Siena 2004

Simultaneous Qualitative and Quantitative Analysis of Whole Proteomes and Complex Protein Mixtures using Accurate Mass and Chromatographic Information

Roy Martin[1], Jeff Silva[2], Keith Richardson[3], Kieran Neeson[3], Scott Geromanos[2] (*Waters Corp.[1]Beverly, MA; [2]Milford, MA, USA. [3]Manchester, UK)*;

Introduction

This poster illustrates the quantitative and qualitative capabilities of the new Waters® Protein Expression System. This system was designed to analyze complex mixtures and quantify changes in protein levels through the relative area difference of accurately measured peptide chromatographic peaks. Since data were acquired in triplicate in alternate scanning low and elevated energy accurate mass LC/MS mode, each peptide measurement has both quantitative statistics and fragmentation information for qualitative assessment. The alternate scan method has considerably higher duty cycle than traditional data-directed analyses which are performed in a serial manner. It also makes full use of the speed and accuracy of an OA-Tof analyzer.

Data presented are from two sources: *E. Coli* cells grown under different carbon source conditions, and commercially available tryptically digested human serum samples spiked with different concentrations of five exogenous internal reference proteins contained in the Waters Protein Expression System Standards. Raw data was processed through a suite of Protein Expression System Informatics tools to produce a list of unique Exact Mass Retention-Time (EMRT) components. The data are processed further to match and compare the intensity of like EMRT components from different sample injections. The accurate masses are then matched with available fragment ions for qualitative assessment. Precursor ions and fragment ions are related by chromatographic peak maxima that they share. Once peptide identities are assigned, the data can be further processed to yield quantitative results based on protein identities.

Data Collection

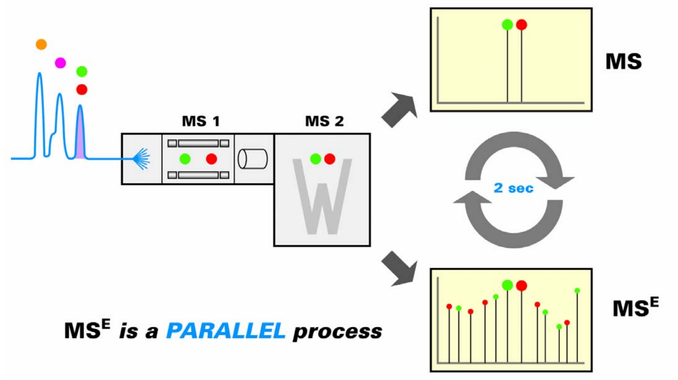


Figure 1. *Data acquisition using the Waters Protein Expression System mass spectrometer (Q-Tof). The instrument acquires data in alternate high and low collision energy scans to generate a comprehensive data where one set of scans contains mostly precursor ions, and a simultaneous set of scans contains mainly fragment ions.*

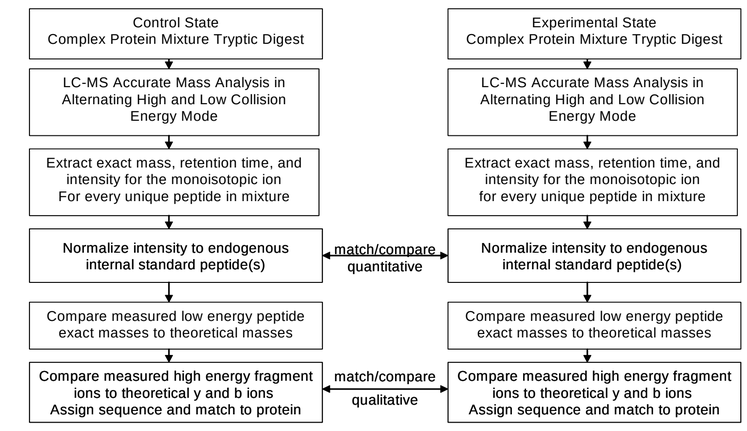


Figure 2*. Flowchart for Quantitative and Qualititative analysis. Two sample states are shown, although the method is not limited and the software can handle up to 300 comparisons.*

Sample Preparation

Human serum (Sigma source) was tryptically digested according to the procedure (C. Dorschel et al., "Protocols to Assure Reproducible Quantitative and Qualitative Analysis of Tryptic Digests of Complex Protein Mixtures for Global Proteomic Experiments"). Five aliquots of this human serum digest were spiked with Waters Mass Prep Digest Standards (MPDS), containing equimolar levels of Yeast Enolase and Alcohol Dehydrogenase, Rabbit Glycogen Phosphorylate, Bovine Serum Albumin and Hemoglobin tryptic digest.

