Skyline Support for Proteome-wide Data Analysis of Bruker timsTOF dia-PASEF Acquisition

Brian S. Pratt[1], Matthew C. Chambers[1], Stephanie Kaspar-Schoenefeld[2], Sven Brehmer[2], Markus Lubeck[2], Ute Distler[3], Stefan Tenzer[3], Michael J. MacCoss[1], Brendan X. MacLean[1]

[1]University of Washington, Department of Genome Sciences, [2]Bruker Daltonik GmbH, [3]Institute for Immunology, University Medical Center Mainz

http://skyline.ms

Overview:

The use of data-independent acquisition (DIA) in quantitative proteomics has been shown to provide SRM-like quantification at proteome-wide scale, with greater selectivity and sensitivity than signal extracted from MS1 survey spectra. Previously, we have shown that alternating low- and high-energy MS spectra with ion mobility separation (IMS) improved quantitative accuracy in both types of spectra, but did not produce quantification with high energy spectra comparable to DIA with quadrupole isolation[1]. Here, we present Skyline support for the new Bruker timsTOF dia-PASEF method which combines parallel accumulation, IMS and quadrupole isolation to increase fragment ion flow and selectivity over traditional DIA with similar quadrupole isolation. The result compares favorably with both quadrupole-only DIA and IMS-only alternating spectra.

Introduction:

Skyline support for targeted filtering of drift tube ion mobility spectrometry (DTIMS – Agilent) and traveling wave ion mobility spectrometry (TWIMS – Waters) was first introduced at ASMS 2015. In JASMS 2018, we published our research showing the ease of use of our implementation, and its quantitative benefits in increasing selectivity in both precursor (low-energy) and fragment (high-energy) spectra. In that study, however, the precursor spectra still outperformed the fragment spectra, contrary to expectations with modern quadrupole isolated DIA. Also in 2018, Skyline added support for trapped ion mobility spectrometry (TIMS – Bruker) and support for MS1 filtering of timsTOF Parallel Accumulation Serial Fragmentation (PASEF) data. After Maier, et al. introduced dia-PASEF[2] in early 2019, we undertook a major effort later that year to support this combination of IMS and modern quadrupole isolated DIA for proteomewide analysis, and tested it on a human, yeast, and E.coli, 3-organism mix, the modern benchmarking standard for this type of analysis system.

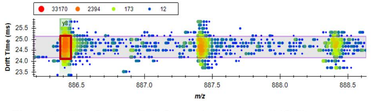

Figure 1: The Skyline Full-Scan plot showing IMS filtering applied to a high energy fragment spectrum.

Methods:

Sample mixtures of 3-organisms were created and separated into 6 fractions consistent with Navarro, Nat. Biotech. 2016[3], along with 5 fractions of each organism separately. The organism-specific fractions were run on a timsTOF in dda-PASEF mode 2 times each and searched with MaxQuant Andromeda. The search result output was imported into Skyline to create a spectral library with normalized retention time (iRT) based on the conserved iRT peptides. The 3-organism FASTA from Navarro, 2016 was imported into Skyline with duplicate peptides removed to create a proteome-wide list of targets. Scripts developed for Navarro, 2016 were used to run the Skyline command-line interface to process the data and determine optimal parameter settings, and run LFQbench for quantitative assessment.

The ProteoWizard data layer used by Skyline and Skyline itself were adapted from the IMS 2D spectrum model in use since 2015 to a new 3D spectrum model which yielded a 50-fold performance improvement for dia-PASEF and at least a 2-fold improvement for most other IMS data.

header		2D spectrum				
precursor m/z	IMS	m/z	m/z	m/z	...	m/z
		intensity	intensity	intensity	...	intensity

63,386,406 per dia-PASEF file

header	3D spectrum				
precursor m/z	m/z	m/z	m/z	...	m/z
	IMS	IMS	IMS	...	IMS
	intensity	intensity	intensity	...	intensity

68,378 per dia-PASEF file

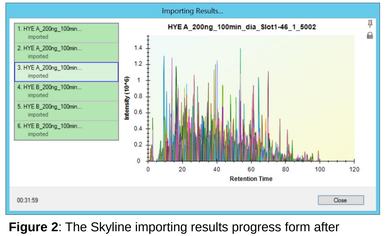

Figure 2: The Skyline importing results progress form after completing proteomewide chromatogram extraction from 6 dia-PASEF files in 32 minutes.

Results:

This resulted in a spectral library with 152,417 unique peptide precursors with stored MS/MS spectra, retention times and ion mobility.

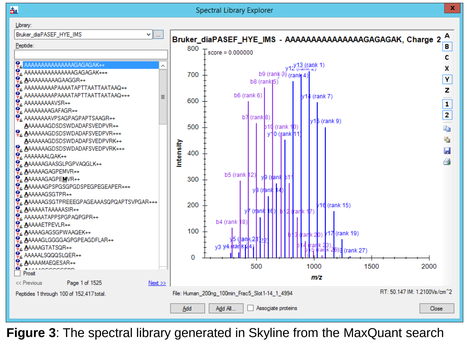

Figure 3: The spectral library generated in Skyline from the MaxQuant search results of 29 dda-PASEF runs of fractionated Human, Yeast and E. coli cell lysate.

The dia-PASEF runs employed a fixed 32-window isolation scheme requiring 8 TIMS accumulations separated into 4 IMS ranges and 4 *m/z* quadrupole isolations each.

