

Isobaric Labeling and Data Normalization without Requiring Protein Quantitation

Phillip D. Kim,* Bhavinkumar B. Patel,* and Anthony T. Yeung[†]

Developmental Therapeutics Program, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111-2497, USA

Isobaric multiplexed quantitative proteomics can complement high-resolution sample isolation techniques. Here, we report a simple workflow exponentially modified protein abundance index (emPAI)-MW deconvolution (EMMOL) for normalizing isobaric reporter ratios within and between experiments, where small or unknown amounts of protein are used. EMMOL deconvolutes the isobaric tags for relative and absolute quantification (iTRAQ) data to yield the quantity of each protein of each sample in the pool, a new approach that enables the comparison of many samples without including a channel of reference standard. Moreover, EMMOL allows using a sufficient quantity of control sample to facilitate the peptide fractionation (isoelectric-focusing was used in this report), and mass spectrometry MS/MS sequencing yet relies on the broad dynamic range of iTRAQ quantitation to compare relative protein abundance. We demonstrated EMMOL by comparing four pooled samples with 20-fold range differences in protein abundance and performed data normalization without using prior knowledge of the amounts of proteins in each sample, simulating an iTRAQ experiment without protein quantitation prior to labeling. We used emPAI,¹ the target protein MW, and the iTRAQ reporter ratios to calculate the amount of each protein in each of the four channels. Importantly, the EMMOL-delineated proteomes from separate iTRAQ experiments can be assorted for comparison without using a reference sample. We observed no compression of expression in iTRAQ ratios over a 20-fold range for all protein abundances. To complement this ability to analyze minute samples, we report an optimized iTRAQ labeling protocol for using 5 μg protein as the starting material.

KEY WORDS: proteome, iTRAQ, optimization, EMMOL, LCM

INTRODUCTION

Mass spectrometry (MS) allows a researcher to identify, quantify, and compare the proteins among samples. Popular approaches include label-free quantitation,^{2–4} isotope-coded affinity tags,⁵ stable isotope labeling by amino acids in cell culture,^{6–8} label-free quantitation,² and isobaric tags for relative and absolute quantitation (iTRAQ).^{9–11}

Although the iTRAQ reagent is recommended for protein quantities of 5–100 μg , typically identical 100 μg samples are used for each channel to allow accurate protein determination using a protein assay. The inaccuracies of protein quantitation for minute samples at the range of 5 μg and the presence of interference substances potentially can mislead the expression ratios and compromise iTRAQ comparative proteomics. Thus, we asked the question of

whether MS can serve as a protein assay during data analysis, made reliable by the accurate quantitation of the higher abundance proteins. Here, we report a novel normalization procedure, in which the exponentially modified protein abundance index (emPAI)¹ quantitation of the relative abundance of the pooled proteins is used with the MWs and iTRAQ reporter ratios to calculate the quantity of each protein in each of the four samples in the pool. Four samples of the same protein lysate varying from 5 μg to 100 μg in total protein were compared in an experiment. This new ability to calculate the quantity of each protein in each sample extends PAI- deconvolution (EMMOL) to comparing multiple separate iTRAQ datasets without the need for a reference channel, provided that one normalizes them all to the same value of total protein in each sample. iTRAQ compression was not observed in our study over a 20-fold range of protein concentrations.

To enhance iTRAQ experiments with precious samples, we report herein an optimized protocol for iTRAQ labeling reaction of small amounts of proteins in the 5- μg range. The EMMOL method is not limited to iTRAQ chemistry. It should be applicable to other methods of

*These authors contributed equally to this research.

[†]ADDRESS CORRESPONDENCE TO: Anthony T. Yeung, Fox Chase Cancer Center, 333 Cottman Ave., Philadelphia, PA 19111-2497, USA (Phone: 215-728-2488; E-mail: anthony.yeung@fccc.edu).

The authors declare no financial support or associations that may pose a conflict of interest.

doi: 10.7171/jbt.12-2301-002

isobaric-labeled quantitative MS, including Tandem Mass Tags of Thermo Scientific (Rockford, IL, USA).

MATERIALS AND METHODS

Reference Sample for Method Evaluation

Escherichia coli NovaBlue(DE3) was grown in Terrific Broth (BD Difco, Franklin Lakes, NJ, USA) for ~19 h at 35.5°C with continuous shaking. Cells (1.24 g) were collected by centrifugation at 6000 *g* and resuspended in PBS. The cells were washed twice in PBS, resuspended, and solubilized in 4 vol 1% SDS at 70°C for 10 min. The solution was centrifuged at 13,800 *g* for 30 min, and a tiny pellet was discarded. An acetone precipitation was performed by adding three times the supernatant volume of cold acetone and centrifuged at 0°C at 3000 *g* for 30 min. The pellet was washed twice each with 15 mL cold acetone and centrifuged. The final pellet was dried and solubilized in 3.8 mL 6 M urea, 2 M thiourea, and 1% 3-[(3-cholamidopropyl)dimethylammonio]1-propanesulfonic acid. The concentration of the protein was found to be 20.3 mg/mL using protein assay dye reagent (Bio-Rad, Hercules, CA, USA) and BSA standard (Bio-Rad) and diluted with the same buffer to 5 mg/mL to be used for all the studies in this report.

Tryptic Digestion of the *E. coli* Lysate

Tris (2-carboxyethyl) phosphine (TCEP; 50 mM) from the iTRAQ kit (yellow cap, Applied Biosystems, Foster City, CA, USA) was added to 500 µg *E. coli* lysate to make the protein solution 5 mM in TCEP and incubated at 30°C for 1 h. Iodoacetamide (2 µL 200 mM) was added for every 10 µL protein sample and incubated in the dark for 30 min. Six volumes of 0.5 M triethylammoniumbicarbonate at pH 8.5 (orange cap, Applied Biosystems) were added. Proteomics-grade porcine trypsin (40 µL 1 mg/mL; Sigma-Aldrich, St. Louis, MO, USA), prepared in 1 mM HCl on ice, was added to

digest the sample at 37°C for 16 h. The resulting peptide mixture was used for all of the iTRAQ reactions in this report.

iTRAQ Labeling of Peptides

The total volume of the solution was determined and divided into five tubes containing 100 µg protein each. The quantities of protein referred to in the iTRAQ labeling reactions in this report do not include the contribution by trypsin, which increased the protein present (see Table 3) by ~12.5 %.

Preparation of a Pooled Sample of Four iTRAQ Labels with 20-Fold Range of Input Proteins

Four tubes of 100 µg trypsin-digested *E. coli* lysate were each concentrated by Speed Vac to a volume of 40 µL. Ethanol (70 µL; red cap, Applied Biosystems) was added to each labeling reagent vial. Each iTRAQ label was added to a corresponding protein digest tube and allowed to incubate for 1 h, and then 100 µL water was added to each tube to quench the iTRAQ reagents. The volume of each labeled digest was determined so that 5%, 25%, 50%, and 100% of the volumes of each respective iTRAQ-labeled sample could be combined to produce the pool for iTRAQ analysis, in which the total protein labeled for the four labels varied as 5 µg (114), 25 µg (115), 50 µg (116), and 100 µg (117; Table 1), respectively. The pooled sample was concentrated by Speed Vac to ~20 µL, at which point 100 µL water was added. This sample was used in Immobilized pH gradient (IPG)-IEF fractionation¹² and liquid chromatography LC-MS/MS analysis, using the traditional two-dimensional (2D) gel IEF condition,^{12,13} as described below.