***E. Coli* Media and Growth Conditions:** Frozen *E. coli* (ATCC10798, K-12) cell stocks were streaked onto Luria-Bertani (LB) plates and grown at 37 degrees Celsius. An individual colony was subsequently streaked onto M9 minimal medium plated supplemented with 0.5% sodium acetate and grown at 37 degrees Celsius. Seed cultures were generated by transferring single colonies into flasks of M9 minimal media supplemented with 0.5% sodium acetate. Seed culture flasks were shaken at 250 rpm at 37 degrees Celsius until mid log phase (OD600 = 0.9-1.1). The seed culture was diluted 1ml:500ml into separate M9 minimal media supplemented with one of three carbon sources (0.5% glucose, 0.5% lactose or 0.5% sodium acetate). Flasks were shaken at 250 rpm at 37 degrees Celsius until mid log phase (OD600 = 0.9-1.1). The *E. coli* cell cultures were harvested by centrifugation, pellets were frozen at -80 degrees Celsius.

***E. Coli* Protein Extract Preparation:** Frozen cells were suspended in 5ml per 1gm biomass in lysis buffer (Dulbecco's Phosphate-Buffered Saline + 1/100 Protease Inhibitor Cocktail [Sigma cat #8340]) in a 50 mL Falcon tube. The cells were lysed by sonication in a Microson XL Ultrasonic Cell Disrupter (Misonix, Inc.) at 4 degrees Celsius. The cell debris was removed by centrifugation at 15,000xg for 30minutes at 4c, and the resulting soluble protein extracts were dispensed into 1 mL aliquots and stored at -80 degree Celsius for subsequent analysis.

LC/MS System: Waters Protein Expression System comprised of the Waters CapLC® System with the Waters Micromass® Q-Tof Ultima API Mass Spectrometer equipped with a NanoLockSpray™ Source operated at 12,000 mass resolving power
Column: Waters NanoEase™ Atlantis™ dC18 Column, 300 µm x 15 cm
Mobile Phase: A = 1% Acetonitrile in Water, 0.1% Formic Acid, B = 80% Acetonitrile in Water, 0.1% Formic Acid
Gradient: 6% to 40% B over 100 min. at 4.4 µL/min, followed by 10 min. rinse (99% B) and 20 min. re-equilibration at initial conditions

Each of the *E. Coli* extract digests and five spiked human serum (HS) digest samples was analyzed in triplicate. Injections were 5 microliters each and were made directly on-column. The total protein load for each injection was ~8.5 micrograms of Human Serum plus either 5.0, 1.0, 0.5, 0.25 or 0.1 picomoles of the spiked MPDS. Exact mass LC/MS data was collected using an alternating low (10eV) and elevated (28eV to 35eV) collision energy mode of acquisition such that one cycle of low and elevated collision energy data was acquired every 4.0 seconds. The NanoLockSpray source was switched every 10 seconds to obtain a reference scan of [Glu[1]]-Fibrinopeptide lockmass calibrant.

Concepts of Data Processing

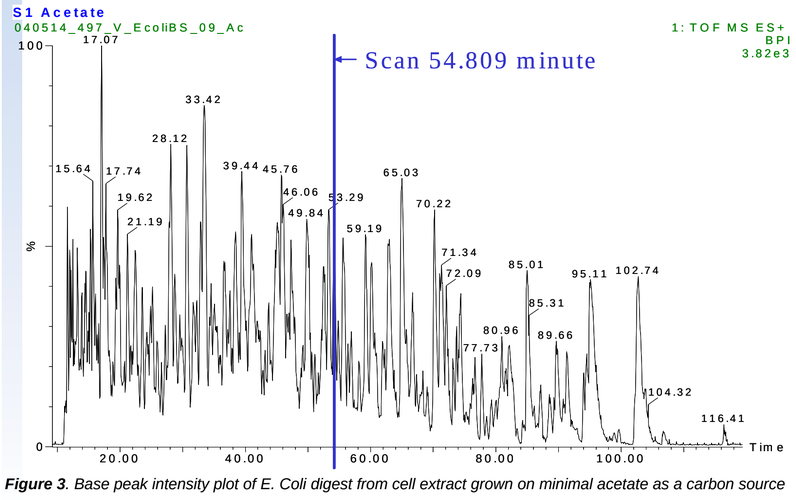


Figure 3*. Base peak intensity plot of E. Coli digest from cell extract grown on minimal acetate as a carbon source*

