Waters™

アプリケーションノート

Label-Free Quantitative and Qualitative Analysis of *Escherichia Coli* Grown on Different Carbon Sources

Jeffrey C. Silva, Richard Denny, Craig Dorschel, Guo-Zhong Li, Keith Richardson, Daniel Wall, S. J. Geromanos

Waters Corporation

Abstract

In this application note *E. coli* was used as a model organism to demonstrate the capabilities of this system for detailed quantitative and qualitative analysis of a complex biological system.

Introduction

Many studies involving protein analysis of complex protein mixtures have been accomplished by combining the well-established separation capabilities of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) with mass spectrometry (MS)-based or tandem mass spectrometry (MS/MS)-based sequence identification of the gel-resolved proteins. Although 2D-PAGE has been used successfully in quantitative proteomics, it is subject to mass range and pI limitations and it requires considerable effort to generate reproducible results. Gel spots may also contain more than one protein which will affect subsequent quantitative analysis.

The subject of this study is an alternative label-free, LC-MS approach which has been developed to enable quantitation and identification of proteins from a single experiment.

The Waters Protein Expression System utilizes the high retention time reproducibility of the Waters nanoACQUITY UPLC System and the exact mass measurement accuracy capability of the Micromass Q-Tof Premier Mass Spectrometer. In addition a novel suite of Informatics tools have been developed, to process and interpret meaningful quantitative/qualitative information from the complex datasets obtained. The Waters Protein Expression System enables the researcher to determine both the changes in relative abundance of peptides across samples and controls, as well as to identify the parent proteins from the same experiment.[1-5]

The patented LC-MS method presented here employs a multiplexed acquisition routine which enables the parallel analysis of the constituent peptides in a complex biological sample, leading to a significant improvement in the sequence coverage obtained for identified proteins.[6]

In previous work, the approach used in the Protein Expression System was shown to yield extensive quantitative and qualitative information from a series of protein mixtures in a background of human serum.[7] The work presented in this application note was designed to use *E. coli* as a model organism to demonstrate the capabilities of this system for detailed quantitative and qualitative analysis of a complex biological system.

Figure 1. The Waters Protein Expression System, incorporating the nanoACQUITY UPLC System, Micromass Q-Tof Premier Mass Spectrometer, Expression Informatics, MassLynx Software, Protein Expression Application Kits, specialized training and operating procedures.

Experimental

Materials and Methods

Media and Growth Conditions

- *E. coli* (ATCC10789, K-12) cells were grown on Luria-Bertani (LB) plates and incubated at 37 °C. An individual colony was plated onto M9 minimal media supplemented with 0.5% sodium acetate.
- Seed culture flasks were shaken at 250 rpm at 37 °C until mid log phase (OD_{600} = 0.9 to 1.1). The seed culture was diluted, 1 mL to 500 mL, into flasks containing M9 minimal media supplemented with one

of three carbon sources (0.5% glucose, 0.5% lactose or 0.5% sodium actetate).

- Flasks were shaken at 250 rpm at 37 °C until mid log phase (OD600 = 0.9 to 1.1). The *E. coli* were harvested by centrifugation and frozen at -80 °C.

Preparation of Soluble Protein

- Frozen cells were suspended in lysis buffer (5 mL per 1 gm biomass; PBS, 1/100 protease inhibitor cocktail [Sigma #8340]) in a 50 mL Falcon tube.
- The cells were lysed by sonication in a Microson XL Ultrasonic Cell Disrupter (Misonix, Inc.) at 4 °C.
- The cell debris was removed by centrifugation at 10,000 g for 30 minutes at 4 °C and the resulting soluble protein extracts were dispensed into 0.25 mL aliquots and stored at -80 °C for subsequent analysis.

Protein Digestion

- Approximately 250 µg of *E. coli* protein were brought up to a final volume of 100 µL with 50 mM ammonium bicarbonate (pH 8.5), containing 0.05% of RapiGest *SF* surfactant to optimize the efficiency of protein digestion.
- The protein sample was reduced in the presence of 10 mM dithiothreitol at 60 °C for 30 min and then alkylated in the presence of 30 mM iodoacetamide at room temperature for 30 min.
- Trypsin digestion was initiated by adding modified trypsin (Promega) at a concentration of 50:1, *E. coli* protein to trypsin, and incubated overnight at 37 °C.
- The tryptic peptide solution was centrifuged at 10,000 g for 10 min and the supernatant was transferred into an auto-sampler vial for LC-MS analysis.

