Maximizing Spectrum Identification Rate in Shotgun Proteomics on the Q Exactive Mass Spectrometer

Yue Xuan, Eugen Damoc, Andreas Kuehn, Catharina Crone, and Markus Kellmann
Thermo Fisher Scientific GmbH, Bremen, Germany

Thermo SCIENTIFIC

Overview

Purpose: Improved throughput and robustness in shotgun proteomics on a Thermo Scientific™ Q Exactive™ mass spectrometer.

Methods: Tryptic HeLa digest was measured with data-dependent top15 experiments on a nanoLC online coupled to a Q Exactive mass spectrometer (MS). Different AGC target values for tandem MS/MS (T = 2e4, 5e4, 1e5, 5e5, 1e6, and 3e6) are investigated.

Results: The number of identified proteins and peptides per single run reaches a maximum at target value 1e5. Good quality of MS2 spectra can already be achieved with 5e4, with a comparable peptide identification score to 1e6. Higher target ions (> 1e5) for MS/MS do not increase the number of identified proteins, but decrease the overall efficiency of the instrument in high-throughput proteomics applications.

Introduction

Due to its high sensitivity and fast scan speed, the Q Exactive MS has been widely applied in the shotgun proteomics field. A typical shotgun proteomics experiment is a data-dependent, Top N experiment in which tens of thousands of MS/MS spectra are acquired. The quality of an MS/MS spectrum, in terms of signal-to-noise ratio, mass accuracy and numbers of fragments, determines whether a peptide is identified from an MS/MS spectrum via database search. The relationship of the number of ions analyzed at a particular AGC target is clearly related to the identification success, and the optimal AGC target value has the potential to increase the efficiency of the instrument for high-throughput proteomics applications. In this application, different AGC values are investigated to optimize the shotgun analysis on the Q Exactive MS.

Methods

Sample Preparation

HeLa whole-cell tryptic digest as powder, dissolved and further diluted with water with 0.1% TFA into a final concentration 500 ng/µl. 2 µl was injected onto the LC column resulting in 1 µg on-column as a final concentration.

Liquid Chromatography

Tryptic HeLa digest (1 µg) was measured on a Thermo Scientific™ EASY-nLC 1000 online coupled to a Q Exactive MS. The sample was loaded onto a Thermo Scientific™ EASY-Spray™ column, (25 cm x 75 µm ID, Pepmap C18 2 µm) and separated with a linear gradient of 8-25% buffer B (0.1% formic acid in acetonitrile) at a flow rate 250 nL/min over 70 min. The total run time is 95 min due to loading and washing steps.

TABLE 1. Top 15 data-dependent method.

Full scan	
Resolution	70,000 (@ *m/z* 200)
AGC target value	3e6
Scan range	350-2000
Maximal injection time (ms)	100
Data-dependent MS/MS	
Resolution	17,500 (@ *m/z* 200)
AGC target value	2e4, 5e4, 1e5, 5e5, 1e6, and 3e6
Maximal injection time (ms)	50
Isolation window width (Da)	3
NCE (%)	27
Data-dependent settings	
Underfill ratio (%)	1%
Charge exclusion	Singly charged
Peptide match	preferred
Exclusion isotopes	on
Dynamic exclusion (s)	30

Mass Spectrometry

The EASY-Spray source was installed on a Q Exactive mass spectrometer. Source settings: spray voltage of 1600 V and capillary temperature of 250 C. MS data were acquired using a data-dependent top15 method, which are listed in Table 1. Each method was analyzed three times. All data are acquired in centroid mode.

Data Analysis

The data files were searched with Thermo Scientific™ Proteome Discoverer™ software v.1.4 with Sequest HT and Mascot™ v. 2.3 (Matrix Sciences Ltd., London, UK) search engines (Table 2).

TABLE 2. Database search details.

