Addressing the reproducibility aspect of LC-MS/MS based protein identification

Matthias Berg[1]*, Lennart Björkesten[2], Harald Pettersen[2], Magnus Isaksson[2] and David Fenyö[2]
GE Healthcare, [1] Freiburg, Germany [2] Uppsala, Sweden

Introduction

A major challenge in proteomics is the lack of reproducibility in protein identification experiments. In LC-MS/MS analyses where the same sample is analyzed in replicates, many of the low abundant peptides that are successfully identified by searching a sequence collection with the MS/MS data are only identified in one or a few of the replicate runs. By using the detection, matching, and visualization approach of DeCyder™ MS were retention time, precursor mass, and the topology of the intensity profile is utilized in combination with the matching of tandem mass spectra, it is possible to achieve repeat analysis with a very high reproducibility.

Methods

In this study we evaluate data from multiple injections of the same protein digest mixture analyzed by 1D LC-MS using an Ettan™ MDLC system (GE Healthcare) in High Throughput configuration directly connected to a Finnigan LTQ™ system (Thermo Electron Corp.). After desalting on a trap column (Zorbax™ 300 SB C18 0.3 x 5 mm, 5 µm particles, Agilent) the bound peptides were separated through a RPC column (Zorbax 300 SB C18 0.075 x 150 mm, 3,5 µm particles, Agilent). The data was acquired using data dependent acquisition with dynamic exclusion.
The peptides were identified by searching the tandem mass spectra from each of the replicate analyses against a sequence collection using X!Tandem (Beavis Informatics) and the reproducibility of identification was studied.
The peptides from the replicate runs were also detected and matched using the DeCyder MS software. The MS/MS information was exported to X!Tandem and the identification results were imported back into DeCyder MS. Peptides showing incomplete or inconsistent identification for the different replicate analyses were assessed in DeCyder MS using the dynamic link to 2D and 3D representations of raw data.
The differences in identification results between replicate analyses were assessed in terms of the variation in the quality of the MS/MS spectra and the variation in peak intensities causing variation in the data dependent acquisition. It was obvious from the visual inspection that most of the peptides were present in all samples (*Fig 3*). This was also supported by the low quantitative variance within the replicate set obtained using DeCyder MS (*Fig 2*).

Results

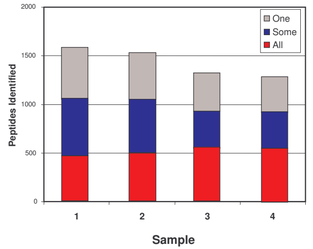

***Figure 1**. The number of peptides identified using X!Tandem with expectation values less than 0.01 in all, some, and only one of the replicates from 4 different Arabidopsis samples.*

Results

	Peptides	Replicate #1	Replicate #2	Replicate #3	Replicate #4	Replicate #5	DeCyder MS
1	AITSTGTKKGDLLLGDVAFQSR	2.00E-03	-	1.50E-04	3.80E-05	2.90E-05	*1.70E-02*
2	ASALIQHEWKPK	-	9.90E-03	-	8.40E-03	-	1.30E-03
3	DSTITVGTGHSLD	-	-	-	-	1.80E-05	7.10E-04
4	DSTITVGTGHSLDPLTSVK	1.80E-09	2.50E-09	1.80E-07	4.60E-09	6.10E-09	4.30E-10
5	EDLIASLTVNDK	1.20E-05	8.30E-06	1.10E-06	1.50E-06	7.30E-07	3.70E-08
6	EWKPKSFTTIGEVDTK	4.50E-03	*1.40E-02*	2.40E-03	*1.30E-02*	*8.10E-02*	-
7	FNTAVGAEVSHK	4.60E-06	1.20E-06	1.30E-04	7.70E-06	4.00E-06	7.10E-08
8	FSITTFSPAGVAITSTGTK	-	7.20E-12	5.90E-11	-	2.10E-04	3.30E-13
9	GDLLLGDVAFQSR	-	-	-	1.70E-05	-	3.90E-07
10	GDLLNASYYHIVNPLFN	1.30E-06	-	2.20E-06	3.70E-06	-	1.40E-07
11	GDLLNASYYHIVNPLFNTAVGAEVSHK	4.20E-11	3.80E-09	3.10E-07	7.50E-09	1.60E-07	1.50E-10
12	GPGLYTEIGK	4.20E-05	9.80E-08	3.60E-07	2.80E-07	2.10E-06	1.00E-07
13	GPGLYTEIGKK	8.20E-03	1.60E-03	1.60E-03	-	-	5.30E-04
14	GTQHSLDPLTSVK	-	-	1.30E-05	-	-	1.20E-04
15	HIVNPLFNTAVGAEVSHK	-	-	-	-	-	2.80E-12
16	ITHPNFNGNTLDNDIMLIK	4.10E-07	*1.60E-02*	*5.50E-02*	2.60E-03	*1.10E-02*	1.50E-10
17	INAGLSFTK	*2.20E-02*	*2.90E-02*	*6.50E-02*	*2.90E-02*	*3.10E-02*	4.40E-03
18	KGDLLLGDVAF	-	-	3.60E-04	-	3.50E-05	5.60E-05
19	KGDLLLGDVAFQSR	3.30E-12	4.00E-11	2.10E-11	9.50E-12	2.60E-11	9.90E-11
20	LGEHNIDVLEGNEQFIN	-	3.30E-08	5.20E-07	6.20E-08	2.40E-08	2.30E-08
21	LGEHNIDVLEGNEQFINAA	-	5.50E-09	6.50E-10	8.50E-09	4.90E-11	1.00E-10
22	LGEHNIDVLEGNEQFINAAK	1.10E-10	8.10E-07	6.40E-13	3.80E-11	2.70E-13	4.70E-14
23	LGEHNIDVLEGNEQFINAAKITHPNFNGN	5.50E-07	1.40E-06	-	-	3.70E-06	-
24	LSSPATLNSR	*2.30E-02*	6.30E-03	*4.50E-02*	*4.40E-02*	*1.00E-02*	1.30E-03
25	PGLYTEIGK	4.00E-04	4.10E-03	3.80E-04	1.30E-03	2.20E-04	2.40E-04
26	SFTTIGGEVDTK	1.30E-09	1.20E-09	1.30E-09	1.10E-10	2.20E-10	2.00E-09
27	SSPATLNSRVATVSLPR	*2.20E-02*	8.20E-03	*3.10E-02*	5.70E-03	-	*1.90E-02*
28	TVQTQHSLDPLTSVK	2.90E-07	-	1.00E-07	7.10E-08	3.00E-08	2.20E-07
29	VATVSLPR	2.50E-03	2.80E-03	5.40E-04	1.70E-03	1.40E-03	7.50E-04
30	VCTDSTFLITATVDEAAPGLR	3.40E-11	-	8.50E-13	2.00E-11	2.20E-11	4.00E-14
31	VELGYLHEY	2.80E-03	-	-	-	-	4.10E-03
32	VKGPGLYTEIGK	4.30E-05	4.20E-05	4.70E-04	8.50E-03	1.60E-04	3.30E-05
33	VNSAGIASALIQHEWKPK	6.00E-04	1.80E-08	3.80E-10	1.40E-09	9.20E-09	1.50E-09
		21	20	23	22	21	29