LC Conditions

- A nanoscale HPLC System was used for separation of the peptide samples.
- The analytical column used was a 300 µm ID X 15 cm Waters NanoEase Atlantis dC18 Column.
- The flow rate through the column was 4.4 mL/min. The mobile phases were A = 1% Acetonitrile in Water, 0.1% Formic acid, B = 80% Acetonitile in Water, 0.1% Formic acid.
- An analytical gradient was applied form 6% B to 40% B over 100 minutes followed by a 10 minute rinse at 99% B with a 20 minute re-equilibration at initial conditions.

MS Conditions

- A Q-Tof Mass Spectrometer was operated in the V-Optics mode of operation at >10,000 FWHM and data was acquired with alternating (low and elevated, MS and MS^E) collision energy in 1.8-second intervals with a 0. 2 interscan delay. During elevated energy data acquisition, the collision energy was ramped from 27eV to 33eV.

- A NanoLockSpray source was used to ensure exact mass accuracy data acquisition, typically less than 3 ppm. The doubly charge ion of Glu-Fibrinopeptide and the singly charged ion of Leucine Enkephalin were used as the lock mass compounds. A NanoLockSpray scan was acquired every 30 seconds.

Results and Discussion

E. coli was grown in the presence of minimal media and one of three primary carbon sources; Glucose, Lactose or Acetate. The digested, constituent proteins from each of the samples were analyzed in triplicate using the Waters Protein Expression System to determine relative peptide expression changes and to identify the parent proteins.

EMRT Signatures – Peptide Detection and Quantitation

The complexity of the three *E. coli* samples is demonstrated in Figure 2, where the average monoisotopic mass for each of the extracted peptide components (MH^+) is displayed against its corresponding average retention time for all three conditions. This is known as the EMRT, or Exact Mass Retention Time, which is used as a specific signature for a given peptide allowing it to be identified in a sample with high specificity, and tracked across sample sets for subsequent quantitative comparison. The peptide ions (EMRT’s) from replicate injections were clustered by their exact mass and retention time. These clustered EMRT pairs can be plotted to display the up and down regulation of peptides between samples. Subsequent databank searching of selected clusters leads to identification of the protein(s).

