

Absolute Quantification of Proteins by LCMS^E

A VIRTUE OF PARALLEL MS ACQUISITION*^[S]

Jeffrey C. Silva^{‡§}, Marc V. Gorenstein[‡], Guo-Zhong Li[‡], Johannes P. C. Vissers[¶],
and Scott J. Geromanos[‡]

Relative quantification methods have dominated the quantitative proteomics field. There is a need, however, to conduct absolute quantification studies to accurately model and understand the complex molecular biology that results in proteome variability among biological samples. A new method of absolute quantification of proteins is described. This method is based on the discovery of an unexpected relationship between MS signal response and protein concentration: the average MS signal response for the three most intense tryptic peptides per mole of protein is constant within a coefficient of variation of less than $\pm 10\%$. Given an internal standard, this relationship is used to calculate a universal signal response factor. The universal signal response factor (counts/mol) was shown to be the same for all proteins tested in this study. A controlled set of six exogenous proteins of varying concentrations was studied in the absence and presence of human serum. The absolute quantity of the standard proteins was determined with a relative error of less than $\pm 15\%$. The average MS signal responses of the three most intense peptides from each protein were plotted against their calculated protein concentrations, and this plot resulted in a linear relationship with an R^2 value of 0.9939. The analyses were applied to determine the absolute concentration of 11 common serum proteins, and these concentrations were then compared with known values available in the literature. Additionally within an unfractionated *Escherichia coli* lysate, a subset of identified proteins known to exist as functional complexes was studied. The calculated absolute quantities were used to accurately determine their stoichiometry. *Molecular & Cellular Proteomics* 5:144–156, 2006.

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 (1, 2). An important technique that is used in these studies to quantify and identify peptides and/or proteins present in simple and complex mixtures is ESI-LCMS.

From the [‡]Waters Corporation, Milford, Massachusetts 01757-3696 and [¶]Waters Corporation, Transistorstraat 18, 1322 CE Almere, The Netherlands

Received, July 26, 2005, and in revised form, September 23, 2005
Published, MCP Papers in Press, October 11, 2005, DOI 10.1074/mcp.M500230-MCP200

To date a majority of the quantitative proteomic analyses have been performed using stable isotope labeling strategies such as ICAT (3), iTRAQTM (4), SILAC (stable isotope labeling by amino acids in cell culture) (5), and ¹⁸O labeling (6, 7). These methodologies require complex, time-consuming sample preparation and can be relatively expensive.

Recently there have been numerous reports applying label-free methods to monitor the relative abundance of protein between different conditions (8–11). Relative quantification provides information regarding specific protein abundance changes between two conditions caused by an induced perturbation (environment-induced, drug-induced, and disease-induced). These studies require comparison of identical proteolytic peptides in each of the two experiments to accurately determine relative ratios of the particular protein(s) of interest. Relative abundance values for each peptide to a given protein can then be obtained to quantitatively characterize the differential expression of proteins between different sample states. Many of these methods are based on determining the ratios of the peak area of identical peptides between different conditions. One critical factor limiting the quantitative reproducibility of these methods includes the ability to efficiently cluster the detected peptides. This in turn relies on the accuracy of the mass measurement and the chromatographic reproducibility. Although relative quantification monitors changes in protein abundance between two conditions, it does not determine the absolute quantity of these proteins.

The ability to determine the absolute concentration of a protein (or proteins) present within a complex protein mixture is valuable for the understanding of the underlying molecular biology guiding the response to an applied perturbation. Cellular responses are often controlled through direct and indirect interactions of proteins present in the cell. These coordinated interactions allow the cell to communicate a response across many cellular compartments. The cell can thereby execute an efficient and expeditious recruitment and production of critical proteins needed for adaptation. 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 understanding of the complicated biological network of cooperative protein interactions that guide cellular responses.

To date a technique capable of determining the absolute concentration of proteins in complex mixtures from a simple LCMS analysis without using specific internal standards for

each protein has not been described. Recently Ishihama *et al.* (12) reported an emPAI¹ value that the authors suggest is directly proportional to the protein content in a protein mixture. The authors reported a quantitative deviation of 63% from the actual abundance using the described emPAI method. The method describes the correlation between the number of observed peptides to a protein and its absolute amount. As the amount of protein increases, the number of observed peptides to the protein also increases. This method is useful within a narrow protein concentration range whereby the observed peptides continue to increase linearly as a function of the amount of protein. However, once a higher protein concentration is reached and all the observable peptides have been identified, the relationship deteriorates to an asymptotic limit. Additionally this method relies on characterizing peptides using traditional data-directed MS/MS and may therefore be sensitive enough only to quantify the more abundant proteins present in a mixture.

A more traditional approach to determine the absolute concentration of a protein (or proteins) in a complex mixture involves the use of stable isotope-labeled peptides spiked into the mixture. This allows direct correlation between the stable isotope-labeled peptide and its naturally occurring analog (13). Kuhn *et al.* (13) have carried out this stable isotope dilution strategy by using synthetic ¹³C-labeled peptides and multiple reaction monitoring as an analytical method for pre-screening candidate protein biomarkers in human serum prior to antibody and immunoassay development.

Typically absolute quantification of proteins requires the use of one or more external reference peptides to generate a calibration-response curve for specific polypeptides from that protein (*i.e.* synthetic tryptic polypeptide product). The absolute quantification of the given protein is determined from the observed signal response for the specific polypeptide in the sample relative to that generated in the calibration curve. If the absolute quantification of a number of different proteins is to be determined, separate calibration curves are necessary for each specific external reference peptide for each protein. Absolute quantification allows one not only to determine changes between two conditions but also to perform quantitative protein comparisons within the same sample.

Gerber *et al.* (14) describe a conventional technique for absolute quantification of proteins and their corresponding modified states in complex mixtures using a synthesized peptide as a reference standard. The reference peptide is chemically identical to one of the naturally occurring tryptic peptides of a given protein. The reference standard is introduced to a complex mixture. The mixture is analyzed using LCMS to measure the corresponding signal intensity for the derivative

peptide along with the endogenous peptide. This intensity signal response is compared with an intensity calibration curve created using the introduced synthetic molecule to determine the amount of the endogenous protein in the mixture. A disadvantage with using synthetic peptides is that extra steps are required to synthesize an authentic sample and to later “spike” the synthetic standard prior to being able to determine the absolute quantity of the protein itself. To perform the absolute quantification for a number of proteins within a mixture would require one to provide a synthetic standard for each protein of interest.

Another technique for absolute quantification of proteins uses radiolabeled amino acids such as [³⁵S]methionine, whose specific activity is known (15, 16). In this type of experiment, an amino acid, such as [³⁵S]methionine, is incorporated in the culture medium of the growing cell(s). As proteins are synthesized, [³⁵S]methionine is incorporated into the cellular proteins. Based on the extent of incorporation of the radiolabel, the absolute amount of the peptide or protein can be determined. These types of experiments are costly, require good standard operating procedures and specialized quantitative techniques, and can also be deleterious to the subject under study. Consequently determining absolute quantification of proteins using radiolabel techniques is limited to expendable biological systems such as microbes, plants, and cell cultures.

In this work we describe a method that provides absolute quantification of proteins from LCMS data of simple or complex mixtures of tryptic peptides without requiring the use of numerous external reference peptide(s) or the implementation of radiolabeling methods. 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.

