# MS1 based quantification optimization on DIA methods on a quadrupole-Orbitrap mass spectrometer

Yue Xuan[1], Roland Bruderer[2], Oliver Bernhardt[2] ,Tejas Gandhi[2] , Lukas Reiter[2] , Andreas F. Huhmer[3], Thomas Moehring[1] , [1]Thermo Fisher Scientific, Bremen, Germany; [2]Biognosys AG, Schlieren, Switzerland; [3]Thermo Fisher Scientific, San Jose, USA

## ABSTRACT

Targeted analysis of data-independent acquisition (DIA) is a powerful mass spectrometric approach for comprehensive, reproducible and precise proteome quantitation, enabling the identification of the majority of expressed proteins. With this achievements in the identification, the reproducibility and quantitative accuracy and precision have become increasingly important. MS1 scans have potentially a higher sensitivity since the peptide precursor is not fragmented in multiple fragments. Here, we optimize data-independent acquisition (DIA) on a Thermo Scientific™ Q Exactive™ HF mass spectrometer for quantification on the MS1 level, and using Spectronaut software for DIA data analysis.

## INTRODUCTION

The performance of a DIA methods is influenced by the number of DIA windows, the scan resolution, the gradient length and the sampling of the chromatographic peaks. Therefore, we evaluated the impacts of all these factors on the DIA performance on a Thermo Scientific™ Q Exactive™ HF mass spectrometer. We assessed MS1 based quantification performance and used the Spectronaut software for the targeted analysis of DIA data. The high resolving power of the Orbitrap mass analyzer for both MS1 and MS2 scans of the DIA method shows an important influence on the performance. Novel methods using high resolution powers of 120,000 in MS1 scan, and 30,000 for DIA MS2 scans enable high precision quantification on MS1, thereby achieving the best peptide identification and quantification performance. With an optimized 2hours DIA method, over 76,000 peptide precursors are identified, and over 66,000 peptide precursors are quantified with a coefficient of variation (CV) better than 20% on MS1 level.

## MATERIALS AND METHODS

### Sample Preparation

Protein samples for HeLa and HEK-293 were both prepared using the FASP protocol. Additionally, 6 high pressure reverse phase (HPRP) fractions of HeLa and HEK were made for building up a comprehensive spectral library, based on which the targeted analysis DIA data analysis is performed. All samples and fractions were acquired using data-dependent acquisition (DDA) methods on the same nanoLC online, coupled to Q Exactive HF MS to build up the spectral library. All the DIA experiments are performed using the HeLa sample only. Biognosys' iRT kit was spiked into the samples before injection.

### Liquid Chromatography

For the spectral library, the 6 HPRP fractions of HeLa and HEK-293 and a triplicate of each unfractionated sample were acquired using a Thermo Scientific™ EASY-nLC™ 1200 nanoLC system equipped with an analytical column (Reprosil-Pur, 75um x 50cm, C18, 1.9um) coupled to a Q Exactive HF mass spectrometer equipped with a Thermo Scientific™ nanoFlex™ nanoelectrospray ion source. Peptides (2 µg of digest) were separated by segmented gradients from 99% buffer A (0.1% FA in 1% ACN) / 1% buffer B (0.1% FA in ACN) to 10% buffer A / 90% buffer B within 75 min.

DIA experiments were performed with the HeLa digest sample. Each sample was analyzed on an EASY-nLC 1200 nanoLC system equipped with an analytical column (Reprosil-Pur, 75um x 50cm, C18, 1.9um) coupled to a Q Exactive HF mass spectrometer. Peptides (2 µg of digest) were separated by segmented gradients from 99% buffer A (0.1% FA in 1% ACN) / 1% buffer B (0.1% FA in ACN) to 10% buffer A / 90% buffer B within 30, 60,120, 240 min. The total run time overhead with loading and washing steps was 50 min. Column oven was set to 50ºC .

### Mass Spectrometry

DDA method on Q Exactive HF MS for the spectral library: The MS was operated in data-dependent TOP15 mode with the following settings: mass range 350 -1650 Th; resolution for MS1 scan 60,000 (FWHM) at *m/z* 200; lock mass: best; resolution for MS2 scan 15,000 (FWHM) at *m/z* 200; isolation width 1.6 m/z; NCE 27; underfill ratio 1%; charge state exclusion: unassigned, 1, >6; dynamic exclusion 30 s.