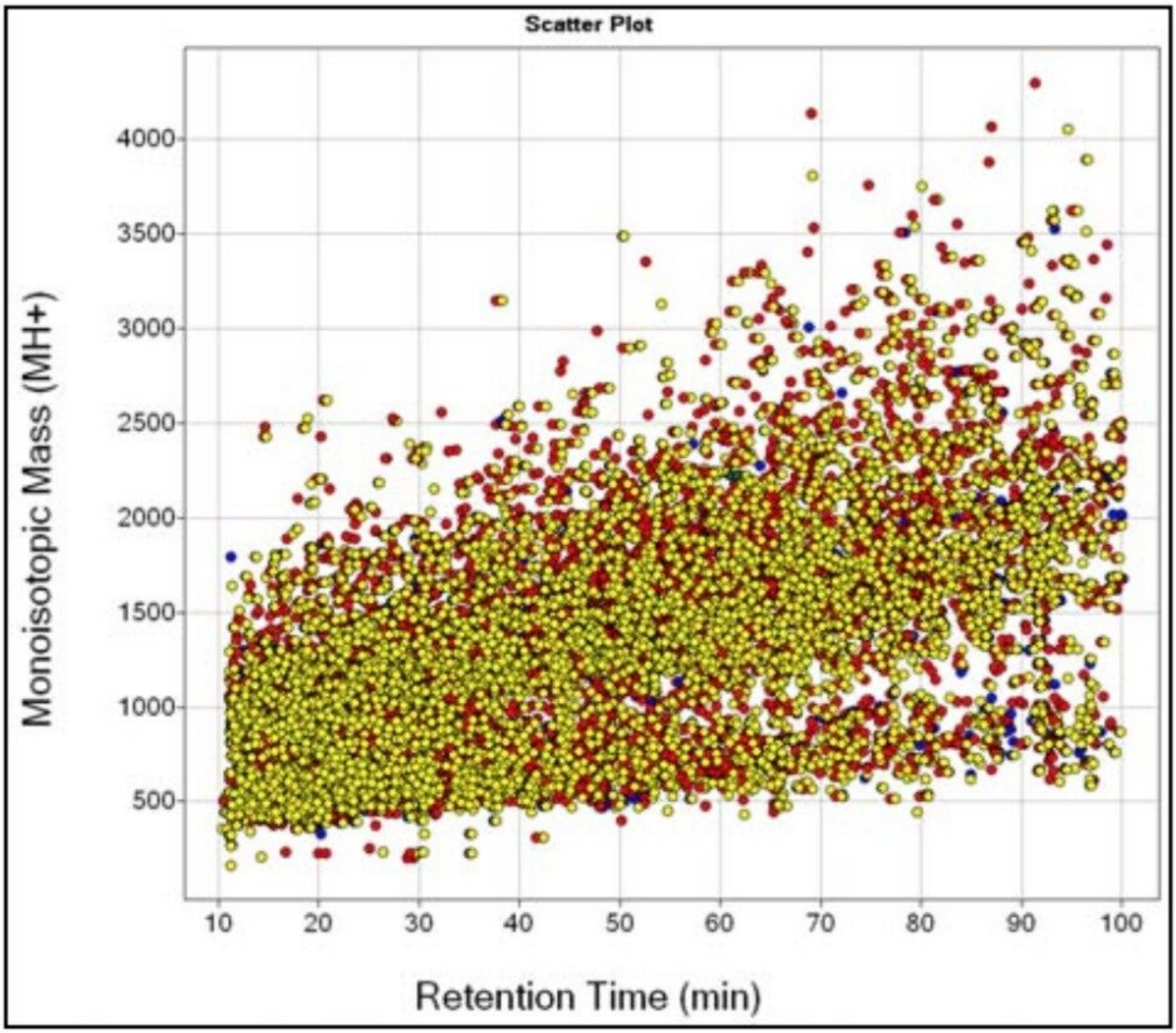

Figure 2. Overlay of the extracted peptide components from E. coli grown on glucose (blue), lactose (red) and acetate (yellow).

Analytical Reproducibility

The analytical reproducibility of the method is shown in Figure 3, which compares two repeat injections of the *E. coli* protein digest sample grown on acetate, demonstrating a coefficient of variation of approximately 13%. Pair-wise comparisons of replicate injections of the *E. coli* grown on glucose and lactose were 17% and 15%, respectively. This degree of analytical performance ensures discovery of small expression level differences between samples.

