PosterREPRINT

Waters

STRATEGIES FOR IDENTIFYING PROTEINS IN COMPLEX MIXTURES BY ELECTROSPRAY MASS SPECTROMETRY

Therese McKenna[1]; James Langridge[1]; Mark Ritchie[1]; Ole Jensen[2]; Allan Stensballe[2]; Thomas Nühse[3]; Scott Peck[3]; Richard Denny[1]; Keith Richardson[1]; Phillip Young[1]; Matthew Willets[1]; Chris Hughes[1]

[1]Waters Corporation, Manchester, UK; [2]University of Southern Denmark, Odense, Denmark; [3]Sainsbury Laboratory, Norwich, UK.

Presented at ASMS, Montréal, Canada, 8th-12th June 2003

OVERVIEW

Five different analytical approaches for protein identification are compared and contrasted.

- Two methods involved fractionation (LC-MALDIprep™ with MALDI Q-Tof™ and LC with automated nanospray) three methods used on-line LC/MS/MS (DDA™, iterative DDA and real time databank searching).
- Initially the five systems were used to analyze a tryptic digest of 14 standard proteins.
- Subsequently these techniques were used on a very complex plasma membrane tryptic digest obtained from Arabidopsis thaliana.
- The individual merits of each system are discussed.

INTRODUCTION

Currently, the most commonly used approach for protein identification and characterization using mass spectrometry involves the detection and subsequent fragmentation of proteolytic peptides. Often the proteins are subjected to some form of pre-fractionation and separation prior to digestion, reducing the complexity of the digest and the dynamic range of protein concentrations.

A greater challenge is the direct analysis of an unseparated digest mixture, of perhaps a cell lysate or sub-cellular fraction. In such samples, the challenge is the analysis of low abundance proteins from the mixture, as the detection and identification of the proteins may be biased towards the larger and most abundant species. The solution may be to use intelligent data gathering, in which the instrument is able to select peptides for MS/MS only from proteins which have not been previously identified. Alternatively, more rigorous interrogation of the sample by either fractionation or variable flow chromatography could be best suited to the characterization of complex digest mixtures.

Here we compare the performance of five different systems: off-line LC separation in combination with automated nanoelectrospray, LC-MALDI with MS/MS analysis using a MALDI Q-Tof and LC/MS/MS on an ESI Q-Tof mass spectrometer. A standard mix of 14 protein digests was used to validate the methods, prior to the analysis of an Arabidopsis thaliana plasma membrane fraction.

METHOD

Samples

Standard 14-Protein mix

14 proteins (Serum albumin (Bovine), Alcohol dehydrogenase (Yeast), Alpha lactalbumin (Bovine), Beta casein (Bovine), Carbonic anhydrase (Bovine), Catalase (Bovine), Cytocrhome C (Horse), Enolase (Yeast), Fetuin (Bovine) Glucose oxidase (Asp. niger), Lactoperoxidase (Bovine), Lysozyme (egg white), Phosphorylase B (Rabbit), and ribonuclease A (Bovine)) were dissolved separately to 100 µg/mL in 0,1M NH_4HCO_3. Each was then reduced with DTT (4hrs room temp) and alkylated with iodoacetemide (4hrs room temp), then digested with sequence grade trypsin (Promega) for 4hrs at 37°C. The solutions were diluted to a convenient concentration then mixed to give a 160 fmol/mL solution. For experiments on the Q-Tof Ultima™ API 1µL (160 fmols) was used. All other experiments used 3 µL (480 fmols).

Arabidopsis thaliana Plasma Membrane sample

Plasma membranes from Arabidopsis thaliana cell culture treated with fungal elicitors were isolated by two-phase partitioning. The membrane preparation was extensively washed to remove peripheral proteins and the microsomes flipped "inside-out" by detergent treatment [Nühse et al, Poster Number: 4:20]. The sample was digested o/n (1:50 ratio trypsin/protein) and acidified with an equal volume of 10% FA and cell debris pelleted by centrifugation at 30,000 g for 30 min.

The supernatant containing the cytosolic domains of membrane associated proteins was analyzed.

