Software for automated differential expression analysis using 2D and 3D representations of LC-MS data for interactive confirmation of results

Lennart Björkesten*, Anders Kaplan, Malin Söderström, Harald Pettersen, Staffan Lindqvist, John Flensburg and David Fenyo
GE Healthcare, Uppsala, Sweden

Introduction

DeCyder™ MS *(Fig. 1)* is a novel software program for fully automated differential expression analysis based on LC-MS/MS data including detection, quantitation, sample to sample comparison and statistical data evaluation. Results can be interactively confirmed against original raw data through 2D and 3D data visualizations.

Methods

LC-MS data is represented as signal intensity maps. Peptides are detected, quantified, and assigned a charge state using a novel image analysis algorithm. Peptides are then matched across signal intensity maps from different samples. This allows for the identification of peptides showing a significant abundance variation between different groups of samples using statistical analysis and including sample-to-sample standardization. Proteolytic peptides can be identified through a link to protein identification software and sorting peptides based on protein identity results in a compilation of abundance information on the protein level. The different analysis steps can either be initiated separately or launched from a batch processing module.

Results

The final result of the analysis is a list of identified and confirmed peptides with a significant variation pattern. The different steps are illustrated with screen dumps from the DeCyder MS software program when applied to LC-MS data sets from two different MS instrument types, Finnigan LTQ™ (Thermo Electron Corp.) and Q-Tof™ (Waters Corp.).

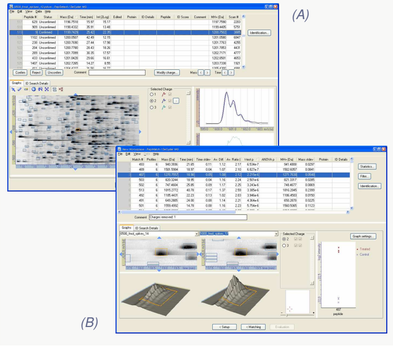

***Fig 1**. DeCyder MS software modules. PepDetect (A) is used for detection and quantitation and PepMatch (B) is used for matching and sample-to-sample comparison*

Data analysis workflow

Visualize LC-MS data as signal intensity maps

LC-MS data is represented as signal intensity maps with the chromatographic separation on one axis and the mass separation on the other. The signal intensity for a certain elution time and mass is represented by a shade in an adjustable grayscale *(Fig 2)*.
Signal intensity maps have shown to be very useful for the overall data quality assessment since chromatographic separation as well as the MS separation can be monitored simultaneously. Phenomena like source fragmentation and adduct formation can easily be revealed as co-eluting features suggesting a need for MS method optimization. The optimization of the MS/MS acquisition method is supported by small red markers indicating the presence of MS/MS scans in the signal intensity map *(Fig 4)*. The signal intensity map can further be used to estimate the *Typical peak width* to use with the peptide detection algorithm as an indication of the chromatographic width of the peptides.

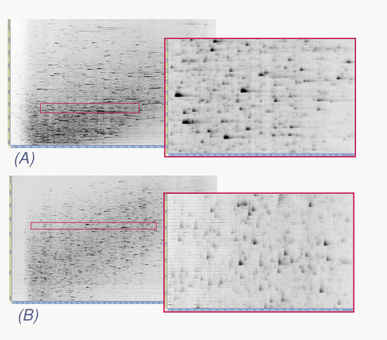

***Fig 2**. Full signal intensity maps (left) and corresponding detail views (right) from complex samples analyzed using LTQ (A) and Q-Tof instruments (B).*

Detect, quantitate and identify peptides

Peptides are detected, quantified, and assigned a charge state using a novel image analysis algorithm. The algorithm reassemblies the signal intensity from the isotopic patterns and from the different charge states present. The resulting peptides are listed in a table ready for an optional interactive confirmation procedure. If MS/MS data is available this data can be exported to an external database search engine and the protein identification results read back. The best candidate is presented for each identified peptide. Data is finally saved on file and made available to the matching and comparison module.
DeCyder MS allows for detection and quantitation of one intensity map at a time allowing for interactive refinement using a set of advanced detection parameters or for detection of a set of intensity maps in batch mode.

