In-Depth Proteome Coverage by Iterative Data Dependent Acquisition on a Benchtop Orbitrap Mass Spectrometer

Mathias Mueller, Andreas Kuehn, Yue Xuan; Tabiwang N. Arrey, Thomas Rietpietsch, Florian Grosse-Coosmann, Catharina Crone, Torsten Ueckert, Markus Kellmann
Thermo Fisher Scientific, Bremen, Germany

Poster Note 64431

Overview

Purpose: Increase the dynamic range for protein identification in complex proteome samples, by applying an iterative acquisition strategy.

Methods: After each LC-MS/MS run the data scenario is analyzed and a strategy regarding LC and MS parameters for a next iteration is determined, based on potential precursor candidates, their timely distribution and ion flux trends. Hereby, already triggered precursors are excluded, while missed candidates are included as targets of interest.

Results: Monitoring the time and flux distribution of available precursors over LC-MS/MS runs can be used to adapt sample loads for iterative LC-MS/MS runs to stay at a high sequencing speed with a maximized number of selectable precursors.

Introduction

Data dependent Top N methods are widely established for large scale bottom-up tandem MS protein sequencing. But protein identification is increased only to a certain point, since this method prefer highly abundant precursors, belonging to highly abundant proteins. Thus, the sequence coverage for highly abundant proteins is increased, while lower abundant peptides were not triggered and their belonging proteins were not improved to identified. Iterative runs with an exclusion of already triggered precursors showed a deeper sequencing of low copy number proteins. [1]

As the precursor distribution is changing with LC separation gradient and sample load, we need to track the total number and time density of accessible candidates.

The detected precursor signals can be clustered by their retention times and the minimum needed ion inject time (IT) for a sufficient quality MS2 spectrum (see FIGURE 2 and TABLE 1).

To adapt the method parameters for the next iteration step, the resulting precursor density over time for different IT bins will give access to determine either

a) the maximum allowed ion inject time (maxIT) to reach a desired number of ddMS2 scans (N) per TopN cycle overall time

b) the maximum number of ddMS2 scans (N) per TopN cycle to keep the ion inject time below a desired threshold, mostly near the transient length of the MS2 scans.

The here described methodology is using rich exclusion/inclusion lists from previous iteration steps and adapts the sample load to stay at a high sequencing speed with a maximized number of selected precursors.

Methods

(Matthias Mann sample –To clarify how to state) HeLa Protein Digest Standard was diluted in HPLC grade H_2O (Fisher Scientific) to a final concentration of 0.5 µg/µL.

Mass Spectrometry

Thermo Scientific ™ Q Exactive ™ HF Mass Spectrometer

Liquid Chromatography
LC Stack: Thermo Scientific™ Dionex™ UltiMate™ 3000 RSLCnano system equipped with nano pump NCS-3500 and autosampler WPS-3000TPL
Mobile Phases: A: 0.1 % FA in water; B: 0.1 % FA in Acetonitrile (Fisher Chemicals)
Gradients: 8–30 % B in of 30, 45, 60, or 90 min
Flow Rate: 250 nL/min
Trapping Column: Thermo Scientific™ Acclaim™ PepMap™100 µCartridge Column C18, 300 µm × 0.5 cm, 5 µm, 100 Å
Separation Column: Acclaim PepMap C18, 75 µm × 50 cm, 2 µm, 100 Å

MS Method Parameter:
[Full MS only] Resolution: 120K, AGC Target: 3e6, maximum IT: 20ms
[Top15] FullMS only settings, ddMS2: Resolution 15K, AGC Target: 1e5, maximum IT: 40ms / 80ms, underfill ratio: 8% (AGC Target 8E3), Isolation width: 1.4 m/z, Peptide match: preferred, Dynamic exclusion: 30s

Data Analysis
All TopN runs have been processed with Thermo Scientific™ Proteome Discoverer™ software 2.0 search engine SEQUEST® HT against IPI fasta database human 3.8. Exclusion and inclusion lists have been generated by a prototype software and partly imported in the Exactive Series instrument TopN method or directly linked to the method.