EXPERIMENTAL PROCEDURES

Preparation of Simple Protein Mixture—A 50 pmol/μl stock of each predigested protein was prepared in 50 mM ammonium bicarbonate (pH 8.5) with 0.05% RapiGest™ to assist in the redissolving the lyophilized tryptic peptides (17). The samples were incubated at 60 °C for 15 min to redissolve the lyophilized samples. The stock solutions of the individual protein digests were used to prepare a single stock solution of the six standard digested proteins (yeast enolase and alcohol dehydrogenase, rabbit glycogen phosphorylase, and bovine serum albumin and hemoglobin). The single stock solution of the six predigested proteins was prepared in 50 mM ammonium bicarbonate with 0.05% RapiGest such that each protein was present at the following concentration: glycogen phosphorylase B, 2.4 pmol/μl; hemoglobin, 4.0 pmol/μl; alcohol dehydrogenase, 4.0 pmol/μl; serum albumin, 5.0 pmol/μl; and enolase, 6.0 pmol/μl. Six additional protein samples were prepared from a dilution series of the following stock solutions in ammonium bicarbonate with 0.05% RapiGest: 2-, 5-, 10-, 20-, and 50-fold. Each sample was diluted with an equal volume of 50 mM ammonium bicarbonate (pH 8.5) to reduce the concentration of RapiGest to 0.025%. The simple protein mixtures were centrifuged at 13,000 rpm for 10 min, and the supernatant was transferred into an autosampler vial for peptide analysis via LCMS^E.

¹ The abbreviations used are: emPAI, exponentially modified protein abundance index; CapLC, capillary LC; PLGS, ProteinLynx Global Server; Cv, coefficient of variation; RSD, relative standard deviation; SR, signal response, peak intensity.

Preparation of Simple Protein Mixture in Human Serum—An additional protein digest stock solution of the six standard proteins was prepared in human serum (~1.5 $\mu\text{g}/\mu\text{l}$ of total serum protein) containing 0.05% RapiGest such that each protein digest was at the following concentration: glycogen phosphorylase B from rabbit, 2.4 pmol/ μl ; hemoglobin from a cow, 4.0 pmol/ μl ; alcohol dehydrogenase from yeast, 4.0 pmol/ μl ; serum albumin from a cow, 5.0 pmol/ μl ; and enolase from yeast, 6.0 pmol/ μl . Six additional samples were prepared from a dilution series of the following stock solution in human serum with 0.05% RapiGest: 2-, 5-, 10-, 20-, and 50-fold.

Preparation of Human Serum for Biological Replicates—Human serum was prepared from seven individuals using BD Biosciences Vacutainers™ with clot activator as suggested by the manufacturer. A total of 300–400 μg of total serum protein (5 μl) was digested according to the procedure outlined under “Protein Digest Preparation.” After digesting the serum proteins, 1 pmol of purified tryptic enolase digest was added to each sample prior to LCMS analysis.

Media and Escherichia coli Growth Conditions—Frozen *E. coli* (ATCC10798, K-12) cell stocks were streaked onto Luria-Bertani (LB) plates and grown at 37 °C. An individual colony was subsequently streaked onto a plate of M9 minimal medium supplemented with 0.5% sodium acetate and grown at 37 °C. Seed cultures were generated by transferring single colonies into flasks of M9 minimal medium supplemented with 0.5% sodium acetate. Seed culture flasks were shaken at 250 rpm at 37 °C until midlog phase ($A_{600} = 0.9\text{--}1.1$). The seed culture was diluted 1 ml:500 ml into separate M9 minimal medium supplemented with 0.5% glucose. Flasks were shaken at 250 rpm at 37 °C until midlog phase ($A_{600} = 0.9\text{--}1.1$). The *E. coli* cell cultures were harvested by centrifugation, and pellets were frozen at –80 °C. Frozen cells were suspended in 5 ml/1 g of biomass in lysis buffer (Dulbecco’s phosphate-buffered saline + 1:100 protease inhibitor mixture (Sigma catalog number 8340)) in a 50-ml Falcon tube. The cells were lysed by sonication in a Microson XL ultrasonic cell disrupter (Misonix, Inc.) at 4 °C. The cell debris were removed by centrifugation at 15,000 $\times g$ for 30 min at 4 °C, and the resulting soluble protein extracts were diluted to 5 mg/ml with lysis buffer, dispensed into 50- μl aliquots (250 μg), and stored at –80 °C for subsequent analysis. The *E. coli* protein extract was spiked with tryptic peptides from yeast enolase to a final concentration of 400 fmol/ μl before storing at –80 °C.

Protein Digest Preparation—A 100- μl aliquot of the human serum samples was reduced in the presence of 10 mM dithiothreitol at 60 °C for 30 min. The protein was alkylated in the dark in the presence of 50 mM iodoacetamide at room temperature for 30 min. Proteolytic digestion was initiated by adding modified trypsin (Promega) at a concentration of 75:1 (total protein to trypsin, by weight) and incubated overnight at 37 °C. Each digestion mixture was diluted to a final volume of 200 μl with 50 mM ammonium bicarbonate (pH 8.5) to reduce the concentration of RapiGest detergent to 0.025%. Each sample was analyzed in triplicate. The tryptic peptide solution was centrifuged at 13,000 rpm for 10 min, and the supernatant was transferred into an autosampler vial for peptide analysis via LCMS. The LCMS^E analysis was performed using 5 μl of the final peptide mixture.

Approximately 250 μg (50 μl) of *E. coli* protein was suspended in a final volume of 100 μl containing 50 mM ammonium bicarbonate (pH 8.5) and 0.05% RapiGest. The protein mixture was reduced and alkylated as described above.

HPLC Configuration—Capillary LC (CapLC) of tryptic peptides was performed with a Waters CapLC system equipped with a Waters NanoEase™ Atlantis™ C₁₈, 300- $\mu\text{m} \times 15\text{-cm}$ reverse phase column. The aqueous mobile phase (mobile phase A) contained 1% acetonitrile in water with 0.1% formic acid. The organic mobile phase (mobile phase B) contained 80% acetonitrile in water with 0.1% formic acid.

Samples (5- μl injection) were loaded onto the column with 6% mobile phase B. Peptides were eluted from the column with a gradient of 6–40% mobile phase B over 100 min at 4.4 $\mu\text{l}/\text{min}$ followed by a 10-min rinse of 99% mobile phase B. The column was immediately re-equilibrated at initial conditions (6% mobile phase B) for 20 min. The lock mass, [Glu¹]fibrinopeptide at 100 fmol/ μl , was delivered from the auxiliary pump of the CapLC system at 1 $\mu\text{l}/\text{min}$ to the reference sprayer of the NanoLockSpray™ source. All samples were analyzed in triplicate.