***Table 1**. The expectation value (e) for peptides with e<0.01 from an arabidopsis porin (Ensembl:At3g01280.1) for five replicate LC-MS/MS runs. There is a variation in the expectation values for the same peptide in the different replicate analyses caused by the varying quality of MS/MS data.*

The results from the combined DeCyder MS and X!Tandem analysis is shown in the last column. Here, the peptides from the replicate runs were also detected and matched using the DeCyder MS software. The MS/MS information associated with the detected peptides was exported and searched with X!Tandem. This resulted in a more complete list of identified peptides as well as faster searches because of the smaller set of MS/MS spectra used in the search..

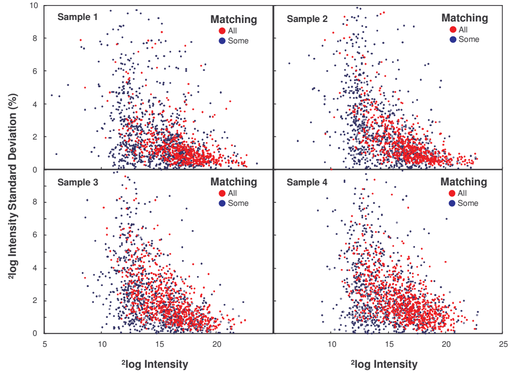

***Figure 2**. The variation in intensity between different replicate runs is shown for 4 different samples. The peptides matching all replicate intensity maps are shown as red dots and the peptides matching some replicate intensity maps (e.g. 1-3) are shown as blue dots. The random variation in ²log peak intensity between repeat analyses is in the range of a few percent, making it straightforward to compare repeat analyses by comparing intensity maps.*

Conclusions

- The variation in intensity between replicate analyses of the same sample causes variation in the data dependent acquisition of MS/MS spectra and in the quality of the MS/MS spectra acquired, leading to variation in the peptides that are identified by database searching.
- It is possible to increase reproducibility of repeat analysis by using the detection, matching, and visualization approach of DeCyder MS were retention time, precursor mass, and the topology of the intensity profile is utilized in combination with the matching of tandem mass spectra.

Results

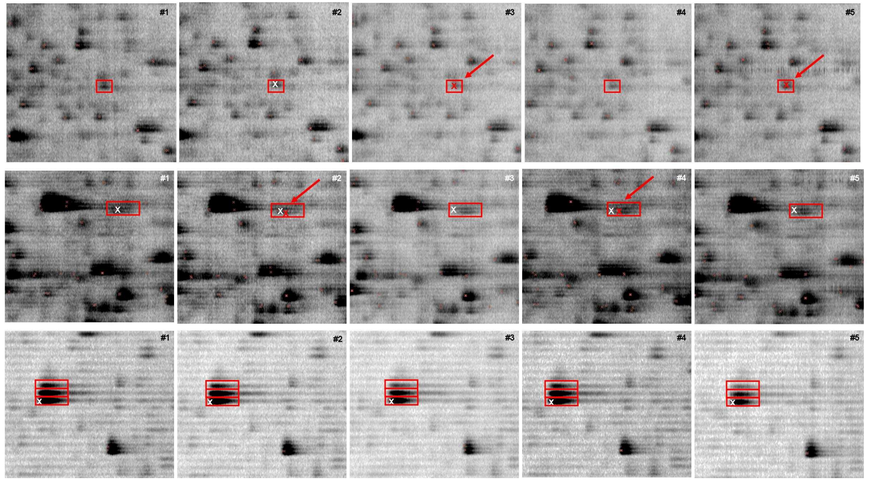

***Figure 3**. Examples intensity maps showing three peptides (From top: KGDLLLGDVAF, ASALIQHEWKPK, INAGLSFTK) from an arabidopsis porin (Ensembl:At3g01280.1) for five replicate LC-MS/MS runs showing that these peptides are present in all samples. The MS/MS spectra that lead to a successful identification are indicated by red markers and the MS/MS spectra that did not lead to a successful identification are indicated by white markers.*

* To whom correspondence should be addressed: +49 761 451 92 35 , matthias.berg@ge.com