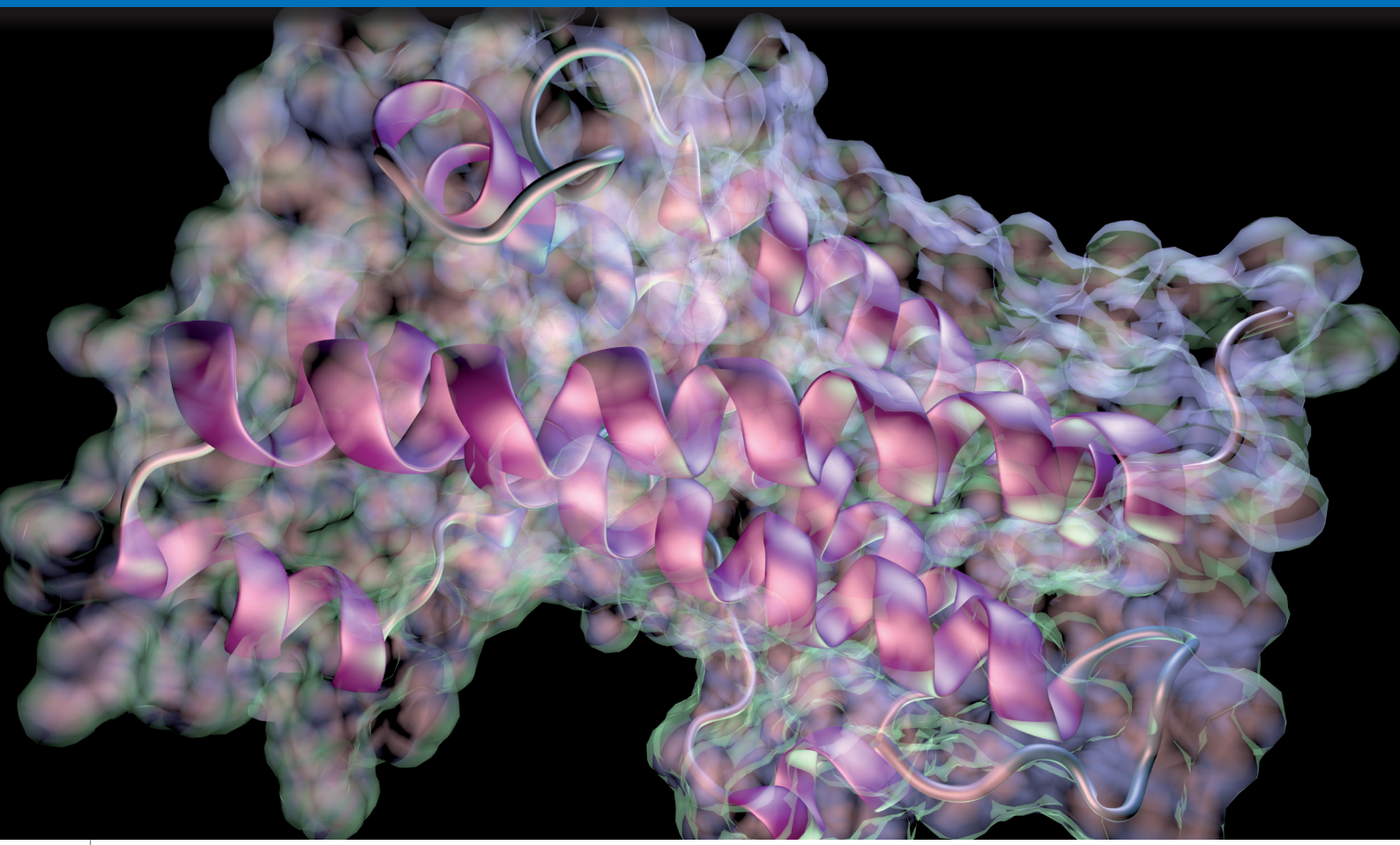

Exploring high complexity proteomes with a DIA based quant and ID workflow

Data Independent Acquisition (DIA) workflows have recently gained in popularity as they overcome the issue of stochastic selection of peptide precursors encountered in typical data-dependent approaches (DDA).

Introduction

The promise of these approaches is to deliver comprehensive and reproducible quantitation of peptides and proteins among pools of samples that will enable better characterization of the dynamics of the proteome.