Mass Spectrometer Configuration—Mass spectrometry analysis of tryptic peptides was performed using a Waters/Micromass Q-TOF Ultima API. For all measurements, the mass spectrometer was operated in V-mode with a typical resolving power of at least 10,000. All analyses were performed using positive mode ESI using a NanoLockSpray source. The lock mass channel was sampled every 30 s. The mass spectrometer was calibrated with a [Glu¹]fibrinopeptide solution (100 fmol/ μl) delivered through the reference sprayer of the NanoLockSpray source. Accurate mass LCMS data was collected in an alternating, low energy (MS) and elevated energy (MS^E) mode of acquisition. The spectral acquisition time in each mode was 1.8 s with a 0.2-s interscan delay. In low energy MS mode, data was collected at a constant collision energy of 10 eV. In elevated MS^E mode, collision energy was ramped from 28 to 35 eV during each 1.8-s data collection cycle. One cycle of MS and MS^E data was acquired every 4.0 s. The R_F applied to the quadrupole mass analyzer was adjusted such that ions from m/z 300 to 2000 were efficiently transmitted, ensuring that any ions observed in the LCMS^E data less than m/z 300 were known to arise from dissociations in the collision cell.

Data Processing and Protein Identification—The continuum LCMS^E data were processed and searched using ProteinLynx Global Server (PLGS) version 2.2. Protein identifications were obtained by searching either a human database to which data from the six standard proteins were appended or an *E. coli* database. The ion detection, clustering, and normalization were processed using PLGS as described earlier (8). Additional data analysis was performed with Spotfire Decision Site version 7.2 and Microsoft Excel.

RESULTS AND DISCUSSION

Method for Absolute Quantification—The method for absolute quantification of proteins requires that a known quantity of intact protein be spiked into the protein mixture of interest prior to digestion with trypsin or that a known quantity of predigested protein be spiked into the mixture after it has been digested. The average MS signal response for the three most intense tryptic peptides is calculated for each well characterized protein in the mixture, including those to the internal standard protein(s). The average MS signal response from the internal standard protein(s) is used to determine a universal signal response factor (counts/mol of protein), which is then applied to the other identified proteins in the mixture to determine their corresponding absolute concentration. The absolute quantity of each well characterized protein in the mixture is determined by dividing the average MS signal response of the three most intense tryptic peptides of each well characterized protein by the universal signal response factor described above.

Analysis of Peptides from the Standard Six-protein Mixture—The serial dilutions of the six standard protein digests were analyzed using an alternate scanning mode of data acquisition (LCMS^E) described previously by us (8). The proc-

TABLE I
Protein coverage from the LCMS^E analysis of the six-protein mixture

These results reflect those peptides that replicated in at least two of the three replicate injections.

Description	Molecular weight	Percent protein coverage, number of peptides, concentration ^a					
		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Enolase	46,624	77, 32, 15.00	62, 29, 7.50	61, 27, 3.00	51, 23, 1.50	37, 17, 0.75	22, 11, 0.30
Serum albumin	69,248	79, 55, 12.50	79, 54, 6.25	77, 53, 2.50	69, 47, 1.25	49, 28, 0.63	31, 18, 0.25
Alcohol dehydrogenase I	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
Phosphorylase B	97,097	69, 62, 6.00	67, 60, 3.00	55, 49, 1.20	34, 32, 0.60	14, 12, 0.30	2, 2, 0.12
Hemoglobin (α)	15,044	84, 9, 5.00	84, 9, 2.50	84, 9, 1.00	59, 7, 0.50	45, 6, 0.25	10, 1, 0.10
Hemoglobin (β)	15,944	91, 14, 5.00	82, 12, 2.50	82, 12, 1.00	71, 10, 0.50	39, 5, 0.25	8, 1, 0.10

^a Concentration is reported as picomoles loaded onto a 300- μ m column.

TABLE II
Identified peptides to P02070 bovine hemoglobin (β chain) (5 pmol) obtained from the LCMS^E analysis of the stock protein mixture in Table I

The table includes the average mass, retention time, and intensity measurements calculated from the triplicate analysis. The residues in parentheses are the N- (left) and C-terminal (right) residues flanking the tryptic peptide as determined from the intact protein sequence.

MH ⁺ ^a	Mass RSD	Rt ^a	Rt RSD	Intensity ^a	Int RSD	Actual MH ⁺	Δ	Sequence
							ppm	
740.3977	1.0	10.12	2.4	35,792	23.8	740.3943	4.6	(K)HLDDLK(G)
821.4150	0.9	12.35	3.4	14,798	16.4	821.4079	-9.3	(-)MLTAEK(A)
950.5161	1.9	51.64	0.7	46,047	8.9	950.5099	6.5	(K)AAVTAFWGK(V)
1,097.5346	0.5	39.59	1.4	22,656	12.4	1,097.5301	4.0	(K)VLDSFSNGMK(H)
1,098.5621	0.9	37.00	1.3	43,353	8.6	1,098.5583	3.4	(K)LHVDPENFK(L)
1,101.5625	0.9	30.26	1.0	47,977	5.4	1,101.5541	7.6	(K)VDEVGGEALGR(L)
1,177.6773	0.4	37.84	1.5	25,346	14.1	1,177.6805	-2.7	(K)VVAGVANALAH(Y)
1,265.8287	3.9	94.01	0.7	623	29.7	1,265.8309	-1.8	(K)LLGNLVVVLR(N)
1,274.7292	0.9	72.38	0.9	112,603	2.2	1,274.7261	2.4	(R)LLVYPWTQR(F)
1,328.7156	0.8	34.91	1.4	39,273	13.1	1,328.7174	-1.4	(K)VKVDEVGGEALGR(L)
1,422.7315	0.8	61.32	1.0	172,329	2.4	1,422.7269	3.2	(K)EFTPLQADFQK(V)
1,448.6856	0.3	50.32	1.2	80,876	4.7	1,448.6844	-1.1	(K)GTFAALSELHCDK(L)
2,089.9633	1.1	77.59	0.9	110,291	3.4	2,089.9541	4.4	(R)FFESFGDLSTADAVMNNPK(V)
2,105.9592	0.6	69.59	1.0	11,327	3.6	2,105.9490	4.8	(R)FFESFGDLSTADAVM*NNPK(V)
Average	1.1 ^b		1.3 ^b		10.6 ^b		4.7 ^c	

^a The relative standard deviation was determined from the replicate mass, retention time, and intensity measurements of each detected peptide. The mass accuracy is reported for each identified peptide. An overall relative standard deviation is reported for the mass RSD, retention time RSD (rt RSD), and intensity RSD (int RSD).

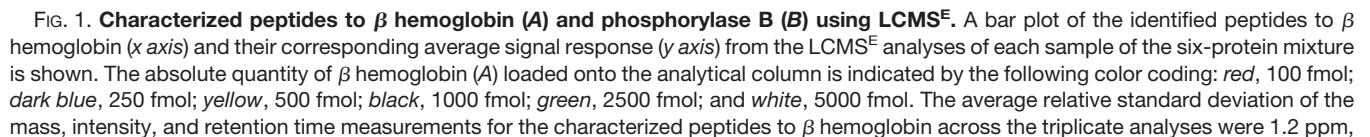
^b An overall root mean square error is reported for all the identified peptides to bovine beta hemoglobin.

^c Oxidized methionine is denoted as M*.

essing software is capable of generating properly integrated peptide signal intensity measurements (deisotoped and charge state-reduced) and accurately mass-measured, peptide ion lists that are used for subsequent qualitative identifications and relative quantification across 3–4 orders of magnitude dynamic range in ion detection. Because this mode of data acquisition does not bias the LCMS analysis by gas phase preselection of candidate peptide precursors, an inventory of all the peptide components (precursor and fragments) above the limit of detection of the instrument is produced. The time-resolved mass measurements provide the ability to properly align detected fragment ions with their respective precursor ions for accurate identification.