Preparation of a Pooled iTRAQ Sample of 5 µg Input Proteins from Different Labeling Conditions

The fifth tube of 100 µg *E. coli* digest was concentrated to a volume of 35 µL by using a Speed Vac. Digest (1.75 µL) and 5 µg protein were used for each of the following three

TABLE 1

iTRAQ Labeling Scheme for the Normalization Experiment

iTRAQ label	Protein quantity to be used in pooling (µg)	Approximate volume (µL) of 100 µg digested proteins after Speed Vac	Ethanol to add to labeling reaction (µL)	Approximate volume of iTRAQ labeling reagent (µL)	Water to add to quench labeling reaction (µL)	Volume of label and digest after quenching (µL)	Volume to match protein quantity desired in pooling (µL)
114	5	40	70	35	100	246	12.3
115	25	40	70	35	100	250	62.5
116	50	40	70	35	100	245	122.5
117	100	40	70	35	100	Combined all labels into this vial	Combined all labels into this vial

TABLE 2

iTRAQ Labeling Scheme for the Comparison of Four Labeling Reaction Compositions

Label	Volume to take of digest to make 5 μg protein (μL)	Dilution with dissolution buffer (μL)	Ethanol to add to label (μL)	Label volume to add after adding ethanol (μL)	Water to add to quench labeling reaction (μL)
115	1.75	38.25	70	All	100
116	1.75	0	70	All	100
117	1.75	0	70	5.08	100

Protein digest (5 μg) was used in each label reaction. Label 114 was taken from experiment in Table 1 as 12.3 μL and was combined to all the other labels at the end.

labeling reaction conditions. iTRAQ label 115: the labeling reaction was as instructed by the reagent manufacturer. Briefly, the sample was adjusted to a sample volume of 40 μL with dissolution buffer; 38.3 μL dissolution buffer (orange cap) was added to make a total volume of 40 μL (Table 2). iTRAQ label 116: to maximize fold excess of iTRAQ label and to minimize the amount of water in competition, no dissolution buffer was added. iTRAQ label 117: to retain the label:protein stoichiometry of the labeling condition for 100 μg protein, yet minimizing the water competition, no dissolution buffer and only 5% of the label reagent was added. After labeling and incubating, the above 115-, 116-, and 117-labeled products were combined with 5 μg iTRAQ 114-labeled product from the experiment of Table 1 and concentrated to ~ 20 μL , at which point, 100 μL water was added. This sample was used in IPG-IEF fractionation and LC-MS/MS analysis as described below.

IPG-IEF Fractionation of iTRAQ-Labeled Peptide Pool¹²

The removal of volatile buffer components by Speed Vac concentration and replacement with 100 μL water were repeated once more, followed by concentration and addition of 300 μL 8 M urea, 1.7 μL IPG buffer, pH 2.5–5.0 (GE Healthcare, Waukesha, WI, USA). Approximately 0.1 μL bromophenol blue saturated in water was added. The sample was applied evenly between the electrodes of an IEF tray. To focus the peptides, a 24-cm Immobiline DryStrip, pH 3.5–4.5 (GE Healthcare), was cut to discard the 6 cm from the positive end (effectively reducing the range to pH 4–4.5) and was subjected to active rehydration at 50 V for 12 h. Next, the strip was focused by a step change in voltage to 250 V for 30 min, followed by a linear change in voltage to 10,000 V for 3 h and then, by a step change to 10,000 V for a total of 90,000 V h, all performed on a Bio-Rad protean IEF cell under a layer of mineral oil. The strip was blotted to remove oil and cut into ~ 60 fractions, from which the peptides were extracted stepwise with 0.1% TFA, then 0.1% TFA in 50% acetonitrile, and then 0.1% TFA in 100% acetonitrile. During each extraction step, the gel pieces were sonicated in the solution with a Branson

1510 sonication bath and allowed to incubate for 30 min. The solution with extracted peptides was then moved to its own Eppendorf microcentrifuge tube, to which the subsequent extracted solutions were combined. The samples were concentrated, and LC-MS/MS [on an Applied Biosystems QSTAR XL time-of-flight (qTOF) mass spectrometer] was performed on each fraction, as described previously in detail,¹⁴ except that no second run using an exclusion list of the first run was performed for the current study. Data processing by Mascot 2.2 software was as described.¹⁴ To facilitate others who wish to adopt this IPG-IEF EMMOL protocol, the primary data of this study with instrument parameters and the photographs and videos of each step in IPG-IEF are available at the website (<http://yeung.fccc.edu>) of Fox Chase Cancer Center (Philadelphia, PA, USA).

Data Normalization

We present two methods of normalization of the samples of iTRAQ experiments. The first is the novel method EMMOL, which enables meta-analysis of multiple iTRAQ experiments and analysis of the proteome rather than one protein at a time. At the end of this report, EMMOL is compared with a second generic method, which we called “Summation of iTRAQ Ratios”. The Excel spreadsheets detailing the normalization process are shown in Table 3–6. The complete spreadsheets with formula are provided as Supplemental Data 1–4 on our web site (<http://yeung.fccc.edu>) and at Tranche Project (Dr. Phil Andrews, Principle Investigator, University of Michigan, Department of Biochemistry and Bioinformatics; <https://trancheproject.org>).

EMMOL Normalization

Mascot software reports iTRAQ quantitation of a protein as ratios of 114/117, 115/117, and 116/117, together with the emPAI score of the protein. This report used “peptide scores >20” and “bold red only”, searched with fixed carboxyamidocysteine, but no “variable methionine oxidation” and no “missed tryptic digestion” to avoid inflation of the emPAI scores.¹ The data are exported from the Mascot

TABLE 3

An iTRAQ Experiment with a 20-Fold Range of Protein Quantities among Four Samples Compared