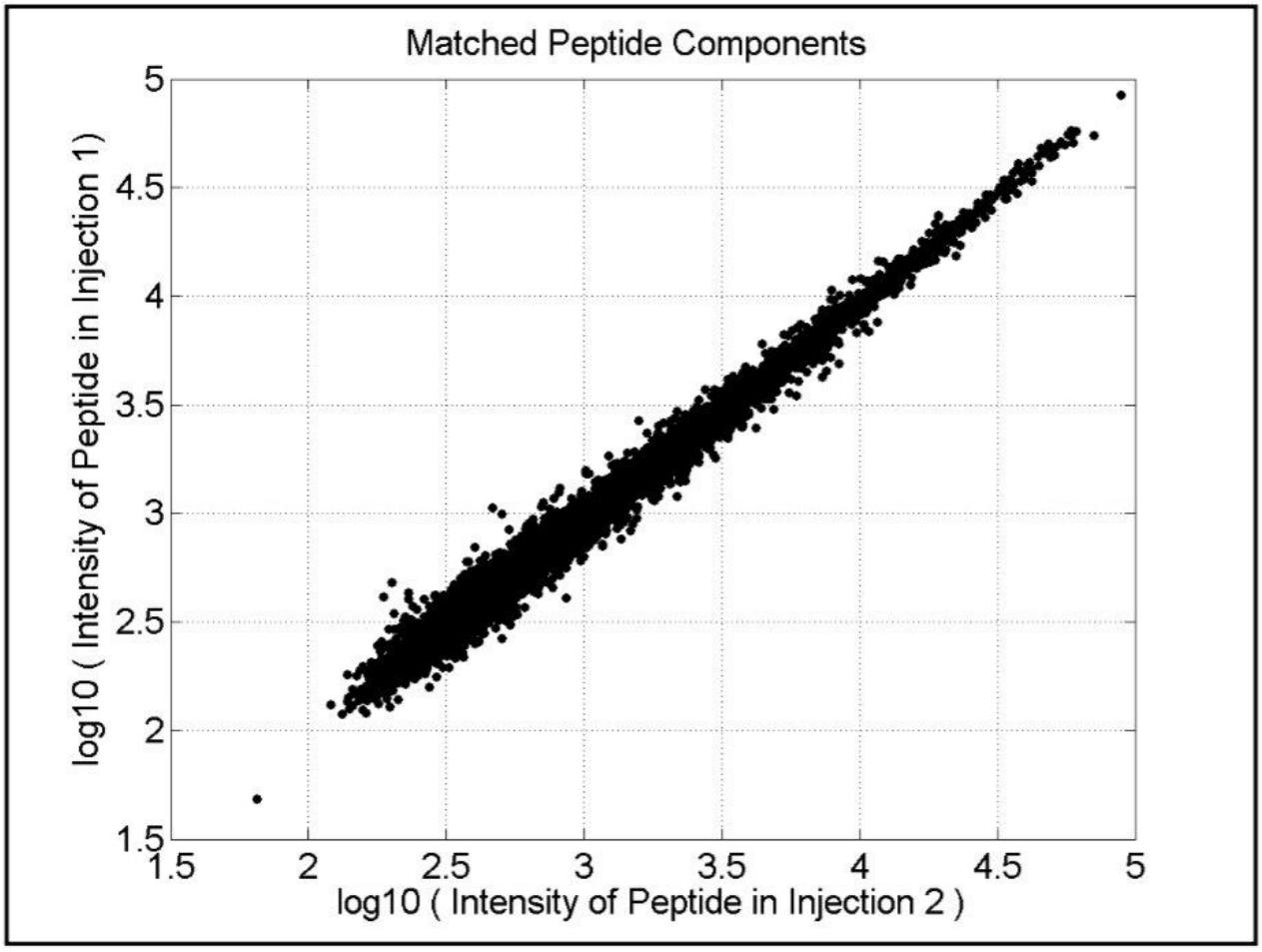

Figure 3. Replicate analysis of Acetate. A scatter plot the matched components (mass precision = 5 ppm, retention time tolerance = 0.25 minutes) between two replicate injections of the acetate growth condition of E. coli. The coefficient of variation among the three replicate injections was approximately 13%.

Normalization of the Data

The discovery of low-level protein abundance changes is also enabled by the Informatics normalization schemes, auto-normalization of the peptide intensities or normalization to an internal standard. Four peptides from protein chain elongation factor, TUFA, were chosen for global normalization throughout the entire experiment. The four TUFA peptides chosen for normalization were found in all replicate injections for the three conditions. The peptides were identified using databank searching against the elevated energy (MS^E) data. Figure 4 shows the annotated MS^E mass spectrum for the AIDKPFLLPIEDVFSISGR peptide (2117.1479 MH^+ at 91.81 min) from TUFA. Peptides were also identified from succinyl-CoA synthetase, isocitrate lyase and citrate synthase (data not shown).