The success of DIA approaches relies both on key instrumental capabilities (DIA requires the repeated acquisition of full-scan MS/MS where resolution, sensitivity, accuracy and dynamic range shall ideally be preserved at a very fast rate) and on the

Keywords:
impact II
Proteomics
DIA
Peptide-centric quantitation

Authors: Stephanie Kaspar-Schoenefeld, Thomas Kosinski, Romano Hebeler, Pierre-Olivier Schmit, Gary Kruppa; Bruker Daltonics

subsequent processing of these data. These requirements are matched extremely well to the impact II's Ultra-High Resolution Q-TOF unique combination of speed, resolution, sensitivity and high dynamic range.

A standard sample set and workflow that enables inter-laboratory comparison and intra-laboratory instrument performance characterization and validation has recently been described [1]. Both to contribute to the further development and testing of such workflows and to benchmark the performance of our DIA solution with the nanoElute LC and impact II QTOF, we employed the workflow as described in reference 1.Briefly, we challenged the instrument by mixing three proteomes in different ratios according to reference 1, acquiring the data in DIA mode and processing them with the Spectronaut™ software suite.

Results

Sample

For evaluation of the quantitative capabilities of the impact II Q-TOF instrument using a DIA approach, we ran two different experiments, both of them consisting of tryptic digests of human, yeast, and E.coli cell lysates (Figure 1). The three proteomes were mixed in defined ratios (according to [1])., either with high (HYE110: 1:1 (A/B) for human, 10:1 (A/B) for yeast, 1:10 (A/B) for E.coli) or moderate fold changes (HYE124: 1:1 (A/B) for human, 2:1 (A/B) for yeast, 1:4 (A/B) for E.coli), with human proteins representing the background proteome for both experiments.

Separation

Chromatographic separation was performed on a nanoElute nano-UHPLC system. A 120 minute curved gradient was used. Solvents were 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The temperature of the separation column was maintained at 50°C and the flow rate was set at 400 nL/min value.

Chromatography (Bruker nanoElute nano-UHPLC)

Elution time	Composition B
0 min	2 % B
75 min	15 % B
105 min	25 % B
120 min	35 % B
130 min	95 % B
140 min	95 % B

Acquisition

Five technical replicates of each sample were measured. The DIA method optimized for the impact II used 28 cycles to iterate through a 400-1000 m/z precursor window range while using a window size of 24 Da at a spectral acquisition rate of 14 Hz. This results in a total cycle time of 2 seconds. Each cycle was started with one MS1 survey scan. The DIA method can be easily adapted from the default DIA acquisition method while using Bruker's Q-TOF acquisition software: mass range; (variable) window size; acquisition speed; and optional inclusion of MS1 scans can be easily selected and adjusted by the user. Furthermore, collision energies are automatically adjusted depending on the precursor window mass range.

Processing

Due to the complexity of the multiplexed spectra generated with DIA approaches, sophisticated data processing software solutions are required. We used Spectronaut (Biognosys) as well as LFQbench (University Medical Center of Johannes Gutenberg University Mainz) software programs for data processing and evaluation respectively, applying the default settings for both solutions. Spectronaut has been specifically developed for DIA-based quantitative proteomics. The processing includes fully automated in-run calibration, automatic interference correction, peak picking and automatic quality control. In combination with the usage of iRT peptides we were able to refer to Biognosys' reference spectral libraries for the peptide and protein identification; therefore reducing the method development efforts as no DDA runs were needed to build up reference libraries. LFQbench was used for the automated evaluation of label-free quantification performance.

Qualitative results: Identified and quantified peptides/proteins

More than 31000 peptides could be identified from both experiments HYE110 and HYE124 (table 1). They yielded the identification of 3654 proteins for HYE124 and of 3877 proteins for HYE110. 91% and 74% of those proteins could be reliably quantified for HYE124 and HYE110, respectively. the use or Spectronaut's internal spectral database, which had been created from the compilation of another type of instrument's DDA analysis, these results compare favorably with what have been achieved from other instrumental platforms [1], [6], particularly when it comes to the number of reliably quantified proteins.

Reproducibility of proteome measurements using DIA

Reproducibility of peptide and protein quantitation determines the confidence in the applied DIA approach.

