A METHOD FOR ESTIMATING THE ABSOLUTE CONCENTRATION OF PROTEINS IN COMPLEX MIXTURES

Waters

Jeffrey C. Silva[1], Craig A. Dorschel[1], Martin Gilar[1], Marc V. Gorenstein[1], Petra Olivova[1], Johannes P. C. Vissers[2], Guo-Zhong Li[1], Scott J. Geromanos[1] and LeRoy B. Martin[1]

[1]Waters Corporation, Milford, Massachusetts 01757-3696, USA and

[2]Waters Corporation, Transistorstraat 18, 1322 CE Almere, The Netherlands

OVERVIEW

We describe a new method of absolute quantification of proteins based on the discovery of a relationship between MS signal response and the absolute protein concentration, wherein the average MS signal response for the three most intense tryptic peptides is constant per mole of protein.

INTRODUCTION

The study of proteins is crucial in understanding and combating disease through identification of proteins, discovering disease biomarkers, studying protein involvement in specific metabolic pathways, and identifying protein targets in drug discovery. ESI-LC-MS is an important technique used in these studies to quantify and identify peptides and/or proteins present in simple and complex mixtures. To date a majority of the proteomic analyses have been performed using various labeling strategies.[1] However, these methods require complex, time-consuming sample preparation and can be relatively expensive.

The ability to determine the absolute concentration of a protein (or proteins) present in a complex mixture is valuable for understanding the underlying molecular biology guiding the response to an applied perturbation or phenotype. A method for determining the absolute quantity of proteins in a complex sample would enable determination of the stoichiometry of proteins within a sample and would facilitate the study of complicated biological networks which rely on the cooperative protein interactions to direct appropriate cellular responses. Current strategies of determining the absolute concentration include the use of isotope labeled peptides for each target protein.[2]

We describe a novel label-free LC-MS method which not only provides simultaneous quantitation and identification of proteins among different samples[3,4,5] but also the ability to determine the absolute quantity of proteins within a sample.[6] The method describes how to obtain a single point calibration for the mass spectrometer that is applicable to the subsequent absolute quantification of all other characterized proteins within the complex mixture.

METHODS

Samples: A dilution series of five proteins (yeast enolase and alcohol dehydrogenase, bovine serum albumin and hemoglobin and rabbit phosphorylase B) was prepared in the presence and absence of human serum as detailed in the **Table 1**. Human serum, soluble *E. coli* and *P. aeruginosa* protein was diluted in ammonium bicarbonate containing 0.1% surfactant (RapiGest SF). The proteins were reduced with DTT, alkylated with iodoacetamide and digested with sequencing grade trypsin.

Data Collection: Data were acquired by ESI-LC-MS using a nanoACQUITY UPLC and a QTof Premier mass spectrometer. Each LC was equipped with a 300 um x 15 cm Atlantis C_{18} column. The mass spectrometer was equipped with a NanoLockSpray source to provide accurate mass reference data. Mass spectral data was collected in alternating low and elevated collision energy using a 1.8 s acquisition at each energy. A lockmass scan was obtained every 30 s to obtain a scan of the accurate mass standard ([Glu¹]-Fibrinopeptide B).

Data Processing: Raw LC-MS data was processed by the ion detection software (ProteinLynx Global Server 2.2.5 and other software). The resulting data was searched against the appropriate host protein database. Further analysis and visualization of the data was carried out with Microsoft Excel and Spotfire Decision Site.

RESULTS AND DISCUSSION

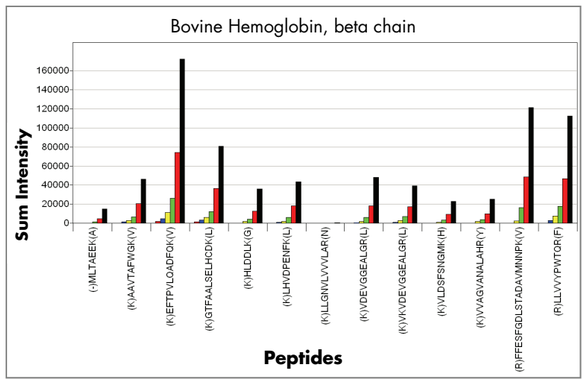

Figure 1.

Figure 1 illustrates the expected change in signal intensity through the dilution series for those peptides identified to bovine hemoglobin (beta). **Figure 2** maps each of the peptides according to their accurate mass and retention time, with spot size proportional to the signal intensity. Relative intensities remain constant through the series while overall intensity increases and new peptides are detected with increasing concentration.

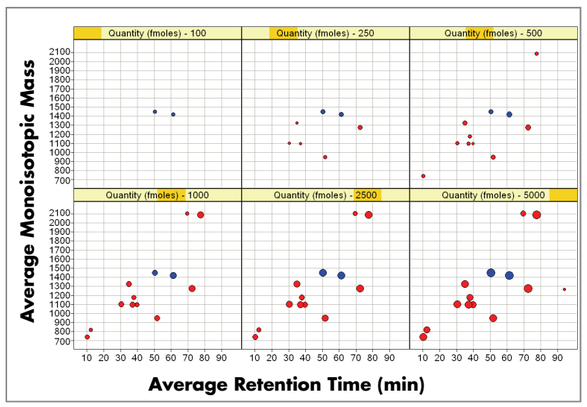

Figure 2.