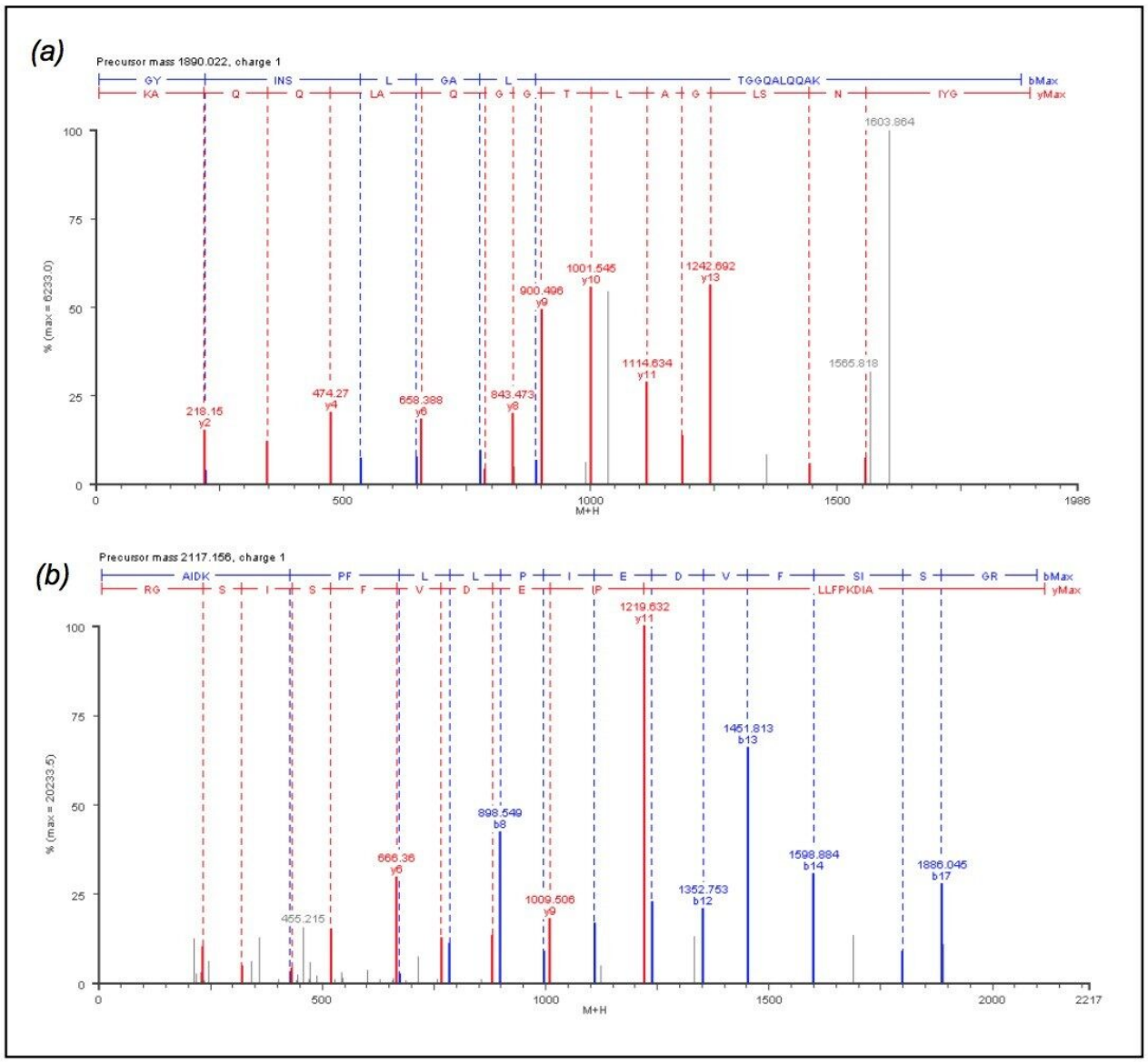

Figure 4. MS^E spectra of two extracted peptides; (a) Isocitrate lyase (b) Protein Chain Elongation Factor, TUFA. This illustrates the quality of the time-resolved elevated energy mass spectrum obtainable by this method in conjunction with the Protein Expression Informatics software.

Quantitative Results

Figure 5a displays the relative peptide abundance observed in the glucose versus lactose growth conditions. Once the matched peptides are plotted according to their relative fold-change, the quantitative comparison of the matched peptides provides a means to quickly identify those specific peptides/proteins that exhibit a noticeable change due to the perturbation.