**DIA method evaluation on Q Exactive HF**: First, the sampling of the LC peak with 5, 8, 11 und 14 data points was evaluated with a 2 hour LC gradient (Figure 1 upper part). MS1 was set to 120k, MS2 is set to 30k. The m/z range covered was 350 to 1650 Th. Based on the average LC peak width 6.6s (FWHM), the number of DIA windows are calculated and adjusted. The calculated numbers of DIA MS2 windows ranged from 10 to 44. Next, DIA methods for different gradients were generated with a fixed sampling of 8 data points per peak. Finally, DIA methods with a fixed cycle time of 2 seconds (8 data points per peak) were generated to evaluate the MS1 resolving power impact. Full scan resolution at 30k, 60k, 120k, and 240k were evaluated with a 2 hour LC gradients (Figure 1 lower part). MS2 scan resolution was set to 30k. The number of DIA windows were calculated and adjusted to match the duty cycle time. The calculated number of DIA MS2 windows ranged from 18 to 25 for each duty cycle.

Full scan AGC target was set to 3e6, IT to 50ms. DIA settings were NCE 27+/- 10%; target value 3e6 and maximal injection time set to "auto", setting max IT to allow the mass spectrometer always operating in the parallel ion filling and detection mode.

### Data Analysis

Database search in MaxQuant software was performed directly with the RAW files using the human UniProt database. Trypsin with up to 2 missed cleavages; mass tolerances set to 20 ppm for the first search and 4.5 ppm for the second search for the Q Exactive data. Oxidation of M and N-terminal acetylation were chosen as dynamic modification and carbamidomethylation of C as static modification. The false discovery rate (FDR) was set to 1% on peptide and protein level.

All DIA data were directly analyzed in Spectronaut 8.0 (Biognosys, Schlieren). Dynamic score refinement and MS1 scoring were enabled. Total peak areas of the isotopic envelope were chosen for quantities. Interference correction and cross run normalization ( based on total peak area) were enabled. All results were filtered by a Q value of <0.01 (equals a FDR of 1% on peptide level).

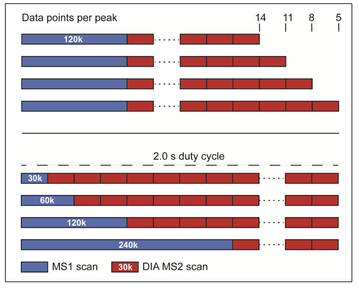



**Figure 1. Data Independent Experiments Design**

Upper illustration: Optimization of data points per peak for identification and quantification. The DIA methods were set to result in 5, 8, 11 and 14 data point per LC peak. The MS1 resolution was set to 120k, MS2 was constant a 30k. Based on the median LC peak width 6.6seconds at FWHM, and 12seconds at base, the number of DIA windows were calculated and adjusted. It resulted in the number of DIA MS2 windows in between 10 to 44. Lower illustration: The duty cycle of DIA method was set to 2 seconds to evaluate the MS1 resolution impact. Full scan resolution @ 30k, 60k, 120k, and 240k were evaluated with a 2hour LC gradient. The MS2 scan resolution was set to 30k. The DIA window numbers were calculated and adjusted based on the duty cycle time, the MS1 scan time and MS2 scan time. It resulted in DIA MS2 windows numbers between 18 and 25.

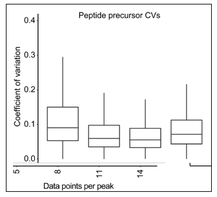
## RESULTS

The performance of DIA methods was influenced by the number of DIA windows, the scan resolution, the gradient length and the sampling of the chromatographic peaks. Additionally, the MS1 scan of the DIA method had an important influence on the performance. To optimize quantification on MS1, we evaluated the following conditions. We established a comprehensive spectral library by measuring triplicates on six high pressure reverse phase fractions of Hela and HEK digestion sample. In total, 18 DDA runs are acquired, which results in > 135,000 peptide precursors, > 94,000 peptides, and >7200 protein groups in the spectral library (Table 1). This spectral library is then applied for the DIA data analysis.

