Fingerprinting Bacterial Strains Utilizing LC/MS-based Profiling of Intact Proteins

Ken Imatani, Carsten Carstens, Bernd Buehler, Dave Horn, Norton Kitagawa and Christine Miller

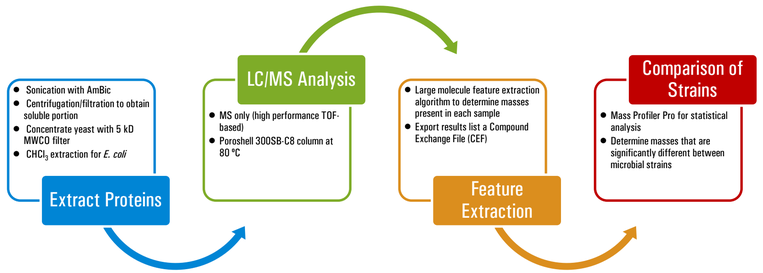
IMSC 2009

Agilent Technologies

Introduction

Combining retention time with accurate mass from LC/MS data is a powerful approach for profiling compounds present in complex biological matrices. This untargeted workflow has been applied to both metabolomic and proteomic samples, but proteomics has been limited to peptide digests as the feature extraction algorithm relied on resolved isotopes. A new algorithm for extracting protein features now allows this profiling approach to be applied to intact proteins. In this study, we explore fingerprinting of bacterial strains by differential analysis of the protein features in cell lysates.

Protein Profiling Workflow



Experimental

Sample preparation:

E. coli cultures were grown for 3.5 – 5 hours to an optical density of 0.75, then washed 2 times with 100 mM ammonium bicarbonate, pH 8.4, containing 5 mM EDTA. The samples were sonicated 2 times with 12 bursts at 50% power and 60% duty cycle, then centrifuged at 13,000 rpm for 10 minutes. The supernatant were stored frozen until use. Unlike yeast lysates which were previously analyzed, the E. coli lysates had a significant amount of lypophilic material which caused blocking of the columns. A simple protocol was developed to remove the lipids using chloroform extraction (50:50 v:v) with subsequent analysis of the aqueous phase.

LC/MS acquisition:

All data was acquired on an Agilent 1200 capillary HPLC system coupled to an Agilent Q-TOF mass spectrometer using a dual ESI interface with a capillary nebulizer. The *E. coli* intact protein separation was done on a 10 x 0.5 cm mRP column using a gradient of acetonitrile in water + 0.1% formic acid. The column temperature was held at 70 degrees C.

Data analysis:

All data was analyzed using Agilent MassHunter BioConfirm software. For molecular feature extraction of the proteins present in the samples, the large molecule feature extraction (LMFE) algorithm was used. This algorithm is described in the next section (above right). Compounds (proteins) with a total abundance of less than 25000 counts were filtered away from the initial set. Extracted features for each LC/MS data file were written out to XML-based Compound Exchange Format (.CEF) files and imported into MassHunter Mass Profiler Professional (MPP) for chemometric analysis. Compound features were binned and aligned across 15 sample files using a nearest neighbor search algorithm based on neutral mass and retention time values using specified tolerances.

Overview of Large Molecule Feature Extraction

LMFE uses a similar methodology to the previously-developed Molecular Feature Extractor (MFE) algorithm:

- Find all peaks in an LC/MS run and creates three dimensional peaks for each species.
- Group the peaks with the same retention time and elution profile into "coelution groups" (Figure 1).
- Background compounds that do not show a true LC elution profile are removed from consideration.
- The peaks within a given coelution group will contain the different charge states of the same protein, which are subsequently grouped together by algebraic charge state deconvolution.
- While algebraic deconvolution can be challenging for very complex spectra, the charge states for a given protein will generally elute at a slightly different time than other eluting proteins even in highly complex mixtures.

LMFE produces a list of compounds within the Mass Hunter Qualitative Analysis software, with links to a compound spectrum containing the different charge states found for a given protein and the extracted compound chromatograms for each compound (Figure 2).