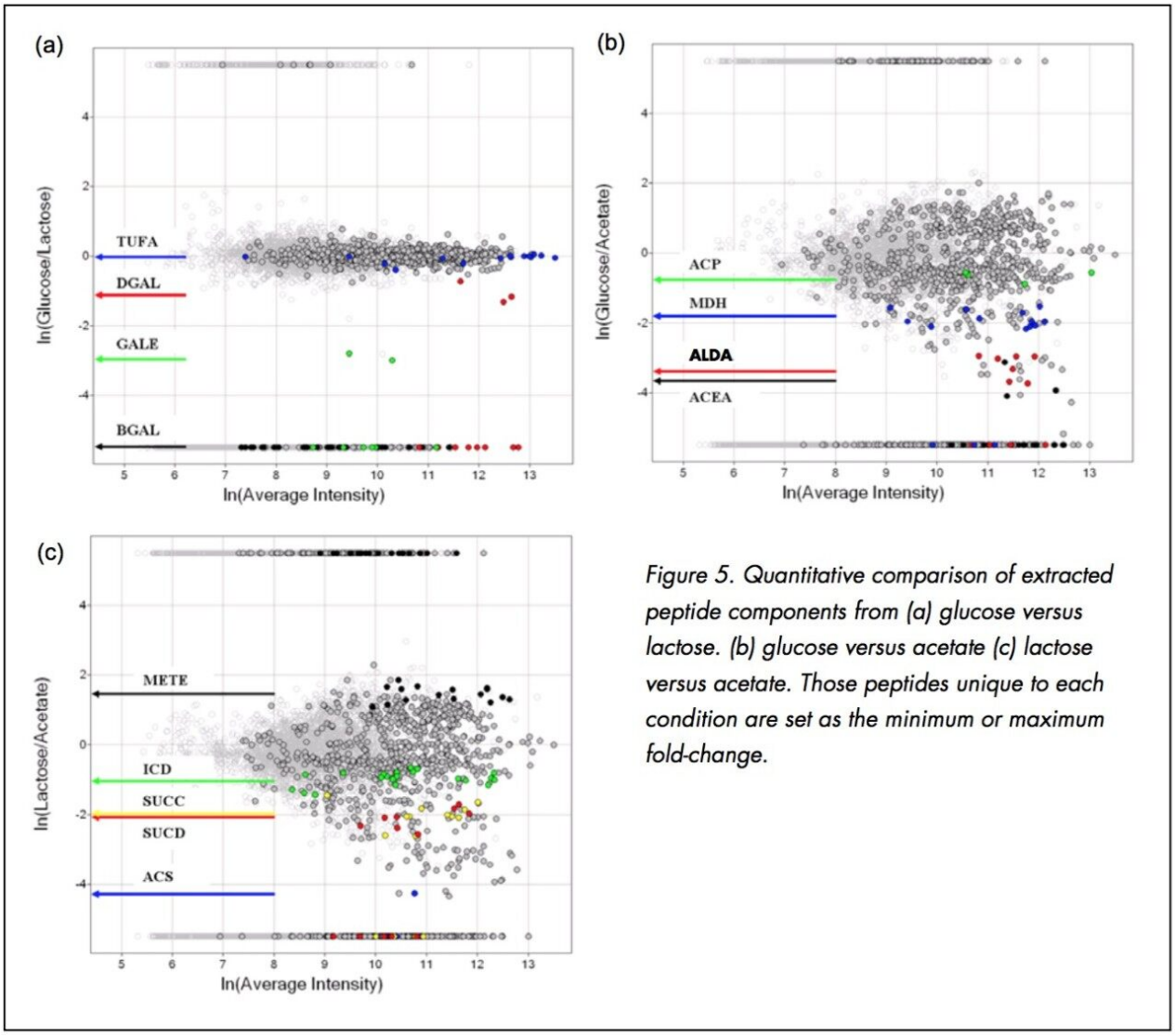

Figure 5. Quantitative comparison of extracted peptide components from (a) glucose versus lactose. (b) glucose versus acetate (c) lactose versus acetate. Those peptides unique to each condition are set as the minimum or maximum fold-change.

Figure 5. Quantitative comparison of extracted peptide components from (a) glucose versus lactose. (b) glucose versus acetate (c) lactose versus acetate. Those peptides unique to each condition are set as the minimum or maximum fold-change.

The comparison in Figure 5a represents the natural log (ln) of the average intensity for the matched peptide components across both conditions (x-axis) versus the ln of the normalized intensity ratio for the matched peptides between the two conditions (y-axis).

Those peptides which were unique to each condition were set to either 5.5 (unique to glucose) or -5.5 (unique to lactose).

The average protein coverage obtained from the individual injections for TUFA was approximately 55%. The inherent redundancy of the tryptic peptides to any particular protein provides multiple

independent quantitative measurements and can be used to determine the relative quantitation of any particular protein between two conditions. The multiple measurements also provide a means to determine a confidence interval for the relative quantitation of any particular protein in a study.

Differential Expression at the Protein Level

Protein chain elongation factor, TUFA, was found in all three conditions and was determined to not be differentially expressed (Table 1). Galactose-binding transport protein (DGAL), UDP-galactose-4-epimerase (GALE), beta-D-galactosidase (BGAL) were also found to be differentially expressed (Figure 5a).

Protein	Description	ln (Glu/ Lac)	95%CI	ln (Lac/ Ace)	95%CI	ln (Glu/ Ace)	95%CI
ACEA	Isocitrate lyase	-0.03	0.16	-3.63	0.54	-3.72	0.59
ACEB	Malate synthase	0.35	ND	-2.37	ND	-2.02	ND
ACEE	Pyruvate dehydrogenase E1	-0.03	0.07	-0.11	0.06	-0.14	0.08
ACNB	Aconitate hydratase 2	0.06	0.07	-1.22	0.11	-1.16	0.10
ACS	Acetyl-CoA synthase	-5.50	ND	-4.27	ND	-5.50	ND
ALDA	Aldehyde dehydrogenase A	-0.07	0.16	-3.13	0.16	-3.23	0.26
ENO	Enolase	-0.17	0.03	0.57	0.10	0.38	0.09
FBAA	Fructose-bisphosphate aldolase II	-0.06	0.08	-0.24	0.13	-0.30	0.17
GALE	UDPglucose-4-epimerase	-2.89	0.20	2.00	0.28	-1.03	0.20
GAPA	Glyceraldehyde-3-phosphate dehydrogenase A	-0.23	0.05	0.63	0.20	0.41	0.19
GLTA	Citrate synthase	0.08	0.11	-2.29	0.16	-2.19	0.23
GPMA	2,3-Bisphosphoglycerate dependent phosphoglycerate mutase	0.17	0.08	-0.82	0.05	-0.65	0.08
ICD	Isocitrate dehydrogenase	0.05	0.06	-0.97	0.09	-0.89	0.08
LACZ	Beta-galactosidase	-5.50	ND	5.50	ND	ND	ND
LPDA	Dihydrolipoyl dehydrogenase	0.10	0.20	-1.79	0.12	-1.66	0.22
MDH	Malate dehydrogenase	0.19	0.10	-2.08	0.12	-1.89	0.13
PGK	Phosphoglycerate kinase	-0.14	0.06	-0.13	0.12	-0.26	0.13
SUCB	2-Oxoglutarate dehydrogenase	0.30	0.23	-2.45	0.47	-2.15	0.26
SUCC	Succinyl-CoA synthase (alpha)	0.16	0.08	-2.02	0.21	-1.86	0.19
SUCD	Succinyl-CoA synthase (beta)	0.00	0.19	-2.12	0.20	-2.00	0.29
TUFA	Protein chain elongation factor	-0.07	0.06	0.07	0.07	0.00	0.10

Table 1. The relative quantitation of the characterized proteins described in this study. The relative quantitation is expressed as the ln of the average intensity ratio for all matching peptides between the two conditions with the appropriate 95% confidence interval. Proteins which were unique to a particular condition have been artificially set to either 5.5 (numerator) or -5.5 (denominator).

