Intact protein profiling of pathogens using LCMS, protein feature extraction software and rigorous differential analysis

ASMS 2010

Denis Andrzejewski[1]; Melinda A. McFarland [1]; Rebecca Bell[1]; John H. Callahan[1]; Steve Musser[1]; Norton Kitagawa[2]; Jose Meza [2]; Greg Kilby[2]; Vadiraja Bhat [2]

[1]FDA-CFSAN, College Park, MD; [2]Agilent Technologies, Santa Clara, CA

Intact protein profiling of pathogens using LCMS, protein feature extraction software and rigorous differential analysis

Introduction

The development of methods for rapid identification and differentiation of bacteria is an important goal in food safety research. Mass spectrometry methods are complementary to DNA based approaches. Such methods include MALDI of intact bacteria, bottom-up and top-down analysis of proteins from bacterial extracts, as well as intact protein liquid chromatography/mass spectrometry. The latter method affords a simple, direct experimental approach in which intact protein mass spectral data from chromatograms are deconvoluted to yield intact mass, retention time maps for proteins extracted from bacteria. However, due to the complexity of the LC/MS data, analysis is difficult and time consuming. Currently, the FDA uses custom software to perform sequential deconvolution of mass spectra in chromatographic time segments, followed by the construction of protein profiles (mass, intensity, retention time) and multivariate stastistical analysis. We present here another approach using Agilent's large molecule feature extraction software to detect and deconvolute the intact proteins followed by multivariate statistical analysis software to provide clustering and PCA analysis.

Experimental

Sample Preparation

Four strains of Salmonella enterica were analyzed - A1, A2 closely related strains from serovar typhimurium, and A39, A40 closely related strains from serovar Heidelburg. A2 is the gene sequenced strain LT2. Bacteria were grown to stationary phase on LB plates. The harvested cell pellets were lysed in a mixture of 50:45:5 acetonitrile:water:formic acid using a PCT Barocycler (Pressure BioSciences, Inc.) at 40 C using repetitive cycling between 35,000 psi and atmosphere for 10min. The extracts were centrifuged and a portion of the clear protein extract analyzed by LC/MS.

LC and MS conditions

Flow – 200uL/min
Stop time – 115min
Post time – 10min
Initial Solvent – 5% B
A: Water with 0.5% acetic
B: ACN with 0.5% acetic
2 Grace Prosphere P-HR 4mm 2.1x150mm columns in series

Time	%B
10	5
16	20
75	50
85	90
97	90
99	5

Table1: Chromatography conditions for the Agilent 1200 HPLC system

Parameter	Setting
Gas temp	325°C
Drying Gas	10 L/min
Nebulizer	45psi
Vcap	4000V
Fragmentor	225V
MS	0.86 scans/sec, 300-3200m/z
Ref correction	1 point @ 922.009798
Acq. mode	Extended Dynamic Mode (2GHz)

Table 2: Instrument settings for the Agilent 6520 QTOF

Large Molecule Feature Extractor (LMFE)

- *This approach first finds all peaks in an LC/MS run, creates three dimensional peaks for each species, and groups the peaks with the same retention time and elution profile into "coelution groups". 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.*

Mass Profiler Professional (MPP)

- *MPP is a comprehensive suite of statistical tools for mass spectrometry-based chemometric data analysis.*
- *The software package allows importing "Large Molecule Feature Extractor" results for rigorous statistical comparison of compounds, which are defined as intact deconvoluted protein masses, summed abundances of all the charge states reflecting those intact masses, and corresponding retention times.*

Comparison with maximum entropy deconvolution data analysis

- *Results were compared with data analysis using ProTrawler (Bioanalyte, Inc), which sums spectra over a designated time interval, removes noise, followed by maximum entropy deconvolution and reconstruction of the mass, intensity, retention time profile for intact proteins.*

Intact protein profiling of pathogens using LCMS, protein feature extraction software and rigorous differential analysis

