ASMS 2005

DIFFERENTIATING PATHOGENIC AND NON-PATHOGENIC STRAINS OF *ESCHERICHIA COLI* USING INTACT PROTEIN LC/MS AND AN AUTOMATED DATA PROCESSING WORKFLOW

Scott J. Berger[1], Tracie L. Williams[2], Ignatius J. Kass[3], Steven M. Musser[2]

[1] Life Sciences R&D, Waters Corporation, Milford, MA [2] U.S. Food and Drug Administration, College Park, MD [3] Waters Corporation, Beverly, MA

OVERVIEW

- The differentiation of microbial strains can be useful for determination of the source and route of infection of a pathogenic bacterium.
- In this poster, we have shown a workflow whereby cellular extracts of bacterial isolates can be analyzed by LC/MS, and the resulting intact protein LC/MS data processed via a novel automated software tool called Auto Maximum Entropy (AutoME).
- Overall reproducibility of the methodolology, from sample culture to data interpretation, is demonstrated using five isolates of a non-pathogenic E. coli strain.
- Differential protein expression analysis from a non-pathogenic strain (K-12), a common pathogenic strain (O157:H7), and a novel pathogenic strain (O157:H–) is presented.

METHODS

Sample Preparation: E. coli strains were grown overnight on tryticase soy agar plates, and harvested as a slurry in 70% ethanol. Cell pellets were extracted by sonication in 50:45:5 water:acetonitrile:HOAc, then cleared by centrifugation.

Sample Analysis: Diluted E. coli extracts were analyzed by LC/ESI-TOF-MS (Waters Q-TOF II with the quad operating in a non-selective pass-thru mode) using a linear gradient of acetonitrile (0.25 ml/min, 10 - 50% acetonitrile) in a background of 5% acetic acid over 50 min. LC/MS data were acquired (ESI+, 600-1800 m/z) using a 0.5 Hz scan cycle using MassLynx 4.0 software.

Automated Maximum Entropy (AutoME) Data Analysis: The AutoME program automates the processing of protein LC/MS data sets (*.raw files) into flat text files containing retention time, deconvoluted mass, intensity files. The following data processing steps were applied to the LC/MS data sets.

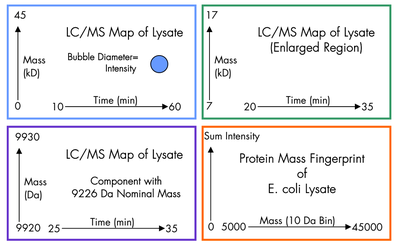
1. Run segmentation: The LC/MS data set was segmented into 6 scan (12 sec) segments
2. Data selection: The entire mass range (300-1800 m/z) was analyzed, and all run segments were selected for processing (No pre-analysis thresholds chosen)
3. MS spectral processing: Baseline subtraction was selected using a15th order polynomial and 1% of signal below baseline.
4. MaxEnt1 Deconvolution: The output mass range was 3000-50000 at 1 Da resolution. Spectra were processed with a 0.7 Da PWHH Gaussian damage model for up to 16 iterations.
5. Lockmass correction: Not selected
6. Data reduction: Deconvoluted spectra were centroided, spectral harmonic artifacts were removed, and the top 40 peaks above a signal threshold of 40 were retained per segment.
7. Data was output as a RT, Mass, Intensity text file (tab delimited), and as a Summary (whole run) file of Mass bin (10 Da increments from 3000-45000 vs. Summed bin intensity).

Data Presentation: LC/MS Maps were created using the output from AutoME and the Bubble Plot feature of Microsoft Excel 2003 where x-axis= Retention Time, y-axis=Deconvoluted Mass, Bubble diameter= Intensity. Protein Mass Fingerprints were created from the Bin file output of AutoME and the scatter-plot feature of Microsoft Excel 2003.

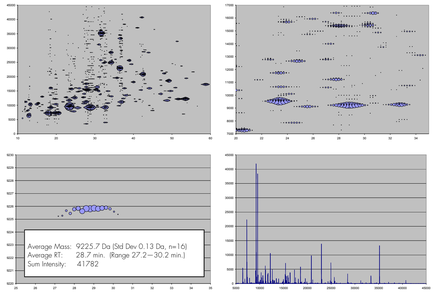
REPRODUCIBILITY OF FIVE E. coli STRAIN K-12 ISOLATES

The resulting data sets represent the combined effects of 1) Biological variability between isolates, 2) Variability in sample preparation and loading, 3) LC/MS analytical variation