Figure 5b and Figure 5c show the differential peptide analysis between those peptides found in the comparison of glucose/acetate and lactose/acetate, respectively. Highlighted in Figure 5b are the peptides identified to acyl-carrier protein (ACP), malate dehydrogenase (MDH), aldehyde dehydrogenase A (ALDA) and isocitrate lyase (ICL or ACEA) from the quantitative comparison of the peptides found in the glucose versus acetate growth conditions.

Figure 5c illustrates the peptides identified to tetrahydropteroyl-triglutamate-homocysteine methyltransferase (METE), Isocitrate dehydrogenase (ICD or IDH), Succinyl-CoA-synthetase (SUCC), Succinyl-CoA-synthetase (SUCD) and Acetyl-CoA synthase (ACS) from the quantitative comparison of the peptides found in the lactose versus acetate growth conditions. Interestingly, SUCC and

SUCD interact to form the heterotetrameric A2B2 complex of succinyl-CoA synthetase. The observed fold-change for SUCC and SUCD is consistent with the structure.

Mapping Differential Expression to the Glyoxylate Shunt

The limited differential expression observed between lactose and glucose can be explained by the metabolic requirements for the utilizing the two different carbon sources. Lactose is a disaccharide of glucose and galactose, and therefore the metabolic differences are manifested in the active transport of the disaccharide carbon source and the conversion and epimerization of the galactose monomer to glucose. Figure 5a highlights the peptides identified to the galactose-specific processing proteins: DGAL, GALE and BGAL. It is not surprising that there is little variation associated with the observed peptide components in Figure 5a, since there is minimal impact to the downstream metabolic activity when supporting growth on either glucose versus lactose. Both carbon sources are ultimately processed through the glycolysis and the citric acid cycle in a similar fashion.

The relative fold-change of a few of characterized proteins from each binary comparison is illustrated in Table 1. Providing acetate as the sole carbon source for *E. coli* requires substantially different metabolic activity to support growth than that of either glucose or lactose. This is evident from the variation associated with the relative foldchange associated with the binary comparisons of the detected peptide components from either glucose or lactose to those of acetate (Figure 5b and 5c). Acetate is a simple carbon source which initially bypasses glycolysis and enters into a modified version of the citric acid cycle, glyoxylate shunt (Figure 6), to provide the necessary primary metabolites and energy to support growth. Before entering the glyoxylate shunt, acetate must first be converted to acetyl-CoA. The conversion of acetate to acetyl-CoA is performed through the activity of acetyl-CoA synthetase (ACS). The evidence accumulated from the peptide analysis of the three conditions indicates that ACS was not detected in the glucose condition, but was present in both the lactose and acetate conditions. The relative quantitation obtained from the comparison of lactose or glucose versus acetate is consistent with the induction of ACS in the acetate condition. Another indication of growth on acetate as the sole carbon source is the activation of the glyoxylate shunt which is accompanied with the relative induction of isocitrate lyase (ACEA) and malate synthase (ACEB) as seen in Table 1.