Results and Discussion

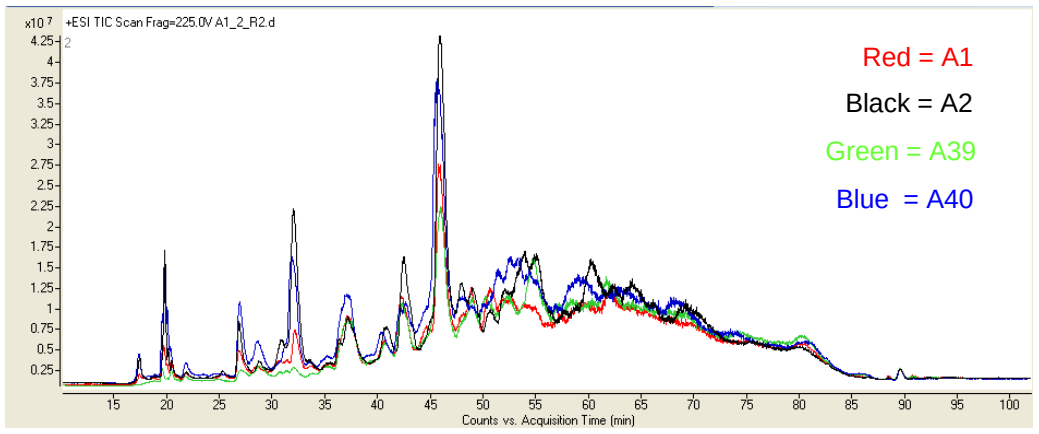

Figure 1: Overlay of representative total ion chromatograms (TIC) of biological replicates of strains A1, A2, A39, and A40

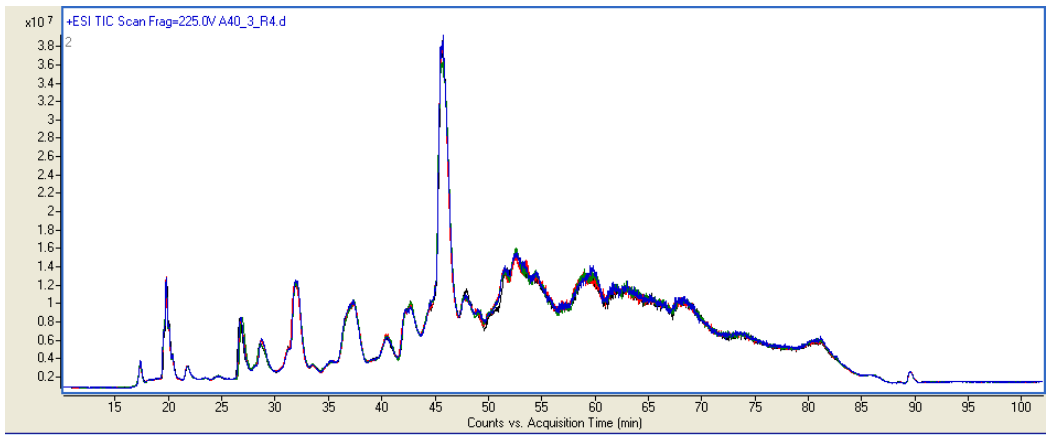

Figure 2: TIC Overlays of 4 technical replicates of strain A40

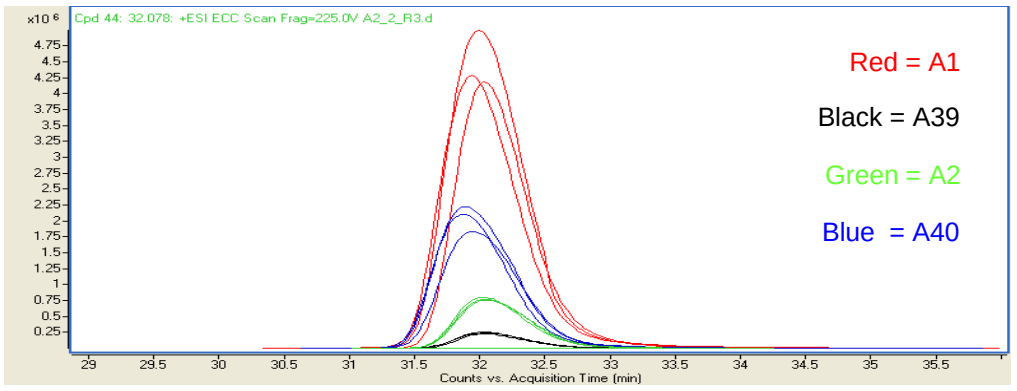

Figure 3: Overlay of extracted compound chromatograms (ECC) for 3 biological replicates of each strain for compound (protein mass) 18,552 at R_T 32.1 min. This particular protein has been identified by top-down proteomics methods as osmY, a periplasmic protein. This protein had the highest statistical discrimination between all 4 strains (refer to Table 3 for statistics)