LC/MS/MS experiments (DDA, RTDS)

On-line Chromatography

- Trapping column; Waters Symmetry300™ C_{18} Opti-Pak™.
- Analytical column; 75 µm x 10 cm Waters Atlantis™ dC_{18}.
- A gradient was run from 0% to 40% B (B=95% ACN 0.1% formic acid) over 90 minutes.

Mass spectrometry and data acquisition

- Mass spectrometry was carried out using a Waters Micromass® Q-Tof Ultima API mass spectrometer, operating with nano ZSpray™ ESI.
- A NanoLockSpray™ source was employed.
- Data from the Q-Tof Ultima API was acquired using three different acquisition modes:

1. Data Directed Acquisition (DDA™); Automatic ion selection for MS/MS by charge state (2,3 or 4 positive charges) in order of intensity
2. Iterative DDA; Ions selected for MS/MS in previous experiments can be written into an exclude list to prevent them from being selected for MS/MS in later experiments
3. Real Time Database Searching (RTDS); Utilizing the database and search engine in ProteinLynx™ Global SERVER 2.0 (PLGS 2). Peptides are identified from analysis of their MS/MS spectrum and once identification of the protein is satisfactory, masses of all tryptic peptides from that protein are excluded from further selection for MS/MS.

For a flow chart of the working of RTDS, see Figure 1.

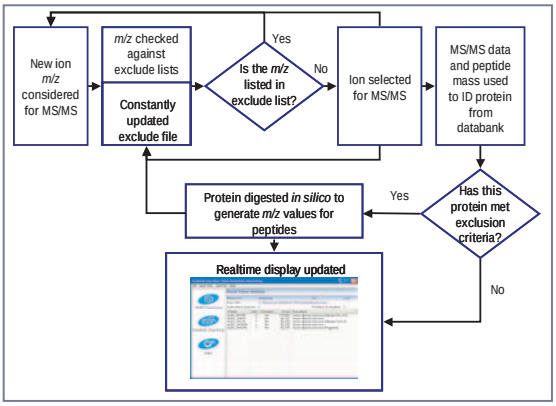

Figure 1

Off-line chromatography and automated nanoESI (NanoMate™)

Off-line chromatography

- Trapping column; Waters Symmetry300 C_{18} Opti-Pak™.
- Analytical column; 320 µm x 15 cm Waters Symmetry® C_{18}.
- Flow rate through both columns was 5 µL/min.
- Gradient from 2% to 40% B (B=95% ACN 0.1% formic acid) over 30 minutes.
- Waters 2757 Sample Manager.
- One-minute wide fractions collected.
- Fractions were evaporated and then reconstituted using 50/50 Methanol / water + 0.1% formic acid.

Mass spectrometry and data acquisition

- Mass spectrometry was carried out using a Waters Micromass Q-Tof micro™ Mass Spectrometer.

- The instrument was run in DDA mode.
- The NanoMate™ (Advion Biosciences, Ithaca, NY) was programmed to deliver 3 mL of each of the fractions for seven minutes.
- The system was operated at a spray voltage of 1.65kV and with a back pressure of 0.2 psi.

LC-MALDIprep fractionation and MALDI Q-Tof MS/MS

Off-line chromatography

- Flow rate 3 µL/min.
- Gradient from 0% to 40% B (B=95% ACN 0.1% formic acid) over 90 minutes.
- Post-column addition of alpha-cyano-4-hydroxycinnamic acid (1mg/mL)
- One-minute wide fractions were sprayed onto a standard 96-well target plate using nitrogen as a nebulizer. Fractions were collected from 11 minutes onwards.

Mass spectrometry and data acquisition

- Waters Micromass MALDI Q-Tof was used for the MALDI experiments.
- The MALDI Q-Tof data was lockmass corrected.
- MALDI MS data was acquired for each one-minute fraction and used to manually select ions for MS/MS.
- Manually selected ions were written into a sample list for automated MS/MS acquisition.