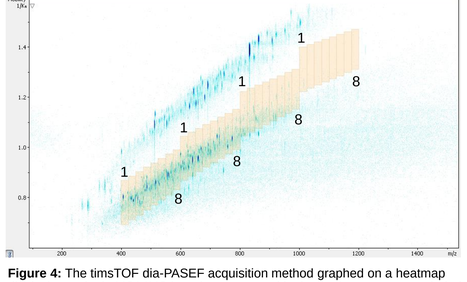

Figure 4: The timsTOF dia-PASEF acquisition method graphed on a heatmap displaying a single MS1 IMS by m/z spectrum with charge 1 precursors excluded above the acquired isolation scheme in displayed as tan bars above.

system	parallel files	import time (min)
Dell i7 4-core 16 GB of RAM and HDD	3	82
Dell i7 6-core 64 GB of RAM and SSD	6	42
Dell Xeon 24-core 192 GB of RAM and SSD	6	32

Table 1: Chromatogram extraction times for 3 systems

The document targets 12,717 proteins and 143,335 precursors with 33 CiRT standards.

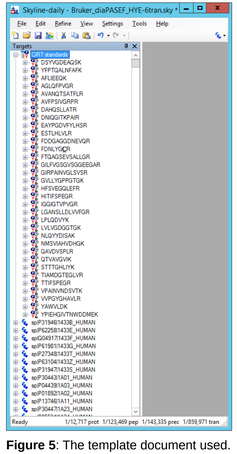

Figure 5: The template document used.

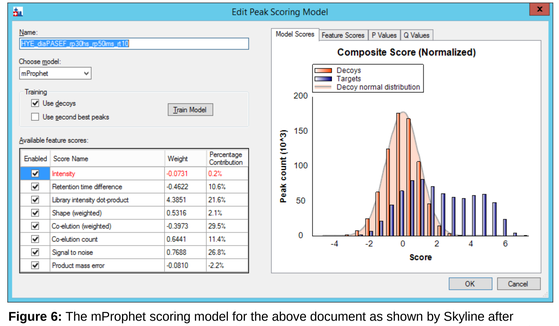

Figure 6: The mProphet scoring model for the above document as shown by Skyline after chromatogram extraction showing a relatively even split between true targets and false targets (scoring similar to decoys) with many features contributing to the final scoring model.

Figure 7: Within condition (A or B) CVs for peptides detected in 3 runs at < 1% FDR.

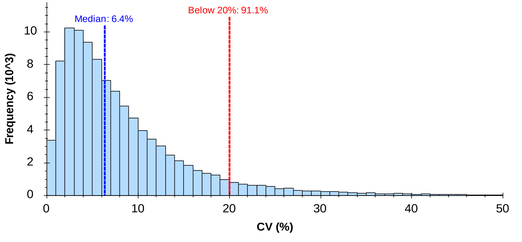

Figure 8:LFQbench quantitative accuracy and precision plots.

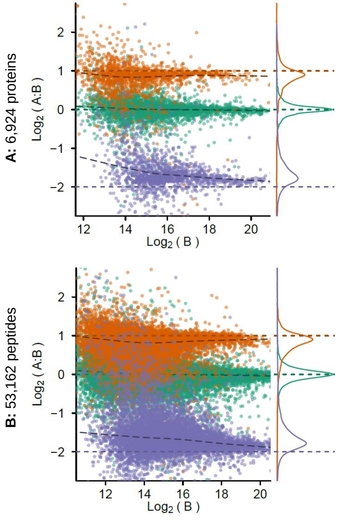

Table 2: Skyline command-line analysis was used to determine the optimal resolving power (50) in the IMS dimension for chromatogram extraction. While detections and quantifiable peptides increased with decreasing RP, and mean accuracy remained relatively stable, precision also declined somewhat.

Resolving Power	Peptide Counts		Accuracy			Precision	Species Overlap	
	Detections	Quantifiable	Human	Yeast	E. coli	Human CV	Yeast-Human (arctanh)	E.Coli-Human (arctanh)
20	58,916	41,684	0	-0.17	0.33	7.00%	1.73	2.16
30	64,241	46,835	0	-0.16	0.32	7.20%	1.76	2.14
40	68,042	50,774	0	-0.15	0.31	7.40%	1.76	2.15
50	70,217	53,162	0	-0.15	0.30	7.70%	1.73	2.07
60	71,732	54,773	0	-0.15	0.30	8.00%	1.70	2.04
70	71,970	55,276	0	-0.15	0.30	8.30%	1.64	2.03
80	71,568	55,079	0	-0.15	0.31	8.70%	1.64	1.96

Conclusions:

- Skyline processing of native dia-PASEF data has achieved exceptional performance, processing a full 3-organism mix comparison with LFQbench for over 100,000 precursors on a modern i7 computer in under 1 hour.
- The dia-PASEF method shows great promise achieving quantitative statistics comparable to or better than observed in the Navarro, Nat. Biotech. 2016 paper with 64-variable windows on a TripleTOF 6600 for over 50% more peptides.

References:

(1) MacLean B., Baker E., et al. JASMS. 2018 Nov;29(11):2182-2188.
(2) Meier F., et al. Biorxiv. 2019 doi: https://doi.org/10.1101/656207
(3) Navarro P., et al. Nature Biotech. 2016 Nov;34(11):1130-1136.