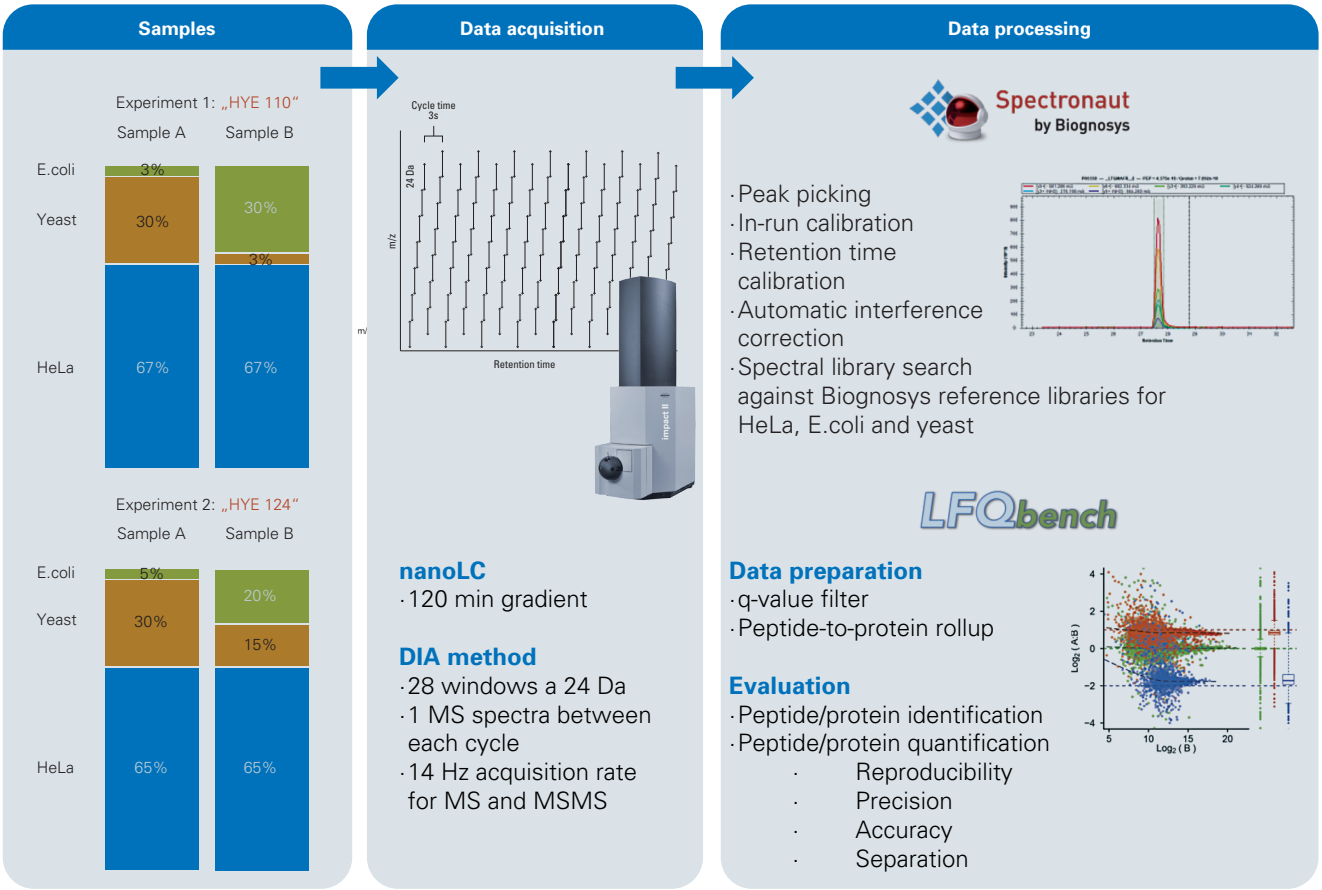

Figure 1: Workflow. Two experiments (HYE110 and HYE124) were prepared containing known quantities of human, yeast and E.coli peptide digests. Both samples differed in their ratios with sample HYE110 having higher ratios.Data acquisition was done using a nanoElute (Bruker Daltonics) nano flow UPLC coupled to an impact II Q-TOF instrument (Bruker Daltonics). Five technical replicates of each sample were analyzed in DIA acquisition mode using 24 Da window size. Data processing was done using Spectronaut software (Biognosys) for advanced peak picking and spectral library search. Subsequently LFQbench was used for in-depth data evaluation of peptide/protein identification and quantification.