## Sampling of liquid chromatographic peaks

The influence of the number of data points per peak for identification and quantification was investigated. The DIA methods were set to result in 5, 8, 11, and 14 data points per LC peak (Figure 1 upper part). MS1 was set to 120k, MS2 was set to 30k. Based on the average LC peak width 12s at base, the numbers of DIA windows were calculated and adjusted. resulted in numbers DIA MS2 windows between 10 and 44. The isolation window widths were calculated based on the target MS2 windows number to cover a mass range from 350-1650 TH. The less number of DIA windows, the wider the each isolation window therefore becomes. The CV of triplicate acquisitions is shown in Figure 2, median CV for all the experiments are better than 10%. With 11 data points per peak was optimal for MS1 quantification to achieve a median CV ~ 5%. With 14 data points per peak, the CV was slightly worse than with 11, which was potentially due to higher complexity of wider DIA isolation indirectly influencing MS1 (peak picking, integration boundaries). With 5 data points, the highest number of protein and peptide ID number were achieved with FDR of 1%. However, to evaluate the reproducibility and precision of quantitation, CV < 10% was applied to the data, the 8 points crossing LC peak acquisition showed the best identification and quantitation result in Table-1 and Figure 3. Therefore, 8 -11 points crossing LC peak gave the best reproducible quantification for the DIA experiments.

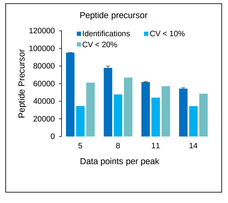
**Figure 2. Data points per liquid chromatographic peaks vs CV of identified peptides**
Sampling of the chromatographic peaks by 5, 8, 11 and 14 data points per peak were evaluated with a 2 hour LC gradient. With 11 points crossing the LC peak gave the best CV results of DIA experiments.

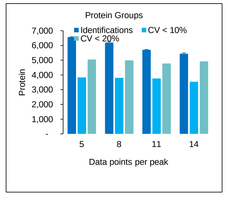



| | Peptide Precursors (Q Value <0.01) | Peptide IDs (Q Value <0.01 and CV < 10%) | Protein Group IDs (Q Value <0.01 and CV < 10%) |
|---|---|---|---|
| Spectral Lib | 135,356 | 94,010 | 7,222 |
| 5 points | 95,223 | 34,558 | 3,828 |
| 8 points | 77,928 | 47,742 | 3,806 |
| 11 points | 61,820 | 44,137 | 3,760 |
| 14 points | 54,326 | 34,451 | 3,533 |

**Table 1. Data points per LC peak for Peptide Precursor with a FDFR of 1%, Peptide / Protein IDs with a FDR of 1% and CV <10%**

**Figure 3. Data points per LC peak Evaluation for Peptide precursors / Protein IDs with a FDR of 1%, CV <10%, and CV <20%.**

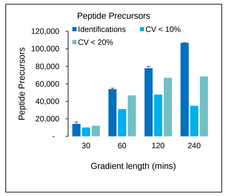



## The LC gradient length

Liquid chromatograph separation performance has a critical influence on DIA performance. Longer LC gradients give better separation resolution, while excessively long LC gradients can cause irreproducible retention time shifting from run to run, as well as consuming much more instrumental time. To find the balance between the LC chromatographic separation and reproducibility/throughput, we evaluated 30mins, 60mins, 120mins, and 240mins LC gradient length. The MS1 resolution was set to 120k and followed with an adapted number of DIA MS2 scans, which are acquired at 30k resolution for constant data points per peak (8). The CV of triplicate acquisitions are shown in Figure 3. median CV for all the experiments are better than 10%. As expected, the identifications of peptides and proteins with FDR of 1% were increasing with the length of the LC gradient from ~12,000 peptides and ~ 3,000 protein groups with 30mins gradient up to ~110,000 peptides and >6700 protein groups with a 4 hour gradient (Figure 4 and Table 2). With CV <20%, 120min gradient acquisitions resulted in similar numbers of peptide precursors and protein groups for quantification as with 240mins (Figure 4). With CV < 10%, 120mins LC gradient acquisitions showed the optimal reproducibility and quantification on MS1 (Figure 5).