5/100 µg (114/117)	25/100 µg (115/117)	50/100 µg (116/117)	SwissProt entry name	pro_t_mass	emPAI	emPAI* protein mass	180 µg total	114 protein (µg)	115 protein (µg)	116 protein (µg)	117 protein (µg)	R ²	Normalized to 100 µg total/label	sd/average CV %			
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
0.06	0.28	0.55	GRCA_ECOBW	15485	6.5	1.00E+05	1.5	0.04	0.22	0.44	0.79	1.00	0.81	0.83	0.85	0.83	1.9
0.06	0.28	0.58	RS10_ECO24	12593	4.2	5.25E+04	0.8	0.03	0.11	0.24	0.41	1.00	0.47	0.43	0.46	0.42	5.0
0.06	0.31	0.59	RL11_ECO24	17229	4.1	7.03E+04	1.0	0.03	0.16	0.32	0.54	1.00	0.58	0.61	0.61	0.56	4.7
0.05	0.25	0.50	HNS_ECOLI	17460	4.0	6.95E+04	1.0	0.03	0.15	0.29	0.58	1.00	0.51	0.54	0.56	0.60	6.2
0.06	0.28	0.50	CSPC_ECOLI	8551	3.8	3.22E+04	0.5	0.02	0.07	0.13	0.26	1.00	0.29	0.27	0.25	0.27	5.1
0.06	0.29	0.55	EFTU1_ECO24	46886	3.6	1.71E+05	2.5	0.08	0.39	0.74	1.34	1.00	1.37	1.46	1.44	1.39	2.9
0.05	0.25	0.58	RL7_ECO24	14305	3.3	4.68E+04	0.7	0.02	0.09	0.22	0.37	0.99	0.34	0.34	0.42	0.39	10.2
0.07	0.28	0.55	TALB_ECOLI	38683	3.0	1.17E+05	1.7	0.06	0.26	0.50	0.92	1.00	1.16	0.95	0.98	0.96	9.8
0.07	0.26	0.53	RL1_ECO24	28173	3.0	8.48E+04	1.3	0.05	0.18	0.36	0.68	1.00	0.86	0.67	0.70	0.71	11.7
0.05	0.27	0.52	TNAE_ECOBW	57462	3.0	1.71E+05	2.6	0.07	0.37	0.72	1.39	1.00	1.37	1.37	1.41	1.44	2.5
0.05	0.24	0.52	CSRA_ECOBW	7428	2.8	2.06E+04	0.3	0.01	0.04	0.09	0.17	1.00	0.16	0.15	0.17	0.18	6.6
0.05	0.32	0.58	AHPC_ECOLI	23024	2.4	5.57E+04	0.8	0.02	0.14	0.25	0.43	1.00	0.40	0.51	0.48	0.44	10.2
0.05	0.27	0.53	DBHA_ECOLI	11258	2.4	2.68E+04	0.4	0.01	0.06	0.11	0.22	1.00	0.19	0.22	0.22	0.23	8.5
0.06	0.27	0.52	RS4_ECO24	26538	2.3	6.18E+04	0.9	0.03	0.13	0.26	0.50	1.00	0.50	0.50	0.50	0.52	2.2
0.07	0.28	0.51	RL9_ECO24	17489	2.3	4.04E+04	0.6	0.02	0.09	0.17	0.33	1.00	0.39	0.34	0.32	0.34	8.1
0.06	0.28	0.54	RS1_ECO57	67575	2.2	1.51E+05	2.3	0.07	0.33	0.65	1.21	1.00	1.21	1.24	1.27	1.26	1.8
0.05	0.30	0.56	OMPA_ECOLI	40174	2.2	8.80E+04	1.3	0.04	0.20	0.38	0.69	1.00	0.67	0.76	0.75	0.72	5.6
0.06	0.24	0.56	ALKH_ECOLI	24458	2.2	5.36E+04	0.8	0.02	0.10	0.24	0.43	1.00	0.45	0.39	0.47	0.45	7.6
0.06	0.27	0.54	ATPB_ECOBW	53377	2.2	1.15E+05	1.7	0.06	0.25	0.49	0.91	1.00	1.07	0.92	0.96	0.95	6.6
0.05	0.27	0.57	RL3_ECO24	25112	2.1	5.27E+04	0.8	0.02	0.11	0.24	0.42	1.00	0.40	0.41	0.46	0.43	6.1
0.06	0.25	0.52	RBSB_ECOLI	34822	2.1	7.28E+04	1.1	0.03	0.15	0.31	0.60	1.00	0.60	0.55	0.60	0.62	4.6
0.06	0.27	0.56	RS3_ECO24	29426	2.0	5.80E+04	0.9	0.03	0.12	0.25	0.46	1.00	0.50	0.46	0.50	0.48	3.7
0.06	0.31	0.55	RL5_ECO24	22795	2.0	4.47E+04	0.7	0.02	0.11	0.19	0.35	1.00	0.37	0.40	0.37	0.36	4.3
0.06	0.27	0.53	G3P1_ECO57	39716	2.0	7.74E+04	1.2	0.03	0.17	0.33	0.63	1.00	0.63	0.62	0.64	0.65	2.2
0.06	0.43	0.56	RS6_ECO24	16186	1.9	3.12E+04	0.5	0.01	0.10	0.13	0.23	0.97	0.23	0.36	0.25	0.24	23.0
0.07	0.29	0.51	RL10_ECO24	19631	1.9	3.77E+04	0.6	0.02	0.09	0.15	0.30	1.00	0.39	0.32	0.30	0.31	12.3
0.06	0.26	0.51	ATPF_ECODH	19616	1.9	3.77E+04	0.6	0.02	0.08	0.16	0.31	1.00	0.32	0.30	0.30	0.32	3.6
0.05	0.26	0.53	EFTS_ECOBW	34697	1.8	6.25E+04	0.9	0.03	0.13	0.27	0.50	1.00	0.50	0.49	0.52	0.52	3.2
0.05	0.26	0.50	THIO_ECOLI	13498	1.8	2.39E+04	0.4	0.01	0.05	0.10	0.20	1.00	0.19	0.19	0.19	0.20	3.4
0.05	0.27	0.53	TIG_ECOBW	53351	1.8	9.34E+04	1.4	0.04	0.20	0.40	0.75	1.00	0.73	0.74	0.78	0.78	3.4
				Sum		1.21E+07	180										
				iTRAQ-emPAI calculated protein amounts*			5.5	26.8	51.5	96.2							
				Original planned amounts**			5	25	50	100							

Only bold red and ion score > 20

Proteins (522) of the *E. coli* lysate were compared by iTRAQ labeling (Supplemental Data file 1), but only the top 30 are shown in this table. The original amounts of protein planned for iTRAQ 114, 115, 116, and 117 were 5, 25, 50, and 100 µg, respectively. *The calculated protein amounts reflect the relative efficiencies of the iTRAQ labeling reaction under the different reagent composition designs. **The original amounts of protein planned for each iTRAQ labeling were 5 g. Only peptides identified by Mascot with bold red and ion score > 20 were used in this analysis. R², R-squared value of the trendline; CV, coefficient of variation.

search results page. It is convenient to first sort the rows of proteins according to decreasing emPAI scores to allow the removal of the lowest values, which represent unreliable data. For the purpose of this demonstration, we cut off at emPAI 0.1 for all proteins with iTRAQ ratios in all channels. Thus, ~522 proteins are considered in this study.

EMMOL Uses the Relationship from Equation 4 of Ishihama et al.¹

$$\text{Protein content (weight \%)} = (\text{emPAI} \times M_r) / \sum(\text{emPAI} \times M_r) \times 100$$

emPAI score is proportional to the fraction of observed peptides/theoretical peptides for a given protein (allowing no methionine oxidation and no missed trypsin cleavage). emPAI is roughly, inversely proportional to the MW of a protein.

Consider an example of the first experiment, in which the total protein in the pool was 180 μg . Although 180 μg is named here to facilitate the illustration that in the end, the total protein amount calculated in each iTRAQ channel by EMMOL matches up with what was initially added to the experiment and therefore, illustrates that the EMMOL method works, it is important to emphasize that one does not need to know the initial protein amounts. Any arbitrary value, i.e., 100% (as is done in the actual equation from the emPAI paper), can be used here.

Each protein in the pool has an emPAI score from Mascot. For each protein, we multiply its emPAI score with its theoretical MW to produce a relative protein amount ($\text{emPAI} \times M_r$) with respect to all other proteins in the pool (column G in Tables 3 and 4). We do this for each protein in the experiment that has iTRAQ ratios, and sum all of the relative protein values $\sum(\text{emPAI} \times M_r)$ in the proteome. For each protein, $[(\text{emPAI} \times M_r) / \sum(\text{emPAI} \times M_r)] \times 180$ gives the total amount of that protein in the pool (column H in Tables 3 and 4). Do this for each protein. The sum of all proteins in the pool should equal 180 μg .

The pool of each protein needs to undergo deconvolution into the protein quantities from each of the four samples. Mascot expresses the four iTRAQ channels relative to one channel. In our example, the denominator is 117; therefore, the ratios of the relative protein quantities for a given protein in the four channels are 114/117, 115/117, 116/117, and 117/117, respectively, where 117/117 = 1.

The amount of a given protein from channel 117 is given in column L in Tables 3 and 4 as: (sum of a given protein in μg from four iTRAQ channels)/(114/117+115/117+116/117+1).

Having obtained the iTRAQ 117 protein quantity for this protein, 114, 115, and 116 for this protein can be calculated from the ratios reported by Mascot (columns I, J, and K in Tables 3 and 4). Repeat this calculation for each protein in the proteome. The sum of proteins in the proteome calculated for each label channel gives the total

amount of all identified proteins in each iTRAQ channel. In practice, one can just cut and paste the iTRAQ ratios and protein identifications of one's experiment into the automated spreadsheet of Table 3, which we provided to generate an analysis similar to Table 3. These protein values calculated from MS for each sample can be used for further normalization of an assortment of iTRAQ channels from an experiment or from separate experiments to perform a comparative proteomics meta-study.

Summation of iTRAQ Ratios Normalization

This approach to normalize the labeled protein quantities of the samples simply sums the values of the ratios of all proteins in each iTRAQ channel, with the denominator channel of iTRAQ 117 set at unity. These sums can then be normalized to an arbitrary number. For example, the value of 100 μg protein for each sample was used in Table 6.

Statistical Evaluation

Two types of statistical evaluation of the results of EMMOL normalization are presented in this study.

First, for each protein detected in the proteome, the value calculated for each protein in an iTRAQ channel is plotted against the sum of all proteins calculated in each channel for the four channels. In the first example, the amounts of protein planned for the experiment for the four channels were 5 μg , 25 μg , 50 μg , and 100 μg , respectively. If EMMOL is perfect, this graph is expected to yield a straight line with a R^2 of ~1 (column M in Table 3). This value actually reflects the performance if an experiment has four replicates. The R^2 for every protein in the proteome was calculated.