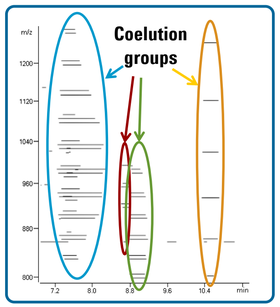
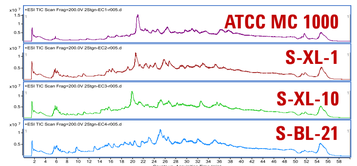


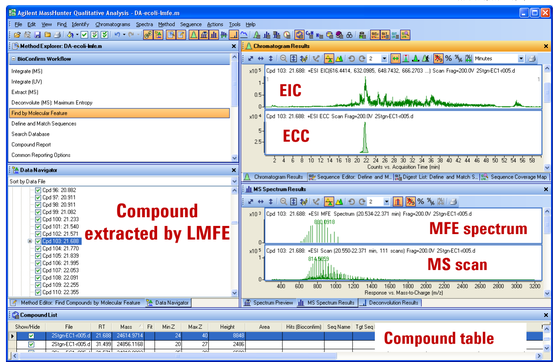
Figure 1. The peaks corresponding to different charge states of the same protein elute at the same time in the LC/MS run. Each oval corresponds to a unique coelution groups, each of which produces one or more protein compounds using algebraic deconvolution.

Results and Discussion

Figure 2 (right). shows typical total ion chromatograms for each species. Five replicate injections were done for each extracted lysate.

Figure 3 (below) shows a a typical result from LMFE in MassHunter Qual.





Results and Discussion

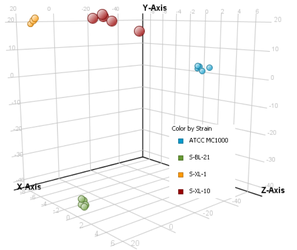


Figure 2. PCA analysis of the different strains was done following a one-way ANOVA on strain (Benjamani-Hochberg multiple testing correction applied). The results demonstrate correct grouping of the technical replicates and clear separation of the strains.

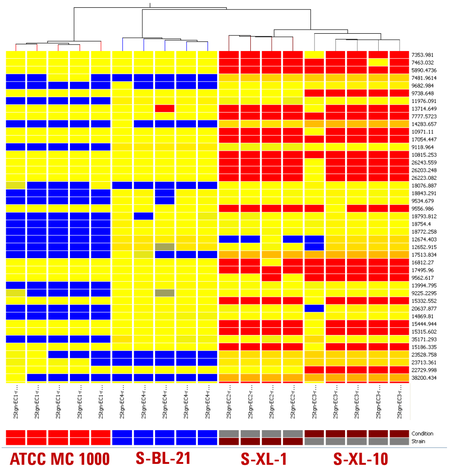


Figure 3. Hierarchical clustering successfully clustered samples From the different strains. The heat map is colored from blue to red, where blue is low abundance and red is high abundance. The first cluster on the left is for the control strain, ATCC MC1000. The 2 strains clustered on the right, XL-1 and XL-10 appear more similar as expected.

Table 1. The top 10 masses, both negative and positive, contributing most to Component 1 from the PCA analysis. Identifying proteins from mass alone is problematic due to post-translational modifications, clipping of initiator Met etc. Future work will involve identification of these proteins by LC fractionation, then tryptic digestion of the fractions, followed by LC/MS /MS analysis and protein database search.

Mass	Retention Time
36699.42	27.329203
24614.97	21.873999
12235.64	35.538998
58407.72	23.999401
18119.2	26.1824
22283.23	36.74207
34357.14	26.593927
17701.36	29.260284
23712.3	33.615215
19727.12	30.241638

Conclusions

- LMFE is a powerful algorithm for extracting features in intact protein data
- A chemometric software package allows importing these results for statistical comparison of proteins extracted from bacterial samples
- Preliminary results show excellent results for fingerprinting 4 different *E. coli* strains.