The protein sequence coverage obtained from the PLGS-processed LCMS^E data of the six protein mixtures is outlined in Table I. The total amount of standard protein loaded onto the 300- μ m column ranged from 100 to 15,000 fmol (scaling

the analysis to a 75- μ m chromatography system would correspond to ~6–900 fmol of total protein for the analysis of the same sample after diluting 16-fold, $1/R^2$). These detection limits are within acceptable levels for existing technologies. A minimum limit of 100 fmol of a single protein was determined for the purpose of this analysis using a 300- μ m chromatography system and a standard nanoelectrospray source at 5 μ l/min. At 100 fmol of a single protein, the top three most intense tryptic peptides to most of the proteins could be identified with a high degree of confidence. The experiment was set up so that each protein spanned a 50-fold dynamic range with an overall dynamic range of ~2.2 orders of magnitude for the entire set of protein standards. The protein sequence coverage of each protein is shown to decrease in a concentration-dependent manner, as one would expect, ranging from 84 to 2% throughout the entire data set. Replicate analysis of each sample produced signal intensity meas-



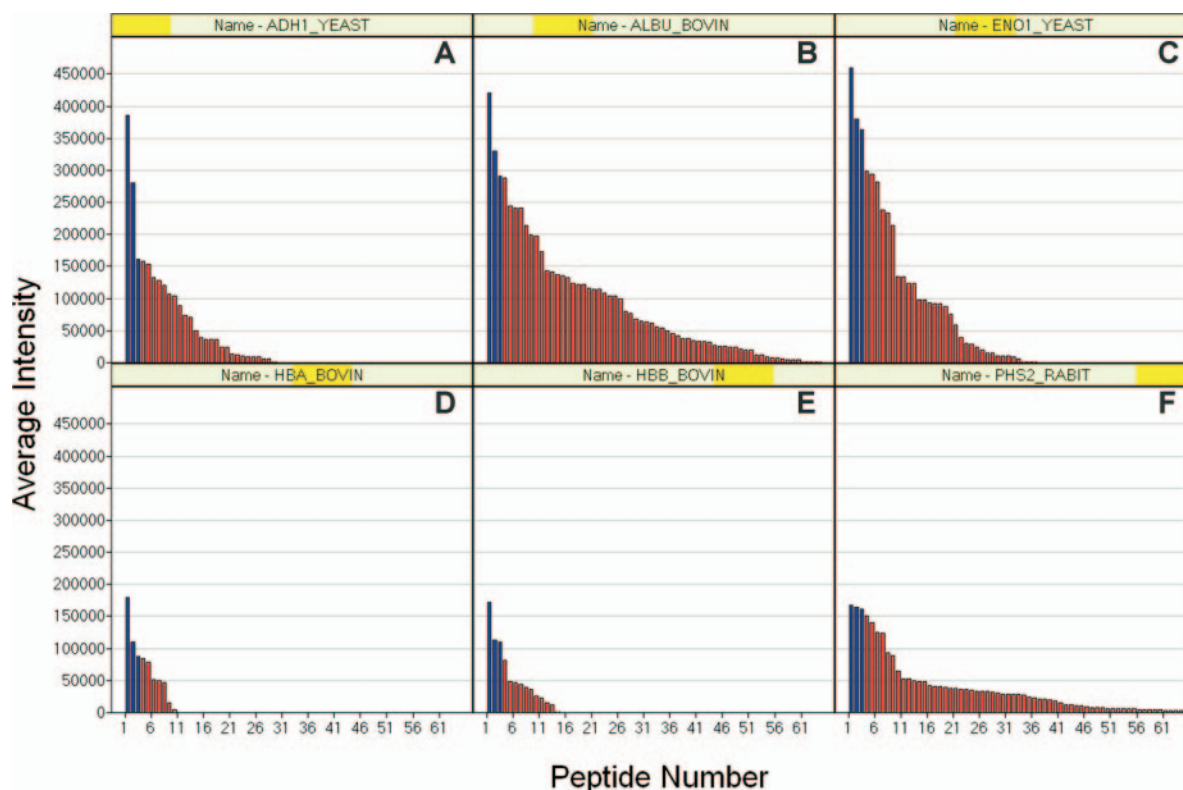


FIG. 2. Signal responses for peptides to the standard proteins. A 5- μ l injection of the following six standard proteins was analyzed by LCMS: glycogen phosphorylase B (6 pmol, 97 kDa) from rabbit, hemoglobin (10 pmol, 31 kDa total, α (15 kDa) and β (16 kDa)) from a cow, alcohol dehydrogenase (10 pmol, 25 kDa) from yeast, serum albumin (12.5 pmol, 70 kDa) from a cow, and enolase (15 pmol, 50 kDa) from yeast. The LCMS data were processed by PLGS, and the identified peptides were organized by decreasing intensity for each protein. The bar plots illustrate the peptides identified to alcohol dehydrogenase (A), serum albumin (B), enolase (C), α hemoglobin (D), β hemoglobin (E), and phosphorylase B (F). The y axis is the average intensity of the detected peptide from the triplicate analysis, and the x axis is the peptide number (sorted by descending, average intensity). The three most intense peptides for each protein are highlighted in blue. The average signal response for the three most intense peptides was calculated from the three tryptic peptides and is shown in A, B, C, D, E, and F for each of the corresponding proteins, respectively. Three ionization efficiency tiers are labeled for yeast enolase (red bar).

measurements with a coefficient of variation (Cv) of less than 30% for each peptide and an average Cv of less than 15% for all characterized peptides. A summary of the replicate analysis of one of the protein standards, β hemoglobin, can be seen in Table II. The 14 characterized peptides to β hemoglobin provided ~91% protein sequence coverage. The replicate analyses typically produced mass measurements with a precision of less than 5 ppm (RSD). An average mass accuracy error of 4.7 ppm (root mean square) was achieved for all of the peptides to β hemoglobin. Similar levels of analytical reproducibility (mass, retention time, and signal response) were observed for the tryptic peptides from the other proteins.

Due to the parallel mode of data acquisition, all tryptic peptides that are of sufficient intensity will produce associated product ion fragmentation data that can be used for

structural identification of that peptide and its corresponding parent protein. From this analysis, the MS data from the tryptic peptides of a protein and the associated sequence information from the MS^E data for each peptide produce a comprehensive analysis for each protein present in the mixture. Fig. 1A illustrates those peptides identified to β hemoglobin (14 kDa) found in each of the six dilutions. An interesting observation obtained from the comprehensive coverage of β hemoglobin shows that the pattern of intensity measurements of the characterized peptides remains preserved throughout the dilution series. As β hemoglobin decreases in concentration, the total number of observed tryptic peptides and their corresponding signal responses decrease in a predictable fashion while maintaining a consistent relative intensity pattern for the remaining tryptic peptides. Fig. 1B illus-

16.7%, and 1.5%, respectively. The median relative standard deviation of the mass, intensity, and retention time measurements across the triplicate analyses were 0.9 ppm, 11.7%, and 1.1%, respectively. The absolute quantity of phosphorylase B (B) loaded onto the analytical column is indicated by the following color coding: red, 120 fmol; dark blue, 300 fmol; yellow, 600 fmol; black, 1200 fmol; green, 3000 fmol; and white, 6000 fmol. Similar statistics were observed for phosphorylase B.