Elevated-Energy Low-Energy Elevated-Energy

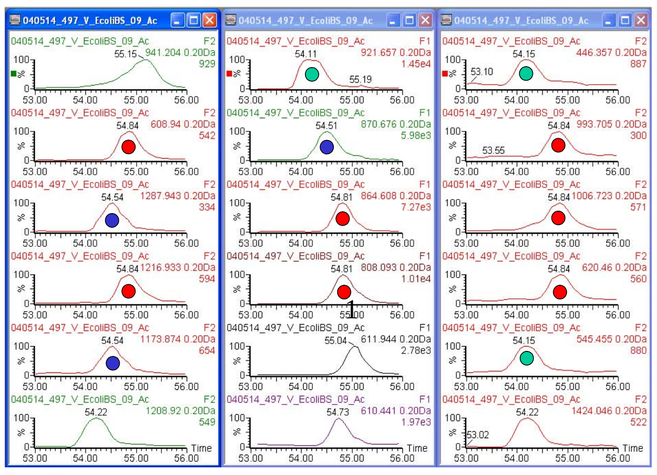


Figure 4. *This shows that while a large number of ions may be present in a single spectrum, the software distinguished between ions that have different peak maxima. In this example, two red ions share the exact same retention time, as do their related fragment ions.*

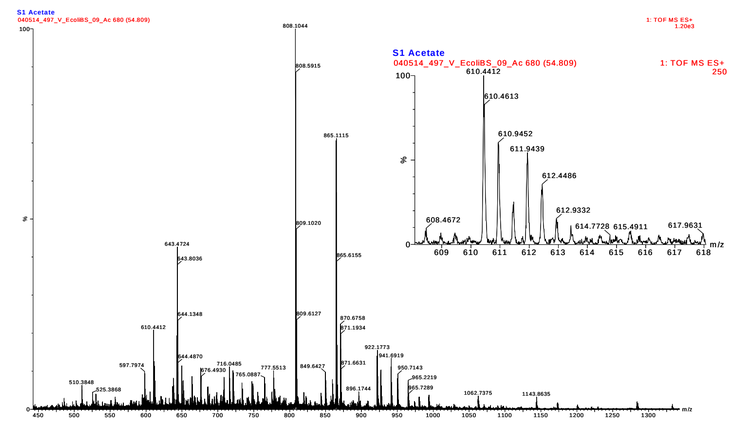


Figure 5*. Single MS scan at 54.8 minutes from E. Coli sample. Insert shows resolution and peak shape, and illustrates an example where a low resolution scan would have difficulty picking the correct precursor mass, and a DDA analysis would most likely pass both precursor ions, confusing the MS/MS interpretation..*

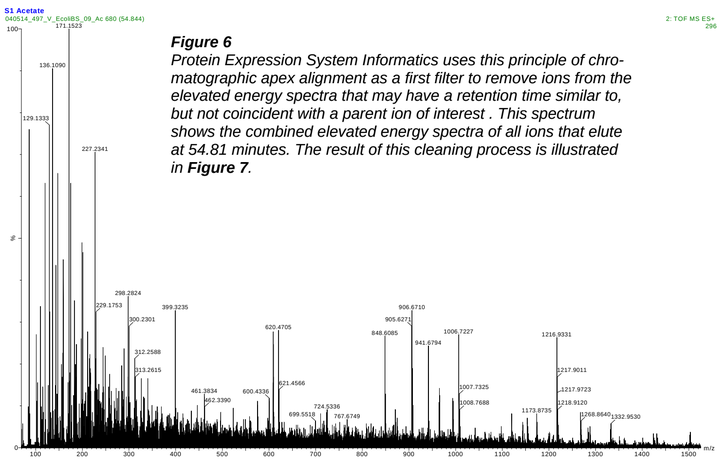
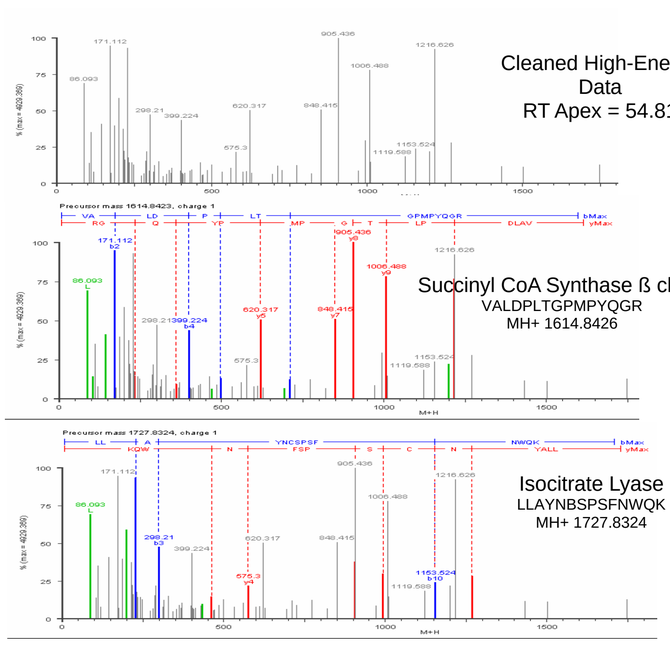


