ASMS 2004

# Protocols to Assure Reproducible Quantitative and Qualitative Analysis of Tryptic Digests of Complex Protein Mixtures for Global Proteomics Experiments

Craig A. Dorschel, Marc V. Gorenstein, Guo-Zhong Li, Jeffrey C. Silva, Scott Geromanos, and Timothy Riley
Waters Corporation, Milford, MA, USA

## Overview

**Purpose:** Demonstrate reproducibility of tryptic digestion of complex protein mixtures for the purpose of quantitative and qualitative proteomics.

**Methods:** Replicate digestions of rat serum were carried out using a recommended practice. LC/MS analysis was performed, again following a recommended practice in a novel alternating low and elevated collision energy exact mass MS mode of acquisition.

**Results**: A high degree of precision of signal intensity, mass measurement, and chromatography was obtained, demonstrating the reproducibility of the digestion.

## Introduction

To obtain meaningful results in qualitative and quantitative proteomics, such as observation of different levels of expression of one or more proteins in a set of samples and identification of the proteins, it is necessary that great care be taken in preparation of samples and in obtaining exact mass LC/MS data. Digestions with proteases such as trypsin should be carried to completion, so that the ratio of concentrations of fragment peptides will accurately reflect the ratio of concentration of the original protein in two samples. Accurate measurement of peptide masses allows more stringent searching of protein databases resulting in faster and more accurate qualitative identification of proteins. Ionization of peptides should be consistent from run to run for valid quantitative comparisons. Maximizing the reproducibility of chromatographic retention times simplifies the overall task of correlating the massive data sets obtained in global proteomic experiments.

We have developed protocols and software tools to enable us to reproducibly digest complex protein mixtures, routinely obtain precise mass measurements on peptides (typically ≤ 5 ppm error), minimize run-to-run retention time drift at the capillary separation scale, and demonstrate the quality and reproducibility of our results. We illustrate this with digestion and analysis of replicate samples of rat serum using the Waters® Protein Expression System.

## Methods

### Digests

Replicate digests of 5 µL (250 – 300 µg protein) aliquots of rat serum (Sigma, catalog number R 9759) were prepared according to a recommended procedure employing a proprietary detergent and in which cystines are reduced and alkylated with iodoacetamide prior to digestion.

### Samples

Seven samples were prepared by combining 75 µL of each of seven replicate rat serum digests and 25 µL of a standard mixture of purified tryptic digests of five proteins. .

### Data Collection

- 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

Three replicate injections of 2.5 µL of each digestion sample solution were made, directly onto the column (no desalting trap was used). Mass spectral data were collected in alternating low and elevated collision energy using a 1.8 second acquisition at each energy. The NanoLockSpray source was switched every 10 seconds to obtain a scan of the accurate mass standard ([Glu$^1$]-Fibrinopeptide B).
The raw data were processed using functions in ProteinLynx™ Global SERVER and Protein Expression System Informatics software tools developed in-house (**Poster TPR 354**, Li et al., "A Novel Algorithm to Track Ions or Peptides Found by LC/MS in Multiple Injections of Multiple Complex Biological Samples"). Additional data analysis was performed using Microsoft® Excel and Spotfire® DecisionSite™ .