Search Platform	Proteome Discoverer v. 1.4
Search Engine	Mascot v 2.3 and Sequest HT
Database	Ipi.HUMAN.v.3.87
Max. Missed cleavage	2
Precursor mass tolerance	20 ppm
Fragment mass tolerance	20 mmu
Static Modifications	Carbamidomethyl (C)
Dynamic Modifications	Oxidation (M)
	Deamidated (N,Q)
FDR	1%

Results

Data-Dependent Top15 Method

1 µg of HeLa tryptic whole cell lysate digest, separated over a 70 min gradient via reversed-phase chromatography on an EASY-nLC 1000. MS data was acquired using a data-dependent top15 method dynamically choosing the most abundant precursor ions from the survey scan (350–2000 Da) for HCD fragmentation.

For most efficient and parallel operation, maximum injection times for MS/MS is set to 60 ms to ensure that the accumulation of ions is not taking longer than the detection of the previous scan, which are acquired at resolution of 17,500 (Figure 1). A full scan cycle of one full MS scan followed by 15 HCD scans is performed in ~1.6 sec (Figure 2).

FIGURE 1. Schematic of the ddTop10 HCD acquisition method comprising a Full MS scan followed by 10 data dependent HCD MS/MS scans. In this example. the Full MS scan set to the resolution of 70,000 which corresponds to a detection time of 256 ms. The following 10 HCD MS/MS scans are set to be detected at the lowest resolution of 17,500 corresponding to a detection time of 64 ms each. Injections of ions for the next scan occurs in parallel to the detection of the ongoing scan as indicated by the blue boxes.

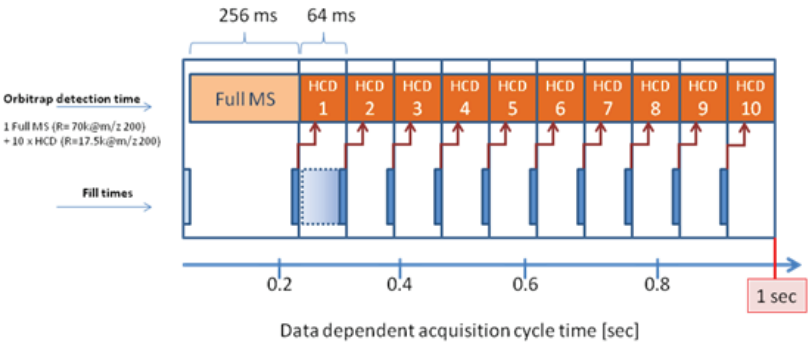
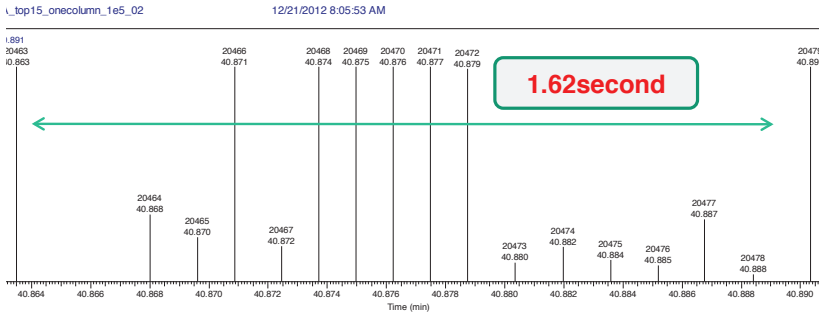


FIGURE 2. One duty cycle of ddTop15 experiment. Each stick represents one scan. The entire duty cycle of a real HeLa run is 1.62 seconds, from one full scan followed by 15 data dependent MS/MS to the next full scan.



Optimized AGC Target Value for Protein IDs

The number of identified proteins and peptides per single run increases while the AGC target value increases until 1e5, where the maximal number of identified proteins (> 3500) and identified unique peptides (~ 17,000) filtered with 1% false discovery rate (FDR) are reached for each single run. Each method was run in triplicate. The sum of protein ID and peptide ID with 1% FDR rate are shown in Figure 3 and Figure 4. Numbers of identified protein and peptides at each target value from each single run are listed in Table 3.