Results

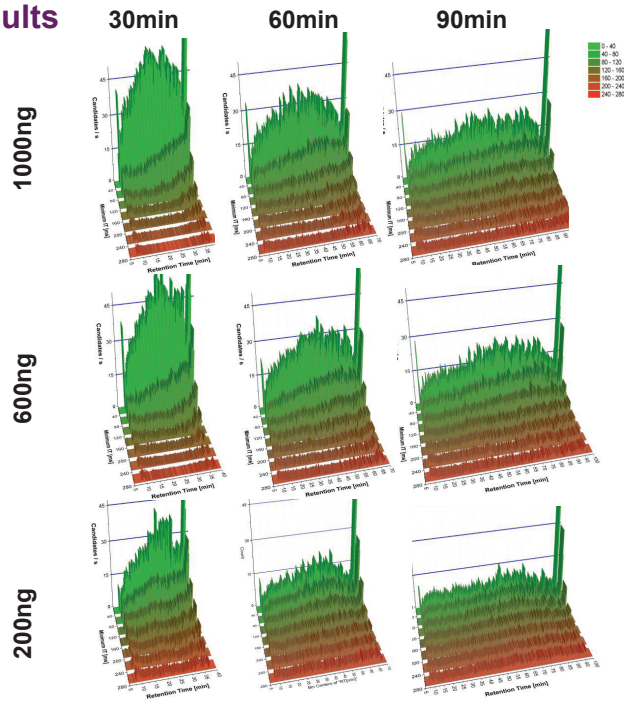

FIGURE 1. Precursors per second for different minimum inject time bands. 2D binning frequency statistics on precursor ion flux over retention time. X-axis: Retention time (10s averaged 1s bins), y-axis: minimum IT (40ms bins) z-axis: frequency (counts). Extracted from Full MS only runs with 30, 60 and 90 min gradients and 200ng, 600ng and 1000ng sample load, each.

Thermo SCIENTIFIC

Gradient [min]	Load [ng]	#Prot Grp	#Peps	PSM	#MS2	Total Precursosr	IT band1 ≤ 40ms	IT band2 ≤ 80ms
30	200	2417	13895	16096	20717	56076	25282	11381
	600	2655	15346	18285	22498	58518	38216	10828
	1000	2713	15425	18545	22672	65267	49934	6780
60	200	3263	20558	24322	34540	93289	24402	15691
	600	3908	24687	30129	40267	102190	48294	20708
	1000	4021	25248	31537	41323	105000	61199	18312
90	200	3362	22196	26953	40142	126126	23951	17263
	600	4466	29983	38467	56347	135281	50619	27556
	1000	4687	31297	40979	58766	139242	62342	32424

TABLE 1. Proteome Discoverer software ID results of the HeLA Top15 runs for different sample load and gradient length. Additional: Extracted total precursor candidates (2 ≤ z ≤ 6, #Isotopes ≥ 2)

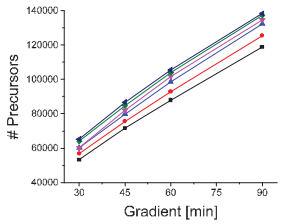

FIGURE 2. Extracted Precursors for a variation of load and gradient

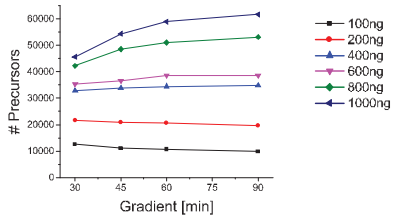

FIGURE 3. Clustered precursors of IT band1 (≤ 40ms) at varying load and gradient

Inter Run Logic: Sample load, Inject time, gradient program

To avoid a decreasing sequencing speed and redundant data generation, the method parameter of following iteration steps can be adapted based on the data scenario of the current iteration. The distribution of the remaining precursor, after removing the already triggered ones, will indicate, if the IT band to run in "parallel mode" is still providing enough candidates over time (Figure 6b). If not, either the sample load or the maxIT can be increased to "refill" the IT bands or accessible precursors. Otherwise, when observing indicaton of high ion suppression effects with decreased dynamic range or coalescense effects, the sample load should be decreased. Thus, a feedback to change the gradient program, like enlonging, is given.

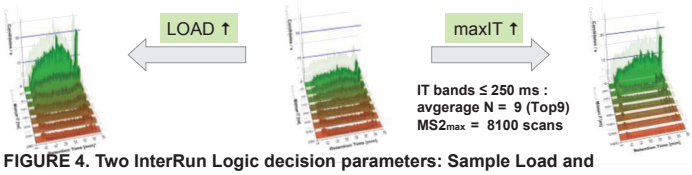

FIGURE 4. Two InterRun Logic decision parameters: Sample Load and Maximum Inject Time (dd-MS2) to "refill" the IT bands or accessible precursors.