### Recommended Procedure for Tryptic Digestion of Sera

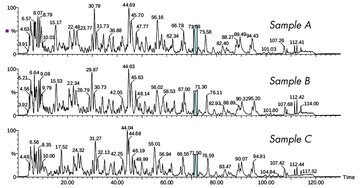
1. Dispense 5 µL of serum (250 – 300 ug protein) in a capped microcentrifuge tube, volume greater than 200 µL.
2. Add 20 µL of 100 mM aqueous[1] ammonium bicarbonate.
3. Add 25 µL of 0.2% solution of RapiGest™ SF in water (add 500 µL of water to a 1mg vial of RapiGest™ SF) and vortex.
4. Place tube in a block heater set at 80° C. Heat for 15 minutes, vortexing periodically.
5. Remove from block and centrifuge to return condensate to bottom of tube.
6. Add 2.5 µL of 100 mM dithiothreitol (15.4 mg/mL in water), vortex.
7. Place tube in a block heater set at 60° C. Heat for 30 minutes.
8. Remove tube from block, allow to cool to room temperature, and centrifuge to return condensate to bottom of tube.
9. Add 2.5 µL of 200 mM iodoacetamide (55.5 mg/mL in water), vortex.
10. Place sample in dark at room temperature and allow 30 minutes reaction time.
11. Add 50 µL of a solution of Promega trypsin in 50 mM ammonium bicarbonate, vortex. (Add 400 µL of 50 mM ammonium bicarbonate to one 20 µg vial of trypsin. Trypsin:protein ratio will be approximately 1:100.)
12. Place tube in a block heater set at 37° C. Incubate overnight.
13. Centrifuge to return condensate to bottom of tube.
14. Add 100 µL of water to the sample, vortex. Final sample concentration is equivalent to 1.25– 1.5 µg protein/µL . Final RapiGest SF concentration will be 0.025%, and the RapiGest is not hydrolyzed. Sample should be refrigerated for short term storage, frozen for long term storage.

[1] Water used in this procedure was purified by a Milli-Q A10 system (Millipore Corporation).

## Results

Tryptic digest of a serum sample can be expected to produce a highly complex mixture of peptides. Waters Protein Expression System Informatics tools have been developed to extract the exact mass, retention time and intensity of each detected peptide or chemical entity in each data set. This exact mass and retention time information serves as a unique signature for each peptide. Additional software matches like peptides from multiple sample injections based on their unique exact mass retention time signature and compares their relative intensity. Peptides matched in this manner are referred to as an exact mass retention time pair or EMRT.
Representative base peak intensity chromatograms from 3 of the 7 samples (**Figure 1**) give a rough indication of the complexity of the samples and the reproducibility of the chromatography.

**Figure 1.** BPI for 3 Samples




The highlighted peak in the chromatograms represents peptide T73 of rat albumin, having sequence TVMGDFAQFVDK. The spectra in **Figure 2** represent the mass and intensity of the doubly charged precursor in each sample. The insert spectrum is the corresponding elevated energy spectrum in each case. Note that the precursor intensity and fragmentation pattern, showing nearly the entire y'' series, are virtually identical. The experimental exact mass of the doubly charged precursor is within 10 ppm of the accurate mass (679.3265).

Selected ion chromatograms for 6 masses in one injection of each sample gives another overview of sample similarity (**Figure 3**).

**Figure 2.** Rat Albumin T73 in 3 Samples

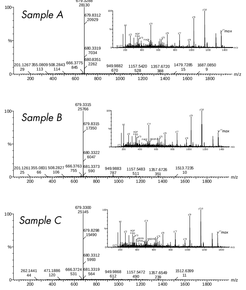
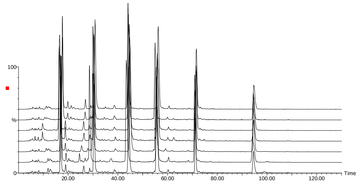
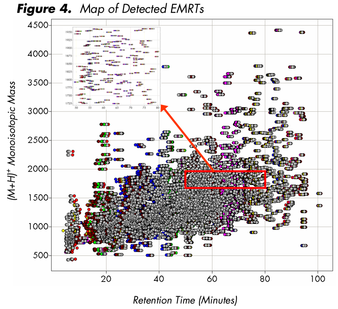



**Figure 3.** Overlaid Selected Ion Chromatograms (6 ions) for 1 Injection of Each Sample




The true complexity of the sample can be appreciated by observing the map of all EMRTs detected in 21 injections (**Figure 4**). The insert is an enlargement of the indicated region.

**Figure 4.** Map of Detected EMRTs




Use of the informatics tools allows a more detailed examination of the data. In 21 injections, a total of 136,117 EMRT components were detected of which 12491 unique EMRTs were found to replicate in at least two or more injections. **Figure 5** illustrates the replication rate of these EMRTs across the 21 sample injections. **Figure 6** illustrates the intensity represented by these replicating EMRTs across the 21 sample injections.