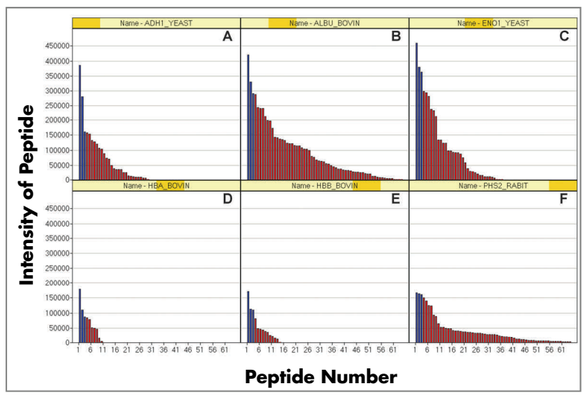

Figure 3.

Figure 3 illustrates those peptides identified to the five standard proteins in Sample 1, sorted by descending intensity. The three most intense tryptic peptides from each protein have been colored blue. The average intensities of the top three peptides are indicated in **Table 2**. Using ADH1 as the internal standard, the relative ratio of each protein has been determined from the average intensity of the top three ionizing tryptic peptides. The correlation between the relative ratio of the proteins provided an indication of the relationship between the absolute quantity of a protein and the average signal response of the three most intense tryptic peptides. Knowing that there was 10 pmoles of ADH1, the relative ratios have been converted to the absolute quantity of protein. The results from this analysis indicated that approximately 26,121 counts corresponds to 1 pmole of protein.

Protein	Average Intensity	Relative Ratio	Theoretical Amount	Calculated Amount	Error (%)	SR/pmol
ENO1	395,716	1.47	15.0	14.7	-2.2	26,381
ALB	337,505	1.25	12.5	12.5	0.1	27,000
ADH1	**269,861**	**1.00**	**10.0**	**10.0**	**0.0**	**26,986**
PB	161,116	0.60	6.0	6.0	-0.5	26,853
HBa	129,280	0.48	5.0	4.8	-4.2	25,856
Hbb	118,244	0.44	5.0	4.4	-12.4	23,649
				Average SR/ pmole		**26,121**
				% CV		**4.9 %**

Table 2.

The correlation between the average signal response of the three most intense tryptic peptides and the molar quantity of protein is valid based on the low coefficient of variation of 4.9% (**Table 2**). To further test this relationship, the peptide/protein results from the remaining samples were organized in a similar fashion to obtain the average intensity of the three most intense tryptic peptides for each of the constituent proteins as a function of their absolute concentration (**Table 1**). The average signal response for the three most intense tryptic peptides was plotted against the absolute concentration for all proteins among the six dilution series. A linear correlation was obtained from 125 to 15,000 fmoles among all the proteins as illustrated in **Figure 4**. A linear fit to the data produces an R^2 value of 0.9939. These results prove that the average MS signal response of the three most intense tryptic peptides is constant for all proteins. As further validation, the signal response from ADH1 spiked into serum was used to determine the absolute concentration of 11 securely identified serum proteins (blue circles and whiskers) from seven individual patient samples (**Figure 5**). These results were compared to the average concentration values obtained from Specialty Laboratories (red circle), which included their expected minimum and maximum values (red whiskers).

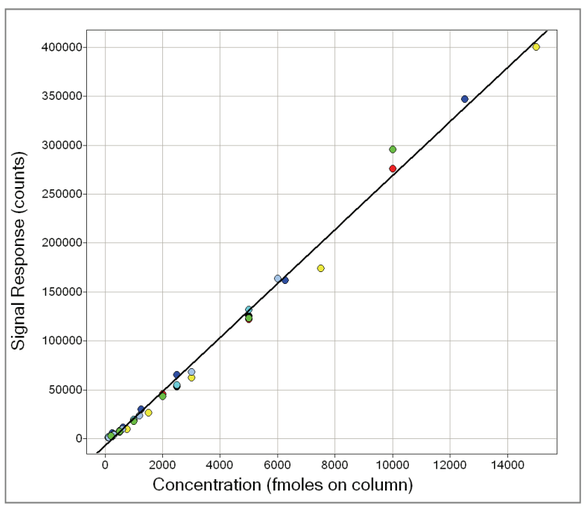

Figure 4.

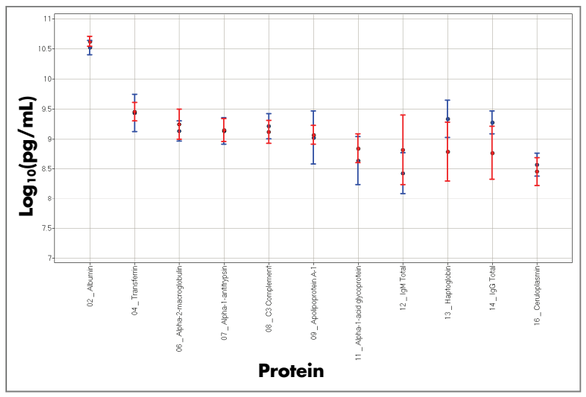

Figure 5.

