# COMPARISON OF ALTERNATE SCANNING LC-MS DATA ACROSS MULTIPLE INSTRUMENTS FOR LARGE-SCALE PROTEOMIC STUDIES

*Craig A. Dorschel, Scott J. Geromanos, Marc V. Gorenstein, Guo-Zhong Li, and Jeffrey C. Silva*
*Waters Corporation, Milford, Massachusetts*

## OVERVIEW

**We demonstrate that it is valid to combine data collected on multiple LC-MS systems into a large-scale quantitative and qualitative proteomic study.**

## INTRODUCTION

Inevitably, future proteomic studies, especially those directed toward biomarker discovery will require analysis of a large number of samples involving perhaps hundreds of LC-MS injections. In order to complete such a study expeditiously, it would be desirable to acquire the data on more than one LC-MS system. Contemplating this, legitimate concerns regarding reproducibility of results from system to system, and the ability of data analysis software to properly correlate data from multiple systems could be expected to arise. The present study is intended to demonstrate that, given reasonable controls, data from more than one LC-MS system can indeed be combined in a single experiment.

## METHODS

**Samples:** Human Serum (Sigma) was diluted in a solution of ammonium bicarbonate and a surfactant (RapiGest SF). The serum was reduced with dithiothreitol, alkylated with iodoacetamide, and digested with sequencing grade trypsin. Standardized digests of yeast alcohol dehydrogenase (ADH) and yeast enolase (MassPrep Digest Standards) were added to portions of the serum digest. One portion (Sample 1) contained 1.25-1.50 ug/uL of serum digest and 1 pmol/uL each of ADH and enolase. The other portion (Sample 2) contained 1.25-1.50 ug/uL of serum digest, 1 pmol/ul of ADH, and 2 pmol/uL of enolase.

**Data Collection:** Data were acquired on two LC-MS systems consisting of a nanoACQUITY UPLC system and a QTof Premier mass spectrometer. Each LC system was equipped with a 300 um x 15 cm Atlantis $C_{18}$ column. Each mass spectrometer was equipped with a LockSpray source to provide accurate mass reference data, and each was tuned to maximize response, though no attempt to equalize response on the instruments was made. A total of 26 injections of 2 uL of each sample was made, sample 1 on instrument 75 and sample 2 on instrument 71. Identical chromatographic and mass spectrometric methods (alternating high/low collision energy scanning) were used on each system.

**Data Processing:** Raw MS data were processed by ion detection software (ProteinLynx Global Server 2.2.5 and other software). The resulting ion lists were searched against a database of the 200 most abundant serum proteins and the added standards, and results from all 52 injections were tracked and correlated. Further analysis and visualization of the data was carried out with Microsoft Excel and Spotfire Decision Site.

## RESULTS AND DISCUSSION

A standard amount of yeast ADH was added to each sample to allow for normalization of peptide intensities between injections and between the two systems. The normalization factor is based on the averaged intensities of the 13 identified peptides.

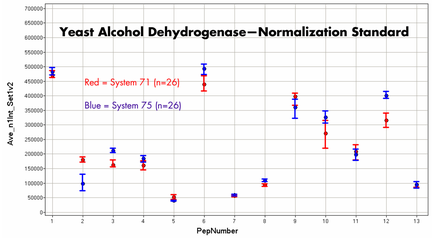
The twofold differential in the yeast enolase concentration tests the ability of the systems and software to return the expected ratio. The figures below illustrate the average intensity for each peptide identified to the standards, and their uncertainty. Data in red were acquired on system 71, those in blue on system 75.

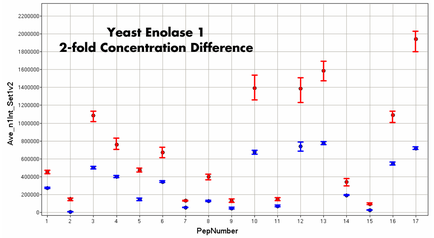
**References**

1. Silva J. C., Denny R., Dorschel C. A., Gorenstein M., Kass I. J., Li G.-Z., McKenna T., Nold M. J., Richardson K., Young P. and Geromanos S. "Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs." *Anal. Chem.* 2005 April; *77*(7):2187-2200.