The database search results are filtered to satisfy a of 1% FDR or better, which were calculated using the Percolator node in Proteome Discoverer software 1.4. The Percolator node uses semi-supervised machine learning to discriminate correct from incorrect peptide spectrum matches and calculates accurate statistics such as the q-value (FDR) and posterior error probability to improve the number of confidently identified peptides at a given false discovery rate. It also assigns a statistically meaningful q-value to each peptide spectrum match (PSM), as well as the probability of the individual PSM being incorrect[1].

FIGURE 3A. Number of Identified Proteins from triplicate runs with 1 % FDR at each target value. 1e5 results in 4285 proteins, which are the most among all target values.

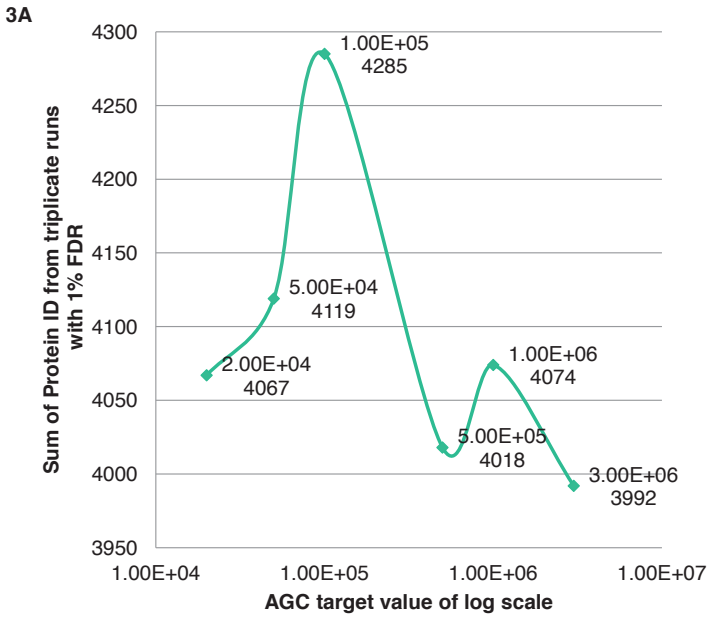


FIGURE 3B. Pie view of identified proteins from triplicate runs at target value 1e5. 3023 proteins are identified by all three runs, each run has around 300 proteins exclusively identified not present in the other two runs. On average, each run has > 3500 proteins identified with 1% FDR.

3B

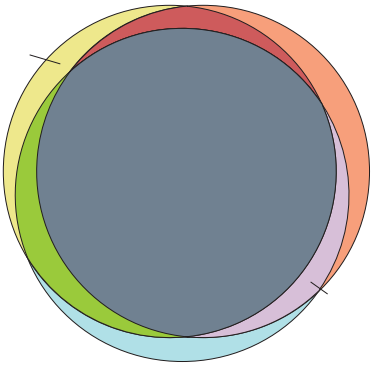
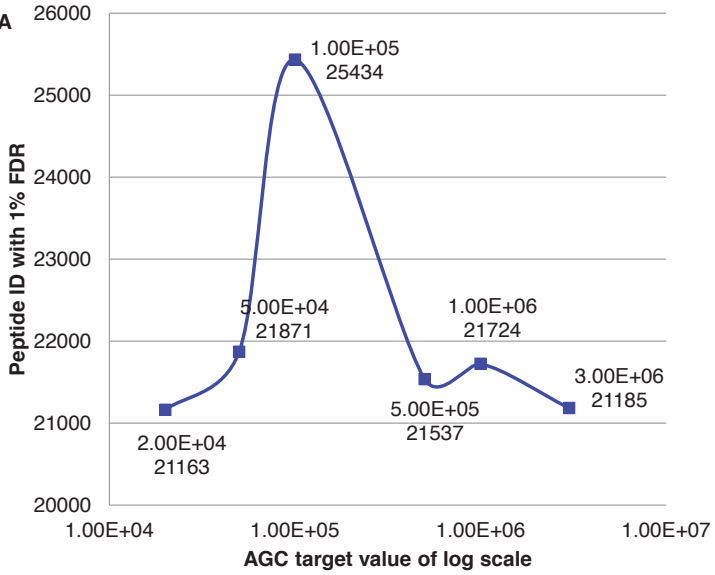
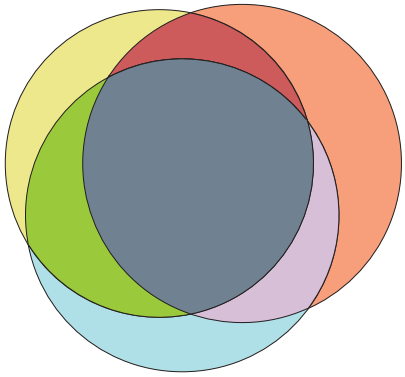