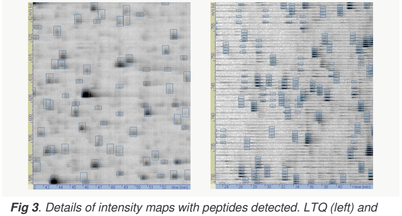

***Fig 3**. Details of intensity maps with peptides detected. LTQ (left) and Q-Tof (right).*

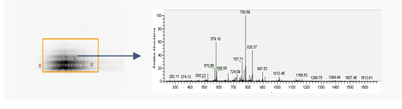

***Fig 4**. Red markers indicate that MS/MS data is available.*

Match and normalize intensity maps

Peptides are matched across signal intensity maps from different samples to facilitate sample to sample comparison. DeCyder MS allows for quantitative comparison using the original intensities or intensities can be normalized according to one of two different methods:

1. If peptides of known amount have been added to all samples and detected, these peptides can be used to standardize peak intensities between different samples.
2. Assuming that the majority of the peaks remain unchanged, the full peptide intensity distributions can be used for standardization.

Fig. 5 and *Fig. 6* show the result of matching peptides detected in technical replicates of complex protein fractions from plant proteomes. The coefficient of variance (COV) for peptides within each replicate set suggests a good basis for scanning for differentially expressed peptides.

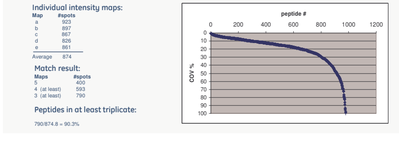

***Fig 5**. Number of peptides detected in each sample and corresponding match statistics for a five-fold technical replicate analyzed in an LTQ system (Left). Corresponding variance plot (Right).*

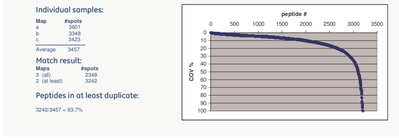

***Fig 6.** Number of peptides detected in each sample and corresponding match statistics for a technical triplicate analyzed in an Q-Tof system (Left). Corresponding variance plot (Right). The feasibility of scanning for 1.5-fold variation among several thousands of peptides using only tree samples in each group is obvious.*

Find significant peptides

Peptides showing a significant abundance variation between different groups of samples can be identified using Student t-test or ANOVA statistical analysis (*Table 1*).