2. Hughes M. A., Silva J. C., Geromanos S. J. and Townsend C. A. "Quantitative Proteomic Analysis of Drug-Induced Changes in Mycobacteria." *J. Proteome Res.* 2006 January; *5*(1):54-63.

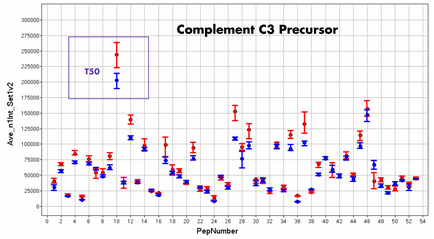
3. Silva J. C., Denny R., Dorschel C., Gorenstein M. V., Li G-Z., Richardson K., Wall D. and Geromanos S. J. "Simultaneous Qualitative and Quantitative Analysis of the E. coli Proteome: A Sweet Tale." *Mol. Cell. Proteomics* 2006 January; M500321-MCP200 MCP Papers In Press.



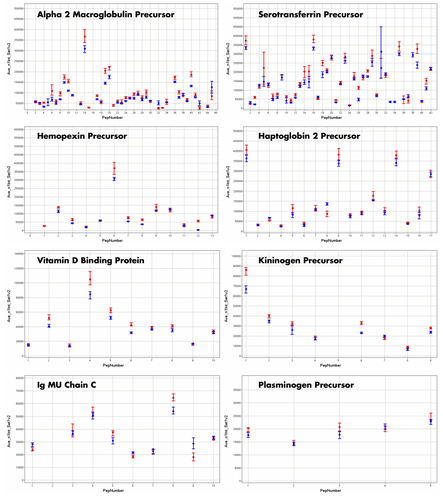




| | | System 71 | System 75 | | |
|---|---|---|---|---|---|
| | Number of | Intensity % RSD | Intensity % RSD | | |
| Protein | Peptides | (Average, Median) | (Average, Median) | Ratio | 95% CI |
| ENO1 | 17 | 13.7 (12.4) | 10.1 (6.0) | 2.35 | 0.35 |
| ADH1 | 13 | 14.3 (12.4) | 9.1 (7.0) | 1.01 | 0.11 |

Since the same serum digest was analyzed on the two systems, the normalized intensities of the peptides matched to serum proteins should be nearly identical on each system. The close tracking of the data for a large majority of the peptides of complement C3 is seen in the next figure (red = system 71, blue = system 75).



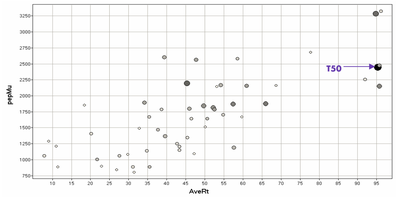

Similar results are seen for the selection of serum proteins illustrated here, including more abundant proteins with many peptides matched per protein, as well as less abundant proteins with fewer peptides.



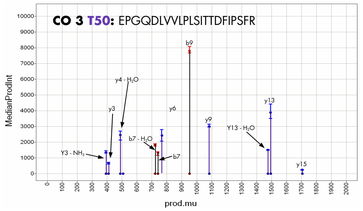

Results for a larger selection of serum proteins are summarized in the following table. The relative standard deviation of the normalized intensity for each peptide matched to the protein was determined for the set of 26 injections on each system, and the average and median of all such RSDs is reported in the table. The third column reports the average and median RSDs for that protein for all injections on both systems. The calculated ratio between the two systems (1.00 expected) and the 95% confidence interval are reported in the final columns. The overall average RSD for all proteins on both systems (18.7%) is somewhat higher than that obtained on a single instrument, but still very reasonable.

