Assessing the Complementarities of MALDI and ESI for Protein Identification in Complex Mixtures

Comparing the AB SCIEX TOF/TOF™ 5800 System with ESI Orbitrap using ProteinPilot™ Software

Sean Seymour¹, Aaron Booy², Rebekah Gundry³, Jennifer Van Eyk³, Christie Hunter¹
¹AB SCIEX, USA, ²AB SCIEX, Canada, ³The Johns Hopkins University, USA

There are two dominant ‘soft’ ionization strategies in proteomics today for biomarker discovery and protein identification in complex mixtures. On-line LCMS using electrospray ionization (ESI) is commonly used with mass spectrometers such as ion traps, quadrupole time of flight, and Orbitrap instruments. Matrix Assisted Laser Desorption Ionization (MALDI) is the other main soft ionization technique and is normally coupled with Time of Flight (TOF) detection systems. With the advent of LC MALDI¹ it has also become widely used in off-line LCMS strategies. While LC MALDI has much more flexibility in acquisition strategies due to the separation of LC from MS analysis, it has traditionally been speed limited relative to the ESI approach.

With the faster laser repetition rate possible on the TOF/TOF™ 5800 System and the intelligent DynamicExit™ Algorithm for efficient acquisition, the speed of LC MALDI is now 10-15X faster² than previous instruments. As LC MALDI begins to approach the speed of on-line ESI strategies, it is worthwhile investigating the complementarities of the techniques to understand how to best leverage the strengths of each and get the most information from complex proteomics samples.

To do the study properly, it was important to assess both the differences between instruments and the differences within multiple acquisitions on each instrument. The strategy used was to measure three replicates of a complex biological sample (E.coli cell lysate) at three different sample loadings on column and assess the differences. Reverse database searching with ProteinPilot™ Software 3.0 was used to measure the numbers of proteins and peptides detected at a fixed false discovery rate (FDR) from each instrument.

![Figure 1. Summary of ESI and MALDI Acquisitions – Protein Level. Proteins identified at 1% global FDR for each of the three replicate injections at each sample loading is shown. Replicate acquisitions were performed on both the ESI-Orbitrap platform (red) and the TOF/TOF™ 5800 System (blue) to enable proper comparison of these orthogonal acquisitions strategies.](image-url)
Materials and Methods

**Sample Preparation:** E. coli lysate was re-suspended in 50% trifluoroethanol/50 mM ammonium bicarbonate buffer at 1.35 mg/mL (by weight). After heat denaturation, reduction (dithiothreitol) and alkylation (iodoacetamide) were performed, the sample was digested with trypsin overnight at 37 °C. Sample was dried using vacuum centrifugation, then re-suspended at 1 mg/mL in 2% acetonitrile / 0.1% formic acid. Sample loadings of 20, 100 and 500 ngs were run in three replicates for each loading.

**LC MALDI Acquisition:** Peptides were separated on a Monolithic CapRod column (100 µm ID x 15cm, Merck, Germany) with a 90 minute linear gradient of 5-35% acetonitrile at 2 µL/min (in 0.1% TFA) using the Tempo™ LC MALDI Spotting System. 720 spots were collected per LC run with online mixing of α-cyano-4-hydroxycinnamic acid matrix (Agilent). MS/MS spectra were acquired on the 25 most intense precursors per spot using a laser frequency of 1000 Hz and DynamicExit™ Algorithm on the TOF/TOF™ 5800 System (for more acquisition details).

**LC ESI Acquisition:** Peptides were separated on a self-packed column (75 µm x 10cm, Magic C18 AQ (Michrom) in a PicoTip emitter (New Objective)) using a 90 min linear gradient of 5-35% acetonitrile at 300 nL/min (Agilent 1200 LC system). The eluent was directed into an LTQ Orbitrap (ThermoFisherScientific). MS scans were acquired in the Orbitrap and the top 7 precursor ions above a threshold of 8000 counts were selected for MS/MS in the LTQ. Data acquisition parameters were optimized according to the vendor’s recommendations, and included monoisotopic precursor selection, 2+, 3+ charge state selection, and a dynamic exclusion repeat count of 1.

**Database searching:** Each dataset was searched with ProteinPilot™ Software 3.0 against a database of all E.coli proteins in UniProt as of Jan. 27, 2009 (4379) plus 77 common contaminant proteins using the Paragon™ Algorithm in thorough mode with biological modifications and substitutions. False discovery rate (FDR) analysis was done by on-the-fly analysis of the reversed sequences using the embedded PSPEP tool. For processing of the Orbitrap data, an mgf file was generated using ReAdw4Mascot2.exe, a NIST derivative of the ISB tool, ReAdW.