Data processing

14-Protein mix

DDA/RTDS

- Data from DDA and RTDS was processed by the use of PLGS 2
- Deisotoping and deconvolution was done using a Maximum Entropy (MaxEnt™) algorithm, and LockSpray™ correction automatically applied.
- Swiss-Prot™ (v.40) database was then searched using tight tolerances for both the parent and fragment ions (20 ppm). The hits returned were often under 10 ppm.

NanoMate

- The NanoMate experimental data was processed using MaxEnt lite algorithm.
- Due to no lockmass being applied to this data the search tolerance was widened to 100 ppm.

Q-Tof Ultima MALDI

- Each MS/MS spectrum was processed separately using MaxEnt3 and a summed data file was created, containing all the MS/MS spectra and was subsequently used to search the Swiss-Prot databank.
- A peptide tolerance of 20 ppm was applied to the database search.

Arabidopsis thaliana plasma membrane sample

- All data was searched against nrdb (NCBI) using an Arabidopsis thaliana species-specific filter.

Results

Analysis of the 14-Protein mixture

DDA, RTDS and the LC-MALDIprep MS/MS approach all identified peptides from 12 out of the 14 proteins (See Figures 2 and 3). In all experiments Catalase and Ribonuclease A remained elusive, perhaps due to poor ionization of their peptides or poor solubility/digestion of the proteins. Further DDA experiments utilizing exclude lists from previous experiments

(Iterative DDA) did not identify any new proteins although some redundant peptides from previously identified proteins were found, thus improving the coverage. The off-line HPLC in combination with NanoMate analysis only identified 10 proteins.

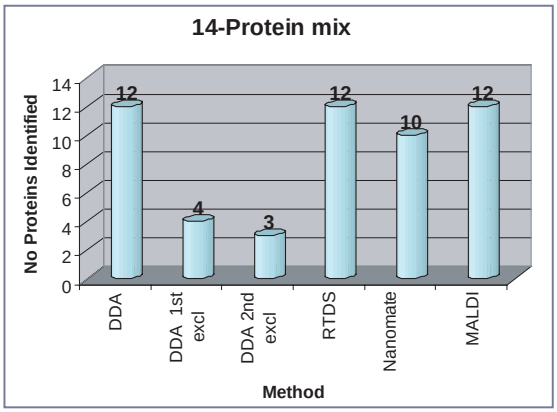

Figure 2: Bar chart showing the number of proteins identified from the 14-protein mix by each method.

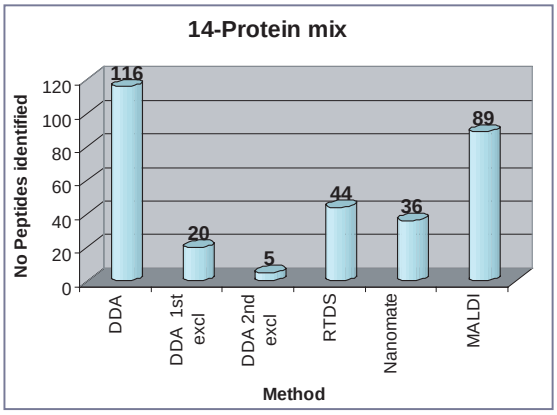

Figure 3: Bar chart showing number of peptides identified from the 14-protein mix by each method.

In the NanoMate experiment, evaporation of the samples was a limiting factor on the performance of the technique. This resulted in the HPLC gradient used for the peptide fractionation being shorter than desired, and therefore less individual discrete fractions were acquired. This resulted in fewer peptides being selected than were required to cover all the proteins present in the sample. In addition, the weaker ion intensities observed during MS/MS and lack of a lockmass may have hindered protein identification.

Although the DDA, RTDS and LC-MALDI MS/MS approaches identified the same number of proteins, it should be noted that the RTDS method did so from around one half to one third of the peptides of the other approaches.

Arabidopsis membrane sample

The plasma membrane fraction analyzed is known to be very complex, being rich in both membrane ATP`ases and protein kinases.The number of proteins identified by each analytical technique is shown in Figure 4.