| | | System 71 | System 75 | All Data | | |
|---|---|---|---|---|---|---|
| | Number of | Average (Median) | Average (Median) | Average (Median) | S-71/S-75 | |
| Protein Name | Peptides | Intensity %RSD | Intensity %RSD | Intensity %RSD | Ratio | 95% CI |
| Complement C3 | 53 | 14.6 (12.8) | 9.6 (9.4) | 18.4 (15.3) | 1.07 | 0.08 |
| Alpha-2-macroglobulin | 45 | 14.9 (13.1) | 10.7 (7.5) | 20.1 (16.4) | 1.05 | 0.09 |
| Serotransferrin | 42 | 16.3 (13.1) | 11.9 (9.1) | 20.9 (17.6) | 1.11 | 0.14 |
| Serum albumin | 41 | 16.0 (14.1) | 9.9 (8.3) | 25.7 (25.2) | 1.18 | 0.25 |
| Alpha-1-antitrypsin | 23 | 17.4 (18.8) | 13.0 (10.7) | 23.2 (22.3) | 1.20 | 0.14 |
| Apolipoprotein A-I | 21 | 15.2 (14.0) | 13.5 (13.1) | 24.8 (23.9) | 1.41 | 0.26 |
| Complement C4 | 19 | 14.2 (12.1) | 11.7 (9.7) | 18.3 (18.0) | 1.14 | 0.17 |
| Haptoglobin-2 | 17 | 15.1 (14.5) | 10.8 (8.3) | 17.4 (14.3) | 1.01 | 0.10 |
| Ceruloplasmin | 15 | 13.0 (10.6) | 10.5 (7.8) | 17.8 (16.8) | 1.17 | 0.21 |
| Hemopexin | 13 | 13.9 (13.8) | 8.8 (6.9) | 22.9 (18.2) | 1.14 | 0.15 |
| Vitamin D-binding protein | 10 | 13.8 (12.1) | 8.0 (7.9) | 13.8 (14.8) | 1.05 | 0.08 |
| Ig MU chain | 10 | 13.5 (9.6) | 9.8 (8.3) | 14.8 (13.8) | 0.92 | 0.13 |
| Kininogen | 9 | 8.9 (8.2) | 11.2 (8.3) | 13.6 (11.7) | 1.11 | 0.13 |
| Alpha-2-HS-glycoprotein | 6 | 18.5 (18.7) | 10.6 (10.7) | 16.1 (15.0) | 0.85 | 0.18 |
| Complement factor B | 6 | 15.0 (12.0) | 9.6 (7.3) | 16.3 (17.4) | 1.11 | 0.13 |
| Plasminogen | 5 | 14.3 (13.7) | 8.8 (8.6) | 14.5 (14.9) | 0.97 | 0.09 |
| Beta-2-glycoprotein I | 5 | 14.4 (12.4) | 11.5 (10.7) | 19.7 (18.0) | 1.23 | 0.20 |
| | | | | | | |
| | | Average System 71 | Average System 75 | **Average overall** | | |
| | | 14.6 (14.6) | 10.6 (10.6) | **18.7 (18.3)** | | |

Finally, armed with the knowledge that peptides are consistently detected on multiple instruments with similar normalized intensities and chromatographic retention times easily tracked, we can construct an "ion map" of the tryptic peptides of any protein, such as the example for Complement C3 here. Peptide mass and retention are mapped on the X and Y axes, and intensity is indicated by the size and darkness of the spots.




For consistent protein identifications, it is important that the elevated collision energy fragmentation data also be reproducible. The following figure shows a representative spectrum for the indicated Complement C3 peptide in 10 injections, showing only a small variation in fragment ion intensity.




## CONCLUSIONS

On the basis of the data presented here, we conclude that data acquired on two LC-MS systems may be combined in a valid proteomic study provided:

The same model mass spectrometers and liquid chromatography instruments are used.

The same chromatography columns, mobile phases and gradients, and connections are used.

A normalization standard at a constant known concentration is added to each sample for relative quantitation.

Instruments are properly tuned, calibrated, and maintained.

These results show that accurate fold-change results may be obtained using data acquired on two instruments.

We have also shown that a consistent set of tryptic peptides corresponding to a given protein are obtained across a large number of injections on two instruments, and that consistent fragmentation data are also obtained. These results assure consistent and valid protein identifications.

720001729EN
©2006 Waters Corporation

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