Second, to evaluate the performance of EMMOL for a single replicate, we presented the CV % (columns R and Q in Tables 3 and 4, respectively). To do this, the four channels are further scaled to the same amount of total protein, say 100 μg . The SD among the four channels for a given protein, divided by the average value among the four channels for that protein time 100, gives the CV %.

RESULTS

EMMOL iTRAQ Analysis of a Pooled Sample of Four iTRAQ Labels with 20-Fold Range of Input Proteins

The result of the quantitation of individual proteins in each of the four samples is shown in Table 3. The complete data for the whole proteome is presented in Supplemental Data file 1. To illustrate the salient features, Table 3 presents the top 30 proteins in this study, which have emPAI ≥ 0.1 and iTRAQ ratios in all lanes. Although the table was sorted in a descending emPAI score for each protein (column F) to facilitate initial removal of all proteins with emPAI < 0.1 , an alternate, useful presentation may be in descending emPAI*protein mass (column G) or the calculated amount of each protein

TABLE 4
An iTRAQ Experiment Testing Four iTRAQ Labeling Conditions

114/117	115/117	116/117	SwissProt entry name	prot_mass	emPAL	emPAL*	20 µg total	114 protein (µg)	115 protein (µg)	116 protein (µg)	117 protein (µg)	Normalized to 5 µg	Normalized to 5 µg	Normalized to 5 µg	Normalized to 5 µg	sd/average CV %
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
0.55	1.13	1.38	EFTU1_ECO24	46886	3.30	1.55E+05	0.23	0.03	0.06	0.08	0.06	0.06	0.06	0.06	0.06	0.84
0.56	1.07	1.31	TNAI_ECOBW	57462	2.98	1.71E+05	0.25	0.04	0.07	0.08	0.06	0.07	0.06	0.06	0.06	3.91
0.58	1.27	1.70	ODP1_ECO57	107009	1.50	1.61E+05	0.24	0.03	0.07	0.09	0.05	0.06	0.06	0.07	0.05	9.67
0.55	1.07	1.33	ATPB_ECOBW	53377	1.95	1.04E+05	0.15	0.02	0.04	0.05	0.04	0.04	0.04	0.04	0.04	2.53
0.57	1.28	1.72	RPOC_EC0DH	168599	0.61	1.03E+05	0.15	0.02	0.04	0.06	0.03	0.03	0.04	0.04	0.03	10.60
0.59	1.12	1.35	EFG_ECOBW	84188	1.17	9.85E+04	0.14	0.02	0.04	0.05	0.04	0.04	0.04	0.04	0.04	4.46
0.52	1.09	1.35	RST_ECO57	67575	1.91	1.29E+05	0.19	0.02	0.05	0.06	0.05	0.05	0.05	0.05	0.05	2.38
0.54	1.05	1.27	TIG_ECOBW	53351	2.61	1.39E+05	0.20	0.03	0.06	0.07	0.05	0.05	0.05	0.05	0.05	3.80
0.54	1.10	1.41	PFLB_ECOLI	92937	1.76	1.64E+05	0.24	0.03	0.07	0.08	0.06	0.06	0.06	0.06	0.06	2.78
0.54	1.65	1.48	OMPA_ECOLI	40174	3.17	1.27E+05	0.19	0.02	0.07	0.06	0.04	0.04	0.04	0.04	0.04	19.18
0.50	1.00	1.09	ENO_ECOBW	51159	1.87	9.57E+04	0.14	0.02	0.04	0.04	0.04	0.04	0.03	0.03	0.04	9.06
0.61	1.26	1.62	TALB_ECOLI	38683	3.02	1.17E+05	0.17	0.02	0.05	0.06	0.04	0.04	0.04	0.05	0.04	7.11
0.59	1.04	1.24	CH60_ECO24	63372	1.79	1.13E+05	0.17	0.03	0.04	0.05	0.04	0.05	0.04	0.04	0.04	7.81
0.55	1.08	1.36	RBSB_ECOLI	34822	2.43	8.46E+04	0.12	0.02	0.03	0.04	0.03	0.03	0.03	0.03	0.03	2.61
0.58	1.22	1.61	DNAK_ECOHS	76495	0.76	5.81E+04	0.09	0.01	0.02	0.03	0.02	0.02	0.02	0.02	0.02	7.15
0.56	1.05	1.13	RPOB_ECOBW	162610	0.60	9.76E+04	0.14	0.02	0.04	0.04	0.04	0.04	0.04	0.03	0.04	8.96
0.55	0.98	1.13	RL7_ECO24	14305	3.27	4.68E+04	0.07	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	9.71
0.59	1.19	1.25	HTPG_ECOHS	78151	1.20	9.38E+04	0.14	0.02	0.04	0.04	0.03	0.04	0.04	0.03	0.03	6.97
0.59	1.32	1.58	CLPB_ECOLI	102182	0.59	6.03E+04	0.09	0.01	0.03	0.03	0.02	0.02	0.02	0.02	0.02	7.28
0.56	1.20	1.54	GRCA_ECOBW	15485	10.72	1.66E+05	0.24	0.03	0.07	0.09	0.06	0.06	0.06	0.06	0.06	5.54
0.59	1.11	1.33	ODP2_ECOLI	73893	1.19	8.79E+04	0.13	0.02	0.04	0.04	0.03	0.03	0.03	0.03	0.03	4.29
0.58	1.05	1.10	PUR9_ECO24	61944	0.79	4.89E+04	0.07	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	10.89
0.56	1.20	1.20	6PGD_ECOLI	55742	1.17	6.52E+04	0.10	0.01	0.03	0.03	0.02	0.02	0.03	0.02	0.02	7.95
0.59	1.23	1.47	KPYK1_FCO57	56083	1.46	8.19E+04	0.12	0.02	0.03	0.04	0.03	0.03	0.03	0.03	0.03	3.83
0.54	1.20	1.56	SYGB_ECOBW	82988	0.84	6.97E+04	0.10	0.01	0.03	0.04	0.02	0.02	0.03	0.03	0.02	7.00
0.56	1.25	1.55	SYP_ECOBW	67889	0.61	4.14E+04	0.06	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	6.36
0.57	1.06	1.12	ATPA_ECOBW	59018	0.96	5.67E+04	0.08	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	9.86
0.57	1.05	1.21	GLYA_ECOBW	49782	1.38	6.87E+04	0.10	0.02	0.03	0.03	0.03	0.03	0.02	0.02	0.03	7.60
0.50	1.01	1.11	SYD_ECOBW	70912	0.94	6.67E+04	0.10	0.01	0.03	0.03	0.03	0.02	0.02	0.02	0.03	8.30
0.59	1.09	1.25	RL9_ECO24	17489	3.04	5.32E+04	0.08	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	6.82
Sum								2.7	5.6	6.7	5.0					6.82
Only bold red and ion score >20								5	5	5	5					5
iTRAQ-emPAL calculated protein amounts*								2.7	5.6	6.7	5.0					6.82
Original planned amounts**								5	5	5	5					5

Proteins (653) of the *E. coli* lysate were compared by iTRAQ labeling (Supplemental Data 2), but only the top 30 are shown in this table. *The calculated protein amounts reflect the relative efficiencies of the iTRAQ labeling reaction under the different reagent composition designs. **The original amounts of protein planned for each iTRAQ labeling were 5 µg. Only peptides with bold red and ion score >20 in Mascot output were used in this analysis.

