UTILITY OF AN ABSOLUTE QUANTIATION METHOD FOR OPTIMIZING ON-COLUMN PROTEIN LOAD IN QUALITATIVE AND QUANTITATIVE PROTEOMICS

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Keith Compson[2] Craig A. Dorschel[1], Jeffrey C. Silva[1,4], Johannes P. C. Vissers[2], Scott J. Geromanos[1], Krisha Panchalingam[3], Christopher Utzat[3], and James L. Sherley[3]
[1]Waters Corporation, Milford, MA; [2]Waters Corporation, Manchester, United Kingdom; [3]Massachusetts Institute of Technology, Cambridge, MA; [4]Current address: Cell Signaling Technology, Danvers, MA

OVERVIEW

The largest number of protein identifications are made when quantity of sample is consistent with column capacity.

Overloading may affect quality of quantitative results.

Absolute quantitation feature enables simple estimation of the quantity of protein present

INTRODUCTION

The best results in LC-MS proteomics experiments are obtained when an optimum quantity of protein digest is applied. Too small a quantity will limit the number of peptides detected and identified, limiting the possible number of protein identifications and quantitative comparisons between samples. An excessive quantity can cause chromatographic distortions or MS problems, such as degraded mass accuracy or distorted isotope distribution. Relative protein quantitation is generally based on the assumption that an equal mass of protein is represented in each sample prepared for analysis. While a number of methods for estimating protein concentration exist, these traditional colorimetric assays can be skewed by sample turbidity. Here we present a novel LC-MS method for estimating protein concentration and demonstrate the impact of column load on protein identifications.

METHODS

Samples

The column loading study was carried out with a standard tryptic digest of *E. coli*, diluted to 0.05, 0.1, 0.25, 0.5, 1.0, and 2.5 µg/µL respectively in 0.1% formic acid.

Lysates of adult stem cells, provided by the MIT group, were reduced (dithiothreitol), alkylated (iodoacetamide) and digested with trypsin at a concentration of 1.5 µg/µL, based on colorimetric assay. Standard digest of yeast alcohol dehydrogenase 1 was added at 250 fmol/µL as a quantitation standard.

A second digest was carried out using additional lysate as determined from the quantitation results.

Data Acquisition and Processing

Chromatography was carried out on 3 µ C_{18} columns, either 150 µ ID x 15 cm (loading study) or 300 µ ID x 15 cm (stem cells). Elution was by means of an acetonitrile/0.1% formic acid gradient. Triplicate injections were made.

Data were acquired on a hybrid quadrupole-time of flight mass spectrometer in alternating low and elevated collision energy scanning mode, utilizing a reference spray of [glu]^1-fibrinopeptide B and erythromycin.

Mass/retention time peak detection, charge state reduction, deisotoping, time alignment, databank searching, and absolute quantitation were carried out with Identity^E informatics software. Additional data analysis was performed with Microsoft Excel.

RESULTS

Results of the loading study show a marked increase in the number of observed protein identifications up to 1µg on column, after which few additional identifications are made (Figure 1). A similar increase in the number of peptide identifications is seen (Figure 2), though these serve to increase sequence coverage rather than identify additional proteins.

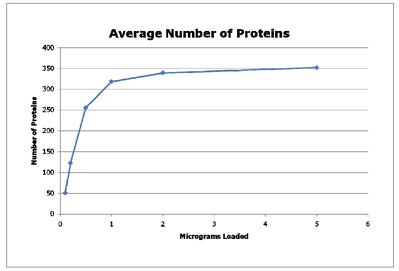

Figure 1. Number of E. coli *proteins (average of 3 injections) identified as a function of quantity of digest on-column.*

Figure 3 shows reconstructed mass chromatograms for an abundant *E. coli* protein, tryptophanase. Band broadening is evident at the higher loads, as is flattening of the peak top at the highest loads. This behavior could adversely affect quantitative results for this protein, suggesting that a column load of 1.0 µg on the 150 µ diameter column would be optimum for quantitation of abundant proteins. A 2.0 µg load might be preferred for qualitative analysis. The gross overload at 5.0 µg should be avoided.

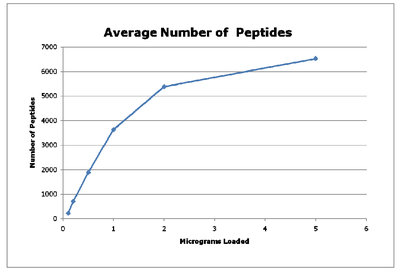

Figure 2. Number of peptides to E. coli *proteins (average of 3 injections) as a function of quantity of digest on-column.*

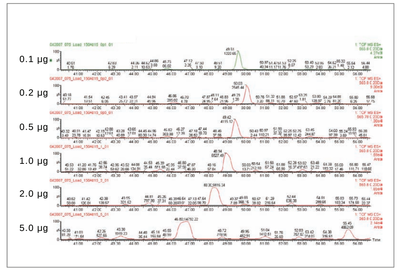

Figure 3. Reconstructed mass chromatograms for an E. coli *tryptophanase peptide at different column loads.*

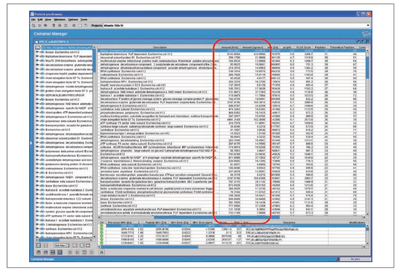

Figure 4. Absolute quantitation results (femtomoles and nanograms) displayed in the search results browser.