FIGURE 4. A) Number of identified peptides from triplicate runs with 1 % FDR at each target value. 1e5 results in 25434 unique peptides, which are the most among all target values. B) Pie view of identified peptides from triplicate runs at target value 1e5. In average, each run has around 17000 peptides identified with 1% FDR.

4A



4B



An average of ion injection time for MS/MS scans is around 30 ms for the AGC target value 1e5. With ten-fold higher target values for MS/MS, 1e6, the maximal injection time of 60 ms is always reached, which slightly increases the duty cycle and results in less MS/MS spectra in total per run. Good quality of MS2 spectra can already be achieved with 5e4, with a comparable peptide identification number to 1e6. 1e6 target ions for MS/MS do not increase the number of identified proteins.

Table 3. Number of identified proteins and peptides at each target value, from each single run and sum of triplicates, with 1% FDR rate.

Runs	Target Value	Protein ID with 1% FDR	Peptide ID with 1% FDR	Sum of Protein IDs from triplicates with 1% FDR	Sum of Peptide IDs from triplicates with 1% FDR
2e4_01	2e4	3394	15872	4067	21163
2e4_02	2e4	3439	15829		
2e4_03	2e4	3466	15805		
5e4_01	5e4	3565	17125	4119	21871
5e4_02	5e4	3573	16822		
5e4_03	5e4	3496	16916		
1e5_01	1e5	3535	16737	4285	25434
1e5_02	1e5	3539	17912		
1e5_03	1e5	3547	17327		
5e5_01	5e5	3481	16777	4018	21537
5e5_02	5e5	3439	16750		
5e5_03	5e5	3418	16663		
1e6_01	1e6	3523	17190	4074	21724
1e6_02	1e6	3501	17980		
1e6_03	1e6	3490	16998		
3e6_01	3e6	3384	16234	3992	21185
3e6_02	3e6	3407	16342		
3e6_03	3e6	3399	16442		

Conclusion

- More than 3400 proteins and > 15000 unique peptides are identified out of the whole HeLa cell lysate digest per a 75 min single experiment by using data-dependent top15 method on a nanoLC-Q Exactive mass spectrometer system.
- Different AGC target values (2e4, 5e4, 1e5, 5e5, 1e6, and 3e6) for MS/MS are evaluated. 1e5 results in the maximum number of identifications of proteins and peptides.
- Good quality of MS2 spectra can be achieved with 5e4, with a comparable peptide identification score to 1e6. 1e6 target ions for MS/MS do not increase the number of identified proteins, while lower target values increase the efficiency of the instrument for high-throughput proteomics applications.

References

1. Käll, L.; Storey, J. D.; MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. J Proteome Res, ***2008***, 7: 40-44.

www.thermofisher.com

Africa-Other +27 11 570 1840
Australia +61 3 9757 4300
Austria +43 1 333 50 34 0
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China +86 10 8419 3588
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland/Norway/Sweden +46 8 556 468 00
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9434
Italy +39 02 950 591
Japan +81 45 453 9100
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Russia/CIS +43 1 333 50 34 0
South Africa +27 11 570 1840
Spain +34 914 845 965
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC

Part of Thermo Fisher Scientific

ASMS13_Th227_YXuan_E 07/16S