TABLE 5

Combination of the Experiments of Tables 3 and 4

SwissProt entry name	Protein mass	4 114	4 115	4 116	4 117	3 114	3 115	3 116	3 117	CV %
EFTU1_ECO24	46886	1.30	1.31	1.34	1.32	1.31	1.40	1.37	1.30	2.6
TNAI_ECOBW	57462	1.53	1.41	1.45	1.51	1.43	1.44	1.47	1.49	2.6
ODP1_ECO57	107009	1.28	1.36	1.53	1.22	1.38	1.32	1.21	1.35	7.5
ATPB_ECOBW	53377	0.90	0.86	0.89	0.91	1.03	0.89	0.92	0.89	5.7
RPOC_ECODH	168599	0.80	0.88	0.98	0.78	0.78	0.82	0.79	0.87	8.3
EFG_ECOBW	84188	0.90	0.83	0.83	0.84	0.82	0.81	0.89	0.83	3.9
RS1_ECO57	67575	1.06	1.08	1.12	1.13	1.10	1.12	1.14	1.10	2.4
TIG_ECOBW	53351	1.22	1.15	1.16	1.25	1.17	1.20	1.24	1.20	2.9
PFLB_ECOLI	92937	1.36	1.35	1.45	1.40	1.23	1.27	1.33	1.44	5.6
OMPA_ECOLI	40174	0.92	1.37	1.03	0.94	0.88	1.01	0.98	1.07	14.7
ENO_ECOBW	51159	0.83	0.81	0.74	0.92	0.93	0.87	0.92	0.83	7.7
TALB_ECOLI	38683	0.99	1.00	1.07	0.90	1.10	0.91	0.92	0.99	7.5
CH60_ECO24	63372	1.07	0.93	0.92	1.02	0.93	0.94	0.95	1.00	5.6
RBSB_ECOLI	34822	0.73	0.70	0.73	0.73	0.71	0.66	0.71	0.74	3.6
RPOB_ECOBW	162610	0.90	0.84	0.75	0.90	0.85	0.87	0.93	0.83	6.5
RL7_ECO24	14305	0.44	0.38	0.37	0.44	0.39	0.40	0.48	0.40	9.3
HTPG_ECOHS	78151	0.86	0.84	0.74	0.81	0.83	0.81	0.78	0.81	4.5
CLPB_ECOLI	102182	0.49	0.54	0.54	0.47	0.48	0.51	0.49	0.50	5.3
GRCA_ECOBW	15485	1.34	1.41	1.51	1.34	1.32	1.36	1.38	1.40	4.4
ODP2_ECOLI	73893	0.80	0.74	0.74	0.76	0.67	0.76	0.75	0.76	4.9
PUR9_ECO24	61944	0.47	0.42	0.37	0.45	0.42	0.45	0.48	0.41	8.4
6PGD_ECOLI	55742	0.58	0.60	0.50	0.57	0.60	0.59	0.56	0.56	5.8
KPYK1_ECO57	56083	0.70	0.72	0.72	0.66	0.69	0.67	0.67	0.69	2.8
SYGB_ECOBW	82988	0.55	0.59	0.64	0.56	0.60	0.55	0.57	0.59	5.7
SYP_ECOBW	67889	0.33	0.36	0.37	0.33	0.23	0.31	0.34	0.36	13.6
ATPA_ECOBW	59018	0.54	0.49	0.43	0.52	0.52	0.50	0.53	0.49	6.8
GLYA_ECOBW	49782	0.64	0.57	0.55	0.62	0.57	0.57	0.62	0.60	5.5
SYD_ECOBW	70912	0.57	0.57	0.52	0.64	0.54	0.65	0.59	0.58	7.8
RL9_ECO24	17489	0.50	0.45	0.43	0.47	0.54	0.47	0.44	0.46	7.2
PURA_ECOBW	50858	0.60	0.67	0.71	0.60	0.61	0.65	0.63	0.64	6.1
Average CV %										16.1

Each column has been normalized to the sum of 100. Proteins (471) were found in common between the two experiments. Only the top 30 abundant proteins are shown in this table to illustrate the feasibility of combining individual, deconvoluted protein quantitation of each sample to enable a new, proteomic analysis without the iTRAQ-labeled samples having been run in the same experiment.

(column H). The calculated sum of all of the proteins in each iTRAQ channel is indicated by an * at the bottom of the table. The originally intended quantity of protein for each iTRAQ channel is indicated by **.

For each protein, the R^2 correlation of the quantities in the four iTRAQ channels versus the total quantities of proteins calculated in each iTRAQ channel is shown in column M and illustrated the relationship of approximately a straight line for most proteins in the proteome.

An Improved iTRAQ Labeling Protocol for Use with Small Quantities of Protein

The result of the quantitation of individual proteins in each of the four iTRAQ channels in this experiment is shown in

Table 4. The complete data for the whole proteome are presented in Supplemental Data file 2. The display of the top 30 proteins in this study of 653 proteins suffices to illustrate the salient features. The result of normalization of the four iTRAQ channels to total protein quantity of 20 μg , in spite of the four channels having different reaction chemistry and labeling efficiency, is shown in columns M–P. The calculated sum of all of the proteins in each iTRAQ channel is indicated at the bottom of the table, as an *. Here, the higher calculated sum of the proteins before normalization in a channel is an indication of higher labeling efficiency produced by that labeling condition.

We also wanted to test the EMMOL normalization effectiveness in this experiment that varied in labeling

TABLE 6
Comparison of EMMOL Precision with Normalization Method Using Only ITRAQ Ratios

5 $\mu\text{g}/100 \mu\text{g}$ (114/117)	25 $\mu\text{g}/100 \mu\text{g}$ (115/117)	50 $\mu\text{g}/100 \mu\text{g}$ (116/117)	100 $\mu\text{g}/100 \mu\text{g}$ (117/117)	SwissProt entry name	114 normalized to 100	115 normalized to 100	116 normalized to 100	117 normalized to 100	(A) sd/avg using only ITRAQ ratios	(B) sd/avg using EMMOL	EMMOL advantage, A/B
A	B	C	D	E	G	H	I	J	K	L	M
0.06	0.28	0.55	1	GRCA_ECOBW	0.18	0.19	0.20	0.19	3.05	1.93	1.58
0.06	0.28	0.58	1	RS10_ECO24	0.21	0.19	0.21	0.19	4.61	5.04	0.91
0.06	0.31	0.59	1	RL11_ECO24	0.19	0.21	0.21	0.19	5.03	4.73	1.06
0.05	0.25	0.50	1	HNS_ECOLI	0.16	0.17	0.18	0.19	7.37	6.24	1.18
0.06	0.28	0.50	1	CSPC_ECOLI	0.20	0.19	0.18	0.19	4.06	5.08	0.80
0.06	0.29	0.55	1	EFTU1_ECO24	0.18	0.20	0.20	0.19	3.58	2.86	1.25
0.05	0.25	0.58	1	RL7_ECO24	0.16	0.17	0.21	0.19	11.39	10.21	1.12
0.07	0.28	0.55	1	TALB_ECOLI	0.23	0.19	0.20	0.19	8.65	9.83	0.88
0.07	0.26	0.53	1	RL1_ECO24	0.23	0.18	0.19	0.19	10.59	11.69	0.91
0.05	0.27	0.52	1	TNAA_ECOBW	0.18	0.18	0.19	0.19	3.64	2.50	1.46
0.05	0.24	0.52	1	CSRA_ECOBW	0.17	0.16	0.18	0.19	7.69	6.65	1.16
0.05	0.32	0.58	1	AHPC_ECOLI	0.17	0.22	0.21	0.19	10.74	10.18	1.05
0.05	0.27	0.53	1	DBHA_ECOLI	0.15	0.18	0.19	0.19	9.64	8.53	1.13
0.06	0.27	0.52	1	RS4_ECO24	0.18	0.18	0.18	0.19	2.99	2.19	1.36
0.07	0.28	0.51	1	RL9_ECO24	0.21	0.19	0.18	0.19	6.88	8.12	0.85
0.06	0.28	0.54	1	RS1_ECO57	0.18	0.19	0.19	0.19	3.15	1.85	1.70
0.05	0.30	0.56	1	OMPA_ECOLI	0.17	0.20	0.20	0.19	6.48	5.61	1.15
0.06	0.24	0.56	1	ALKH_ECOLI	0.19	0.17	0.20	0.19	8.12	7.59	1.07
0.06	0.27	0.54	1	ATPB_ECOBW	0.21	0.18	0.19	0.19	5.71	6.63	0.86
0.05	0.27	0.57	1	RL3_ECO24	0.17	0.18	0.21	0.19	7.30	6.11	1.19
0.06	0.25	0.52	1	RBSB_ECOLI	0.18	0.17	0.19	0.19	5.17	4.56	1.13
0.06	0.27	0.56	1	RS3_ECO24	0.20	0.18	0.20	0.19	3.57	3.69	0.97
0.06	0.31	0.55	1	RL5_ECO24	0.19	0.21	0.20	0.19	4.19	4.33	0.97
0.06	0.27	0.53	1	G3P1_ECO57	0.18	0.18	0.19	0.19	3.28	2.23	1.47
0.06	0.43	0.56	1	RS6_ECO24	0.18	0.29	0.20	0.19	22.78	22.96	0.99
0.07	0.29	0.51	1	RL10_ECO24	0.23	0.19	0.18	0.19	10.98	12.32	0.89
0.06	0.26	0.51	1	ATPF_ECODH	0.18	0.17	0.18	0.19	3.87	3.64	1.06
0.05	0.26	0.53	1	EFTS_ECOBW	0.18	0.18	0.19	0.19	4.32	3.16	1.37
0.05	0.26	0.50	1	THIO_ECOLI	0.17	0.18	0.18	0.19	4.37	3.44	1.27
0.05	0.27	0.53	1	TIC_ECOBW	0.17	0.18	0.19	0.19	4.71	3.44	1.37
30.5	147.5	278.7	522.0		100.0	100.0	100.0	100.0			