Strain	Abundance Avg	Std. Dev.	%CV	Mass Avg	Std. Dev.	Std. Dev. (PPM)	RT Avg	Std. Dev.	Std. Dev. (Secs)
A1	185918920	17307388	9.31	18552.85	0.0241	1.30	32.04	0.095	5.73
A2	31015638	1337616	4.31	18552.83	0.0106	0.57	32.12	0.036	2.19
A39	9140272	700361	7.66	18552.83	0.0191	1.03	32.10	0.045	2.72
A40	87962664	7393685	8.41	18552.84	0.0127	0.69	31.97	0.055	3.32
Combined	**78509373**	**6684762**	**7.42**	**18552.84**	**0.0167**	**0.90**	**32.06**	**0.058**	**3.49**

Table 3: Statistics by individual strain (n=12) and all data combined (n=48) for compound 18,552 (osmY) at R_T 32.1 min, including precision of abundance, mass measurement, and retention time

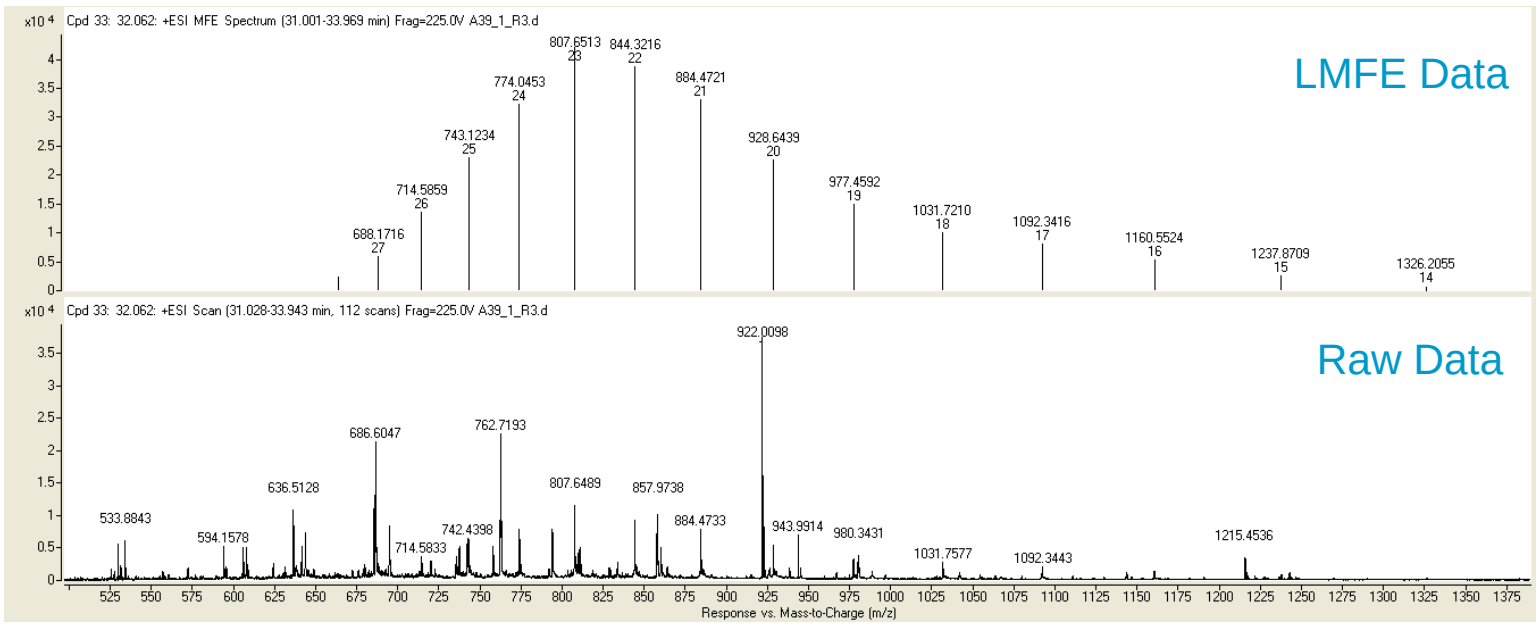

Figure 4: Extracted mass spectra for compound 18,552 (osmY) at R_T 32.1 min. The bottom panel is the raw unprocessed spectrum over the chromatographic peak, the top panel is the extracted LMFE spectrum for that discrete compound, which is one of several proteins present at this retention time.