Table 1: Qualitative results obtained from the HYE110 and HYE124 experiments. Identifications have been obtained by searching Spectronaut's internal spectral library, from the 120 min gradient analyses.

	Sample	Number of IDs	Valid quantification ratios	Median CV human*
Peptides	HYE 124	33667	27431	0,08
	HYE 110	34857	23168	0,07
Proteins	HYE 124	3654	3320	0,0600
	HYE 110	3877	2848	0,0400

For evaluation of the presented DIA method, we checked the reproducibility between technical replicates and determined coefficients of variation (CV%) for both identified proteins and peptides. Reproducibility between single runs was quantified by calculating R2 values. A comparison of all single replicate runs for sample A of experiment HYE110 is shown in Figure 2a for all quantified signals. Very high R2 values have been found for this complex sample on the peptide as well as on the protein level (Figure 2a and b).

The median CV for background species among the replicate runs was well below 10% for both sample sets, both on the peptide and the protein level, illustrating the extremely good reproducibility of the impact II's MSMS level quantification (Figure 2 b). For the experiment HYE110, 2178 human proteins had a median CV of 4%, and 18423 peptides had a median CV of 7%. For the HYE124 experiment the median CV levels were of 5% for proteins (2104 proteins), and 7% for peptides (17998 peptides).