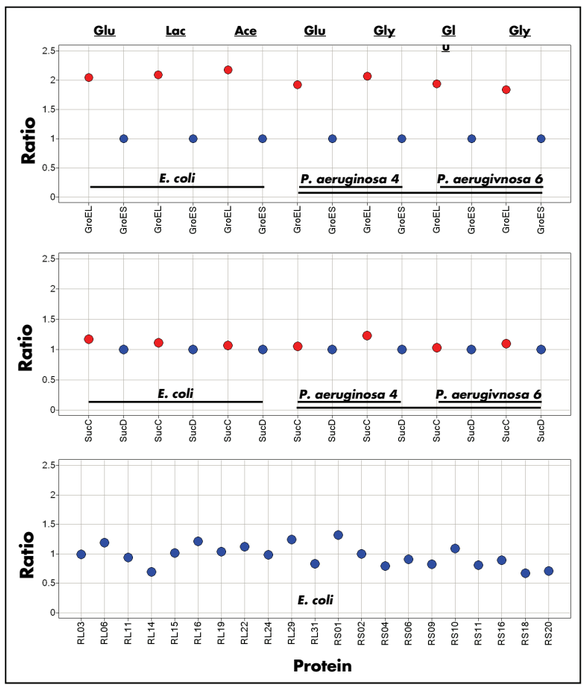

Figure 6.

Having the ability to determine the absolute quantity of proteins in a complex sample allows one to determine the stoichiometric relationship of proteins within the same sample. The top three most intense tryptic peptides to GroEL and GroES were compared among several microbes from a variety of culture conditions (Glucose, Lactose, Acetate and Glycerol) to determine the stoichiometry of the known complex within each sample (**Figure 6**, top panel). These results were consistent with the known structure of molecular chaperonin (2:1). The stoichiometry of the α and β components of succinyl-CoA synthetase (SucD and SucC, respectively) was determined as well. The results were consistent with its known heterotetrameric A2B2 (1:1) structure. The relative stoichiometry of twenty ribosomal proteins from *E. coli* was determined after normalizing their corresponding average signal response of the three most intense tryptic peptides to that of ribosomal protein, RS02. These results were consistent with the stoichiometry of known ribosomal structure (1:1).

		Percent protein coverage, number of peptides, concentration (pmoles)					
Description	**Mw**	**Sample 1**	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**	**Sample 6**
ENO1_Sc	46,624	77, 32, 15.00	62, 29, 7.50	61, 27, 3.00	51, 23, 1.50	37, 17, 0.75	22, 11, 0.38
ALB_Bov	69,248	79, 55, 12.50	79, 54, 6.25	77, 53, 2.50	69, 47, 1.25	49, 28, 0.63	31, 18, 0.32
ADH1_Sc	36,669	72, 27, 10.00	71, 26, 5.00	71, 26, 2.00	47, 20, 1.00	31, 12, 0.50	25, 8, 0.25
PB_Rb	97,097	69, 62, 6.00	67, 60, 3.00	55, 49, 1.20	34, 32, 0.60	14, 12, 0.30	2, 2, 0.15
HBa_Bov	15,044	84, 9, 5.00	84, 9, 2.50	84, 9, 1.00	59, 7, 0.50	45, 6, 0.25	10, 1, 0.13
HBb_Bov	15,944	91, 14, 5.00	82, 12, 2.50	82, 12, 1.00	71, 10, 0.50	39, 5, 0.25	17, 2, 0.13

Table 1 summarizes the protein sequence coverage and number of peptides identified to each standard protein obtained from the analysis of the sample mixtures without human serum.

CONCLUSION

The ability to determine the absolute concentration of proteins in based on the discovery that the average MS signal response for the three most intense tryptic peptides is constant per mole of protein of average molecular weight.

The label-free LC-MS method described in this work is ideally suited for determining the absolute concentration of proteins present in both simple and complex mixtures.

The ability to determine the absolute concentration of proteins in complex mixtures provides a means to study stoichiometric relationships among proteins within a sample.

References

1. Gygi S. P., Rist B., Gerber S. A., Turecek F., Gelb M. H. and Aebersold R. "Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags." Nat. Biotechnol. (1999) 17:994-999.
2. Gerber S. A., Rush J., Stemman O., Kirschner M. W. and Gygi S. P. "Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS." Proc. Natl. Acad. Sci. U.S.A. (2003) 100(12)6940-6945.
3. 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) 77(7):2187-2200.
4. 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) 5:54-63.
5. 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, MCP Papers In Press.
6. Silva J. C., Gorenstein M. V., Li G-Z., Vissers J. P. C. and Geromanos S. J. "Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel MS Acquisition." Mol. Cell. Proteomics (2006) 5:144-156.

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

720001515EN