TABLE III

Summary of the absolute quantification results obtained from the analysis of the six-protein mixture described in Table I

Alcohol dehydrogenase (bold) was used as the internal reference standard in both studies as discussed in the text.

Relative ratio	Average SR of top three peptides	Protein	Theoretical	Calculated	Error	SR/pmol
			<i>pmol</i>	<i>pmol</i>	%	
1.47	395,716	Enolase	15.0	14.7	−2.2	26,381
1.25	337,505	Serum albumin	12.5	12.5	0.1	27,000
1.00	269,861	Alcohol dehydrogenase	10.0	10.0	0.0	26,986
0.60	161,116	Phosphorylase B	6.0	6.0	−0.5	26,853
0.48	129,280	Hemoglobin (β)	5.0	4.8	−4.2	25,856
0.44	118,244	Hemoglobin (α)	5.0	4.4	−12.4	23,649
	269,861	Normalization		Average SR/pmol Cv		26,121 4.9

trates those peptides identified to phosphorylase B (97 kDa) found in each of the six dilutions, showing a similar behavior throughout the dilution series. The identified peptides to the other proteins in the sample also illustrate the same phenomenon.

The characterized tryptic peptides from each protein in Sample 1 (Table I) were sorted by descending intensity as shown in Fig. 2 (A–F) to illustrate the observed signal response for all detected tryptic peptides to each protein. The bar plot illustrates the varying signal responses obtained from the observed tryptic peptides to each protein. The comprehensive analysis of the protein samples obtained from the parallel MS acquisition provided the ability to determine a relationship between the observed signal response of the three most intense peptides of a protein and the absolute protein concentration. Alternative LCMS methods, such as data-dependent analysis or even targeted MS/MS, would only provide partial sampling of this sample and would therefore not provide the comprehensive inventory of peptides required to perform this quantitative analysis. The details of this relationship have been summarized in Table III. The average signal response of the three most intense tryptic peptides was determined for each protein. The relationship between the average MS signal response of the three most intense tryptic peptides and the absolute quantity of protein can be immediately inferred from the relative ratio of the average MS signal responses. The relative ratios of the average MS signal responses are proportional to the absolute quantities of each protein present in the sample. An average signal response of 26,121 counts was consistently associated to 1 pmol of protein on column with a Cv of ~4.9%. Because the proteins spanned a wide range of molecular masses (14–97 kDa), the relationship appears to be independent of the protein molecular mass. Alcohol dehydrogenase was treated as the internal standard protein for this analysis. The universal signal response factor (counts/mol, SR/pmol) was determined from the three most intense tryptic peptides to alcohol dehydrogenase. The absolute quantity of the remaining proteins was calculated using the normalized signal response obtained

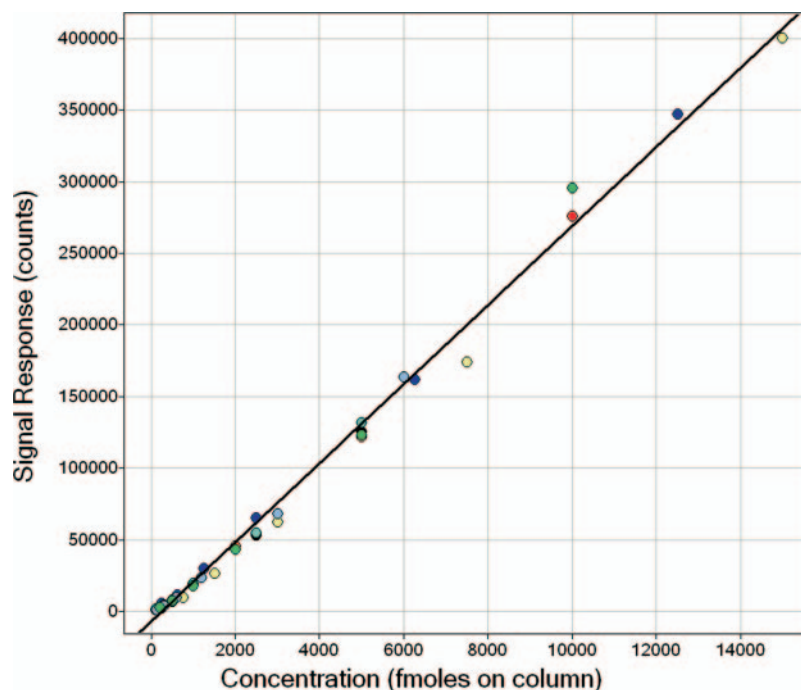
from alcohol dehydrogenase. From these results we observed that the average MS signal response for the three most intense tryptic fragments is constant per unit quantity of protein. From these observations, we propose that the average signal response for the three most intense tryptic peptides can be used to estimate the absolute quantity of other well characterized protein within the same mixture.

The six serial dilutions were analyzed by LCMS^E and subsequently processed using PLGS version 2.2. The relationship between the absolute quantity of protein loaded onto the column and the average signal response of the three most intense tryptic peptides from each of the six proteins was used to generate the results illustrated in Fig. 3. The data points from the dilution series of the proteins were determined to lie on a linear response curve ranging from 100 to 15,000 fmol (*x* axis). The average signal response of the proteins extends up to 400,000 counts (*y* axis). A linear curve 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 the proteins. Because the response curve is independent of the protein, it can therefore be used as a universal means to obtain an absolute quantity of any other well characterized protein present in the mixture.

Although six proteins represent a small sample size, the response curve illustrated in Fig. 3 illustrates a linear correlation between the three most intense tryptic peptides of each protein and its corresponding absolute concentration. Taken together with the results obtained later from both human serum proteins (Fig. 4) and *E. coli* proteins (Fig. 5), the data support the foundation of this hypothesis. The information provided by the sequences of the three most intense tryptic peptides in these studies and those obtained in future studies will further our understanding of this phenomenon. Future studies with well defined protein complexes will help identify proteins that do not fit the proposed model. A thorough analysis of the highest and lowest ionizing peptides may help define possible exceptions. It is our intent to include an explanation of the foundation of the absolute quantification

FIG. 3. Universal signal-response curve for the absolute quantification of the standard proteins.

The average signal response for the three most intense tryptic peptides to each of the six proteins was obtained for all proteins found in each of the six samples. A single scatter plot of the average signal response (y axis) and the corresponding protein concentration (x axis) was produced for all the proteins in all six samples and found to be linear for over 2 orders of magnitude. The data from each of the proteins are color-coded as follows: red, yeast alcohol dehydrogenase; dark blue, bovine serum albumin; yellow, enolase; black, bovine α hemoglobin; light blue, bovine β hemoglobin; green, bovine α/β hemoglobin; and gray, rabbit phosphorylase B. A linear curve fit was calculated for the entire data set ($y = 27.6 \times (X) - 7401$, $R^2 = 0.9939$).



method and address possible exceptions to the method in a manuscript in progress.² A list of the three most intense tryptic peptides from a subset of proteins identified in this study have been provided in Supplemental Table 1.