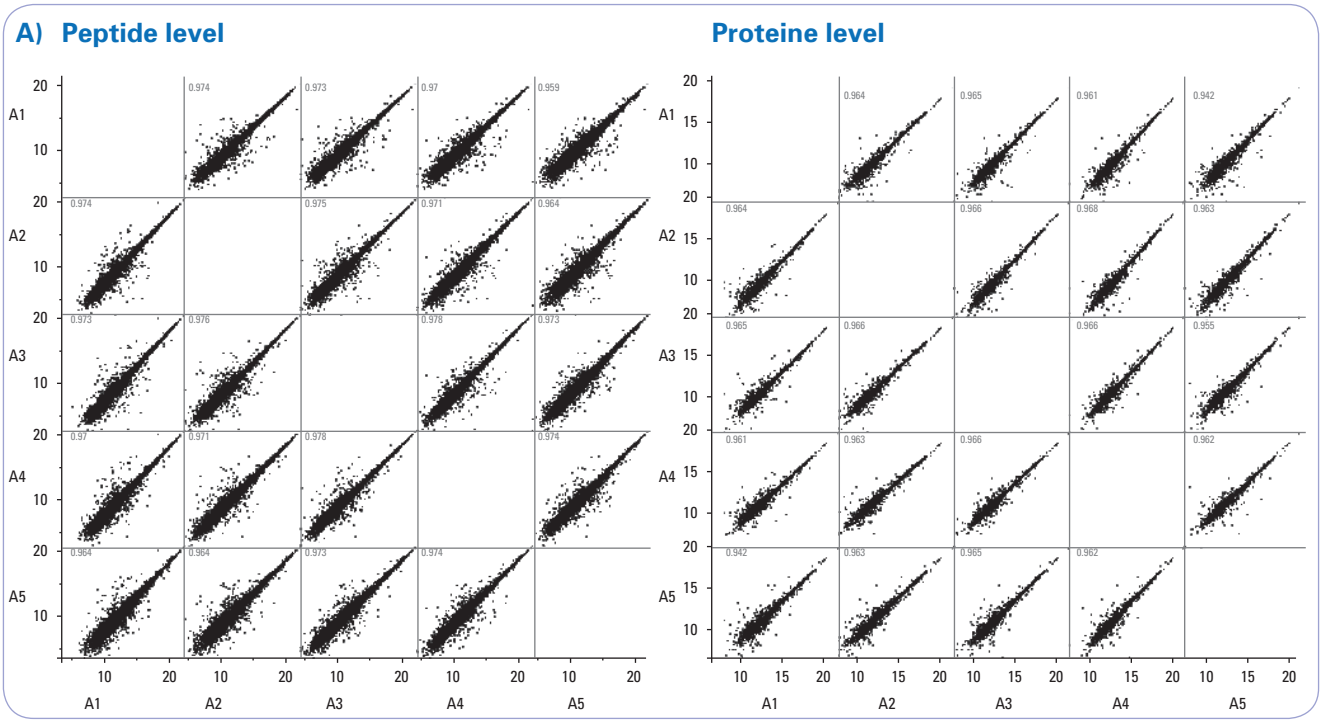

Figure 2: Reproducibility of DIA measurements. (a) Intensity correlation between all five technical replicates on peptide and protein level for Sample A (Experiment HYE124). (b) Summary of the reproducibility for the presented DIA measurements. R2 was calculated as average of all possible pairs for sample A, separately for both sample sets (HYE110 and HYE124). The "Median CV – human" represents the median CV for the background human species among the replicate runs.

		R2 value	Median CV - human
Peptides	HYE110	0.97 (±0.004)	7%
	HYE124	0.97 (±0.005)	8%
Proteins	HYE110	0.96 (±0.004)	4%
	HYE124	0.96 (±0.007)	6%

Accuracy of proteome quantification by DIA

The design of the analyzed proteomics samples enabled the evaluation of the quantitation accuracy of the applied DIA approach while used with high-complexity protein mixtures. Background human proteins were spiked in equal amounts resulting in a theoretical ratio of 1:1. For both experiments, HYE110 and HYE124, we found human proteins to be centered at a log2 ratio of sample A versus sample B of 0, which corresponds to the theoretical ratio of 1:1. This ratio could be measured across the complete dynamic range (Figure 3). This results in an extremely good global precision (standard deviation of log ratios) of 0.27 for HYE110 and of 0.21 for HYE124, respectively.

For yeast and E.coli proteins and peptides global accuracy was better for experiment HYE124 compared to HYE110. Ratio determination in HYE110 was more challenging due to the higher fold change in the experiment, forcing ratio calculation from signals sometimes closer to noise threshold (Table 2). This outcome is consistent with behavior observed on other instrument types for the described DIA approach [1].

In total, we reliably identified and quantified more than 33,000 peptides in the single experiments (both for HYE110 as well as HYE124) covering easily five orders of magnitude for human proteins (Figure 4) without any pre-fractionation. The dynamic range covered for E. coli and yeast proteins was slightly reduced due the lower complexity of both proteomes (compared to the human proteome) as well as the reduced amounts injected. The results confirm that the usage of the described DIA approach on the impact II Q-TOF instrument results in simultaneous very good proteome coverage, reproducibility and accurate quantitation despite the challenge constituted by sample complexity, dynamic range and even for low-fold changes in concentration of a factor of two or four.

Methods

Samples were prepared for two different experiments, HYE110 and HYE124 each consisting of two different samples, by mixing of three proteomes (Figure 1, "sample preparation") according to reference [1]. Human cell