Agilent Technologies

Intact protein profiling of pathogens using LCMS, protein feature extraction software and rigorous differential analysis

Results and Discussion

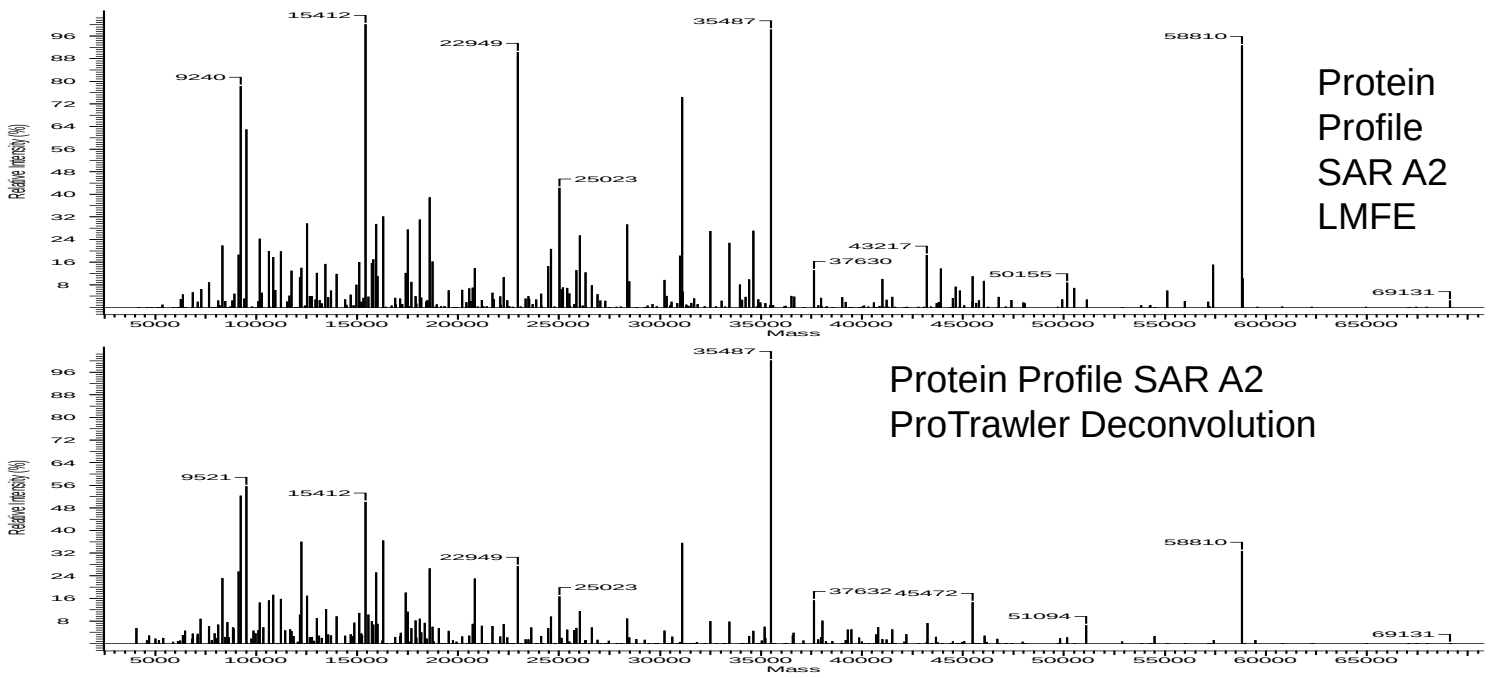

Figure 5. Comparison of protein profile for strain A2 (LT2) for LMFE (top) and ProTrawler deconvolution (bottom) for top 300 most abundant proteins.

There is overall agreement between the profiles obtained via the two data analysis methods, as shown in Figure 5. Of the top 100 most abundant masses, 85 are found using both methods, 8 are only found using LMFE, and 7 are only found using maximum entropy methods. Agreement between the two data analysis methods decreases with decreasing protein abundance.