It is worth highlighting at this point that because the quantitation relies on correctly identifying the top three most intense tryptic peptides, larger errors will occur with smaller proteins. The results illustrated in Figs. 1 and 2 demonstrate this dependence. The magnitude of the error is dependent on the size of the protein because there are fewer peptides to choose from within the highest intensity region. Smaller proteins will have fewer tryptic peptides that may have a wide range from the most intense to the next most intense tryptic peptide. With larger proteins there are many more tryptic peptides of higher intensity so that if one (or more) of the three most intense tryptic peptides is not accurately identified, several other peptides will be found close to the same intensity, keeping the altered average intensity value close to the true average intensity value of the three most intense tryptic peptides.

Analysis of the Standard Six-protein Mixture in Human Serum—A second series of samples contained identical concentrations of the standard digest with each sample containing equivalent amounts of human serum ($\sim 8.75 \mu\text{g}$ of serum protein in a $5\text{-}\mu\text{L}$ injection). These complex protein mixtures were analyzed by LCMS^E and processed with PLGS as before to obtain corresponding protein identifications. The three most intense peptides to each of the six spiked proteins were

identified, and the corresponding average signal responses were calculated. These results are outlined in Table IV and were found to be similar to the results obtained from the analysis of the simple protein mixtures. These results again showed that the relative ratio of the average signal response for the three most intense tryptic peptides was consistent with the relative ratio of the absolute quantity of the individual proteins with the complex protein mixture. Although there was approximately a 20% decrease in the resulting signal response factor (counts/pmol), the results are internally consistent within the same dataset. The variability (Cv) of the normalized signal response per picomole increased slightly from 4.9 to 8.4% when obtained from the more complex sample. The error associated with the absolute quantification of the six standard proteins increased slightly as well. These results show that the determination of the absolute concentration of the six proteins was not affected by the additional complexity of the digested serum protein sample matrix and that the quantification method is capable of determining the absolute concentration of all properly characterized proteins present in a complex sample.

Absolute Quantification of Serum Proteins—The signal response of 20,597 counts/pmol (Table IV) was used to determine the absolute concentration of 11 identified serum proteins from the average intensity of the three most intensely ionizing tryptic peptides. The concentration of the 11 proteins was determined from a single serum sample to illustrate the variability associated with the analytical method (Fig. 4A). The 11 proteins were a subset of the 46 well characterized proteins described by Anderson and Anderson (18) and Anderson *et al.* (19). The results obtained from the replicate analysis of

² M. V. Gorenstein J. C. Silva, G. F. Li, and S. J. Geromanos, manuscript in preparation.

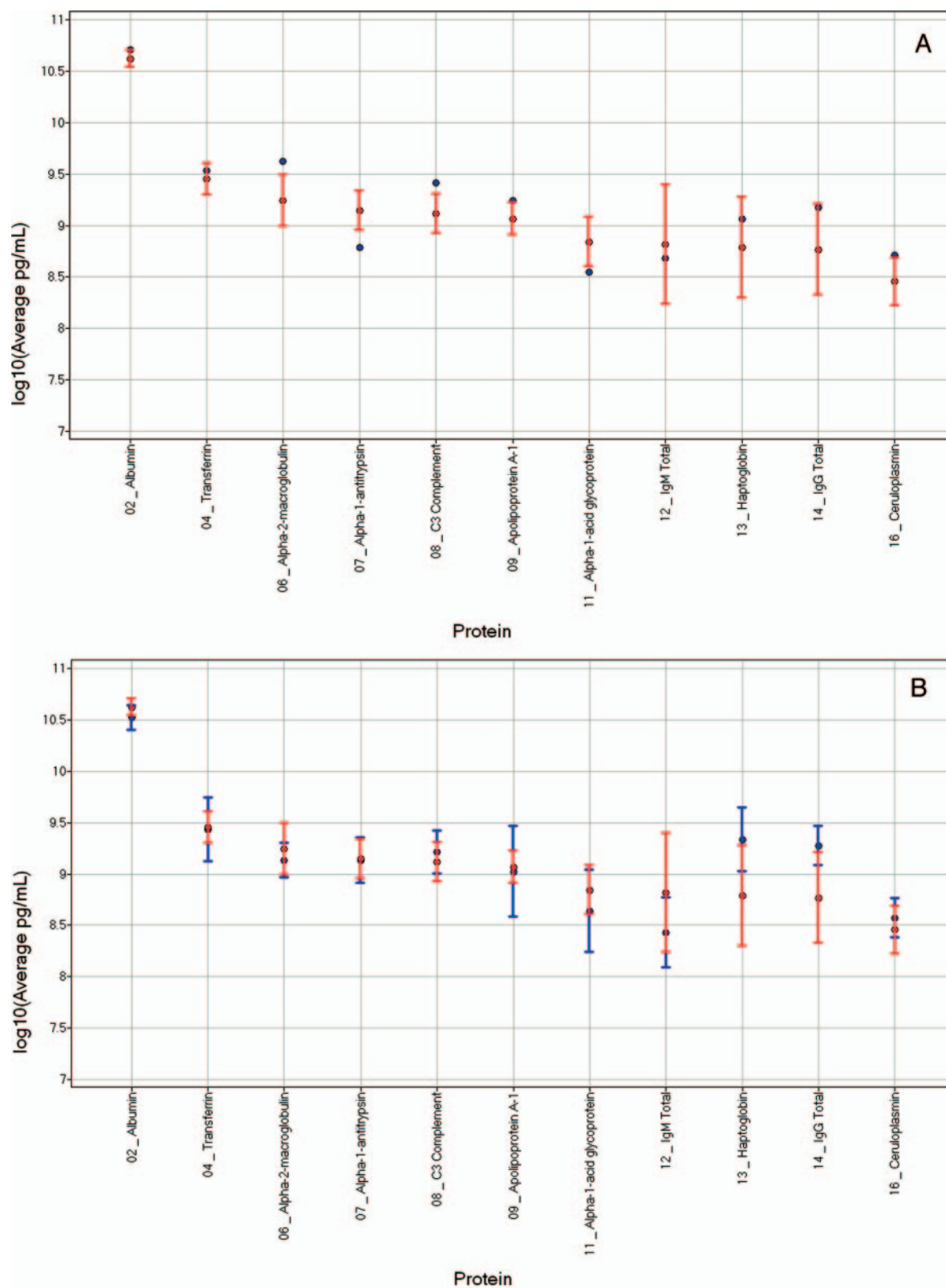


FIG. 4. **Absolute quantity of human serum proteins from analytical (A) and biological (B) replicates.** The absolute quantity of 11 well characterized human serum proteins was calculated using the described method to produce the following average concentration measurements ($\log_{10}(\text{pg/ml})$, blue circle). The average concentration values (red circle) for the human serum proteins were obtained from Specialty Laboratories and were plotted along with their expected minimum and maximum values (red whiskers). The absolute quantities obtained from