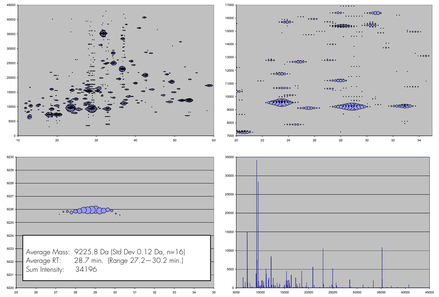
FIGURE KEY



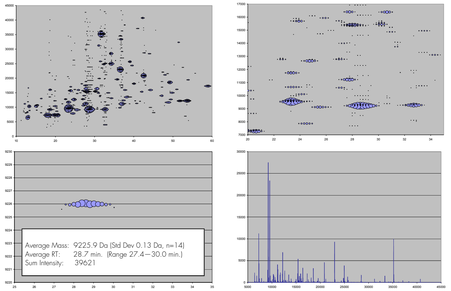
Strain K-12 Isolate #1



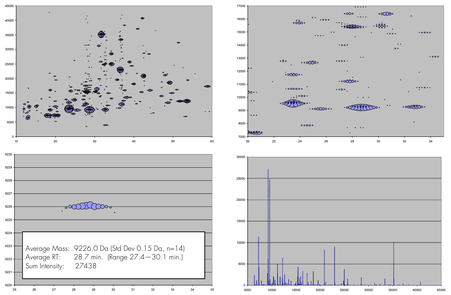
Strain K-12 Isolate #2



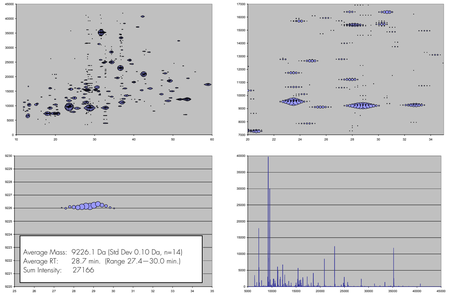
Strain K-12 Isolate #3



Strain K-12 Isolate #4



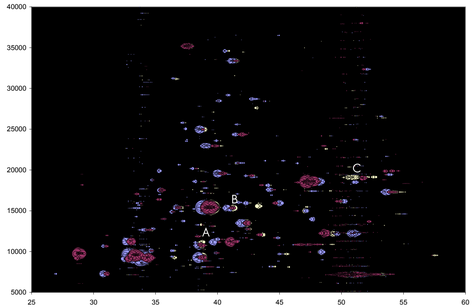
Strain K-12 Isolate #5

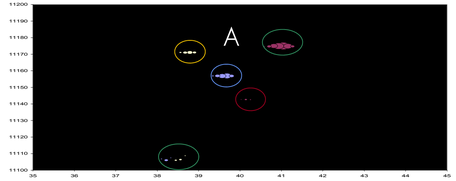


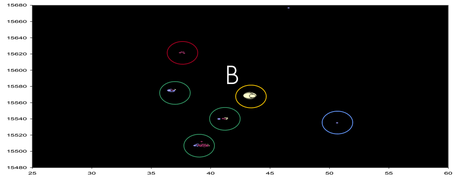
IDENTIFICATION OF DIFFERENCES UNIQUE TO A NOVEL PATHOGENIC STRAIN

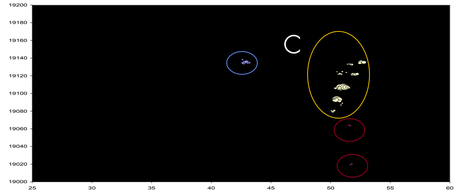
Identification of strain differences using overlays of LC/MS protein maps

Key: K-12 (Non-Pathogenic) O157:H7 (Common Pathogen) O157:H- (Novel Pathogen)
Zoom Plots: Unique components in colored circles, Common components = Green circles









Identification of strain differences using Protein Mass Fingerprint differences

ID Mass Bins where MS signal s greater than 10X above background signal (20), and the other two strains are less than 2X above background.

Unique to K-12 E. *coli*

10 Da Bin	K-12	O157:H7	O157:H-
11160	2329	0	0
13970	921	0	0
28700	893	0	0
18500	550	0	0
15100	503	25	0
16020	471	0	0
12780	439	0	0
29180	378	0	0
22990	355	25	0
18490	347	26	0
13740	336	0	0
25580	310	0	0
28480	296	0	0
28470	293	0	0
19290	272	0	0
15930	270	0	0
17790	264	0	0
25040	231	0	0
31230	224	0	0
13270	217	24	22
40710	202	0	0

Unique to O157:H7 E. *coli*

10 Da BIN	K-12	O157:H7	O157:H-
9740	0	10003	0
12250	0	4788	26
12240	0	1686	35
19000	0	1304	0
18600	29	725	0
9190	0	665	0
19870	24	531	0
16380	0	459	0
11850	0	381	0
10920	24	366	0
9840	0	366	0
13680	0	348	0
18470	0	334	0
37990	0	327	0
9780	0	308	0
26350	0	308	0
9760	0	304	0
18140	0	259	0
22960	0	258	0
18630	29	246	0
13180	0	240	0
10880	0	223	26

Unique to O157:H– E. *coli*

10 Da BIN	K-12	O157:H7	O157:H-
19110	0	0	2595
24970	0	0	1411
19090	0	0	1180
19120	0	0	689
27610	20	0	411
31160	0	0	308
9530	0	31	295
28460	0	21	244

CONCLUSIONS

- Mycrobial lysates can be reproducibly prepared and characterized by LC/MS analysis.
- The resulting LC/MS data sets can be processed using an automated workflow to permit component level characterizations.
- Using a LC/MS processing segment width at a fraction of the chromatographic peak width, a protein map can retain chromatographic characteristics of components.
- Replicate analysis of isolates shows comparable component patterns despite biological, processing, and analytical variation.
- Using this approach, a non-pathogenic strain (K-12), a common pathogenic strain (O157:H7), and novel pathogenic strain (O157:H–) could be differentiated by multiple features of their LC/MS data.

Lit Code Number 720001235EN ©2005 Waters Corporation

TO DOWNLOAD A COPY OF THIS POSTER VISIT WWW.WATERS.COM/POSTERS