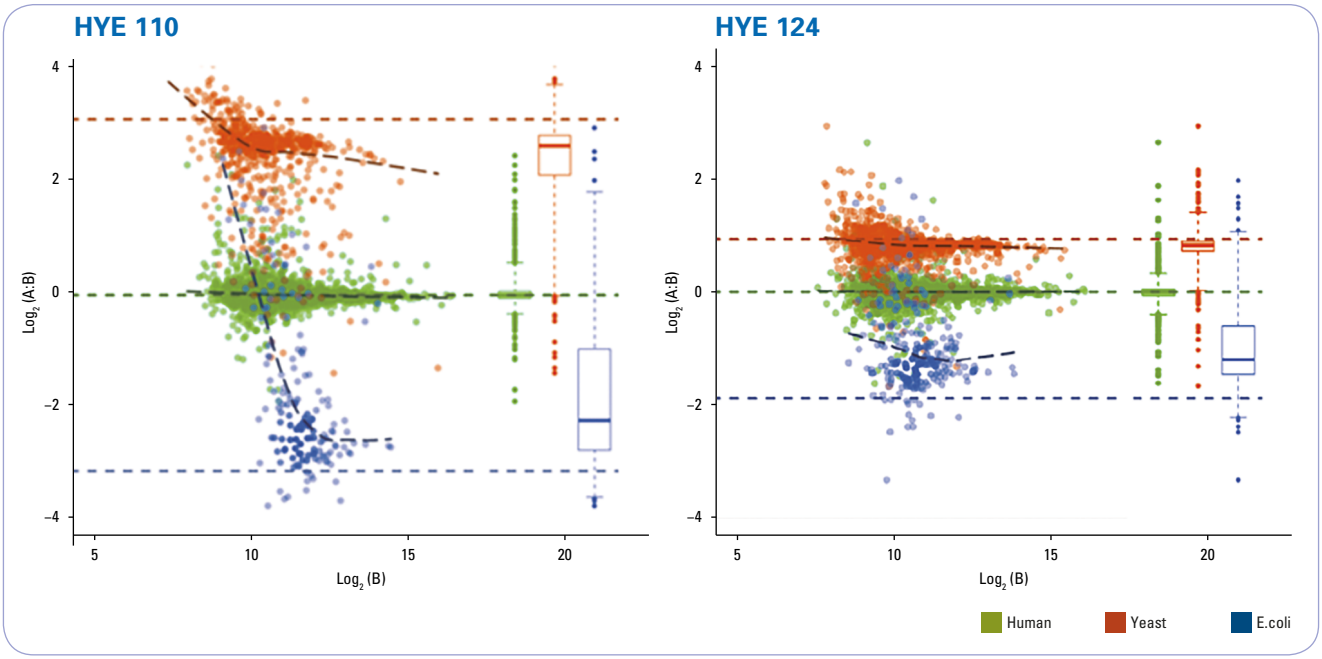

Figure 3: Protein level quantitative results for HYE110 and HYE124 sample analyzed after processing using LFQbench. Log-transformed ratios (log2(Sample A/ Sample B)) of yeast (orange), human (green) and E. coli (purple) proteins are plotted over log-transformed intensity for sample B (2834 proteins plotted for HYE110 and 3320 for HYE124). Colored dashed lines represent expected values.

Table 2: Qualitative results obtained from the HYE110 and HYE124 experiments. Identifications have been obtained by searching Spectronaut's internal spectral library, from the 120 min gradient analyses.

			Human	Yeast	E.coli
Global Accuracy	Protein level	HYE110	0,0000	-0,5001	0,9535
		HYE124	0,0000	-0,1140	0,7298
	Peptide level	HYE110	0,0000	-0,5024	0,8510
		HYE124	0,0000	-0,1217	0,7586
Global Precision	Protein level	HYE110	0,2683	0,9665	1,6007
		HYE124	0,2137	0,3493	0,8649
	Peptide level	HYE110	0,3353	0,9966	1,7651
		HYE124	0,2932	0,4224	1,0612

line and E.coli digest were prepared according to ref [3]. Predigested yeast sample was re-suspended in 0.1% trifluoroacetic acid. To generate samples for the experiment HYE110 tryptic digests of the three proteomes were mixed in the following ratios: sample A was composed of 67% w/w human, 30% w/w yeast, and 3% w/w E.coli proteins; sample B was composed of 67% w/w human, 3% w/w yeast, and 30% w/w E.coli proteins. To generate samples for the experiment HYE124 tryptic digests of the three proteomes were mixed in the following ratios: sample A was composed of 65% w/w human, 30% w/w yeast, and 5% w/w E.coli proteins; sample B was composed of 65% w/w human, 15% w/w yeast, and 20% w/w E.coli proteins. The iRT retention time kit (Biognosys, Switzerland) was spiked at a concentration of 1:20 v/v in all samples. Sophisticated retention time calibration using iRT (indexed Retention Time) enables usage of smaller extraction windows and results in higher precision of peptide quantification across several runs [4].

Peptides (1µg/µl) were separated using a nanoLC column and subsequently detected using a DIA approach