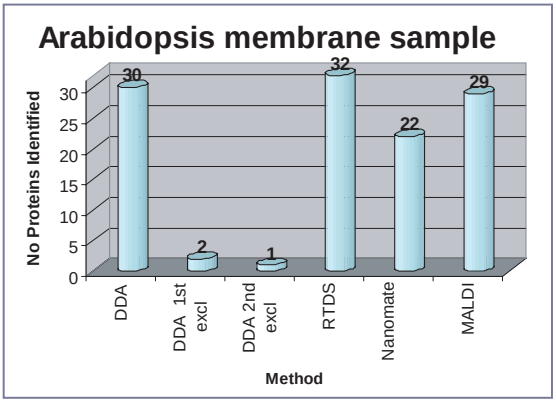

Figure 4: Bar chart showing the number of proteins identified from the Arabidopsis membrane sample by each method.

The performance of DDA, RTDS and the LC-MALDI MS/MS approaches appear to be comparable with respect to the number of protein identified in each case. The NanoMate was hindered by the limitations discussed in the previous section, hence identified fewer proteins. Closer analysis of the data provides some further discussion points.

Discussion

• RTDS compared to DDA

The fragment ion MS/MS spectra were individually inspected for validation purposes. It was noted that the RTDS data contained fewer ambiguous results, due to more intense fragment ion information, possibly due to better parent ion intensity. This may be the product of more conservative switching in the RTDS method compared to the standard DDA method.

MS/MS sequence data from up to four peptides is normally sufficient for identification of a protein. Any additional data could therefore be considered redundant for identification purposes. Figure 5 shows a plot of the number of peptides per protein identified as a percentage of all peptides identified from both DDA and RTDS.

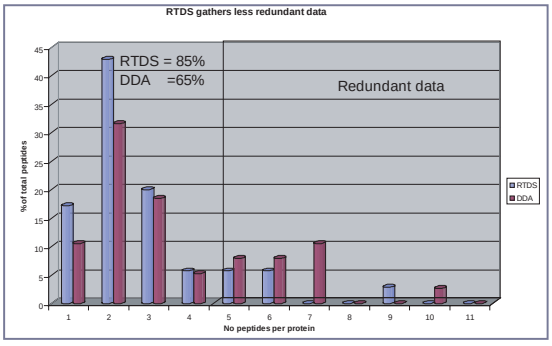

Figure 5: Comparison of the amount of redundant data acquired during CapLC® ESI MS/MS steered by DDA and RTDS.

If a cutoff is drawn on this plot at four peptides, to indicate a redundancy threshold, it can be seen that the RTDS approach shows a greater number of proteins below this line than DDA. In total, 85% of proteins identified by RTDS fall below this threshold where as only 65% identified by DDA do. Around 25% of peptides identified by DDA are from a single plasma membrane protein kinase, which was present in great abundance. RTDS identified this protein early in the experiment and steered the acquisition away from switching on any more of the theoretical tryptic peptides, allowing more acquisition time for peptides originating from other proteins. This provides a more even spread of data in the RTDS experiment.

Additional runs of DDA using exclude lists generated by the previous experiment (iterative DDA) identified only a few new proteins (two proteins in the next round and one following). The majority of peptides identified in these runs only served to increase the coverage of proteins identified by previous DDA experiments.

• ESI compared to MALDI

The ESI and MALDI strategies appear to work equally well in terms of the number of proteins identified, even though MALDI fragmented less peptides and had lower coverage. On inspection of this data however, it is apparent that the techniques do not identify or even select the same peptide related ions for MS/MS. Out of the total list of proteins identified by each method only 17 are common to both techniques. This may be in part due to the different ionization characteristics of the peptides present. An example of this is shown for a peptide derived from Ubiquitin, with a molecular weight of 1038.6. In the MALDI spectrum (Figure 6) the 1+ ion is shown clearly,

and was selected for MS/MS. In the equivalent ESI spectrum (Figure 7) the 2+ ion is very weak and the 1+ barely visible. Consequently this ion was not selected for MS/MS, due to other more intense species being present. This resulted in the protein not being identified in the ESI DDA experiment. Similar examples could be seen in the converse situation.