Data from Table 3 are normalized without EMMOL (columns C–K), and the sd/average in percent is shown in column K. Data from Table 3 using EMMOL are shown in column L. Values in column K/column L give the EMMOL advantage as ratios ≥ 1 in two-thirds of proteins.

efficiency. Thus, the originally intended quantity of protein for each iTRAQ channel is indicated by **. An estimate of the effectiveness of normalization is shown as the CV % value in column Q.

EMMOL iTRAQ Proteomes Obtained from Assorted Individual Samples can be Combined for a Proteomic Meta-Analysis

The above two experiments produced deconvoluted proteomes for eight iTRAQ reactions under different conditions of chemical labeling and different total protein amounts during MS. In Table 5, the proteins identified in common in the two experiments were compared. The complete data for the whole proteome are presented in Supplemental Data file 3. Each proteome was normalized to a total protein value of 100 μg . The CV % of the eight reactions for the top 30 proteins were calculated and presented in Table 5.

Summation of iTRAQ Ratios Normalization of the Data in the Experiment of Table 3 and Comparison with Using EMMOL

The simplest method of normalization of the protein quantities in the samples that we have validated in this study is shown in Table 6, by summing the values of the ratios of all proteins in each iTRAQ channel, giving 30.5, 147.5, 278.7, and 522, respectfully. These sums can then be normalized to an arbitrary number. The value of 100 μg protein for each sample was used in Table 6. The normalized quantity of each protein for each iTRAQ channel is shown in columns G–J. The precision of the relative quantitation for each protein was shown in column K as the CV % (SD of values in columns G–J divided by their average value time 100). This number is compared with the value presented in Table 3 obtained using EMMOL. To estimate the advantage conferred by using EMMOL, column M shows the ratio of column K to column L. A value of 1 is returned for a given protein when both methods are identical in precision. A number greater than 1 indicated that EMMOL is more precise than the summation of the iTRAQ ratios. A number less than one indicated that EMMOL is less precise than normalization by summation of the iTRAQ ratios.

DISCUSSION

The novelty of EMMOL comes from applying emPAI values of each protein in a pool to iTRAQ information obtained from the Mascot output. Using the iTRAQ reporter ratios, one can calculate how much of this protein weight percent is distributed among the individual iTRAQ channels. Once this is done for all of the proteins present, the normalization of the total protein in each sample becomes more accurate than without using the EMMOL procedure. One can look at each sample individually down the entire protein list and obtain their relative expression values more accurately than without

using EMMOL. The total amount of protein for each iTRAQ channel can be normalized to the same number (say 100 μg but does not have to be 100 μg), regardless of how much protein was originally in that channel. This method can then be extended to comparing multiple, separate iTRAQ datasets without the need for a reference channel, provided that one normalizes them all to the same value.

iTRAQ chemistry for comparative MS has many virtues. The multiplexed isobaric iTRAQ labeling scheme allows the same peptide in the four samples to share the same physical properties, so as to cofractionate, thereby overcoming the problem of parent ion under-sampling inherent in LC-MS/MS sample comparison. Under-sampling refers to the chance appearance or not of a given ion in consecutive LC-MS/MS runs, a problem that requires a statistically valid number of replicates to be performed or different peptides to be used to represent the same protein in different samples. It also refers to the possibility of ion sampling at different positions and amplitudes of the elution peak of a parent ion. iTRAQ multiplexing of four samples in one peptide MS/MS event enhances the ion statistics and detection sensitivity. Looking to the future, chemical labeling-multiplexed MS is an approach that can increase the throughput of expensive mass spectrometers by four- to eightfold using available reagents.

iTRAQ Analysis of a Pooled Sample of Four iTRAQ Labels with 20-Fold Range of Input Proteins

We performed this experiment using a cell lysate of a popular strain of *E. coli* using a commercially provided growth media, so this standard can be reproduced by any laboratory. By using 100 μg aliquots of the same lysate for each of the four iTRAQ labels under standard reaction conditions, we removed the variable of labeling efficiency from the experiment in Table 3, while we created a pool of labeled samples with a range of 20-fold difference in labeled peptides. During various steps of sample evaporation-concentration in the iTRAQ protocol, we measured the volumes of each step using a variable volume pipette to ensure the presence of correct amounts of proteins and peptides in each reaction.

Effect of the EMMOL Normalization Approach that Considers Protein Mass and the Deconvolution of the Observed Proteins to Appropriate Samples in the Pool

Normalization of iTRAQ experiments, mostly at the level of individual peptides, and their summation to individual proteins have been discussed in literature,^{15–19} and it is also supported by popular software, including ProteinPilot of AB Sciex (Foster City, CA, USA) and Scaffold 3 with Q+ of Proteome Software (Portland, OR, USA). This report compares some normalization approaches and illustrates an experimental design that can facilitate the studies of precious biological samples. Our approach is facilitated by the

EMMOL normalization and deconvolution approach, which may be of interest to others. As Mascot's iTRAQ analysis presents the ratios of the intensities of the iTRAQ reporters detected, instead of retrieving the peak intensities of the labels, it is convenient to use the Mascot-reported iTRAQ reporter ratios to compute the share of each protein in each iTRAQ channel. Indeed, biochemists have long used chemical labeling as a means of protein quantitation by labeling the primary amines moieties or the sulfhydryl groups. iTRAQ reagents enable protein assay using MS.

The sum of the proteins in each column of label gives a good approximation of the relative protein determination for that sample (Table 3) to facilitate normalization to the physical reality of the sample type. An example may be normalization to the average total protein concentration of serum. Alternatively, just normalizing to 100 still corrects for the differences in protein in each of the four samples. In this report, where the starting material was an identical digest of an *E. coli* lysate, the ratios are anticipated to be 1:1:1:1 for a perfect outcome, which is illustrated by the R^2 values in column M in Table 3.

Traditionally, one begins an iTRAQ experiment using 100 μg of each sample that has been carefully quantified by a reliable protein assay. The apprehension not to deviate from this popular approach creates a concern when one wishes to apply iTRAQ chemistry to samples that are difficult to obtain in quantity. This study shows that there is, in fact, no requirement for performing an accurate protein assay prior to an iTRAQ study. Variations in iTRAQ labeling efficiency as a result of different protein quantities can be effectively corrected by the EMMOL normalization procedure, as demonstrated in this report. It is thus feasible to combine precious quantities of proteins with a more ample sample for the latter to facilitate effective identification in an iTRAQ experiment.

It should be noted that proteins that are highly glycosylated or otherwise difficult for trypsin to digest will yield artificially low emPAI scores. Also, proteins low in abundance will yield low emPAI scores. Fortunately, the EMMOL normalization process accounts for protein mass differences, wherein the normalization is weighed in favor of the abundant proteins, which are quantified more accurately than the low-abundance proteins.