Inter Run Logic: Static Exclusion

To avoid redundant data, all precursors with an acquired MS2 scan with sufficiently spectral quality should be excluded, as well, as all of their existing charge members. To avoid ClusterTop hopping, due to statistical variations, a set of up to three most abundant isotopes can be added to the exclusion list. The "timed" exclusion entries will get an additional time buffer to compensate for potential time variations.

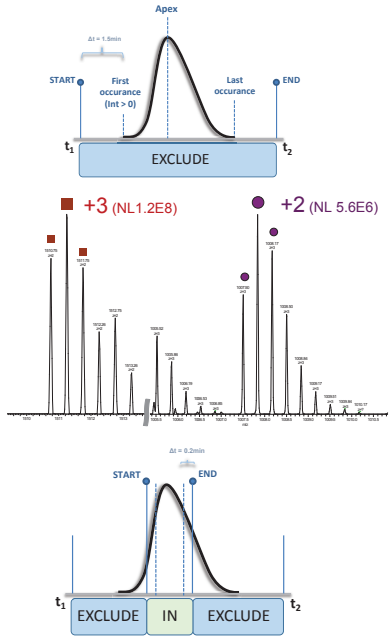

FIGURE 5. Top: Exclusion time span with time buffer. Middle: Isotope clusters for two charge members. Bottom: Inclusion time span with time buffer time and optional adjacent exclusion time spans

Ex

mz	z	start	end
1510.7572	2	40.637	44.759
1511.2557	2	40.637	44.759
1511.7563	2	40.637	44.759
1007.5027	3	40.637	44.759
1007.8365	3	40.637	44.759
1008.1704	3	40.637	44.759

TABLE 2. Exclusion entries, Three most abundant isotopes of each charge member

In

mz	z	start	end
1510.7572	2	42.500	42.800

Ex

mz	z	start	end
1510.7572	2	40.637	42.50
1510.7572	2	42.800	44.759

TABLE 3. Inclusion (top) and optional adjacent exclusion entries (bottom) for the example of a single precursor isotope.

Inter Run Logic: Static Inclusion (IT optimization)

Including precursors can be used to control single charge state picking at a set of charge members or applying a specific selection strategy for a lower IT band.

Additionally, when the elution profile is wide enough or if the experimental setup shows a high retention time stability, a combination of exclusion and inclusion ranges can be applied to focus on the higher flux region and reduce the required inject time.

Iterative Data Dependent Acquisition

Three iterative LC-MS/MS runs of 200ng proteolytic HeLa digest were performed, using a 30min separation gradient with a variable sample load.