**Figure 5.** EMRT Replication

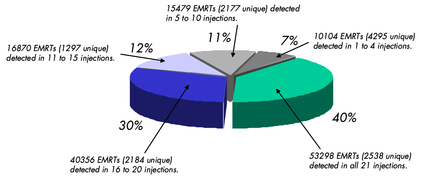
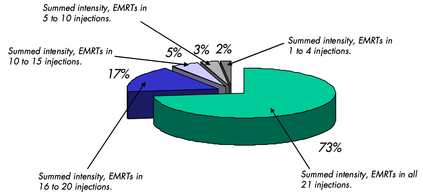


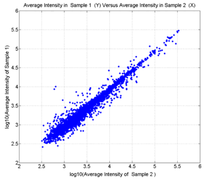

**Figure 6.** EMRT Intensity Fraction




EMRTs detected in 16 to 21 of the 21 injections account for 90% of the summed MS signal intensity for all the EMRTs. The low intensity of the less frequently detected EMRTs suggests these are near the limit of detection or are random events of chemical noise.

Off-diagonal plots correlating intensities of individual EMRTs in pairs of samples show a strong correlation in each instance. **Figure 7** is a representative example for sample 1 and sample 3.

**Figure 7.** Correlation of EMRT Intensity in Two Samples




The coefficients of variation for the intensities of each replicating EMRT were calculated, first for each sample (3 injections) and then for the entire experiment (21 injections). Within a sample (**Figure 8**), CV values are low, and reflect the fact that injection volume and ionization efficiency are very reproducible. CV values for the entire experiment (**Figure 9**) include such factors as sampling of the serum and digestion efficiency. These CV values also indicate excellent reproducibility. (Note that, for example, an EMRT found in all 21 injections will contribute 7 values to **Figure 8** but only one value to **Figure 9**.)

**Figure 8.** Intensity CV Frequency for EMRTs in Samples

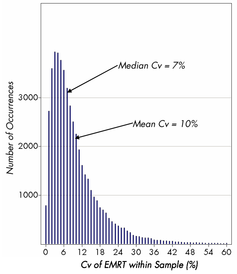
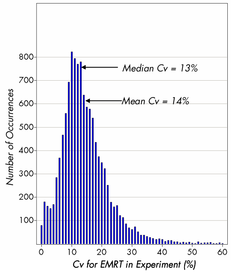
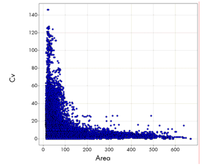



**Figure 9.** Intensity CV Frequency for EMRTs in Experiment




The highest EMRT intensity CV values are associated with EMRTs of low intensity (**Figure 10**).

**Figure 10.** CV vs. Intensity (Area)




Mass measurement precision of replicating EMRTs is excellent. A great majority of such measurements have a precision within ± 5 ppm (**Figure 11**). The histogram of summed EMRT intensity for each mass precision bin (**Figure 12**) shows that the least precise mass measurement is associated with trace components.

**Figure 11.** EMRT Mass Precision Frequency

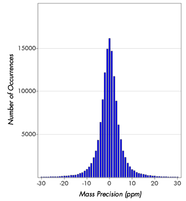
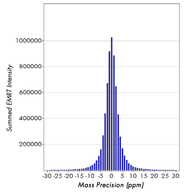



**Figure 12.** Mass Precision/ EMRT Intensity Sum




While great care was taken to assure that the gradient composition was consistent from injection to injection, some run-to-run variability in retention time was observed. The effect of even small compositional "ripples" in the gradient profile is a shift in retention time of all peptides eluting at that composition. The informatics tools used to correlate EMRTs from injection to injection are able to accurately account for this retention time shift, in effect greatly reducing the retention time variability in the data set and increasing the confidence in the correlated data. This is explored in much greater depth in **Poster TPR 354**.

## Conclusions

Through the use of the Waters Protein Expression System and Informatics tools, we have demonstrated that:

- Tryptic digestion of complex protein mixtures (sera) can be reproducibly performed such that peptides produced vary in intensity with coefficients of variation typically well below 30%.
- Electrospray ionization efficiency of peptides is reproducible from run to run and sample to sample.
- Exact masses of peptides can be determined with precision typically less than 5 ppm throughout a series of samples.


720000894EN May 2004