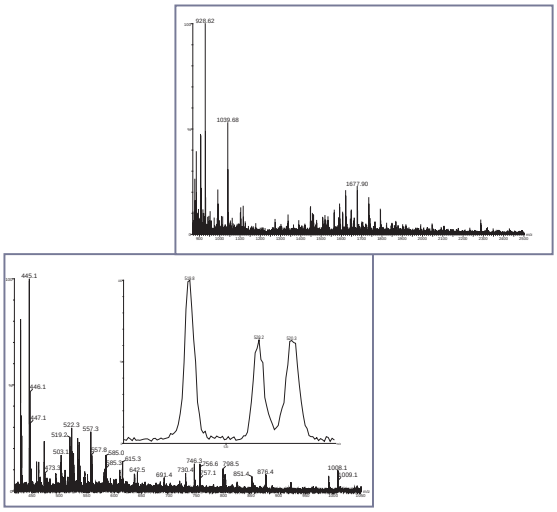

Figure 6 and Figure 7: ESI and MALDI ionize peptides with different efficiencies, resulting in identification of different proteins by each technique. In the top spectrum (Figure 6), acquired by MALDI, the ion of m/z 1039 is clearly visible. MS/MS of this resulted in the identification of Ubiquitin. The equivalent spectrum by electrospray (Figure 7) shows this ion to be considerably lower in abundance, hence was not selected for MS/MS and the protein was not identified by this technique.

• LC-MALDIprep compared to NanoMate

The NanoMate system was hampered by sample evaporation, as described previously. Fractions redissolved in the microtitre plate wells required analysis within a short time, hence the number of fractions that could be automatically analyzed was restricted. This resulted in a faster HPLC gradient being run, resulting in less chromatographic separation and a greater number of peptides per fraction. As a result there were peptides that were detected in the MS survey that were not selected for MS/MS because of experimental time constraints. This resulted in fewer proteins being identified by this method.

Solubilizing the sample immediately prior to analysis, in the case of the NanoMate, would allow longer fractionation gradients to be run and acquisition time per sample to be increased from 7 mins to 30 mins. Both would greatly increase the extent of characterization of a complex sample.

LC-MALDIprep is unaffected by evaporation issues as it sprays the column eluent directly onto a MALDI target plate. This allowed a longer, shallower gradient to be run during the HPLC separation and therefore the number of peptides per spot was reduced. This method was able to identify as many proteins as the DDA/RTDS experiments.

PosterREPRINT

Waters

Conclusion

- All methods performed well in the two samples studied.
- In a sample of medium complexity (i.e. the 14-protein mix) with very little variety in dynamic range, DDA with a suitably long gradient is sufficient for identification of tryptic peptides representative of the proteins in the sample. This also applies to the LC-MALDI MS/MS approach.
- In a more complex sample (30+ proteins) with a large inherent dynamic range the quality of data can be improved by the use of Real Time Database Searching (RTDS). This is an attractive addition to DDA in that it requires no additional hardware or experiment time, and it reports the databank search results in real time during the experiment.
- The NanoMate system has the potential to be improved if the evaporation problems can be addressed.
- LC-MALDIprep fractionation followed by MALDI Q-Tof analysis performs almost as well as the ESI LC/MS/MS methods in terms of the number of proteins identified. The different ionization technique provides complementary information, which has led to the identification of a greater number of proteins.

Acknowledgements

Thanks to Mark Allen, Alistair Sterling and Mark Baumert (Advion Biosciences, Norwich UK) for the use of a NanoMate System.

To whom all correspondence should be addressed:
e-mail: Jim_Langridge@waters.com

WATERS CORPORATION
34 Maple St.
Milford, MA 01757 U.S.A.
T: 508 478 2000
F: 508 872 1990
www.waters.com

Waters

MICROMASS MS TECHNOLOGIES

Opt-In! My Profile www.waters.com

LLOYD'S REGISTER QUALITY ASSURANCE · ISO9001 · UKAS QUALITY MANAGEMENT 001