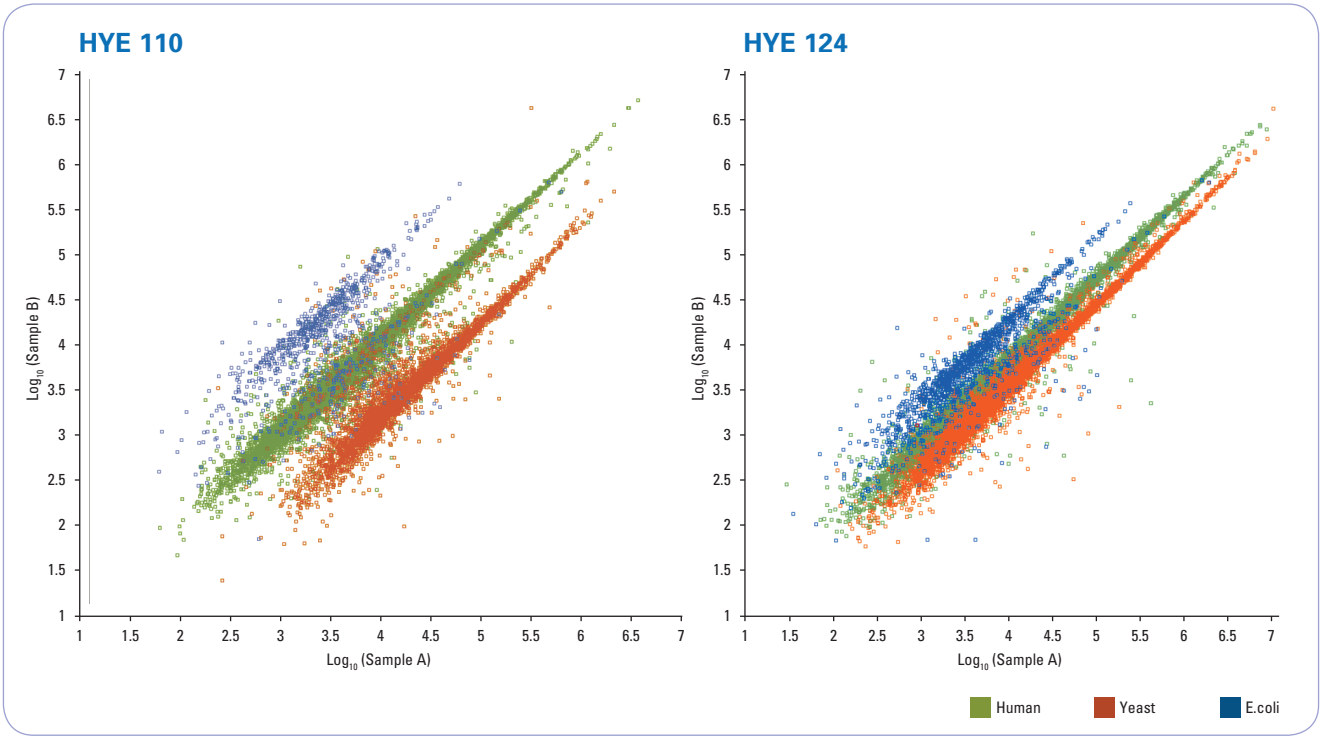

Figure 4: Accuracy in peptide level quantification. Log-transformed intensity (log10) of sample B is displayed over log-transformed intensity of sample A for all detected and quantified yeast (orange), human (green) and E. coli (purple) peptides (23036 peptides plotted for HYE110 and 27400 for HYE124). Results show that accurate quantification over > 4 orders of magnitude is possible using the applied DIA approach.

on the impact II QTOF-MS (Bruker Daltonics). Each sample was measured in five technical replicates.

Data processing was done using Spectronaut software (version 9, Biognosys) applying default settings. To increase the precision of quantification the interference detection algorithm and cross runs normalization as implemented in the software was used. Reference libraries as provided by Biognosys were used as spectral libraries (HeLa: 28674 unique peptide sequences, yeast: 19101 unique peptide sequences, E. coli: 13074 unique peptide sequences). Peak picking results were subsequently exported to LFQbench for further processing as described in [1] for data evaluation. LFQbench (https://omictools.com/lfqbench-label-free-quantification-bench-tool) is an open-source R library for the automated evaluation of label-free quantification results of hybrid proteome data sets [5].

Settings nanoElute™ (Bruker Daltonics)

Column	Nano Trap Column, 100 µm i.d. × 2 cm, packed with Acclaim PepMap100 C18, 5 µm, 100 Å, nanoViper and Acclaim PepMap RSLC C18, 2 µm, 100 Å, 75 µm i.d.×50 cm, nanoViper (Thermo Fisher Scientific)
Flow rate	400 nL/min
Mobile phase	*A = water + 0.1% formic acid, B = acetonitrile + 0.1% formic acid*
Gradient	t = 0 min, 2% B; t = 2 min 5% B; t = 90 min, 35% B; t = 100 – 110 min 95% B; t =111 – 125 min 5% B
Source	CaptiveSpray ionization source
Ionization:	ESI(+)
Mode:	DIA (details see Table 2)
Scan range:	m/z 400–1000
Window size:	24 Da
Acquisition rate:	14 Hz

Table 3: Details about window placement used for described DIA method, including m/z ranges and collision energies.