**Protein and Peptide ID Results**

Three different sample loadings of E.coli were analyzed by ESI Orbitrap and LC MALDI 5800 system and compared using the protein and peptide identifications at the 1% global FDR. The processing of the three data acquisition sets is summarized in Figure 1 and 2. At the protein level (Figure 1), the results at the 20 ng level were quite comparable. At the higher sample loading of 100 and 500 ngs, a higher number of proteins were routinely seen using the LC MALDI strategy. The same observation was seen at the peptide level (Figure 2), with the LC MALDI approach finding ~2x more peptides at the highest loading amounts.
Using the 500 ng sample loading level, the intersection between the best ESI-Orbitrap acquisition was compared to the best LC MALDI 5800 acquisition at the protein and peptide level (Figure 3, left), using the 1% global FDR values. Out of a total of 778 proteins identified, 466 proteins (60%) were shared with the 5800 system finding about 30% more proteins. Even more significant differences were observed when comparing the peptide level intersection, where only 25% of peptides were shared between acquisitions. Here, LC MALDI on the 5800 system found about 108% more peptides than the ESI-Orbitrap system, meaning more proteins and more sequence coverage for the detected proteins.

Protein and Peptide Level Intersection

The value of higher numbers of peptides for the purpose of protein identification can be determined by aligning the protein level identifications and comparing the Unused Prot Score (UPS - a measure of protein confidence in ProteinPilot™ Software) from each acquisition strategy. As a rough approximation, each 2 units of UPS correspond to one high confidence peptide sequence segment (excluding multiple modification states). The additional peptide coverage by LC MALDI clearly translates into more confident protein identifications (Figure 3, right, log scaled) as seen by the higher proportion of data above the diagonal.

Assessing the Peptide Level Intersection

A comparison using only one replicate is insufficient and provides only limited insight. It is important to assess both intra- and inter-instrument reproducibility. The degree of similarity between each pair of replicates – both within and across instruments – was calculated as the ratio of the intersection to the union of identifications for the replicate pair (Figure 4). The average intersection at the peptide level between replicates on the TOF/TOF™ 5800 system at the 500 ng loading was high at 64% (Figure 4, Samples 1-3). The average intersection between the Orbitrap replicates was also high at 61% (Samples 4-6). However, if the peptide comparison is done between platforms, the average intersection decreases to only 25% (Figure 4, red boxes). Comparison of replicates within each instrument clearly show more similarity than comparisons of replicates from different instruments. The source of this difference is an interesting investigation.
Previous work has suggested that there is a slight bias towards specific types of peptides detected between ESI and MALDI experiments. However, without proper assessment of the random nature of acquisition, no real trends can be assessed. Here, using three replicate acquisitions as the basis for our analysis, we can assess real acquisition based differences. As shown in Figure 5, there are higher proportions of Arginine (Arg)-terminating peptides detected in LC MALDI and this is consistent across replicates. In the ESI acquisition, the distribution of peptides between Lysine (Lys) and Arg is roughly equal. Therefore, when the unique peptides found by only LC MALDI were analyzed, there was a very high proportion of Arg peptides found (~67%) vs only ~33% by ESI (Figure 5, right).

The mass of the peptides found were also analyzed, to understand if there were any real differences in peptide size. Figure 6 shows the distribution of precursors found across all replicates. No strong differences are seen between acquisition strategies, although there are small differences that appear consistent across replicates.
Conclusions

By analyzing both the intra- and inter-instrument acquisition replicates, a much clearer picture into the complementary nature of ESI and MALDI can be ascertained. The LC MALDI using the AB SCIEX TOF/TOF™ 5800 system reproducibly obtains more proteins and peptides per acquisition over the ESI-Orbitrap acquisition at the higher samples loadings. At lower amounts on column, the acquisition results are comparable.

The analysis of the peptide intersection demonstrates that distinct peptide populations are detected when using LC MALDI vs. ESI acquisition (Figure 3). This suggests that a complementary strategy combining both ESI and MALDI analysis could provide deeper proteome coverage. Consideration of combinations of replicates (Figure 7, 500ng) shows that multiple replicate acquisitions on an ESI-Orbitrap does not equal the peptide or protein yield from a single LC MALDI run. This comparison is of interest because of the acquisition time consideration, 3 LCMS runs takes ~7.5 hours of instrument time and 1 LC MALDI run takes ~7 hours2.

Another interesting comparison is that the combination of two 5800 system runs is nearly equivalent to a combination of 1 ESI-Orbitrap run and 1 LC MALDI run. Since one ESI and one MALDI run can be done in less time and the differences in the peptide populations detected, the combined use of these two platforms might be a valuable strategy for deep proteome coverage.

References

2. Effects of Intelligent Data Acquisition and Fast Laser Speed on Analysis of Complex Protein Digests. AB SCIEX Technical Note 0921810-01.