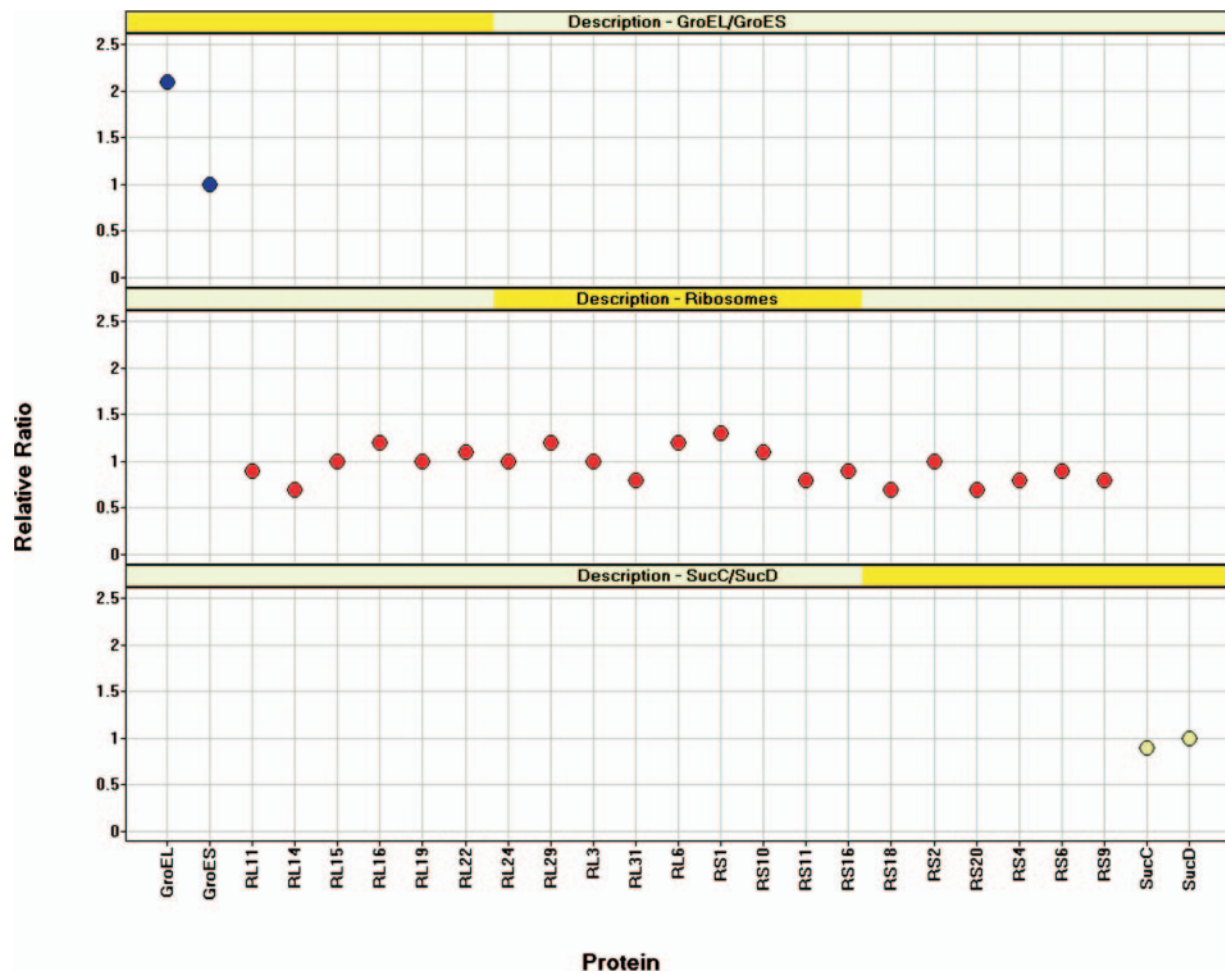


FIG. 5. **Relative levels of estimated absolute protein abundance within a single sample of *E. coli*.** The absolute quantity of protein was determined for a subset of *E. coli* proteins known to exist as multiprotein complexes. A relative ratio was calculated for each of the proteins listed in the following set of protein complexes (GroEL-GroES, ribosomes, and SucC-SucD). GroES, RS2, and SucD were used as the reference (or normalization) proteins for each of the protein complexes, respectively.

TABLE IV

Summary of the absolute quantification results obtained from the analysis of the six-protein mixture described in Table I in human serum. Alcohol dehydrogenase (bold) was used as the internal reference standard in both studies as discussed in the text.

Relative ratio	Average SR of top three peptides	Protein	Theoretical	Calculated	Error	SR/pmol
			<i>pmol</i>	<i>pmol</i>	%	
1.36	287,764	Enolase	15.0	13.6	−9.3	19,184
1.29	273,241	Serum albumin	12.5	12.9	3.3	21,859
1.00	211,572	Alcohol dehydrogenase	10.0	10.0	0.0	21,157
0.65	137,933	Phosphorylase B	6.0	6.5	8.7	22,989
0.47	100,208	Hemoglobin (β)	5.0	4.7	−5.3	20,042
0.43	91,745	Hemoglobin (α)	5.0	4.3	−13.3	18,349
	211,572	Normalization		Average SR/pmol		20,597
				Cv		8.4

the analytical replicates were typically less than 15%. Because the y axis is presented as a log₁₀ scale, the analytical error is not shown. The average (blue circle), minimum, and maximum concentration values obtained from the analysis of the biological replicates are indicated (blue whiskers) in B.

TABLE V
Absolute quantification of human serum proteins

Protein	kDa	Intensity ^a	pmol ^b	ng ^c	μg/μl ^d	pmol/ml ^e	pg/ml	log10 (pg/ml)
Albumin	70.0	1,896,530	92.08	6,445.46	51.56	736,624	51,563,664,611	10.7
α ₂ -Macroglobulin	163.0	66,801	3.24	528.65	4.23	25,946	4,229,184,056	9.6
Transferrin	77.0	113,485	5.51	424.25	3.39	44,078	3,394,026,315	9.5
C3 complement	187.0	35,593	1.73	323.15	2.59	13,825	2,585,188,523	9.4
Apolipoprotein A-I	31.0	146,274	7.10	220.15	1.76	56,814	1,761,225,033	9.2
IgG total	36.0	107,714	5.23	188.27	1.51	41,837	1,506,123,804	9.2
Haptoglobin	38.0	77,994	3.79	143.89	1.15	30,293	1,151,147,060	9.1
α ₁ -Antitrypsin	47.0	33,501	1.63	76.45	0.61	13,012	611,563,626	8.8
Ceruloplasmin	122.0	10,789	0.52	63.91	0.51	4,191	511,242,608	8.7
IgM total	49.5	25,064	1.22	60.24	0.48	9,735	481,882,993	8.7
α ₁ -Acid glycoprotein	23.5	38,420	1.87	43.84	0.35	14,923	350,680,196	8.5

^a Average signal response of the three most intense tryptic peptides.

^b The calculated mole quantity of protein loaded onto the analytical column (pmol).

^c The calculated mass quantity of protein loaded onto the analytical column (ng).

^d The calculated concentration of protein in the original sample (μg/ml).

^e The calculated concentration of protein in the original sample (pmol/ml).

the single serum sample are outlined in Table V. The analytical variability associated with these measurements was typically within a relative error of 15%. These errors were consistent with what has been described previously using this method (8). The results in Table V indicate that albumin and the immunoglobulins account for the majority of the mass of the total protein loaded onto the column (~71%) for LCMS analysis. This value is similar to the mass fraction for these proteins as determined from the concentration values provided by Specialty Laboratories (~67%). The absolute concentrations were found to be close to the expected concentration values for a number of the characterized serum proteins with four proteins outside the typical range as indicated by Specialty Laboratories.³ This variability can be attributed to the lack of proper sampling statistics from a single sample of human serum.