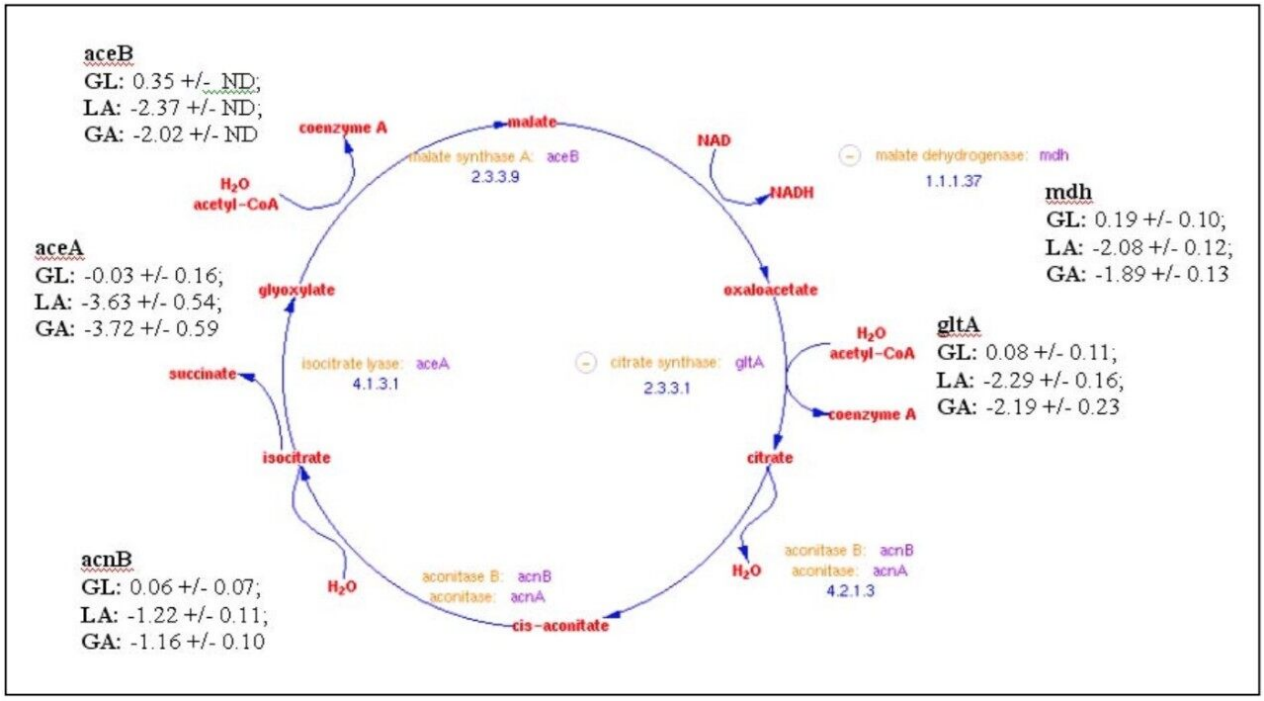

Figure 6. Metabolic scheme for glyoxylate shunt and the associated protein quantitation. The quantitation results from the identified proteins in this study are reported at each step of the biosynthetic pathway. The fold-change for each of the comparisons is provided as follows: GL = glucose versus lactose, LA = lactose versus acetate and GA = glucose versus acetate. The fold-change is reported as the average ln ratio along with the 95% confidence intervals.

Interestingly, over 60% of the ribosomal proteins as well as a few translation factors were identified from the three different growth conditions. The average protein coverage obtained from these abundant proteins was approximately 40%. The quantitation results indicated that the level of the ribosomal machinery in acetate was approximately 2.5 times lower than the levels observed in either glucose or lactose. These results are in agreement with the consequences to *E. coli* when subjected to poor growth conditions.[8-10]

Conclusion

- Expression profiling of proteins in *E. coli* grown on three different carbon sources has been shown using a labelfree LC-MS based method. A comprehensive report of this data will be presented as the topic of future work.[10]

- Relative quantitation of the key up- and down-regulated proteins between the conditions could be made with high sequence coverage and high confidence.
- The results generated were in good agreement with the literature and demonstrate the power of this methodology for qualitatively and quantitatively characterizing changes in a complex proteome.
- This approach provides a comprehensive method of relative protein profiling suitable for time course or multi-condition studies
- This LC-MS approach has also been extended to the analysis of other biological systems, such as Mycobacterium (*M. bovis*) to study proteomic profiles under different drug treatments in an effort to determine mechanism of action of novel drugs.[11]

References

1. Geromanos, S.; Dongre, A.; Opiteck, G.; Silva, J. C., U. K. Patent 2, 385, 918 A, 2003.

2. 3. 4. 5. 11. Scientific posters from Proceedings of the 52nd ASMS Conference on Mass Spectrometry and Allied Topics, 2004, Nashville, TN, USA. Posters can be downloaded at www.waters.com < http://www.waters.com> .

6. Bateman, R. H; Hoyes, J. B., U. K. Patent 2, 364, 168 A, 2002.

7. Silva, J. C.; Denny, R.; Dorschel, C.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P. and Geromanos, S. (2005) *Analytical Chemistry*, April 1; 77(7): 2187–2200.

8. Marr, A. G. (1991) *Microbiology Reviews*; 55(2):316–333.

9. Toa, H.; Bausch, C.; Richmond, C.; Blattner, F. R. and Conway T. (1999) *J. Bacteriol.* 181(20):6425–6440.

10. Silva, J. C.; Denny, R.; Dorschel, C.; Li, G. Z.; Richardson, K.; Wall, D. and Geromanos, S. Simultaneous Qualitative and Quantitative Analysis of the *E. coli* Proteome: A Sweet Tale “. *Manuscript in Preparation*.

11. Hughes, M.; C.; Silva, J. C.; Dorschel, C.; Geromanos, S. and Townsend, C. A. In Proceedings of the 53nd Annual ASMS Conference on Mass Spectrometry and Allied Topics, “Target Identification Studies: Application of a New LC-MS and Data Analysis Software System to Identify Drug-Induced Changes in Mycobacteria”, 2005, San Antonio, TX, USA.

Featured Products

ACQUITY UPLC M-Class System <https://www.waters.com/134776759>

MassLynx MS Software <https://www.waters.com/513662>

720001310, August 2005