Match #	Profile	Mass Cal.	Time (m)	Time stdev (mins)	Av. Ratio	Anal-2	Protein	Comment
456	6	2109.46	11.85	0.05	3.66E07	2.4E-04	Myoglobin (2)	
448	6	1981.15	12.18	0.05	2.58179	1.8E-04	Myoglobin (1)	
605	5	2019.24	12.19	0.06	2.679095	3.7E-04	Myoglobin (1)	Protein. added (+39) to 1981.15
443	6	1853.04	12.67	0.05	2.533043	6.6E-05	Myoglobin	
450	6	1915.09	12.92	0.07	2.672192	5.4E-06		
602	5	1981.62	13.62	0.09	2.166393	3.5E-04		
493	5	2810.2	13.66	0.06	1.31869	9.7E-04		
440	6	1559.5	14.78	0.08	2.234635	4.5E-06		
603	5	1485.68	16.71	0.06	2.274373	3.2E-05		
388	6	896.73	18.27	0.10	1.705768	1.6E-05	Angiotensin	
601	5	867.3	18.94	0.04	2.453148	2.3E-04		
430	6	1501.6	18.97	0.04	2.098855	5.1E-07	Myoglobin	
428	6	1270.76	18.98	0.05	2.121608	1.8E-06	Myoglobin	
589	5	705.35	18.98	0.05	1.760416	8.0E-05	Myoglobin	fragment MS/MS
425	6	1267.87	20.36	0.09	2.079678	1.0E-05		
453	6	962.77	21.65	0.11	2.577532	2.5E-05	Myoglobin(1)	Sodium adduct (+22) to 940.39
434	6	940.39	21.65	0.11	2.172717	5.5E-07	Myoglobin (1)	
433	6	1105.44	22.23	0.13	2.03094	4.2E-06		
418	6	1660.78	22.25	0.13	2.027738	7.1E-06	Myoglobin	
429	6	1605.82	23.07	0.06	2.240612	1.8E-05	Myoglobin	
438	6	1078.51	23.08	0.08	2.305784	5.0E-05		fragment
357	6	1157.4	23.09	0.08	1.160205	3.9E-04		fragment
435	6	1007.49	23.09	0.07	2.257523	3.5E-05		
458	6	1643.77	23.09	0.07	2.435117	7.1E-04	Myoglobin	Protein. added (+39) to 1605.82
461	6	1628.88	23.09	0.07	2.448136	9.2E-05	Myoglobin	Sodium adduct (+22) to 1605.82
389	6	1414.91	23.82	0.10	1.673244	1.4E-05		
432	6	649.28	24	0.08	2.210419	3.8E-06	Myoglobin	
872	4	1944.88	24.69	1.81	0.035579	6.7E-04		
317	6	1333.85	24.93	0.06	0.952446	7.8E-04		
441	6	747.46	25.85	0.09	2.252896	3.3E-06	Myoglobin	
412	6	1569.63	26.39	0.08	2.037748	1.8E-05	Fibropeptide B	
421	6	1607.72	26.41	0.07	2.093579	3.1E-04	Fibropeptide B	Protein. added (+39) to 1569.62
420	6	1591.73	26.41	0.07	2.175007	2.7E-04	Fibropeptide B	Sodium adduct (+22) to 1569.62
445	6	1505.97	27.38	0.09	2.325633	1.4E-05	Myoglobin	
21	6	2007.72	29.63	0.09	1.179554	2.9E-05		
407	6	2161.82	30.21	0.09	1.842533	3.5E-05		
392	6	985.9	30.65	0.13	1.71584	4.0E-04		
442	6	1377.81	30.78	0.13	2.378475	1.6E-05	Myoglobin	
128	6	2140.95	32.49	0.10	1.379527	7.8E-04		
639	4	2592.81	32.69	0.05	1.045267	8.7E-04		
451	6	2548.91	35.54	0.13	2.685274	6.1E-05		
394	6	1253.06	36.02	0.11	1.802362	2.1E-04		
449	6	1815.28	40.78	0.17	2.589119	3.4E-06	Myoglobin	
478	6	1862.82	40.81	0.21	6.808039	9.3E-04		

***Table 1**. Horse myoglobin peptides, angiotensin and fibropeptide B were spiked in triplicate on an eight protein background mix to mimic 2-fold variation (500/1000fm) 45 significant peptides were found by the application of Student t-test ($p<10^{-3}$). Co-eluting peptides are indicated, supporting the identification of adducts and fragments*

Confirm results

Results can be interactively confirmed against original raw data through 2D and 3D data visualizations (*Fig 7*).

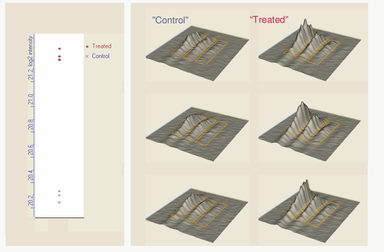

***Fig 7**. Interactive confirmation of a significantly varying peptide in a "Control/Treated" dataset run on an Ettan™ MDLC-LTQ system (Peptide 434 in Table 1). The corresponding abundance ratio was 2.17 and the Student t-test p-value was $5.5 \cdot 10^{-7}$.*

Conclusions

DeCyder MS provides tools for:

- assessment of LC-MS raw data quality.
- automatic detection, comparison and presentation of peptides with significant variation between groups of samples for large datasets.
- visualization, verification and adjustment of detection and comparison down to a peak-by-peak level.

* To whom correspondence should be addressed: +46 18 61 20557, lennart.bjorkesten@ge.com