A DETAILED STUDY ON THE SAMPLE LOADING REQUIREMENTS FOR QUALITATIVE AND QUANTITATIVE LC-MS BASED PROTEOMICS

Craig A. Dorschel, Martha D. Stapels and Scott J. Geromanos
Waters Corporation, Milford, MA

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

INTRODUCTION

The goal of many proteomic experiments is the identification of proteins from complex biological mixtures. In many such experiments, particularly LC-MS/MS experiments, chromatography systems have been intentionally overloaded in the belief that a significant increase in the number of protein identifications will be obtained.

We have previously reported (ASMS 2007) that the number of proteins identified and the extent of sequence coverage observed when a dilution series of an *E. coli* digest standard is analyzed by LC-MS^E, a data-independent alternating low and elevated collision energy acquisition mode, and a novel ion accounting databank search algorithm rises sharply with increasing sample load until an optimum is reached. Injection of further material gives a small increase in sequence coverage but few additional protein identifications.

Here we further consider the effects of the quantity of protein digest loaded on the LC column, the chromatographic behavior of the peptides and how this affects quantitative and qualitative proteomics.

METHODS

Sample, Data Collection, and Data Processing

The Waters Identity^E High Definition Proteomics system was used to acquire and process the data.

A dilution series of a standard *E. coli* digest was prepared such that 2 µL injection aliquots contained 100 ng, 200 ng, 500 ng, 1000 ng, 2000 ng, and 5000 ng, respectively, based upon the amount believed to be in the vial. Each aliquot contained 500 fmol of yeast alcohol dehydrogenase digest as a quantitation standard. Injections were made directly onto a 150 µm x 10 cm chromatography column packed with 1.7 µm diameter C_{18} particles. (nanoACQUITY BEH, Waters Corp.) Peptides were eluted by a 110 min. gradient from 1% to 40% acetonitrile in 0.1% formic acid (nanoACQUITY UPLC, Waters). Three injections were made at each load level.

Data were collected using a QTof Premier (Waters) mass spectrometer operating in MS^E mode, collecting alternate low and elevated collision energy spectra each 700 msec. A reference mass spectrum ([glu]^1-fibrinopeptide B) was collected every 30 seconds.

Data were processed with Identity^E informatics. The minimum criteria for identification of a protein from an *E. coli* database with an equal number of random sequences appended were 3 fragment ions per peptide, 7 fragment ions per protein and 1 peptide per protein. A false discovery limit of 4% was set based on identification of random protein sequences.

RESULTS

Search Results

The number of proteins identified at each loading level is summarized in the table below.

Amount Loaded (ng)	Number of *E. coli* Proteins in 2 or 3 Injections	Number of *E. coli* Proteins in 1 Injection Only	Total *E. coli* Proteins Identified
5000	366 (1.1% FDR)	103	469
2000	338 (0.6% FDR)	107	445
1000	284 (1.0% FDR)	109	393
500	153	34	189
200	58	9	67
100	16	8	24

As in the previous report, the number of identified proteins is beginning to plateau at the 2000 ng load level, suggesting that the optimum loading for the 150 µm is in the 2000 to 5000 ng range.

At the three lowest levels, the databank search was exhausted before any random protein sequences were identified. At the higher levels, the number of random sequences identified in more than one injection was quite low, suggesting a very low false discovery rate, as indicated in the table. The random sequences identified in any one injection at a given loading level served to terminate the search when 4% of the total identified proteins were randoms in each case. Roughly 77% of the protein identifications at each level are found in at least 2 injections.

Chromatographic Behavior of Peptides

The peptide VGEEVEIVGIK derived from Elongation Factor Tu was chosen to study the effect of loading on the chromatographic behavior of a peptide across all levels. As this is a strongly ionizing peptide from an abundant protein, it represents a worst case of peptide behavior.

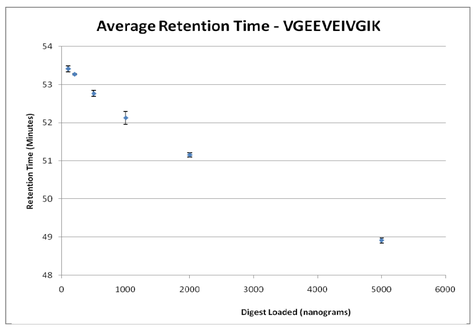

Figure 1. Retention time shift with increasing total load of digest.

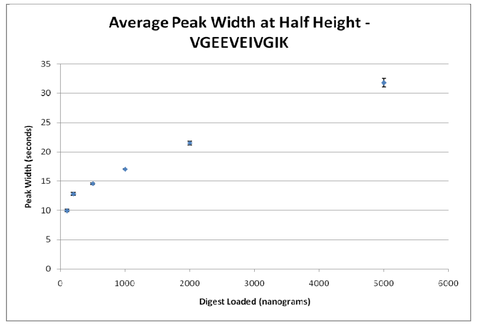

Figure 2. Peptide chromatographic peak widths from selected ion chromatogram for +2 charge state.