Another study was performed using serum from six individuals to determine the quantity of endogenous serum proteins within this small sample set. The concentration of the 11 proteins was determined from seven different serum samples to demonstrate the observed biological variability of these proteins in human serum (Fig. 4B). A plot of the calculated concentrations, log₁₀(pg/ml), of 11 well characterized human serum proteins is illustrated in Fig. 4B along with the typically observed concentration values available from Specialty Laboratories.³ Fig. 4B illustrates the range of protein concentration calculated from the six different serum samples and reflects the associated errors inherent to the analytical method and to the biological variability. These results better illustrate the correlation between the literature values obtained from Specialty Laboratories, which incorporate the

biological variability from a much larger population. The majority of the concentration values calculated on the basis of the LCMS^E results for the identified proteins in Fig. 4B were found to lie within the expected concentration ranges, providing further validation to the absolute quantification method (18, 19).³ Accumulation of data from more samples would provide a better correlation with the literature values due to the improved sampling statistics.

The absolute quantification method outlined in this work provides a means to carry out a mass balance analysis as a useful accounting mechanism that can be applied to the inventory of peptides from any given LCMS^E analysis. The total amount of protein present in human serum is ~60–80 mg/ml. Using an estimated concentration of 70 mg/ml of total serum protein, a 5-μl sample of human serum would contain ~350 μg of total protein. According to the digestion protocol described under “Experimental Procedures,” the 350 μg of total protein was digested with trypsin in a final volume of 200 μl to produce a digested protein solution containing ~1.75 μg/μl. A total of 5 μl of digest was loaded onto the chromatography column (~8.75 μg of total digested protein). The results from the absolute quantification accounts for ~10.0 μg of protein digest from the 11 identified proteins as indicated in Table V; this is in good agreement with the theoretical value.

Absolute Quantification of *E. coli* Proteins—Having the ability to determine absolute quantification of a protein allows one to determine the stoichiometric relationship of proteins within the same sample. To explore this possibility a yeast enolase protein digest of known concentration was spiked into the whole cell lysate of *E. coli*. The sample was analyzed in triplicate using the LCMS method described above. The average intensity value of the top three ionizing peptides to yeast enolase was used to convert the average intensity of the top three ionizing peptides for a number of well characterized

³ Directory of Services and Use and Interpretation of Tests, Specialty Laboratories, Santa Monica, CA (www.specialtylabs.com/default.htm).

E. coli proteins to the corresponding absolute quantity of protein loaded on column. The relative levels of the estimated absolute concentrations of a number of these proteins were found to be consistent with known quaternary structural information of these proteins. The results obtained from this quantitative assessment are outlined in Fig. 5. A number of identified ribosomal proteins were found to exist at the same relative abundance (1:1), consistent with the structure of the ribosomal complex (20). The stoichiometry of GroEL and GroES was also consistent with the known structure of the molecular chaperonin (2:1). GroEL exists as two stacked heptameric rings of 14 identical 57-kDa monomers to form a cylindrical structure. GroES exists as a single heptameric ring of seven identical 14-kDa monomers that reside at one end of the GroEL structure (21). Another example is illustrated by comparing the relative level of the estimated absolute quantity obtained for the α and β chains of succinyl-CoA synthetase (SucC and SucD). These proteins were identified, and the corresponding stoichiometry was also consistent with its known heterotetrameric A2B2 (1:1) structure (22). These examples provide additional validation to the method described in this study for the determination of the absolute concentration of proteins using the signal response of the highest ionizing tryptic peptide fragments of identified proteins.

Conclusion—The label-free method described in this work is ideally suited for determining the absolute concentration of proteins present in both simple and complex mixtures. The described method takes full advantage of the recently introduced LCMS^E mode of data acquisition and its ability to comprehensively reduce tens of thousands of ion detections to a simple inventory list of peptide precursors along with their time-resolved fragment ions. The specificity afforded by the accurate mass measurements of both the precursors and associated fragment ions (typically less than 5 ppm) provides the ability to identify, with high confidence, a large number of proteins with high sequence coverage. The ability to collect the MS data across the entire chromatographic peak width for all peptides above the limit of detection of the instrument allows for accurate quantification of peptides/proteins from the deconvoluted signal intensities (deisotoped and charge state-reduced). These three attributes of LCMS^E data acquisition in association with the correlation between the average MS signal response of the three best ionizing peptides to a protein provide a means to determine the absolute concentration of any well characterized protein present in a sample. Future studies will be performed with known complexes to further validate and understand the guiding principles of this general methodology.

Acknowledgments—We thank the valuable contributions of Timothy Riley throughout the development of this work. We also thank Dr. Sheryl Silva for collecting the serum samples and the Waters employees who donated blood for the purpose of this study. We also acknowledge Jeanne Li, Scott Berger, and Martin Gilar for contributions throughout the editing of this manuscript.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

§ To whom correspondence should be addressed: Waters Corp., 34 Maple St., Milford, MA 01757-3696. Tel.: 508-482-3005; Fax: 508-482-2055; E-mail: jeff_silva@waters.com.

REFERENCES

1. Stoughton, R. B., and Friend S. H. (2005) How molecular profiling could revolutionize drug discovery. *Nat. Rev. Drug Discov.* **4**, 345–350
2. Kramer, R., and Cohen, D. (2004) Functional genomics to new drug targets. *Nat. Rev. Drug Discov.* **3**, 965–972
3. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999
4. Ross, R. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplex protein quantification in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169
5. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386
6. Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G., Thomas, B., Wilm, M., and Mann, M. (1997) Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **11**, 1015–1024
7. Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fensalau, C. (2001) ¹⁸O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **73**, 2836–2842
8. 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. (2004) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **77**, 2187–2200
9. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **75**, 4818–4826
10. Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T. G., Foss, E., Mao, Y., and Emili, A. (2004) Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* **3**, 984–997
11. Wiener, M. C., Sachs, J. R., Deyanova, E. G., and Yates, N. A. (2004) Differential mass spectrometry: a label-free LC-MS method for finding differences in complex peptide and protein mixtures. *Anal. Chem.* **76**, 6085–6096
12. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4**, 1265–1272
13. Kuhn, E., Wu, J., Karl, J., Liao, H., Werner, Z., and Guild, B. (2004) Quantification of c-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and ¹³C-labeled peptide standards. *Proteomics* **4**, 1175–1186
14. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6940–6945
15. Sirlin, J. L. (1958) On the incorporation of methionine 35S into proteins detectable by autoradiography. *J. Histochem. Cytochem.* **6**, 185–190
16. Gupta, N. K., Chatterjee, B., Chen, Y. C., and Majumdar, A. (1975) Protein synthesis in rabbit reticulocytes: a study of Met-tRNA^{fMet} binding factor(s) and Met-tRNA^{fMet} binding to ribosomes and AUG codon. *J. Biol. Chem.* **250**, 853–862

17. Yu, Y. Q., Gilar, M., Lee, P. J., Bouvier, E. S., and Gebler, J. C. (2003) Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **75**, 6023–6028
18. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867
19. Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A. (2004) The human plasma proteome: a nonredundant list of developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326
20. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896
21. Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578–586
22. Fraser, M. E., James, M. N. G., Bridger, W. A., Wolodko, W. T. (1999) A Detailed Description of the Structure of Succinyl-CoA Synthetase from *Escherichia Coli*. *J. Mol. Biol.* **285**, 1633–1653