**Figure 4. LC gradient length Evaluation for Peptide precursors / Protein IDs with a FDR of 1%, CV <10%, and CV <20%.**



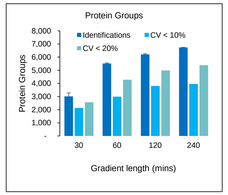




**Table 2. LC gradient length Evaluation for Peptide Precursor / Peptide / Protein IDs with a FDR of 1%**

| | Peptide Precursors (Q Value <0.01) | Peptide IDs (Q Value <0.01) | Protein Group IDs (Q Value <0.01) |
|---|---|---|---|
| Spectral Lib | 135,356 | 94,010 | 7,222 |
| 30 min | 14,300 | 12,064 | 3,015 |
| 60 min | 54,065 | 42,270 | 5,516 |
| 120 min | 77,928 | 59,082 | 6,205 |
| 240 min | 106,848 | 79,058 | 6,738 |

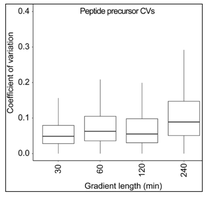



**Figure 5. LC gradient length vs. media CV of identified peptides:** DIA acquisitions with a LC gradient length with 30 mins, 60 mins, 120 mins, and 240 mins were evaluated. 120 mins resulted in the best media CV results.

## High Resolving Power of MS1

The high resolution of the MS1 scan can distinguish the interferences from the analyte of interest. Therefore, resolution has a high impact on the identification and quantification of peptides at the MS1 level. We evaluated the MS1 resolution @ 30k, 60k, 120k and 240k, following with the DIA MS2 scans, in which the resolution was set to 30k. We fixed the duty cycle of the DIA method to 2 seconds and a 2hours LC gradient. The DIA window numbers were calculated and adjusted based on the duty cycle time, MS1 scan time, and MS2 scan time. This approach resulted in numbers of DIA MS2 windows from 18 to 25 for each duty cycle. The results of identified peptides and proteins were calculated and are listed in Table 3 with a FDR of 1%. with 120k MS1, the most number of peptides (>59,000) and proteins (~6200) are identified (Table 3). The reproducibility of the quantitation is shown in Figure 6. 120K MS1 also showed the best media CV among all the resolution settings. The combination of 120k MS1 with 30k MS2 resolution resulted in optimal quantification for the 120 min gradient (Figure 7 and Table 2, and Table 3).

**Table 3. MS1 Resolution Evaluation for Peptide Precursor / Peptide / Protein IDs with a FDR of 1%**

| | Peptide Precursors (Q Value <0.01) | Peptide IDs (Q Value <0.01) | Protein Group IDs (Q Value <0.01) |
|---|---|---|---|
| Spectral Lib | 135,356 | 94,010 | 7,222 |
| 30k MS1 | 70,656 | 54,373 | 6,033 |
| 60k MS1 | 76,533 | 58,104 | 6,166 |
| 120 MS1 | 77,928 | 59,082 | 6,205 |
| 240 MS1 | 76,485 | 57,889 | 6,151 |

**Figure 6. MS1 resolution vs. CV of the identified peptides**

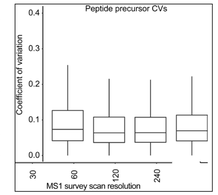
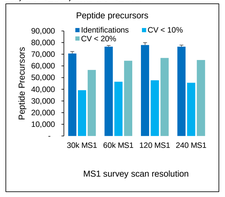
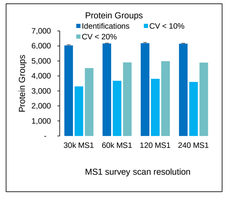



**Figure 7. MS1 Resolution Evaluation for Peptide precursors / Protein IDs with a FDR of 1%, CV <10%, and CV <20%.**




## CONCLUSIONS

The performance of a DIA methods were evaluated for different scan resolutions, gradient lengths and the sampling of the peaks on a Q Exactive HF mass spectrometer. Using the Spectronaut software, identification and quantification of total peak areas on the MS1 level were used as performance measures.

A novel DIA method using a resolving power of 120,000 for MS1 scans and 30,000 for DIA MS2 scans gave the best identification and quantitation results.
•With a 4 hour LC gradient over 100,000 peptide precursors, >79,000 peptides, and >6,700 proteins were identified with an FDR of 1%.
•With a 2 hour LC gradient, more than 80% of the identified peptides (~ 67,000) and proteins (~ 5,000) are quantified on MS1 with a CV% better than 20%.

The median CVs of all the DIA methods are lower than 10%. This result demonstrates that the high resolving power of the Orbitrap mass analyzer enables highly reproducible precision quantitation on MS1.

## TRADEMARKS/LICENSING