Figures 1 and 2 show that as the overall sample load and the load of the individual peptide increases the sample/stationary phase equilibrium is affected resulting in substantial broadening of the chromatographic peaks and a reduced retention time. Because of the peak broadening, the peak height measurements are distinctly non-linear with respect to the loading (Figure 3).

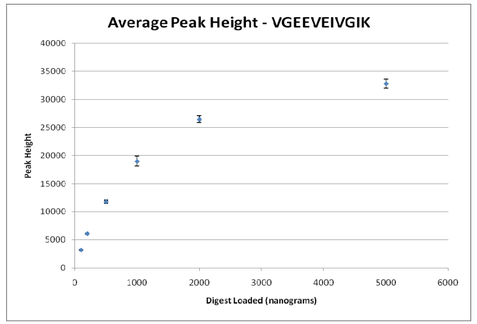

Figure 3. Peptide chromatographic peak height from selected ion chromatogram for +2 charge state.

The Identity^E software determines peptide intensities by combining all detected charge states and all naturally occurring isotopes into a single integrated value. Figure 4 shows that these intensities are linear for the VGEEVEIVK peptide (r^2 = 0.997) despite the heavy load, with an RSD no worse than 6.4% at any level.

Also presented in Figure 4 are the intensities for an equimolar, but less intensely ionizing peptide (TVGAGVVAK) from Elongation Factor Tu. RSDs are also low and linearity is excellent (r^2 = 0.9999).

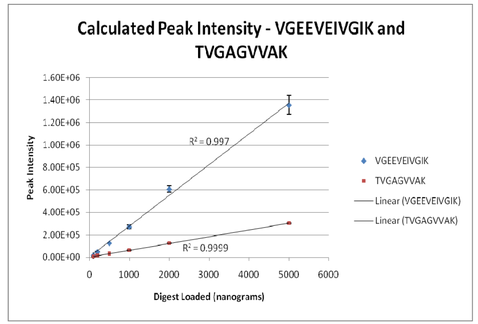

Figure 4. Peptide intensities determined by Identity^E, based on peak areas for all charge states and isotopes.

Protein Identifications at Different Load Levels

Figure 5 represents the relative concentrations of each reproducibly identified *E. coli* protein. The dots are color coded to indicate the loading level at which the protein is first identified in at least 2 of the 3 injections. While there is considerable overlap, it should be noted that only abundant proteins are first identified at a low sample load (red dots) and that the least abundant proteins (green and cyan dots) are identified reproducibly only at the highest sample loads.

These observations indicate that the experiments are "well-behaved" from an analytical perspective. If the identification of an abundant protein is first made from the injection of a high sample load; or the opposite — a scarce protein at a low sample load, then this would indicate erroneous results.

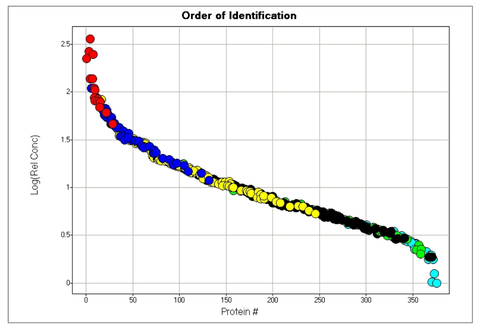

Figure 5. Relative concentrations of identified E. coli *proteins, colored by load level at which the protein was identified in at least 2 of 3 injections. Red = 100 ng, blue = 200 ng, yellow = 500 ng, black = 1000 ng, green = 2000 ng, cyan = 5000 ng.*

DISCUSSION

Retention times for peptides examined in this study are extremely reproducible at a constant sample load level, though they vary considerably when column overload occurs. Shifts in retention time for peptides as the load varies can be expected to be seen when data from samples representing multiple expression levels of a protein are compared. Software tracking peptides between such samples must be capable of detecting these retention time shifts. Such is the case with the Track 3D routine incorporated in the Expression^E Software.

Increasing peak widths will result in more peptide elution overlap, which may increase "distraction" in the data. This problem becomes more severe as sample loading on the chromatography column is increased beyond the optimum.

Because of the poor linearity, peak heights of extracted ion chromatograms are seen to be inferior to integrated peak areas incorporating all charge states and isotopic forms of the peptides of interest for the purpose of quantitation.

In summary, chromatographic quality cannot be sacrificed in a proteomic experiment without degradation in the identification and quantification of the proteins.

CONCLUSION

- **An optimum quantity sample must be applied to an LC-MS system to obtain the greatest possible number of reproducible protein identifications.**
- **The quantity of sample affects a number of chromatographic parameters:**
- **Increasing load shortens retention times.**
- **Increasing load causes peak broadening.**
- **Chromatographic peak height measurements are likely non-linear and could result in poor quantitation.**
- **Integrated peak areas are preferable for quantitation.**
- **In a properly performed experiment, a low sample load will result in identification of only abundant peptides.**

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

720002498EN