We have found that the absolute quantities of proteins in a complex mixture can be estimated by comparison if the averaged intensity of the three most intense tryptic peptides of a protein with those of an added standard protein of known quantity (reference 1). This algorithm has been incorporated in the latest informatics version, where the results, calculated in femtomoles and micrograms, are displayed in the protein search results table (figure 4) This algorithm may be used to estimate the amount of protein digest applied to a column, as we illustrate here in the case of adult stem cells.

The lysate of cell line 7 Xn was estimated to contain 59 mg/mL of protein based on a colorimetric method. A digest was prepared to a nominal concentration of 1.5 µg/µL, so that a 3 µL injection would fall in the expected optimum loading range for the 300 µ diameter column used. This injection volume contained 750 fmol of standard yeast alcohol dehydgogenase digest as a quantitation standard.

The results, as shown in the browser (figure 5) show that 201 proteins were identified and the calculated quantity for each is shown in the table. The data were copied into a Microsoft Excel spreadsheet in order to easily sum the mass quantities of the identified proteins, which we assume to be >95% of the total protein.

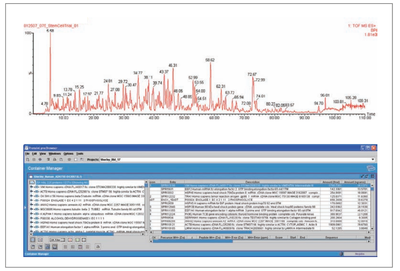

Figure 5. Base peak intensity chromatogram and results table for initial digest of cell line 7 Xn.

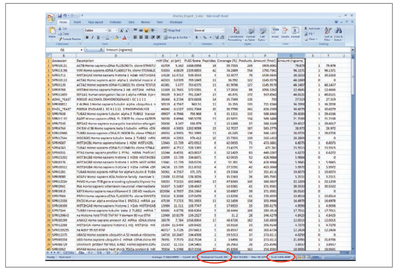

Figure 6. Absolute quantitation results for first digest of cell line 7 Xn in Microsoft Excel.

The calculated mass of 1.46 µg was substantially less than the expected 4.5 µg. In order to optimize protein detections, a second digest was prepared, using approximately 3x the quantity of lysate. The results of this second injection may be seen in figures 7 and 8.

It will be seen that the maximum of the base peak intensity chromatogram is much greater in the second digest, and that 337 proteins were identified, a 68% increase in protein identifications. The calculated quantity of protein on-column is now estimated at 6.44 µg, though this reduces to 5.26 µg if possible duplicates are eliminated.

Four additional stem cell line lysates were analyzed. Three of these were also determined to be significantly less concentrated than determined by the colorimetric method.

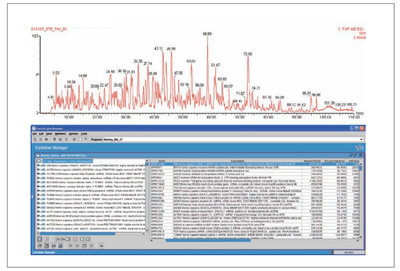

Figure 7. Base peak intensity chromatogram and results table for second digest of cell line 7 Xn.

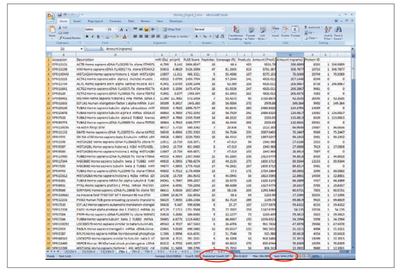

Figure 8 Absolute quantitation results for second digest of cell line 7 Xn in Microsoft Exce..

Preparing and analyzing a second digest at an appropriate concentration caused the number of protein identifications to increase by 22%, 90%, and 38%, respectively.

In circumstances where it is not feasible to prepare a second digest at optimum concentration based on a survey run, the volume of sample injected can be adjusted to provide the appropriate column load.

CONCLUSIONS

- **Chromatography column must be properly loaded with sample to obtain the maximum number of protein identifications.**
- **Increasing the sample load beyond optimum increases sequence coverage without significantly increasing new identifications, but quantitation may be adversely affected.**
- **Optimum on-column load of protein digest was found to be 1-2 ug for a 150 u diameter column.**
- **Optimum loads for 75 u and 300 u columns, by extension, are 250-500 ng and 4-8 ug, respectively.**
- **Traditional protein concentration assays may be inaccurate, overestimating the amount of protein and therefore limiting the number of protein identifications.**
- **Absolute quantitation based on an added standard digest provides a simple method to more accurately estimate the amount of protein digest injected.**
- **The more accurate estimate of protein concentration can result in a dramatic increase in the number of proteins identified.**

Reference

1. J. C. Silva, M. V. Gorenstein, G-Z. Li, J. P. C. Vissers and S. J. Geromanos, "Absolute Quantification of Proteins by LCMS^E," *Molecular and Cellular Proteomics*, **5**, 144–156 (2006).

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

720002228EN