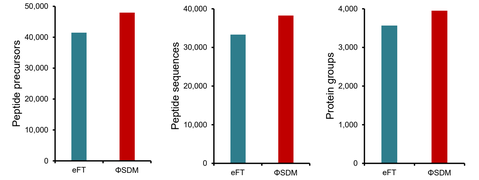
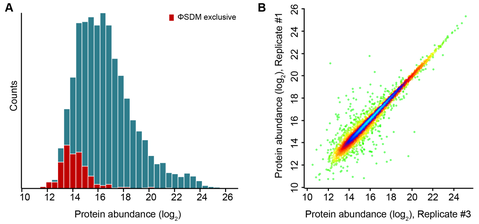


Figure 9. Quantification of the yeast proteome in 20 min.

A) Protein abundance range in single runs. Proteins exclusively quantified with ΦSDM but not eFT are highlighted in red.

B) The short cycle time enabled accurate quantification with a median CV of 8% and a data completeness of 99% in triplicates.



CONCLUSIONS

ΦSDM achieves the same mass resolution as standard eFT signal processing in half or less the transient time. We developed an auxiliary computing unit that enables full mass range signal processing with ΦSDM in real-time without introducing any scan time overhead. Application to rapid single run DIA analysis of yeast approaches complete proteome coverage in 20 min. At a fixed cycle time, ΦSDM enabled us to distinguish otherwise unresolved and in particular low-abundance precursor and fragment ions. Overall, this increased the coverage of the yeast proteome to about 4,000 proteins. We conclude that super-resolution FTMS signal processing has a great potential for proteomics research with its increasing demand for high-throughput technologies, for example in cell biology and biomedical research.

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. A. Kulak *et al.*, Nature Methods 11, 319-324, **2014**.
5. C. Wichmann *et al.*, Mol. Cell. Proteomics, **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

PO65491-EN0419S