Figure 6
*Protein Expression System Informatics uses this principle of chromatographic apex alignment as a first filter to remove ions from the elevated energy spectra that may have a retention time similar to, but not coincident with a parent ion of interest . This spectrum shows the combined elevated energy spectra of all ions that elute at 54.81 minutes. The result of this cleaning process is illustrated in **Figure 7**.*

Figure 7
*The spectra below show the effect of the cleaning process. The top panel shows the deconvoluted masses that apex at 54.81 minutes in **Figure 4**. The second and third panels show matches to two different peptides.*



Qualitative Principle

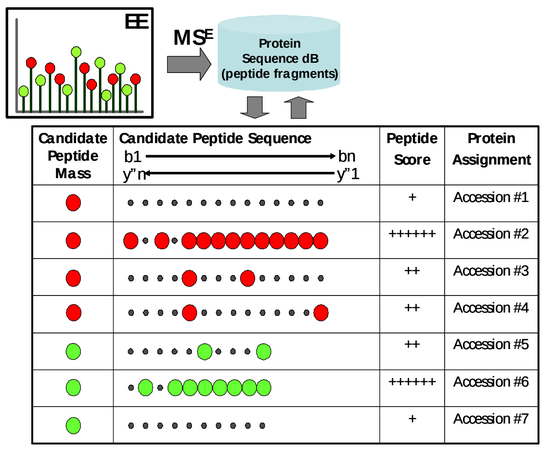


Figure 8
Candidate Peptides are identified by matching accurately measured masses against a database containing all possible tryptic fragments with no or one missed cleavages. This prescreen generates a smaller table of candidate peptide sequences. For positive identification, the ions observed in the co-apexing cleaned high energy data are compared against a model fragmentation pattern generated from the database peptide sequence. At the mass accuracies typically used (<5ppm) the correct sequence is usually clearly distinguished from incorrect sequences.

Quantitative Principle

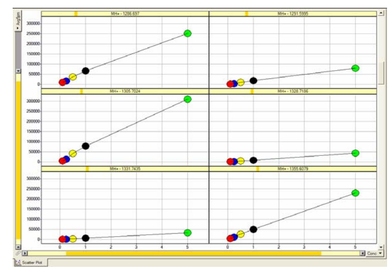
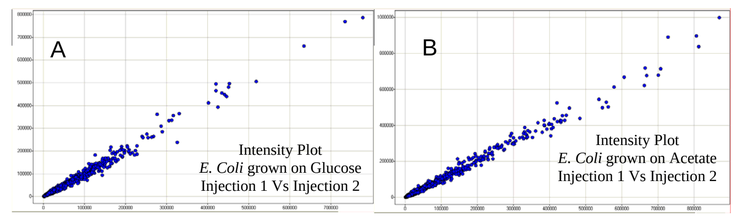


Figure 9
Peptides are clustered based on mass and retention time pairs. Quantification is based on accurately measure chromatographic areas of the deconvoluted molecular mass. The example at left shows the response curves for 6 of the EMRT' associated with the MPDS over five spiked levels. Those EMRT's associated with serum had a slope of zero. Peptides can be quantified on their own, or as sets of peptides identified in databases then quantified using protein ID's as a basis for comparison.

Results and Discussion
E. Coli Data



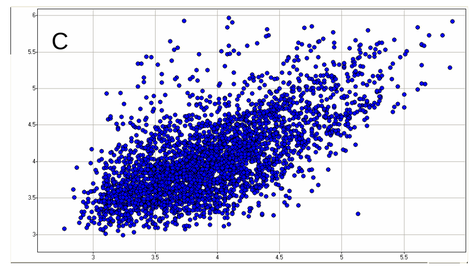


Figure 10
The three panels above illustrate the result of comparing E. Coli grown on glucose with those grown on Acetate. Data points indicate log intensity of compared peptides in each sample. Panels A and B show how different injections of the same lysate digest compare with each other. The straight diagonal line shows little variation in intensity between replicate injections. Panel C is the result when cells grown in acetate are compared with those grown in glucose. The wide dispersion of peptide intensities, as would be expected when E. Coli are grown under conditions that involve a different set of metabolic pathways.