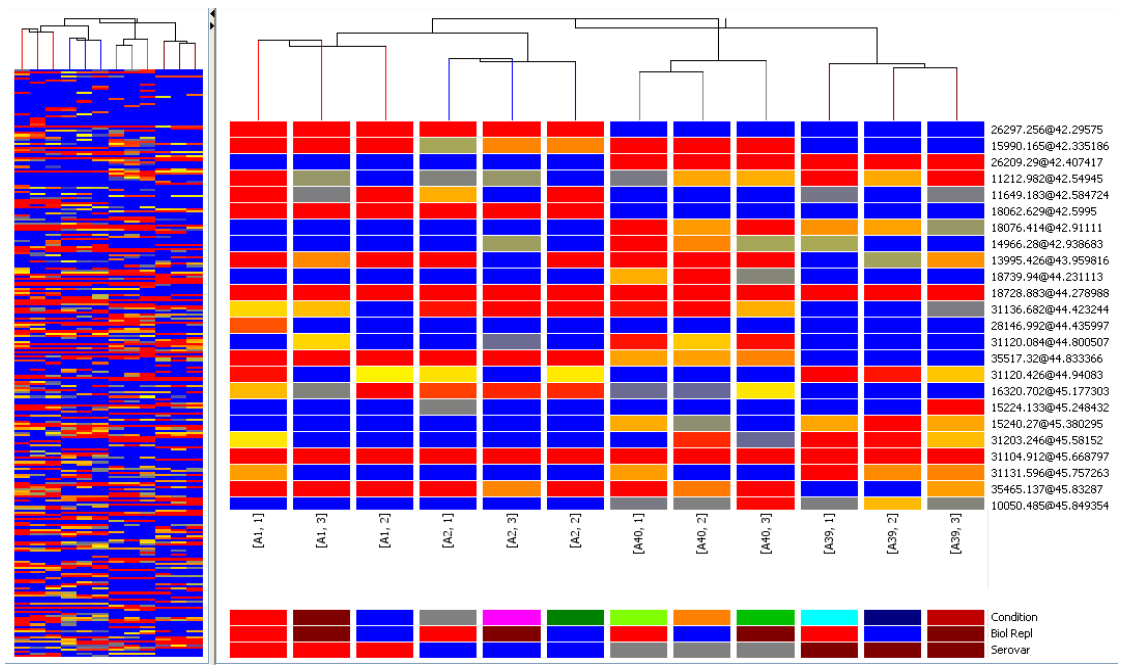

Figure 6. Hierarchical clustering analysis of the 3 biological replicates of the 4 strains. Blue represents lower abundance compounds shading to red representing higher abundance compounds. As expected, strains A1 and A2 cluster closer together as do A39 and A40.

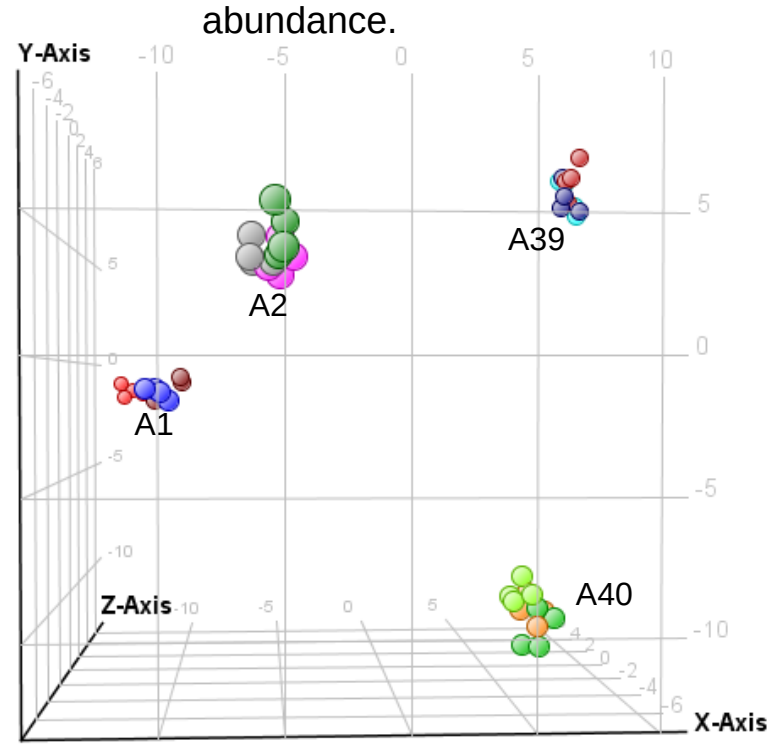

Figure 7. PCA analysis of the different serovars was done following a one-way ANOVA on serovars (Benjamani-Hochberg multiple testing correction applied). The results demonstrate grouping of the 4 technical replicates and 3 biological replicates of each individual strains and clear discrimination of the

Conclusions

- *The methodology presented in this poster allows for the successful profiling of intact proteins from closely related Salmonella strains and produces results similar to approaches previously employed for generating bacterial protein profiles.*
- *The combination of: 1) a simple extraction method with no clean-up; 2) a highly reproducible chromatographic method giving consistent retention times and abundance; 3) excellent mass accuracy, retention time and abundance reproducibility, (table 3); and 4) intelligent LMFE algorithm and profiling software (MPP) allows for:*
  - *A) the differentiation of 4 strains of Salmonella enterica (figure 7)*
  - *B) the correct clustering of the closely genetically related strains (figure 6), namely strain A1 clusters closer to strain A2 (both are from serovar typhimurium) and strain A39 clusters closer to strain A40 (both are from serovar Heidelburg)*