An Improved iTRAQ Labeling Protocol for Use with Small Quantities of Protein

The experiment of Table 4 compared four protocols, which a given laboratory may have used to label 5 μg protein digest with the iTRAQ reagent. iTRAQ label 114 (a) is an aliquot of the same label reaction of the experiment in Table 3, an experiment that already validated our reproducibility. iTRAQ label 114 (a) repre-

sents the reaction conditions recommended by the reagent manufacturer. The second, iTRAQ label 115 (b), is simply 5 μg peptides in the normal sample volume of 40 μL when added into the iTRAQ reaction. The third, iTRAQ 116 (c), concentrated the peptide solution to 5% the volume (2 μL) of the normal sample volume before adding to the iTRAQ reaction. The fourth, iTRAQ 117 (d), added the sample as 2 μL but proportionally decreased the amount of iTRAQ label reagent to maintain the same label:peptide stoichiometry as in a. The lower calculated total protein value for iTRAQ 114 before normalization (which is not an indication of an incorrect value) does not indicate that EMMOL is at fault but instead, demonstrated the relative inefficiency of the traditional iTRAQ labeling protocol for 100 μg protein in our hands compared with the protocols using 5 μg protein.

Both c and d achieved lowering the amount of water competition against the decreased peptide primary amine moieties, whereas c provided 20-fold more iTRAQ reagent excess. Reaction kinetics may be slower in some of these situations than in a, but the important concern appeared to be reaction completion when ample time is allowed. It is no surprise that in the presence of increased iTRAQ reagent excess, one accomplishes more thorough peptide labeling in a 5- μg sample compared with the routine 100 μg sample labeling. This fact was illustrated by the summed, calculated emPAI score for the sample in c to be the highest of the four labeling conditions (Table 4 column K, indicated by *). Thus, the most effective labeling condition was the sample that used iTRAQ 116, in which we minimized the input protein volume and water competition but retained the maximum iTRAQ reagent excess. Higher labeling efficiency yields more robust MS data than lower labeling efficiency. The performance improvement is also reflected in the identification of 653 proteins in the "5 μg protein per label" experiment of Table 4 compared with 522 proteins in the "100 μg per label" experiment of Table 3 analyzed under identical conditions.

It is important to note that the MS conditions dictate the amount of labeled peptides that can be applied to the LC-MS/MS system so that the excess of labeled peptides from 100 μg /sample labeling is not necessarily an advantage. Chemical labeling reaction increases in efficiency, as reagent excess increases when target molecules decrease in number. Labeling of lower protein quantities is further favored by decreasing the amount of water competing for the reaction. Comparison of reactions b to c to d illustrates that reducing water competition and maintaining a high ratio of iTRAQ reagent to peptides are the optimal conditions for iTRAQ labeling. The condition for reaction c will facilitate future studies of samples with minute quantities

of proteins. Importantly, EMMOL corrects for the differences in labeling efficiency among samples to lead to meaningful normalization. Admittedly, the experiments in this report are not exhaustive. No attempts have been made to perform the experiments multiple times nor with label-flipping, as is customary for a DNA microarray study. However, our data are internally consistent with our conclusions and serve to form the foundation from which other core facilities can further optimize their iTRAQ studies.

iTRAQ Proteomes Obtained from Assorted Individual Samples Can Be Combined for a Proteomic Meta-Analysis

Table 5 combines the data of Tables 3 and 4 to illustrate that in spite of the 20-fold range of starting protein quantities for the iTRAQ labeling chemistry and the use of different labeling conditions, different amounts of proteins in each iTRAQ channel, and combining the data of the two experiments, EMMOL can compare and normalize this data without using a common reference channel for normalization, as is required in literature and in Scaffold Q+ (<http://www.proteomesoftware.com/posters/Q+Poster.pdf>). The EMMOL results are surprisingly comparable across the eight reactions. The CV % is adequate if relative expression of more than onefold is used to indicate significant expression change for a protein. The test in Table 5 illustrates that it is possible to combine proteomes obtained by multiplexed iTRAQ chemistry in separate experiments into a meta-analysis that is not anticipated in the initial experimental plans. As a result, the study may achieve greater statistical significance.

Comparison of Methods of iTRAQ Data Normalization

iTRAQ reporter peak intensities can be extracted from the MS files by software developed by investigators to enhance the quality of the reporter ratios.^{15,19,20} However, The Association of Biomolecular Resource Facilities Proteomics Research Group PRG2011 survey showed that amongst 242 participants of the survey, Mascot was the most popular search engine used in proteomics laboratories (<http://www.abrf.org/ResearchGroups/Proteomics/Studies/ABRF2011PRGSurveyPresentation.pdf>). Mascot returns iTRAQ reporter intensities as ratios of different iTRAQ channels. Therefore, it is useful for the average laboratory to have an easy, accurate routine, such as EMMOL, which requires only the iTRAQ channel ratios and the emPAI scores as provided by Mascot.

Basic iTRAQ data analysis assumes that the amounts of protein in the four channels and their labeling efficiencies are identical.^{18,21–23} Indeed, most iTRAQ experiments in literature begin with 100 μg protein for each channel.^{19,23} The “normalization” of iTRAQ data addressed in literature concerned mainly the ratios of individ-

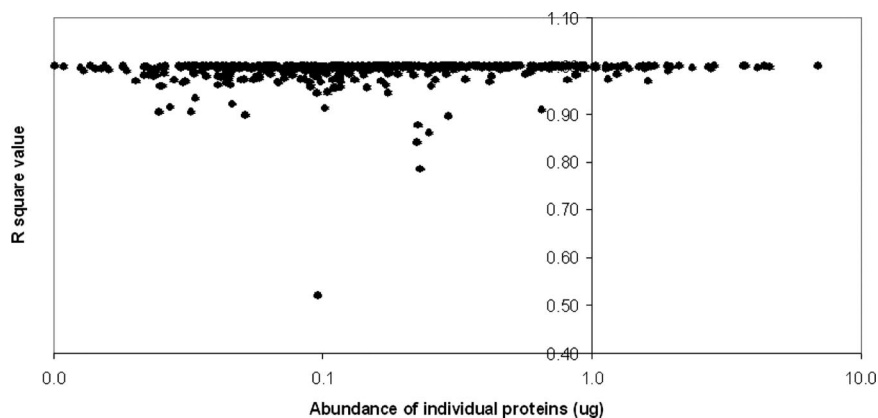
ual peptides and the summation of these peptide ratios to individual proteins^{19,22} but seldom the normalization of the total protein of the samples in the pool. However, the ratio for each protein depends on the quantity of total protein in each sample and its labeling efficiency. Thus, normalization of the total labeled protein is an important step. It can be seen in Table 6 that the “summing iTRAQ ratios” method results in protein ratios for the four samples of the experiment in Table 3 that are approximately correct, namely, 5.8, 28.2, 53.4, and 100 μg , respectively (normalized from sum of the channels of 30.5, 147.5, 278.7, and 552 for iTRAQ 114, 115, 116, and 117, respectfully), compared with the carefully planned amounts of 5, 25, 50, and 100 μg , respectively. However, this approach assumes that the data accuracy is independent of protein abundance for each protein and the length of the peptides, which is not always true and can lead to slightly incorrect measurements of total protein quantity in samples. Moreover, as pointed out by others previously (<http://www.proteomesoftware.com/posters/Q+Poster.pdf>), this method is sensitive to outliers and extreme ratios unless the data are filtered.

On the other hand, the EMMOL approach is less susceptible to technical variability and extreme iTRAQ ratios by uncoupling the measurement of protein abundance (emPAI) from the protein assays performed prior to protease digestion or the iTRAQ labeling efficiency. The EMMOL approach puts greater emphasis on the abundant proteins for normalization, as they have more and stronger peptide measurements than those of the low-abundance proteins. Hence, by comparing columns K and L in Table 6, the EMMOL normalization proved itself an improvement over the approach of summing iTRAQ ratios. Column M of Table 6 is one way of illustrating the EMMOL advantage by expressing the ratio of the values in columns K and L. For the first 30 proteins, as well as for all 522 proteins, two-thirds of the ratios of column K/column L is greater than one, meaning that for two-thirds of the proteins in the study, the EMMOL method returned quantitation with higher precision than the method of summing the iTRAQ ratios.

