LC-MS Based Detection of Molecular Genetic Modifications in Bacteria

Patricia K. Lankford, Tse-Yuan S. Lu, Bruce A. Tomkins, Anthony V. Palumbo, W. Hayes McDonald

Oak Ridge National Laboratory, Oak Ridge Tennessee 37831-6131

OVERVIEW

- The ability to detect and identify genetically modified bio-threat agents represents a difficult analytical challenge
- We have begun to explore this challenge using a standard molecular biology strain of *E. coli*
- A shotgun proteomic analysis allows relatively rapid detection and characterization of expressed genetically modified elements

INTRODUCTION

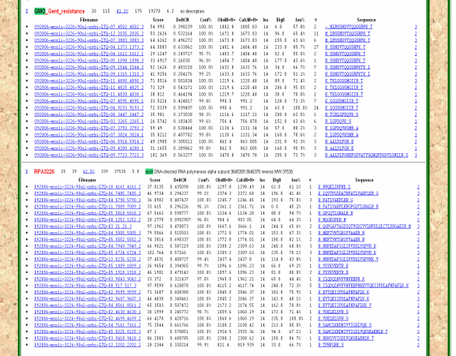
- Genetic modification of microorganisms is the cornerstone of any modern molecular biology laboratory. In fact, these techniques are so routine and "cookbook" that even labs possessing little or no expertise can take full advantage of the technology. Obviously, with even relatively minor precautions, these manipulations can be performed without endangering workers or the environment. Unfortunately, given the appropriate starting materials and sufficient resources, these same basic molecular manipulations could be used to engineer novel threat agents that posses either emergent properties or are designed to elude current testing, diagnosis, and treatment regimes.
- The major challenge with detecting genetically engineered threats is that it is necessary to target the modified element as opposed to the organism as a whole.
- However, Genetic engineering of bacteria to either enhance the pathogenicity of non-threat agents or disguise the identity of threat agents should leave signatures that could be targeted for identification. These targets include virulence factors, toxin genes, insertion elements, and antibiotic resistance markers.

EXPERIMENTS

- As a surrogate for a genetically modified threat we are using a standard lab *E. coli* strain (BL-21) with a broad host range plasmid (based on pBBR1MCS5) expressing a subunit of *R. palustris* RNA polymerase (RpoA, RPA3226). We are attempting to detect two elements from within the background of the *E. coli* proteome:
  - Gm-resistance (antibiotic selection marker)
  - *R. palustris* RpoA (heterologously expressed protein)
- Two sample preparation protocols were used (~ 50 ul cell pellets)
  - A trifluoroethanol (TFE) trypsin-digest-based shotgun proteomic preparation modified from Wang et al. 2005
  - Microwave assisted acid hydrolysis modified from Swatkoski et al 2006. Pellet solubolized in 300 ul 96% formic acid followed by microwaving on high for 90 s.
- Data dependent shotgun analysis of preparations using MudPIT LC/LC-MS/MS and LC-MS/MS acquired using an LTQ-orbitrap. MudPIT samples were loaded via pressure cell and 1D separations were autosampled.
- Peptide spectra were searched against protein database containing *E. coli*, *R. palustris*, antibiotic resistance proteins, common contaminants, along with reversed versions of each protein using DBDigger. Data were filtered and aggregated using DTASelect 2.0

RESULTS AND DISCUSSIONS

Figure 1: MudPIT analysis of tryptic digest. Modified proteins are able to be detected from a background of ~1000 identified *E. coli* proteins.

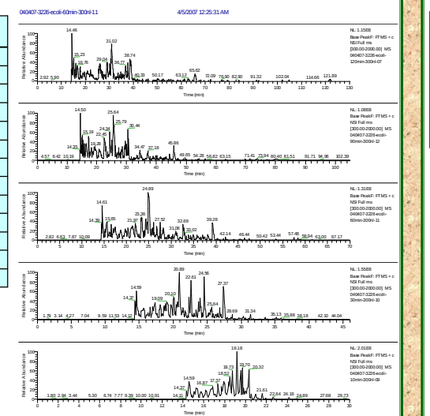


	Proteins	Peptide IDs	Spectra
Unfiltered	17125	138673	152850
Filtered	1067	7810	15110
Forward matches	1023	7721	15017
Decoy matches	44	89	93
Forward FP rate	4.3%	1.15%	0.62%

12-step, 22 hour MudPIT analysis of 90 ul of tryptic digest (~600 ul total). Using a decoy database, filtering was performed to yield approximately 5 % false positive peptide rate. 2 peptides per protein were required. Shown are DTASelect outputs for the two targeted genetically modified elements and summary table. Gm-resistance and RpoA are the 14th and 85th proteins respectively as sorted by total spectral counts matched.

Figure 2: 1D LC analysis of tryptic digest. Genetically modified elements are able to be detected using relatively short gradient separations.