Spiked Serum Data

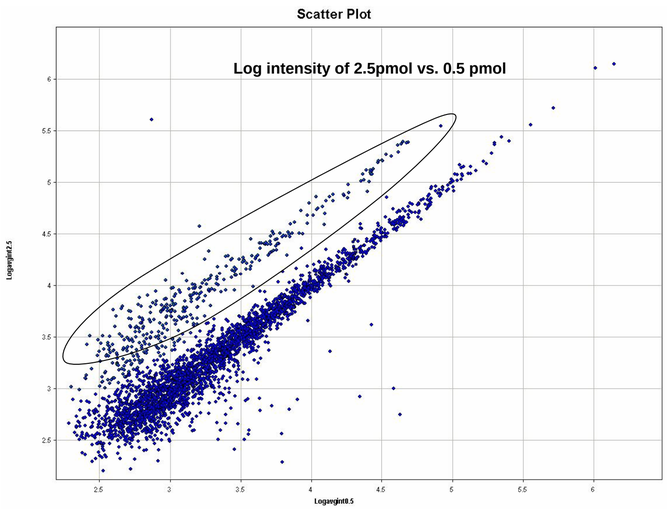


Figure 11
This plot is the comparison of two of the different spiked serum samples. Serum spiked with 2.5 pmol of protein standard is plotted on the vertical axis, and serum spiked with 0.5 pmol protein standard is spiked on the horizontal axis. The circled points off the diagonal axis correspond to the peptides from the differentially spiked proteins. The accurate masses of these peptides can be used to identify the proteins alone, or they can be combined with the high energy data for further confirmation as shown below.

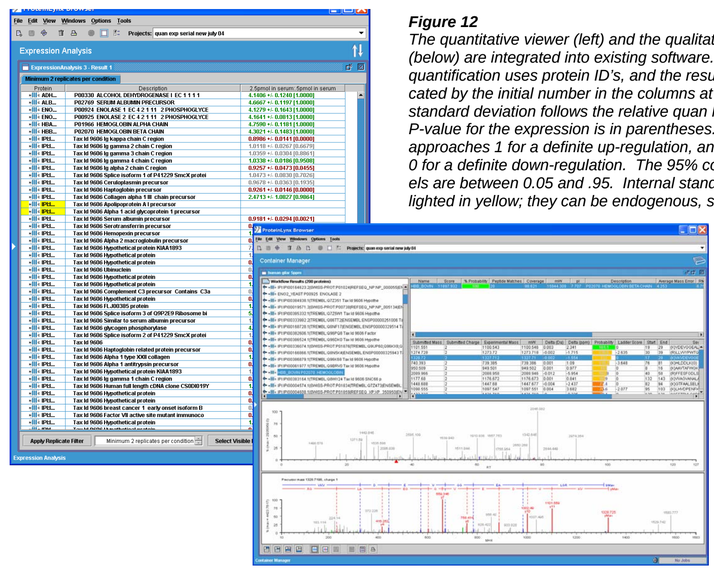


Figure 12
The quantitative viewer (left) and the qualitative viewer (below) are integrated into existing software. Relative quantification uses protein ID's, and the results are indicated by the initial number in the columns at left. The standard deviation follows the relative quan level, and the P-value for the expression is in parentheses. A P-value approaches 1 for a definite up-regulation, and approaches 0 for a definite down-regulation. The 95% confidence levels are between 0.05 and .95. Internal standards are highlighted in yellow; they can be endogenous, spiked or both.

Summary

The protein digestion protocols employed in this study provide quantitatively reproducible results even on complex protein mixtures as complex as human serum.

The analytical protocols employed in this study are capable of reproducibly measuring the intensity of peptides in complex protein digest mixtures over three to four orders of magnitude.

Waters Protein Expression System Informatics is capable of extracting peptide accurate mass, chromatographic retention time, and intensity information from complex protein digest mixtures in a quantitatively reproducible manner.

The analytical protocols employed in this study demonstrate that the combination of exact mass and chromatographic retention time can provide a very unique signature for each peptide contained in a complex protein digest mixture.

Waters Protein Expression System Informatics uses peptide exact mass and retention time signature information to match and quantitatively compare the intensity of identical peptides contained in protein digestion mixtures of a comparable control and experimental state.

Comprehensive data sets are generated which can be extensively mined for further qualitative and quantitative information.