#	m/z start	m/z end	Collision Energy	#	m/z start	m/z end	Collision Energy
MS			7	MS			7
1	400,4319	424,4428	27	27	388,4264	412,4374	27
2	424,4428	448,4537	27	28	412,4374	436,4483	27
3	448,4537	472,4646	27	29	436,4483	460,4592	27
4	472,4646	496,4756	27	30	460,4592	484,4701	27
5	496,4756	520,4865	27	31	484,4701	508,481	27
6	520,4865	544,4974	31	32	508,481	532,4919	31
7	544,4974	568,5083	31	33	532,4919	556,5028	31
8	568,5083	592,5192	31	34	556,5028	580,5138	31
9	592,5192	616,5301	31	35	580,5138	604,5247	31
10	616,5301	640,541	39	36	604,5247	628,5356	39
11	640,541	664,552	39	37	628,5356	652,5465	39
12	664,552	688,5629	39	38	652,5465	676,5574	39
13	688,5629	712,5738	39	39	676,5574	700,5683	39
14	712,5738	736,5847	42	40	700,5683	724,5792	42
15	736,5847	760,5956	42	41	724,5792	748,5902	42
16	760,5956	784,6065	42	42	748,5902	772,6011	42
17	784,6065	808,6174	42	43	772,6011	796,612	42
18	808,6174	832,6284	45	44	796,612	820,6229	45
19	832,6284	856,6393	45	45	820,6229	844,6338	45
20	856,6393	880,6502	45	46	844,6338	868,6447	45
21	880,6502	904,6611	45	47	868,6447	892,6556	45
22	904,6611	928,672	48	48	892,6556	916,6666	48
23	928,672	952,6829	48	49	916,6666	940,6775	48
24	952,6829	976,6938	48	50	940,6775	964,6884	48
25	976,6938	1.000,70	48	51	964,6884	988,6993	48
26	1.000,70	1.024,72	48	52	988,6993	1.012,71	48

Conclusions

- The presented study illustrates the benefits of the impact II's unique combination of speed, resolution and sensitivity which are allowing for high DIA performance while working with average sized windows. The resulting very short cycle time guarantees a good chromatographic resolution; Together with the Chromatographic performance of the nanoElute, and the spray reproducibility brought by the CaptiveSpray nanoBooster, it yields market-leading performance for accurate ratio determination.
- Extremely high reproducibility of MS/MS quantitation with CV levels well below 10% enables the accurate quantitation of small as well as large fold changes.
- The older maXis impact had already proven great potential for targeted proteomics operation, including DIA analysis [2]. The impact II/Spectronaut solution now takes these capacities a step forward by enabling direct DIA identification and the most accurate ratio determination from complex samples from the same acquisition.
- Those results are perfectly competitive with what could be obtained also from platforms designed for DIA acquisition, with no compromise made on the impact II's versatility for other applications like Glycoproteomics of Proteoform Profiling.

Learn More

You are looking for further Information? Check out the Link or scan the QR Code.

www.bruker.com/literature-proteomics

References

[1] Pedro Navarro, et al. (2016) A multicenter study benchmarks software tools for label-free proteome quantification. Nature Biotechnology 34, 1130-1136

[2] Application Note LC-MS89: High quantification efficiency in plasma targeted proteomics with a full-capability discovery Q-TOF platform https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/ApplicationNotes/LCMS-89_Targeted_Proteomics_ebook.pdf

[3] Wang, H., et al. (2005) Development and evaluation of a micro- and nanoscale proteomic sample preparation method. Journal of proteome research 4, 2397-2403

[4] Bruderer, R., et al. (2016) High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantification. Proteomics 16, 2246-2256

[5] Kuharev, J., et al. (2015) In-depth evaluation of software tools for data-independent acquisition based label-free quantification. Proteomics 15, 3140-3151

For research use only. Not for use in diagnostic procedures.

Bruker Daltonics GmbH & Co. KG

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com – www.bruker.com