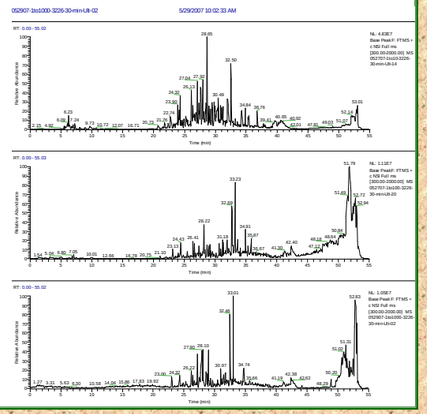
	Gm-resistance	*R. palustris* RpoA
120 min gradient #1	7, 9 , 34.9 %	3, 3, 11.8 %
120 min gradient #2	8, 8, 34.9 %	5, 5, 20.1 %
90 min gradient #1	7, 10, 21.7 %	4, 4, 18.0 %
90 min gradient #2	8, 9, 34.9 %	3, 3, 11.8 %
90 min gradient #3	6, 6, 34.9 %	3, 3, 11.8 %
60 min gradient #1	9, 14, 36.0 %	5, 5, 16.5 %
60 min gradient #2	7, 8, 34.9 %	5, 5, 16.5 %
60 min gradient #3	6, 6, 27.4 %	5, 5, 17.4 %
30 min gradient #1	7, 10, 21.7 %	2, 2, 7.4 %
30 min gradient #2	6, 7, 21.7 %	1, 1, 5.3 %
30 min gradient #3	7, 7, 21.7 %	3, 3, 14.5 %
10 min gradient #1	6, 7, 18.3 %	-
10 min gradient #2	3, 3, 10.9 %	-
10 min gradient #3	6, 7, 18.3 %	-
10 min gradient #2	3, 3, 10.9 %	-



10 ul injections (~ 600 ul total sample preparation). Injection and column re-equilibration times were kept constant and gradient lengths were varied from 10 to 120 minutes. Shown are tabular result for *peptide #*, *spectral #*, and *percent sequence coverage* for the two targeted elements. Representative base-peak chromatograms for each of the gradient lengths are also shown.

Figure 3: 1D LC analysis of tryptic digest (continued). Gm-resistance protein is able to be detected at much lower injection levels.

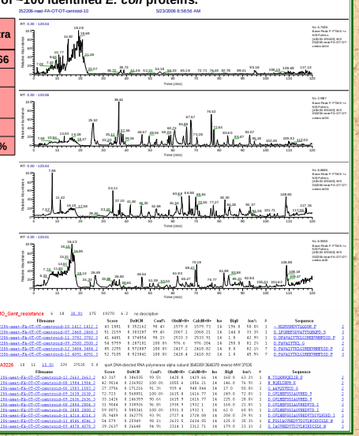
	Gm-resistance	*R. palustris* RpoA
1/10th #1	7, 7, 20.0 %	-
1/10th #2	5, 5, 14.3 %	-
1/10th #3	5, 5, 13.5 %	-
1/10th #4	4, 4, 14.3 %	-
1/100th #1	4, 4, 14.3 %	-
1/100th #2	4, 4, 10.3 %	-
1/100th #3	4, 4, 14.3 %	-
1/100th #4	2, 2, 10.3 %	-
1/1000th #1	-	-
1/1000th #2	3, 3, 14.3 %	-
1/1000th #3	1, 1, 9.1 %	-
1/1000th #4	3, 3, 10.3 %	-



Dilutions were prepared to give fractional percentages of the 10 ul injections shown in Figure 2 – 1/10th, 1/100th, and 1/1000th. 30 minute gradients were used. Shown are tabular result for *peptide #*, *spectral #*, and *percent sequence coverage* for the two targeted elements. Representative base-peak chromatograms for each of the dilutions are also shown. 1/1000th dilution represents proteins from ~ 1,000,000 initial cells.

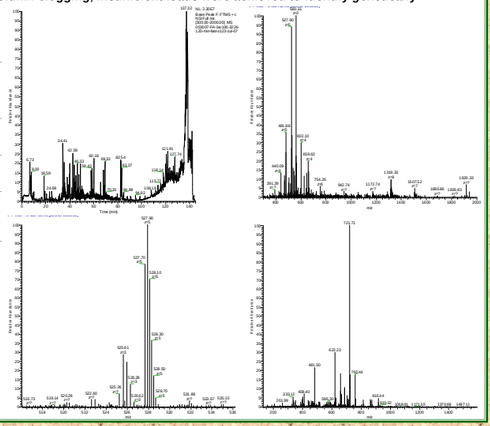
Figure 4: MudPIT analysis of microwave assisted acid digest. Genetically modified elements are able to be detected from a background of ~100 identified *E. coli* proteins.

	Proteins	Peptide IDs	Spectra
Unfiltered	16930	84832	105766
Filtered	115	501	745
Forward matches	114	499	743
Decoy matches	1	2	2
Forward FP rate	0.88%	0.4%	0.27%



12-step, 22 hour MudPIT analysis of 10 ul of microwave assisted formic acid digest (~600 ul total). Using a decoy database, filtering was performed to yield approximately 5% false positive peptide rate. Two peptides per protein were required. Shown are selected base-peak chromatograms, DTASelect outputs for the two targeted genetically modified elements, and summary table. Gm-resistance and RpoA are the 11th and 19th proteins respectively as sorted by total spectral counts matched.

Figure 5: 1D LC analysis of microwave assisted acid digest. Due to column clogging, insufficient loads were achieved to identify genetically modified elements.



1 ul injections of 1/100 dilution of initial sample preparation (~ 300 ul total in sample preparation). 120 minute gradient separation was used. Shown are DTASelect results for 6 of the identified proteins. Also shown are base-peak chromatogram, full-scan spectrum, zoomed spectrum, and MS/MS spectrum identified as -.MNKSQLIDKIAAGADISKAAAGRALD.A. Most identification were from early in our gradient and there were numerous larger peptide fragments of higher charge states present.

CONCLUSION

- Using shotgun proteomic LC-MS/MS techniques we are able to detect genetically modified elements from within the background of another proteome.
- Both standard tryptic-digest based and microwave assisted acid hydrolysis sample preparations allow detection of these elements
- However, additional development is needed to allow for more reproducible and more easily loaded microwave assisted acid hydrolysis protocol
- Ultimately, for use in detection regimes, targeted peptide identifications will be preferred over data-dependent LC-MS/MS analysis

ACKNOWLEDGMENTS

This research sponsored by the Office of Biological and Environmental Research, U.S. Department of Energy. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725