Peptide fractionation prior to MS is a key to minimize the distortion of iTRAQ reporter intensity coming from unintended parent ions of similar masses. To do this, strong cation-exchange chromatography is a popular method for the fractionation of the iTRAQ-labeled peptides. In this report, we used the IEF fractionation method similar to the popular OFFGEL method¹³ (Agilent Technologies, Santa Clara, CA, USA), because of the availability of equipment and our experience from running hundreds of 2D gels using the same procedures. The standard operating procedures used in our laboratory are posted on the website (<http://yeung.fccc.edu>)

FIGURE 1

Effect of individual protein abundance on the linearity of iTRAQ labeling efficiency. The data of relative protein abundance from Table 3 are plotted against the R^2 value of the linearity of labeling efficiency over a 20-fold range of protein abundance for each protein presented in log scale.



for Fox Chase Cancer Center. Peptides theoretically fractionate into three main pH ranges: pH 3.5–4.5, pH 5.5–6.3, and pH 7.3–9.3.²⁴ Thus, our choice of 60 fractions from only the pH 4–4.5 range effectively divided our total peptides into as many as 180 fractions, which may be the reason for cleaner iTRAQ quantitation.

A common concern about iTRAQ-derived expression ratios is the possibility of compression, meaning the observed fold change may be smaller than the true fold change in the sample.^{4,25–27} Curiously, we illustrated a R^2 of close to 1 for almost all proteins in our study (Table 3), meaning we observed no significant compression for the 20-fold range of protein quantitation of our proteome. Moreover, as shown in Fig. 1, the iTRAQ quantitation is accurate for this range over 3 logs of protein relative abundance in the proteome, except for a handful of outliers. As explained above, the absence of compression in iTRAQ quantitation in this study is likely a result of the use of highly fractionated peptide samples, which allow less contribution from nontargeted ions that have ratios of 1:1:1:1.²²

The current study made no attempt to obtain as deep a proteome coverage as possible. For example, we only used a moderately sensitive qTOF (QSTAR XL) using a single gradient elution of LC-MS/MS without performing a second gradient elution using an exclusion list of the ions sequenced in the first gradient elution. The use of current new generations of mass spectrometers can greatly enhance the depth of proteome coverage and iTRAQ comparison statistics, especially for the lowest abundance proteins.

Our conclusions are not meant to trivialize the difficulties of working with tiny quantities of protein material. The quality of data decreases with decreasing protein abundance for a given MS situation. However, improvements in mass spectrometers and miniaturization and automation of the systems for sample handling will surely come about. We have illustrated that iTRAQ experiments can be performed without prior knowledge of protein quantities and without requiring equivalent amounts of proteins for the different

iTRAQ labels. We showed that a more abundant control sample may be used to drive the overall performance of proteome identification for samples with low protein quantities. This report also suggests that higher performance using iTRAQ reagents can be obtained by dividing a sample to provide replicates of smaller protein quantities so as to obtain deeper proteome coverage and greater label linearity from higher labeling efficiency. We also showed that proteomes of separate experiments analyzed by EMMOL at the level of the protein composition in each sample can be merged and compared without a common reference channel. Thus, it is hoped that the EMMOL workflow may facilitate comparative proteomic experiments of precious biological samples to contribute to the understanding of biology.

ACKNOWLEDGMENTS

Support for this work was provided by the Driskill Foundation, National Cancer Institute, Work Assignment #16 of Contract No. NO1-CN-15103, Institutional Core Grant P30CA06927, RC2 HL101713, and Tobacco Settlement Funds from the Commonwealth of Pennsylvania, the Pew Charitable Trust, and the Kresge Foundation. The authors are grateful for critical reading of the manuscript by Brian Searle and quality input from the reviewers of *JBT*.

REFERENCES

- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann N. 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* 2005;4:1265–1272.
- Old WM, Meyer-Arendt K, Aveline-Wolf L, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 2005;4:1487–1502.
- Steen H, Jebanathirajah JA, Springer M, Kirschner MW. Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proc Natl Acad Sci USA* 2005; 102:3948–3953.
- Braisted JC, Kuntumalla S, Vogel C, et al. The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* 2008;9:529.
- Griffin TJ, Lock CM, Li XJ, et al. Abundance ratio-dependent proteomic analysis by mass spectrometry. *Anal Chem* 2003;75:867–874.

6. Waanders LF, Hanke S, Mann M. Top-down quantitation and characterization of SILAC-labeled proteins. *J Am Soc Mass Spectrom* 2007;18:2058–2064.
7. Gronborg M, Kristiansen TZ, Iwahori A, et al. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol Cell Proteomics* 2006;5:157–171.
8. Hanke S, Besir H, Oesterhelt D, Mann M. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J Proteome Res* 2008;7:1118–1130.
9. DeSouza L, Diehl G, Rodrigues MJ, et al. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cCAT with multidimensional liquid chromatography and tandem mass spectrometry. *J Proteome Res* 2005;4:377–386.
10. Zhang Y, Wolf-Yadlin A, Ross PL, et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 2005;4:1240–1250.
11. Chong PK, Gan CS, Pham TK, Wright PC. Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: implication of multiple injections. *J Proteome Res* 2006;5:1232–1240.
12. Lengqvist J, Uhlén K, Lehtio J. iTRAQ compatibility of peptide immobilized pH gradient isoelectric focusing. *Proteomics* 2007;7:1746–1752.
13. Chenu J, Michelland S, Sidibe J, Seve M. Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. *Proteome Sci* 2008;6:9.
14. Ke E, Patel BB, Liu T, et al. Proteomic analyses of pancreatic cyst fluids. *Pancreas* 2009;38:e33–e42.
15. Boehm AM, Putz S, Altenhöfer D, Sickmann A, Falk M. Precise protein quantification based on peptide quantification using iTRAQ. *BMC Bioinformatics* 2007;8:214.
16. Hill EG, Schwacke JH, Comte-Walters S, et al. A statistical model for iTRAQ data analysis. *J Proteome Res* 2008;7:3091–3101.
17. Oberg AL, Mahoney DW, Eckel-Passow JE, et al. Statistical analysis of relative labeled mass spectrometry data from complex samples using ANOVA. *J Proteome Res* 2008;7:225–233.
18. Lippert DN, Ralph SG, Phillips M, et al. Quantitative iTRAQ proteome and comparative transcriptome analysis of elicitor-induced Norway spruce (*Picea abies*) cells reveals elements of calcium signaling in the early conifer defense response. *Proteomics* 2009;9:350–367.
19. Rajcevic U, Petersen K, Knol JC, et al. iTRAQ-based proteomics profiling reveals increased metabolic activity and cellular cross-talk in angiogenic compared with invasive glioblastoma phenotype. *Mol Cell Proteomics* 2009;8:2595–2612.
20. Mahoney DW, Therneau TM, Heppelmann CJ, et al. Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J Proteome Res* 2011;10:4325–4333.
21. Lin WT, Hung WN, Yian YH, et al. Multi-Q: a fully automated tool for multiplexed protein quantitation. *J Proteome Res* 2006;5:2328–2338.
22. Karp NA, Huber W, Sadowski PG, et al. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics* 2010;9:1885–1897.
23. Kassie F, Anderson LB, Higgins L, et al. Chemopreventive agents modulate the protein expression profile of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone plus benzo[a]pyrene-induced lung tumors in A/J mice. *Carcinogenesis* 2008;29:610–619.
24. Eriksson H, Lengqvist J, Hedlund J, et al. Quantitative membrane proteomics applying narrow range peptide isoelectric focusing for studies of small cell lung cancer resistance mechanisms. *Proteomics* 2008;8:3008–3018.
25. Ow SY, Salim M, Noirel J, Evans C, Wright PC. Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. *Proteomics* 2011;11:2341–2346.
26. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, Lilley KS. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics* 2010;9:1885–1897.
27. DeSouza LV, Romaschin AD, Colgan TJ, Siu KW. Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer. *Anal Chem* 2009;81:3462–8470.