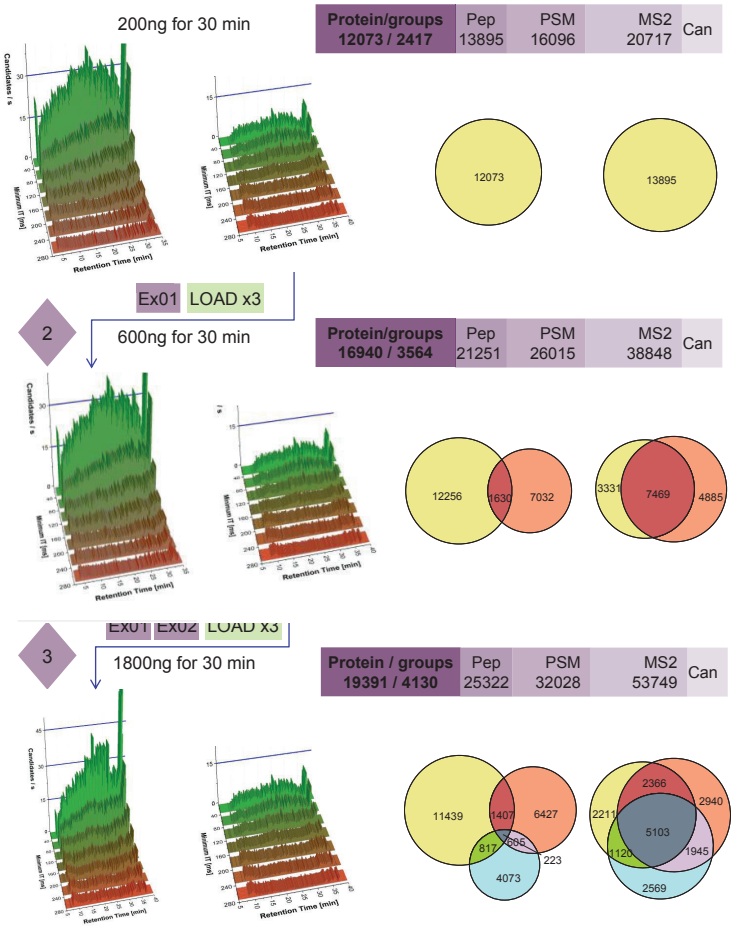

FIGURE 6. Example A including all iteration steps. Left: Precursors per second for different minimum inject time bands, with and without triggered and excluded precursors. Right: Resulting Venn diagrams on peptides and protein levels.

	#Run, load, gradient	Venn Diagram #Peptides	Vann Diagram #Proteins
B	1: 400ng in 45min 2: 1200ng in 45min		
C	1: 400ng in 60min 2: 800ng in 60min		

Table 4. Vann diagrams on peptides and protein levels for Example B and C with the used parameters

Experimental results. Gradient load variation for optimized ID rate

Iterative runs with different combinations of sample loads of a proteolytic HeLa digest were analyzed. Most of the accessible precursors with a required IT ≤ maxIT (IT Band) have been triggered (Cover rate). The significant increase in protein and peptide IDs are shown in Table 5. Example D shows the strategy to adapt the maximum IT (maxIT) to access lower abundant precursors (marked with mIT)

Example	Run	Load [ng]	#Protein groups	#Prots	#Unique Peptides	#PSM	#MS2	IT Band	Cove
A 30min ExList	1	200	2417	12073	13895	16096	20717	25282	82%
	1+2	+600	3564 (+47%)	16940 (+%)	21251 (+%)	26015	38848	38216	100%
	1+2+3	+1800	4130 (+70%)	19391 (+%)	25322 (+%)	32028	53749	55000	96%
		600	*3100*	*13225*	*8866*	*9648*	*18127*		
		1800	*2830*	*11430*	*5821*	*6158*	*14901*		
B 45min ExList	1	400	3291	15778	19911	23962	30548	38924	78%
	1+2	+1200	4343 (+32%)	20205	27743	34800	53996	65000	83%
	2	*1200*	*3649*	*15237*	*10255*	*11097*	*23448*		
C 60min ExList	1	400	3849	18282	24160	29813	39916	39807	100%
	1+2	+800	4772	22128	31940	41116	68788		100%
	2	*800*	*3894*	*16117*	*10867*	*11680*	*28872*		
D 90min ExList maxIT80	1	800	4458	21050	29369	38192	55695	59580	
	1+2	+800 maxIT80	5202	24307	36457	48851	87601	93600	93%
	2	*800*	*4446*	*20258*	*23922*	*30114*	*44403*		

Table 5. ID rates for the different load to gradients experiments # given for each experiment

Conclusion

Monitoring the time and flux distribution of available precursors over LC-MS/MS runs can be used to adapt sample loads for iterative LC-MS/MS runs to maintain high sequencing speeds with a maximum number of selectable precursors.

Alternatively or in combination with a variation of the maximum inject time for data dependent MS2 scans, a significant increase of protein and peptide IDs can be achieved, even with a reduced amount of injected sample.

An extended exclusion strategy of related charged precursor species and isotopes avoids redundant data acquisition, while including precursors can be used to control specific selection strategy for lower inject time band precursors.

References

1. Chen, H.; Rejtar, T.; Andreev, V.; Moskovets, E.; Karger, B. L. *Anal. Chem.* **2005,** *77,* 7816–7825.
2. Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. *Nat. Methods* **2014,** *11,* 319–324.
3. Scherl, A.; Shaffer, S. A.; Taylor, G. K.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R. *Anal Chem.* **2008,** *80(4),* 1182–1191.

www.thermofisher.com

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic)
400 650 5118 PN64431-EN 0616S
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 45 453 9100
Korea +82 2 3420 8600
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
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand