

Differential Protein Expression Analysis Using Stable Isotope Labeling and PQD Linear Ion Trap MS Technology

Jenny M. Armenta,^a Ina Hoeschele,^{a,b} and Iulia M. Lazar^{a,c}

^a Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

^b Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

^c Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

An isotope tags for relative and absolute quantitation (iTRAQ)-based reversed-phase liquid chromatography (RPLC)-tandem mass spectrometry (MS/MS) method was developed for differential protein expression profiling in complex cellular extracts. The estrogen positive MCF-7 cell line, cultured in the presence of 17 β -estradiol (E2) and tamoxifen (Tam), was used as a model system. MS analysis was performed with a linear trap quadrupole (LTQ) instrument operated by using pulsed Q dissociation (PQD) detection. Optimization experiments were conducted to maximize the iTRAQ labeling efficiency and the number of quantified proteins. MS data filtering criteria were chosen to result in a false positive identification rate of <4%. The reproducibility of protein identifications was ~60%–67% between duplicate, and ~50% among triplicate LC-MS/MS runs, respectively. The run-to-run reproducibility, in terms of relative standard deviations (RSD) of global mean iTRAQ ratios, was better than 10%. The quantitation accuracy improved with the number of peptides used for protein identification. From a total of 530 identified proteins ($P < 0.001$) in the E2/Tam treated MCF-7 cells, a list of 255 proteins (quantified by at least two peptides) was generated for differential expression analysis. A method was developed for the selection, normalization, and statistical evaluation of such datasets. An approximate ~2-fold change in protein expression levels was necessary for a protein to be selected as a biomarker candidate. According to this data processing strategy, ~16 proteins involved in biological processes such as apoptosis, RNA processing/metabolism, DNA replication/transcription/repair, cell proliferation and metastasis, were found to be up- or down-regulated. (J Am Soc Mass Spectrom 2009, 20, 1287–1302) © 2009 American Society for Mass Spectrometry

Recently, two-dimensional liquid chromatography (2DLC) with tandem MS detection has emerged as an attractive technology for quantitative proteomic profiling of complex cellular extracts [1–4]. Stable isotope labeling and label-free quantitation strategies have been explored [4–13]. Isotope labeling approaches rely on the covalent attachment of stable isotope tags to specific amino acid residues of proteins or peptides during metabolic, enzymatic, or chemical processes. Label-free quantitation methods rely on measuring peak areas, intensities, or spectral counts, and benefit from not having to chemically alter the sample. Throughput, however, is lower, and quantitation errors are higher, as the samples are processed independently. Among chemical labeling techniques, isotope tags for relative and absolute quantitation (iTRAQ) has received much attention [14–30]. In this approach, peptides are labeled with isobaric tags at the N-terminus and the lysine side chains. MS/MS fragmentation produces

signature ions that can be used to obtain quantitative information. Perhaps the most attractive advantage of this approach is that it can be used for the simultaneous quantification (relative or absolute) of up to four/eight different samples. Simplicity, of course, is an added benefit.

A number of publications have addressed the performance and challenges associated with the iTRAQ labeling strategy. Using 2D-gel or LC sample fractionation and matrix assisted laser desorption ionization (MALDI)-time-of-flight (TOF)/TOF detection, Wu et al. have reported a comparative study of three proteomic quantitative methods, namely, difference gel electrophoresis (DIGE), cleavable isotope coded affinity tagging (cICAT), and iTRAQ [5]. All three methods exhibited relatively good accuracy (experimental protein ratios for standards were within 81%–122% of the expected theoretical values), and were found to be complementary in nature. DeSouza et al. used iTRAQ and cICAT to identify potential cancer markers in endometrial tissues [26]. A total of 63 and 68 proteins were identified and quantified with iTRAQ and cICAT, respectively, and nine combined putative markers that

Address reprint requests to Dr. I. M. Lazar, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Washington St., Bio II/283, Blacksburg, VA 24061, USA. E-mail: lazar@vbi.vt.edu

met the >2-fold change threshold in their differential expression level were found. The RSD of differential expression ratios was in the range of ~1%–67%. Currently, there has been growing interest in developing proteomic quantitative protocols for a variety of biological applications that involve, for example, the study of protein–protein interactions [16], the monitoring of temporal changes in perturbed signaling pathways [18], and the discovery of novel disease biomarkers in samples of biological origin [27–30]. For example, Keshamouni et al. performed differential protein expression analysis of lung cancer cells undergoing epithelial-mesenchymal transition using iTRAQ followed by 2DLC-MS/MS [29]. Out of 325 identified proteins, ~29 were found to be up-regulated and 22 down-regulated. To account for technical errors, replicate experiments were conducted and data normalization was imperative.

As iTRAQ quantitation is based on detecting and measuring the intensity of low m/z fragment ions generated by tandem MS [21], most iTRAQ-LC-MS/MS based research has been carried out on TOF/TOF-MS instruments. More recently, a novel approach for precursor ion activation/dissociation in ion trap mass spectrometers that allows for the trapping of low m/z ions, termed pulsed Q dissociation (PQD), has been developed. Very few studies, however, describe the performance of the iTRAQ labeling/PQD approach for quantitative proteomics [31–33]. The main focus of these studies was on optimizing key PQD parameters such as collision energy (CE), activation Q, and delay time (T) to improve the quantitation accuracy and detection limits. These studies have shown that PQD operation has a much narrower range of optimal CE values than CID, and that this range has no universal settings for all LTQ instruments [32, 33]. Further work has demonstrated that the quantitation accuracy of isotope labeled peptides can be much improved if a high-resolution/high mass accuracy instrument, such as the Orbitrap, is used for detection (due to the capability to distinguish between ion species with very close m/z) [33, 34]. In the present work, we have developed an iTRAQ-RPLC/MS/MS strategy using PQD detection on a low-resolution linear ion trap mass spectrometer, and evaluated the performance of this approach for the analysis of complex cellular extracts in terms of reproducibility and accuracy of protein identifications and quantitation, respectively. This MS sample processing strategy was complemented by an effective statistical approach for the selection of putative biomarker candidates (based on the calculation of four iTRAQ ratios for each protein), and demonstrated for the analysis of breast cancer cells.

Experimental

Reagents

MCF-7 breast cancer cells and common cell culturing reagents (Eagle's minimum essential medium-EMEM,

fetal bovine serum-FBS, insulin, trypsin/EDTA) were purchased from ATCC (Manassas, VA). Phenol red-free DMEM (Dulbecco's modified Eagle's medium) was from Invitrogen (Carlsbad, CA), charcoal/dextran treated fetal calf serum from HyClone (Logan, UT), and trypsin (phenol red free) from SAFC Biosciences (Lenexa, KS). β -Estradiol (E2), tamoxifen (Tam), L-glutamine, protease inhibitors (NaF, Na_3VO_4), buffers and denaturing reagents (trifluoroacetic acid, acetic acid, formic acid, TrisHCl, sodium chloride, urea, dithiothreitol-DTT), and all bovine protein standards (hemoglobin α/β , serum albumin, cytochrome *c*, α -lactalbumin, carbonic anhydrase, α -casein, β -casein, and fetuin) were purchased from Sigma (St. Louis, MO). RIPA lysis buffer was from Upstate (Lake Placid, NY) and sequencing-grade modified trypsin from Promega Corp. (Madison, WI). iTRAQ reagents were purchased from Applied Biosystems (Foster City, CA). Ammonium bicarbonate was from Aldrich (Milwaukee, WI). HPLC-grade methanol and acetonitrile were obtained Fisher Scientific (Fair Lawn, NJ). All aqueous solutions were prepared using D.I. water from a MilliQ Ultrapure water system (Millipore, Bedford, MA).

MCF-7 Cell Culture

For initial optimization studies, MCF-7 breast cancer cells were cultured according to a procedure described in detail elsewhere [35]. Briefly, the cells were grown to 70% confluence in EMEM containing 10% FBS and 10 $\mu\text{g}/\text{mL}$ bovine insulin (maintenance medium), at 37 °C, in a humidified atmosphere of 5% CO_2 . The cells were rinsed with phosphate buffered saline (PBS, pH 7.4), and a solution of trypsin/EDTA (0.25% trypsin/0.53 mM EDTA) was added for cell detachment. Following incubation for 5 to 10 min, culture medium was added to stop the digestion. The cells were centrifuged/rinsed with PBS, harvested and stored at –80 °C. For protein differential expression analysis, MCF-7 breast cancer cells were first cultured in maintenance medium for approximately 2 wk (see above). To precondition the cells before experimental treatment, the medium was changed to a 3:2 mix of DMEM red-free (complemented with 10% charcoal stripped fetal calf serum, 1 $\mu\text{g}/\text{mL}$ insulin and L-glutamine 4 mM) and EMEM (complemented with 10% FBS and 10 $\mu\text{g}/\text{mL}$ insulin) for 1 d, and then to complete DMEM red-free (complemented with 10% charcoal stripped fetal calf serum, 1 $\mu\text{g}/\text{mL}$ insulin and L-glutamine, 4 mM) for 6 d. [Note: Charcoal treated fetal calf serum has reduced levels of hormones and growth factors, while red-free DMEM is missing the phenol-red pH indicator, which is an estradiol mimic. By eliminating the influence of other growth hormones or estradiol mimics, the effect of estradiol stimulation can be more accurately evaluated]. Next, DMEM red-free cultured cells, at ~35%–40% confluence, were divided into four batches and further stimulated with (A) E2 (1 nM), (B) E2 (1 nM)/Tam(1 μM), (C) E2 (10 pM)/Tam (1 μM), and (D) Tam (1 μM). Cells

were cultured for an additional 3 d, harvested, and stored at -80°C . The confluence level for the four culturing conditions before harvesting was $\sim 70\%$ – 80% for condition A, $\sim 65\%$ – 75% for condition B, and $\sim 45\%$ – 55% for conditions C and D, demonstrating that in the presence of very low E2 concentrations (i.e., 10 pM), or complete absence of E2, tamoxifen suppresses the proliferation of MCF-7 breast cancer cells. For differential protein expression studies, the same amount of protein extract from each cell state was used for analysis.

Cell Lysis and Protein Extract Processing

Cells were thawed and lysed following a procedure described in previous work [35]. The cell lysis solution was prepared from 1 mL RIPA buffer (500 mM TrisHCl pH 7.4, 1.5 M NaCl, 10% NP-40, 2.5% deoxycholic acid, 10 mM EDTA), 100 μL protease inhibitor cocktail (104 mM AEBSF, 0.08 mM aprotinin, 2 mM leupeptin, 4 mM bestatin, 1.5 mM pepstatin A, 1.4 mM E-64), 100 μL NaF (~ 100 mM) and 50 μL Na_3VO_4 (~ 200 mM) as phosphatase inhibitors, and 8.75 mL of ice cold water. Cell and lysis buffer were mixed in a ratio of $\sim 1:10$, incubated/rocked for ~ 2 – 3 h at 4°C , and centrifuged for ~ 15 min at 13,000 rpm at 4°C . The protein content in the supernatant was measured at 595 nm (Bradford assay) using a SmartSpec Plus Spectrophotometer (Bio-Rad, Hercules, CA), and the samples were stored at -80°C . For further processing, the samples were thawed, the soluble protein extract was treated with urea (8 M) and DTT (4.5 mM) for 1 h at 60°C , diluted $10\times$ with 50 mM NH_4HCO_3 , and digested with trypsin (24 h) at 37°C (substrate:enzyme ratio was $\sim 50:1$). The protein digest solution was cleaned up from salts and buffer components with SPEC-PTC18 solid-phase extraction pipette tips (Varian Inc., Lake Forest, CA). Insulin treated MCF-7 cell extracts were spiked after tryptic digestion with a 9-protein mix digest solution, before SPEC-PTC18 clean-up and iTRAQ labeling. E2 and Tam treated MCF-7 cell extracts were spiked with a ~ 5 μM solution of 8 standard bovine proteins before tryptic digestion (to result in a final concentration, after digestion, of 100 $\mu\text{g}/\text{mL}$ in protein extract and of ~ 0.05 μM in standard proteins).

iTRAQ Labeling

A protocol for iTRAQ labeling was developed and evaluated by using standard proteins (hemoglobin α/β , bovine serum albumin, cytochrome *c*, α -lactalbumin, α -casein, β -casein, albumin, and fetuin) and MCF-7 cells cultured in maintenance medium. The method was subsequently applied to protein differential expression analysis of MCF-7 cells cultured in the presence of E2 and Tam. MCF-7 protein extract tryptic digest solutions (4–125 μg protein content), each processed separately with a SPEC-PTC18 cartridge, were concentrated to ~ 5 – 10 μL with an Eppendorf Vacuufuge (Eppendorf

AG, Hamburg, Germany), re-dissolved in 25–30 μL iTRAQ dissolution buffer (provided in the iTRAQ kit), and treated each with iTRAQ reagent solution for 2 h at room temperature (the iTRAQ reagents being each dissolved in 70 μL of ethanol). The labeled samples were combined in various ratios, cleaned up with SPEC-PTSCX solid-phase extraction pipette tips (Varian Inc.) to eliminate compounds that may interfere with MS analysis, brought to dryness, and ultimately re-dissolved in LC buffer system A. For optimization studies, four aliquots (72 μg protein content each) of an MCF-7 (EMEM/insulin culture) protein extract digest solution, spiked with standards after digestion but before SPEC-PTC18 clean-up (see above), were labeled with iTRAQ reagents 114, 115, 116, and 117, and mixed in a ratio of 0.2:1:1:5. Alternatively, different amounts (4, 20, 20, and 100 μg) of the same cell extract were labeled with iTRAQ reagents and mixed in a ratio of 1:1:1:1, to generate the same final protein ratios of 0.2:1:1:5. In addition, three aliquots (5, 25, and 127 μg) of a standard protein mix digest solution (0.5 μM) were labeled with iTRAQ reagents 114, 116, and 117, and combined 1:1:1 to generate protein ratios of 0.2:1:5. E2 and Tam treated samples (80–100 μg each), spiked with standards before digestion (see above), were labeled with iTRAQ reagents and mixed in A:B:C:D ratios of 1:1:1:1 (4-plex experiment), or in A:A:C:C ratios of 1:1:1:1 (double 2-plex experiment). One 4-plex and one double 2-plex experiment were conducted for optimization work, and a second double 2-plex experiment was conducted for the final differential protein expression study.

RPLC-ESI-MS/MS

RPLC-MS/MS analysis was performed using a micro liquid chromatography system (Agilent Technologies, Palo Alto, CA) and an LTQ ion trap mass spectrometer (Thermo Electron Corp., San Jose, CA). Coupling of the LC system to the LTQ was accomplished via an on-column/no-split injection set up described in detail elsewhere [35]. The separation column was a 100 μm i.d. \times 12 cm capillary packed with 5 μm Zorbax SB-C18 particles (Agilent Technologies). A nanospray emitter was generated by inserting a ~ 1 cm long (20 μm i.d. \times 90 μm o.d.) capillary into the separation column. Mobile phase A was $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ (95:5 vol/vol) and mobile phase B was $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ (20:80 vol/vol), each containing 0.01% CF_3COOH . The volumetric flow rate in the separation column was set to ~ 160 – 180 nL/min, with a 3 h long separation gradient running from 0% to 100% B.

MS data were acquired using data-dependent acquisition conditions: each MS event was followed by zoom/MS² scans on the five top-most intense peaks; zoom scan width was ± 5 *m/z*; dynamic exclusion was enabled at repeat count 1, repeat duration 30 s, exclusion list size 200, exclusion duration 60 s, and exclusion mass width ± 1.5 *m/z*; PQD parameters were set at

isolation width 3 m/z , normalized collision energy 35%, activation Q 0.7, and activation time (T) 0.1 ms; the threshold for MS/MS acquisition was set to 100 counts. Protein identification was performed with the Bioworks 3.3 software (Thermo Electron Corp., San Jose, CA) using a minimally redundant human protein database downloaded from the ExPASy/SwissProt website on January 22, 2007 (37,690 entries, including 12 bovine proteins that were used for spiking the sample) [36]. The database search parameters included the followings settings: number of allowed missed tryptic cleavage sites was set to 2, the peptide tolerance was 2 u, the fragment ion tolerance was 1 u, and only fully tryptic fragments were considered for peptide selection. Five iTRAQ related dynamic modifications (144.1 Da) at the N terminus and at four additional lysine residues were allowed, and all peptides were assigned to unique protein references. Peptides that carried additional amino acid modifications were observed in our study (~13% of all internal tyrosines were found to be labeled with iTRAQ reagents, and ~4% of all lysines were carbamylated), however, these peptides typically displayed low quality tandem mass spectra. To avoid false positive peptide identifications and a distortion of protein iTRAQ ratios, such modifications were not allowed in the final quantitation analysis. The sensitivity threshold and mass tolerance for extracting the iTRAQ ratios were set to 1 and ± 0.5 , respectively (Note: the Bioworks software, to the authors' best knowledge, does not correct for isotope overlap between iTRAQ reporter ions). Data filtering parameters were chosen to generate false positive protein identification rates of <4%, as calculated by searching the MS² scans against a forward-reversed database of proteins (compiled from the original SwissProt database). At the peptide level, mass spectral filtering was accomplished with the Xcorr versus charge state parameter set at minimum 1.5, 2.0, and 3.0 for singly, doubly, and triply charged peptides, respectively. Specific settings for various experiments are described at appropriate locations in the text. At the protein level, only top matching proteins with $P < 0.001$ were considered for analysis. The P value represents the probability of a random match for a peptide, as generated by the Bioworks software from the parameters that characterize the quality of a tandem mass spectrum. For proteins identified by a single peptide, the P value of the protein has the same value as the P value of the matching peptide. For proteins matched by several peptides, the P value is adjusted to reflect increased confidence in the identification of that protein.

Results and Discussions

Breast cancer cell lines are commonly used as model systems to investigate the pathways that lead to the development of cancer [37–45]. In a previous work, we generated a full proteome profile of the MCF-7 cell line using 2DLC-MS/MS technology [35]. The study resulted in the identification of ~2000 proteins ($P <$

0.001), of which 200 were correlated with cellular processes relevant to cancer. The objective of the present study was to develop and assess the effectiveness of a one-step iTRAQ-LC-MS/MS approach, using ion-trap PQD technology, for protein differential expression analysis in complex cellular extracts. A systematic evaluation of iTRAQ labeling efficiency, PQD parameter settings, reproducibility of protein identifications by LC-MS/MS, and reproducibility and accuracy of iTRAQ quantitation, was performed. The identification and quantitation of a large number of proteins, as facilitated by the analysis of cellular extracts, has enabled the implementation of a global normalization process and of a statistical approach for selecting differentially expressed proteins.

iTRAQ Labeling Reaction Efficiency

The iTRAQ reagents are isobaric components used for tagging peptides for MS quantitation. They consist of a reporter group (*N*-methylpiperazine derived), a balance group (carbonyl), and a reactive group (*N*-hydroxysuccinimide ester) that links to peptides via the N-terminus and the lysine side chains [21]. To improve detection limits and quantitation accuracy, all peptides should be fully labeled. Because other sample components having primary amino groups in their structure may interfere with the labeling reaction, proper sample preparation is imperative for the success of the technique. Of specific concern for our experiments were the typical Tris and ammonium bicarbonate buffers that were used during cell extract preparation. The labeling efficiency was investigated with standard mixtures of proteins and MCF-7 extracts that enabled the counting of ~400 and ~2000–4000 peptide hits, respectively. Samples containing ~50–100 μg protein were labeled with iTRAQ reagent 114, and processed as described in the experimental section. The percentage of all N-terminal amino acids, internal lysines and C-terminal lysines that were chemically modified with the 144.1 m/z iTRAQ tag, out of all eligible peptide hits, was determined. To achieve a >90% labeling efficiency, the samples were cleaned-up before iTRAQ labeling with SPEC-PTC18 cartridges to remove the amine-containing buffers, and the labeling reaction was allowed to proceed for 2 h, instead of 1 h, as suggested by the manufacturer. Without effective sample clean-up measures the labeling efficiency was significantly compromised, affecting especially the N-terminal residues where the percentage of labeled amino acids dropped to ~50%. By using optimized conditions for sample processing, the experiment that involved the analysis of E2/Tam treated MCF-7 cells resulted in the labeling of >93% of all lysines and N-terminal amino acids. Specifically, we found that only ~2.4% of C-terminal lysines (out of 2365), ~6.7% of internal lysines (out of 2840), and ~2% of N-terminal lysines (out of 100), as well as ~7.3% of all N-terminal amino acids (out of 4304), were not labeled with iTRAQ reagents. A total of

4304 peptides were considered in this analysis. We note that: all peptide hits matched proteins with $P < 0.001$, but not all peptide hits had $P < 0.001$, per se; peptides were filtered with the Xcorr versus charge state parameter set at 1.5, 2.0, and 3.0; and the peptide level false positive rate was 1.6–2%, i.e., below the ~7%–10% of nonlabeled peptides.

PQD Operation

The PQD method in an LTQ-MS instrument relies on activating the precursor ions at high Q values for a very short time (T), and then performing ion fragmentation/daughter ion collection at low Q values to enable the trapping of low m/z ions. As relatively few results have been reported so far with this novel ion dissociation method, the Q and T parameters, and the collision energy, were varied in an attempt to maximize the ion fragmentation/trapping efficiency, and, therefore, increase the number of identified/quantified proteins (see PQD optimization in Supplementary material, which can be found in the electronic version of this article). As the Q = 0.7, T = 0.1 ms, and CE = 35% conditions generated some of the largest number of identified proteins ($P < 0.001$), all future experiments were performed using these parameter settings. Overall, however, PQD detection enabled the identification of only ~50%–65% of proteins that were detected with conventional CID. This outcome was a result of less efficient peptide fragmentation in the PQD operation mode of the ion trap (many tandem mass spectra being dominated by the undissociated parent ion). In this study, best PQD performance was observed at CE settings of 31%–35%, close to the typical values that are used for CID. Over time, however, changes in the optimal CE values have been observed, thus, CE optimization before performing a new set of experiments was necessary. Nevertheless, a strong and recurring distortion of iTRAQ reporter ion intensities with increased collision energy settings was not observed (i.e., consistently lower intensity m/z 114 versus m/z 117 [32]). The values of the global iTRAQ ratios confirmed the lack of a consistent bias in the detection of 114–117

reporter ions. Detailed explanations on how the global iTRAQ was calculated are provided in the following sections.

Reproducibility of Protein Identifications

Before conducting quantitative analysis of unknown samples, a qualitative evaluation of the iTRAQ-LC-MS/MS protocol, in terms of number of identified proteins, percent of false positives, and reproducibility of protein identifications, was carried out. For this purpose, MCF-7 cell extracts (~72 μg protein content in each batch) were labeled with iTRAQ reagents 114, 115, 116, and 117, and mixed in a ratio of 0.2:1:1:5 (i.e., ~3 μg :15 μg :15 μg :72 μg) to cover a dynamic range of 25 (see the Experimental section). Batch number three was labeled with reagent 116, and was used as a reference for quantitation. Further discussions in the text will analyze the outcome of this experiment as a whole, or in terms of mixing ratios 0.2:1, 1:1, or 5:1 alone. Depending on the amount injected on the LC nano-column, the MS/MS analysis of labeled MCF-7 extracts resulted, typically, in the identification of 100–500 proteins. Three consecutive injections of a sample containing ~8 μg protein extract (according to the initial protein concentrations measured with the Bradford assay, and ignoring possible losses during sample processing) resulted in the identification of 272–305 proteins per LC-MS/MS run. The combined results of all three runs summed up to 472 proteins, and will be referred from now on as “multiconsensus” results, as defined by the Bioworks software (Table 1). These proteins were matched by a total of 2378 peptides (MS^2 scans), of which 1179 were unique. Only proteins with $P < 0.001$ that were matched by peptides with Xcorr versus charge state values of 1.9, 2.2, and 3.8 for $z = 1, 2,$ and $3,$ respectively, were considered for comparison. Using such conditions, with either PQD or CID, the typical protein overlap between two consecutive runs was ~60%–67%, and between three runs ~50% (Figure 1). Such low overlap between protein I.D.s relates to the low abundant proteins that generate very few and low intensity peptides. As a result of small changes in

Table 1. Identification/quantitation of MCF-7 proteins, and run-to-run reproducibility of global iTRAQ ratio measurements. Conditions: MCF-7 cells were cultured in EMEM/insulin, labeled with iTRAQ reagents 114, 115, 116, and 117, and mixed in a ratio of 0.2:1:1:5. iTRAQ ion 116 was used as a reference. The protein concentration in the final sample subjected to LC-MS/MS analysis was ~1 $\mu\text{g}/\mu\text{L}$. LC injection volumes were 8 μL . Proteins that were counted as detected/quantified had $P < 0.001$, and were matched by peptides that passed the Xcorr versus charge state filter (1.9, 2.2, 3.8). Multiconsensus results were generated from three LC-MS/MS runs.

Run #	Theoretical iTRAQ ratios						
	0.2:1			1:1		5:1	
	# Detected proteins	# Quantified proteins (%)	Global iTRAQ ratio	# Quantified proteins (%)	Global iTRAQ ratio	# Quantified proteins (%)	Global iTRAQ ratio
Run 1	272	199 (73%)	0.37	230 (85%)	1.03	223 (86%)	4.57
Run 2	305	222 (73%)	0.38	254 (83%)	1.16	256 (84%)	5.16
Run 3	296	214 (72%)	0.35	249 (84%)	1.01	250 (84%)	4.45
Multiconsensus	472	389 (82%)	0.36	421 (89%)	1.04	424 (90%)	4.65

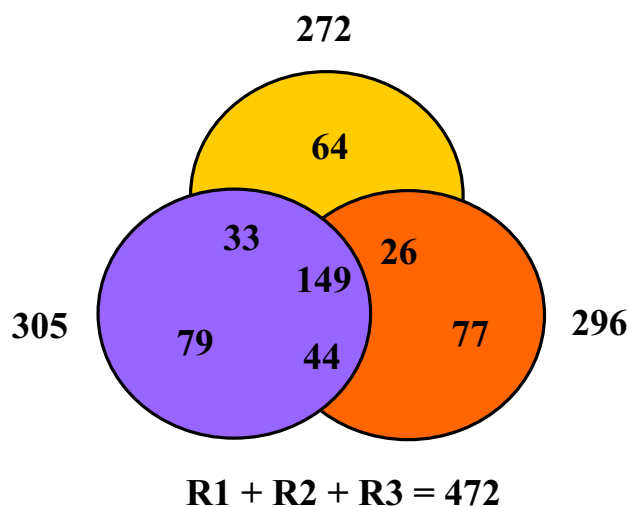


Figure 1. Reproducibility of protein identifications across replicate LC-MS/MS runs. Conditions: MCF-7 cells were cultured in EMEM/insulin, labeled with iTRAQ reagents 114, 115, 116, and 117, and mixed in a ratio of 0.2:1:1:5. The protein concentration in the final sample subjected to LC-MS/MS analysis was $\sim 1 \mu\text{g}/\mu\text{L}$. LC injection volumes were $8 \mu\text{L}$. Only proteins with $P < 0.001$, and that were matched by peptides that passed the Xcorr versus charge state filter (1.9, 2.2, 3.8), were considered in the analysis. Total unique proteins identified in all three runs (R1, R2, R3) was 472.

retention times, such peptides can co-elute with different background ions during chromatographic analysis, and may, or may not be selected for fragmentation during data dependent MS (typical intra-column reproducibility of retention times was $< 2\%$). Thus, the number of identified proteins can change from one chromatographic run to another (typical variations were $\sim 5\%$ – 10%). We have shown, however, that if the stringency of the data filtering parameters is high, and if only proteins that are matched by ≥ 2 unique peptides are considered for comparison, the reproducibility between consecutive runs can be as high as $\sim 90\%$ – 98% [35].

Reproducibility of iTRAQ Quantitation

The quantitation reproducibility was evaluated in terms of number of proteins that were quantified in replicate

LC-MS/MS runs, and of individual or global RSD values of iTRAQ ratios. Certain peptides that generated good quality tandem mass spectra for identification purposes did not produce iTRAQ reporter ions in the low m/z region of the mass spectrum, presumably because the PQD fragmentation was not efficient. Alternatively, the intensity of the reporter ions in the mass spectrum was very low. As a result, these peptides and the corresponding proteins (if matched by such peptides only) could not be quantified. We note that the protein iTRAQ ratios were calculated as an average of all contributing peptide iTRAQ measurements. For the example provided in the previous section, the run-to-run variability, in terms of quantified proteins in three consecutive LC-MS/MS analyses, is also summarized in Table 1. For any single analysis, the number of quantified proteins represented $\sim 72\%$ – 86% of the identified proteins, while for the multiconsensus results, this number increased to 82% – 90% . The injection of $\sim 8 \mu\text{g}$ sample ($\sim 0.22 \mu\text{g}:1.1 \mu\text{g}:1.1 \mu\text{g}:5.5 \mu\text{g}$) enabled the quantitation of ~ 199 – 222 , 230 – 254 , and 223 – 256 proteins for sample ratios 0.2:1, 1:1, and 5:1, respectively. An advantage of performing repetitive LC-MS/MS analyses is that complementary information can be obtained from several injections, thereby, increasing not only the total number of identified or quantified proteins, but also the protein sequence coverage and quantitation accuracy. The effect of protein sequence coverage, i.e., of the number of unique peptides/protein, on protein identification and quantitation, is shown in Table 2 (generated with the same MS/MS data that were used for generating Table 1). As expected, the more peptide matches required for a protein I.D., the smaller the subset of proteins that could be identified. The percentage of quantified proteins (out of total identified) increased, however, from 82% – 90% to 99% – 100% , as the number of matching peptides per protein was increased from one to five, respectively. Most importantly, for quantitation purposes, $\geq 94\%$ of the proteins that were identified by at least two peptides (as typically required for confident protein identification) generated measurable iTRAQ ratios for quantitation.

The RSD of iTRAQ ratios for individual protein standards spiked into the MCF-7 extracts was in the range of $\sim 5\%$ – 50% (across three repetitive runs). This

Table 2. Identification and quantitation of proteins in the MCF-7 cell line as a function of the number of unique peptides used for protein identification. Conditions: MCF-7 cells were cultured in EMEM/insulin, labeled with iTRAQ reagents 114, 115, 116, and 117, and mixed in a ratio of 0.2:1:1:5. The protein concentration in the final sample subjected to LC-MS/MS analysis was $\sim 1 \mu\text{g}/\mu\text{L}$. LC injection volumes were $8 \mu\text{L}$. Proteins that were counted as detected/quantified had $P < 0.001$, and were matched by peptides that passed the Xcorr versus charge state filter (1.9, 2.2, 3.8). Results were generated from a multiconsensus file prepared from three LC-MS/MS runs.

# Unique peptides/protein	# Detected proteins	iTRAQ ratio 0.2:1	iTRAQ ratio 1:1	iTRAQ ratio 5:1
		# Quantified proteins (%)		
1 peptide	472	389 (82%)	421 (89%)	424 (90%)
2 peptides	268	251 (94%)	262 (98%)	263 (98%)
3 peptides	169	166 (98%)	169 (100%)	169 (100%)
4 peptides	127	125 (98%)	127 (100%)	127 (100%)
5 peptides	95	94 (99%)	95 (100%)	95 (100%)

relatively broad range was attributed to the large variance associated with the measurement of iTRAQ ratios for individual peptides. While it was beyond the purpose of this study to develop a statistical method that would incorporate the contribution of peptide variance into the protein-level variance, and into the accuracy of iTRAQ quantitation, it is worth mentioning the complexity of the problem. The possible individual contributors to the iTRAQ ratio variance of a protein are: the number of unique matching peptides, the various charge states of the same peptide, the number of tandem mass spectra per unique peptide and per charge state, the number of iTRAQ tags per peptide, the number of LC-MS/MS replicates that are performed to generate multiconsensus results, and the intensity of the signal. Overall, in this study, variations as high as 300%–500% were occasionally encountered for repetitive measurements of a peptide iTRAQ ratio, and the variations were generally higher for low intensity signals (the results are, however, in agreement with the coefficients of variation reported for PQD detection and for label-based quantitative methods [25, 30–32, 46–48]). For example, cytokeratin 18 from one of the MCF-7 extracts was matched by a total of 190 tandem mass spectra (34 unique peptides), of which 105 generated measurable iTRAQ ratios. The RSD of all contributing peptide iTRAQ ratios was as high as 146%. For the case of only one cytokeratin peptide (V*K*LEAEIATYRR, “*” indicating the iTRAQ tag) that was observed as a doubly and triply charged ion, and that was matched by 36 tandem mass spectra of which 18 generated measurable iTRAQ ratios, the RSD was smaller, i.e., 55%. While the variance across all peptides was, obviously, larger than for a single peptide, a larger number of total measurements for any given protein resulted in a more reproducible (and ultimately more accurate) protein iTRAQ ratio. We note that the variation in peptide iTRAQ ratios was not introduced by sequence redundancy between different keratins (i.e., peptide contributions from various keratins with different abundance), but rather by random differences between iTRAQ ratios generated for the same peptide by different tandem mass spectra. Only one of the keratin 18 matching peptides (IVLQIDNAR) was also identified in keratin 19, however, the contribution of this peptide to the iTRAQ variance was minimal (1 out of 105 measurements). Low *m/z* contaminants that could overlap with the iTRAQ reporter ions are not observable on a low-resolution ion trap mass spectrometer, thus, it was not possible to assess whether such interferences affected the outcome of these experiments or not.

Due to the relatively broad range of individual protein RSD values, to gain a better understanding of how well an experiment evolved, we introduced the calculation of a “global iTRAQ ratio”, defined as the average of all protein iTRAQ ratios within a given dataset that passed certain data filtering criteria (for example, Xcorr versus charge state and P-threshold).

We hypothesized that this global iTRAQ could provide a better measure for a preliminary assessment of overall quantitation reproducibility. As the calculation of the global iTRAQ involved averaging iTRAQ ratios for hundreds of proteins, the global RSDs were indeed much smaller than that of individual proteins, i.e., ~4%–8% (as calculated from the three replicate analyses shown in Table 1). Experiments performed months later on the same sample demonstrated similar consistency in the values of the global iTRAQ, and confirmed its utility for evaluating not only global quantitation reproducibility but also accuracy.

Accuracy of iTRAQ Quantitation

Relative quantitation of proteins in cellular extracts relies on the following assumptions: (1) the total amount of protein extract considered for analysis is the same for each cell state, (2) the expression level of most proteins does not change in response to the perturbation considered in the study, and (3) the change in expression level of some proteins, if any, as a result of the perturbation, will have a negligible impact on the overall quantitation/normalization protocols. As a result of these assumptions, we speculated that (1) the value of the experimentally determined global iTRAQ ratios should be close to the theoretical values, i.e., to the mixing ratios of the proteins from each cell state, (2) any major departure of the experimental global values from the theoretical values should be the result of a global bias that could be corrected by a normalization process, and (3) the global iTRAQ ratios could be used for (global) data normalization, such use being supported by the reproducibility of data shown in Table 1. Major contributors to a global bias could be factors related to: (a) sample processing steps (for example, inaccurate protein concentration measurements, non-uniform recovery of entire extracts from C18 clean-up cartridges, loss of sample during precipitation/redissolution, etc.); (b) data processing artifacts (for example, it was observed that the iTRAQ values were somewhat dependent on the threshold setting parameter that was used by Bioworks for their calculation); (c) dependence of iTRAQ reporter ion intensities on collision energy (as reported by Griffin et al. [32], even though in our data we did not observe such an effect); and, (d) isotope contamination between iTRAQ reagents (consistent contamination, however, was not observed in this work). To correct for this bias, individual protein iTRAQ values were calculated by normalizing the experimental data with the aid of correction factors, according to the following equations:

$$\text{iTRAQ}_{\text{PRON}} = \text{iTRAQ}_{\text{PRO}} / \text{CF} \quad (1)$$

$$\text{CF} = \text{iTRAQ}_{\text{C}} / \text{iTRAQ}_{\text{THEOR}} \quad (2)$$

$$i\text{TRAQ}_G = \left(\sum i\text{TRAQ}_{\text{PRO}}/n \right) \quad (3)$$

where $i\text{TRAQ}_{\text{PRO}}$ is the normalized $i\text{TRAQ}$ ratio of a protein, $i\text{TRAQ}_{\text{PRO}}$ is the experimentally determined $i\text{TRAQ}$ ratio of a protein (i.e., the average $i\text{TRAQ}$ ratio of all corresponding peptides), $i\text{TRAQ}_G$ is the global $i\text{TRAQ}$ ratio (i.e., the average of all protein $i\text{TRAQ}$ ratios within a given dataset), $i\text{TRAQ}_{\text{THEOR}}$ is the theoretical $i\text{TRAQ}$ ratio (i.e., the mixing ratio), CF is the correction factor (for theoretical mixing ratios of 1:1:1:1, the correction factor is equal to the global $i\text{TRAQ}$), and n is the number of proteins in the dataset.

Table 3 displays global $i\text{TRAQ}$ ratios for various datasets, and the corresponding correction factors that were generated for normalization purposes. Data are provided for six independent $i\text{TRAQ}$ experiments involving the labeling of different sample amounts and mixing of samples in different ratios, as well as different cell culturing conditions: one standard protein mixture, two sets of MCF-7 cells cultured in the presence of insulin, and three sets of MCF-7 cells cultured in the presence of E2/Tam. Three to five LC-MS/MS injections were performed for each experiment to generate multiconsensus results. Rows 6 and 7 in Table 3 refer to the same $i\text{TRAQ}$ experiment, but involve different data filtering parameters. Most global correction factors had values in the range of ~ 0.5 – 2.2 , with a few extremes in the range of ~ 0.3 – 4 . The cell states that were used as a reference for quantitation were assigned an $i\text{TRAQ}$ ratio of 1. Generally, the errors were higher for the lower end of the quantitation scale, with the least sample considered for analysis (note one case of a correction factor of 3.95 for one of the smallest batches of labeled standard proteins). The effect of the global normalization process on individual $i\text{TRAQ}$ ratios for standard proteins spiked into MCF-7 extracts is shown in Table 4. Results are shown for three independent $i\text{TRAQ}$ experiments involving cell extracts 3, 4, and 6 from Table 3, using unique or different cell culturing conditions. For insulin-only cultured cells, and $i\text{TRAQ}$

mixing ratios of 0.2:1:1:5 (Table 4/column 2), the normalized experimental measurements were in the range of (0.14–0.3):(0.82–1.60):1:(2.92,–6.25). For cells cultured using two different conditions, in the presence of E2 and Tam, and $i\text{TRAQ}$ mixing ratios of 1:1:1:1 (Table 4/columns 3 and 4), the normalized measurements were in the range of (0.48–2.68). The quantitation errors seemed to be somewhat higher when additional biases were introduced by culturing and processing the cells independently, processes that involved separate cell harvesting, lysing, protein concentration measurement through the Bradford assay and tryptic digestion (note a range of 0.48–2.68 for E2/Tam cultures versus a range of 0.82–1.60 for insulin cultures at 1:1 mixing ratios). This observation was evident from the global correction factors, as well, shown in Table 3 (column 8): for E2/Tam cultures the global correction factors were in the range of 0.31–3.03, while for the insulin cultures in the range of 0.84–1.8. Overall, at the individual protein level, the global normalization process reduced the quantitation errors to less than 2- to 3-fold.

Once a global average ($i\text{TRAQ}_G$) was determined for an LC-MS/MS experiment, the percent variation for each individual protein was calculated according to eq 4:

% Variation

$$= [(i\text{TRAQ}_{\text{PRO}} - i\text{TRAQ}_G)/i\text{TRAQ}_G] * 100 \quad (4)$$

For the previously described LC-MS/MS experiment, with $i\text{TRAQ}$ mixing ratios of 0.2:1:1:5, the distribution of the $i\text{TRAQ}$ % variations for all quantified MCF-7 proteins is shown in Figure 2. Overall, combined datasets provided better results than any individual set. Specifically, for a mixing ratio of 1:1, the multiconsensus results revealed that $\sim 60\%$ of proteins could be quantified with \pm (0%–30%) accuracy (Figure 2a). In addition, for mixing ratios that involved larger amounts of proteins and generated more intense signals, the quantitation was more accurate, i.e., a larger percentage

Table 3. Global $i\text{TRAQ}$ ratios and corresponding correction factors measured for various samples and experimental conditions.

Sample	Amounts labeled (μg)	Injection amount/volume	# Inj.	# Proteins ($P < 0.001$)	Mixing ratios	Global $i\text{TRAQ}$	Correction factors
						ratios ^c 114:115:116:117	
1. Standard 9 mix ^a	5:25:127	$\sim 4 \mu\text{g}/8 \mu\text{L}$	2	9	0.2:1:5	0.79:1:6.27 ^d	3.95:1:1.25
2. MCF-7/insulin ^a	72:72:72:72	$\sim 8 \mu\text{g}/8 \mu\text{L}$	3	472	0.2:1:1:5	0.36:1.04:1:4.65	1.80:1.04:1:0.93
3. MCF-7/insulin ^a	4:20:20:100	$\sim 3 \mu\text{g}/8 \mu\text{L}$	3	468	0.2:1:1:5	0.32:0.87:1:4.20	1.60:0.87:1:0.84
4. MCF-7/E2/Tam/4plex ^a	100:100:100:100 ^e	40 μL	3	145	1:1:1:1	1:0.82:0.54:0.31	1:0.82:0.54:0.31
5. MCF-7/E2/Tam/2plex ^b	80:80:80:80 ^e	40 μL	3	154	(1:1):(1:1)	1:1.61:3.03:2.21	1:1.61:3.03:2.21
6. MCF-7/E2/Tam/2plex ^a	100:100:100:100	$\sim 16 \mu\text{g}/8 \mu\text{L}$	5	407	(1:1):(1:1)	1:2.23:1.88:1.44	1:2.23:1.88:1.44
7. MCF-7/E2/Tam/2plex ^b	100:100:100:100	$\sim 16 \mu\text{g}/8 \mu\text{L}$	5	255	(1:1):(1:1)	1:2.14:1.84:1.43	1:2.14:1.84:1.43

^aPeptides were filtered with the Xcorr versus charge state filter set at 1.9, 2.2, and 3.8, and all proteins with $P < 0.001$ were counted.

^bPeptides were filtered with the Xcorr versus charge state filter set at 1.5, 2.0, and 3.0, and only proteins with $P < 0.001$ and matched by two peptide $i\text{TRAQ}$ measurements were counted (see additional selection criteria in the text).

^cThe reference ions have an $i\text{TRAQ}$ value of 1.

^dIn this experiment, $i\text{TRAQ}$ labels 116, 114, and 117 were used for labeling 5, 25, and 127 μg protein, respectively (label 114 was the reference).

^eVery little sample was available for analysis, and part of the sample was lost during precipitation; as a result, the exact injection amount is not known, and a much smaller number of proteins were identified/quantified.

Inj. stands for LC-MS/MS injections.

Table 4. Normalized iTRAQ ratios for standard proteins spiked into MCF-7 cell extracts. Conditions: MCF-7 cells were grown either in the presence of insulin (10 $\mu\text{g}/\text{mL}$), or E2/Tam, using the following conditions: (A) E2 (1 nM), (B) E2 (1 nM)/Tam (1 μM), (C) E2 (10 pM)/Tam (1 μM), and (D) Tam (1 μM). For the insulin stimulated cells, different aliquots of the same cell state were labeled with iTRAQ reagents 114, 115, 116, and 117, and mixed in a ratio of 0.2:1:1:5. For the E2/Tam 4-plex experiment, cell conditions A, B, C, and D were labeled with reagents 114, 115, 116, and 117, respectively, and mixed in a ratio A:B:C:D of 1:1:1:1. For the E2/Tam double 2-plex experiment, cell condition A was labeled with reagents 114 and 115, cell condition C was labeled with reagents 116 and 117, and the samples were mixed in a ratio (A:A):(C:C) of (1:1):(1:1). Peptides were filtered with the Xcorr versus charge state filter set at 1.9, 2.2, and 3.8, and all proteins with $P < 0.001$ were counted in the analysis.

Standard proteins	MCF-7/insulin (0.2:1:1:5)	MCF-7/E2/Tam-4plex (1:1:1:1)	MCF-7/E2/Tam-2plex (1:1):(1:1)
	Correction factors		
	1.60:0.87:1:0.84	1:0.82:0.54:0.31	1:2.23:1.88:1.44
	Normalized iTRAQ ratios		
Hemoglobin α	0.30:0.93:1:5.17	1:1.84:1.56:2.29	1:1.33:1.58:1.74
Hemoglobin β	0.24:0.90:1:4.76	1:1.20:1.22:1.35	1:1.04:1.06:1.15
α -2-HS-glycoprotein	0.26:1.60:1:6.25	1:1.11:1.15:1.26	1:1.16:2.11:2.16
α -S1-casein	0.26:1.10:1:5.02	1:0.93:0.96:0.77	1:1.15:1.48:1.37
α -S2-casein	0.15:0.92:1:4.48	1:0.80:1.44:0.68	1:0.85:0.94:1.19
β -Casein	0.24:0.89:1:2.92	1:1.34:1.57:1.29	1:0.52:1.06:1.27
Carbonic anhydrase	0.24:1.21:1:4.99	1:1.12:1:00:1.45	1:0.92:1.16:0.87
Cytochrome <i>c</i>	0.22:1.02:1:4.77	1:2.02:1.46:2.68	1:0.84:1.03:1.17
α -Lactalbumin	0.14:0.82:1:4.04	1:0.78:1.07:0.48	1:1.4:0.72:0.61
BSA	0.28:0.98:1:4.85	N/A	N/A

of proteins could be quantified with better accuracy (see Figure 2b, 0.2:1 versus 1:1 and 5:1 ratios). Similar trends were obtained for the standard proteins that were spiked into the MCF-7 extract. However, as the accuracy of the analytical method is sensitive to the complexity of the sample [5], the analysis of standards

alone, for example at 5:1 mixing ratios, resulted in the quantification of all proteins within $\pm 30\%$ accuracy of their theoretical ratios.

The % variation of iTRAQ values (1:1 mixing ratio) as a function of the number of unique peptides that matched any given protein is shown in Figure 3. Evidently, the accuracy of the method increased with the number of peptides that were required to identify/quantify a given protein. The percentage of proteins that was quantified with $\pm(0\%–30\%)$ accuracy, increased from $\sim 60\%$ to $\sim 85\%$, as the number of unique peptides/protein was increased from 1 to 5. The number of quantifiable proteins dropped significantly, however, if such stringent conditions were used (see also Table 2). The standard deviations (SD) of iTRAQ ratio distributions for proteins measured by 1 to 5 peptides were 0.47, 0.38, 0.30, 0.27, and 0.23, respectively, (we

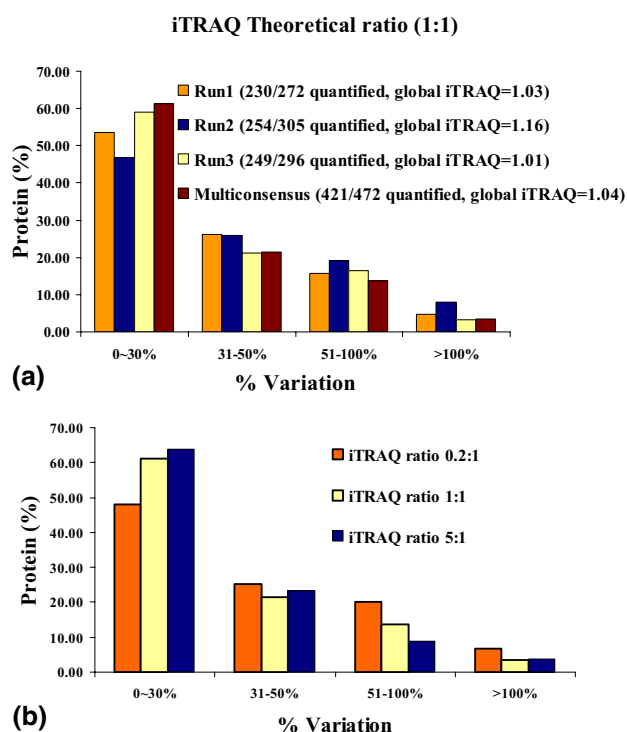


Figure 2. Distribution of experimental iTRAQ values for quantified proteins in the MCF-7 cell extract. Conditions were the same as provided in Table 1. (a) % Variations for individual LC-MS/MS runs and multiconsensus results are provided for the 1:1 protein mixing ratio. (b) % Variations for multiconsensus results are provided for 0.2:1, 1:1, and 5:1 protein mixing ratios.

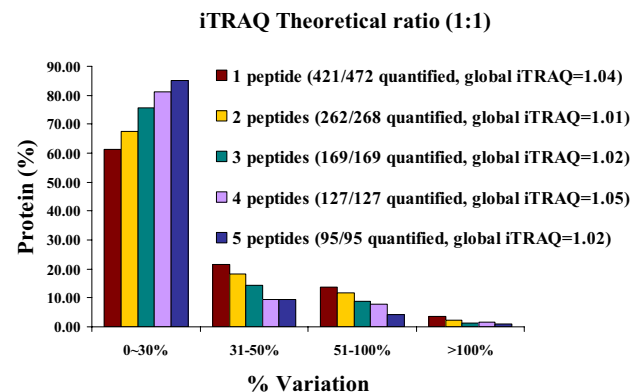


Figure 3. Distribution of experimental iTRAQ values for quantified proteins in the MCF-7 cell extract as a function of unique peptides/protein (1:1 mixing ratio). Conditions were the same as provided in Table 1. Multiconsensus data from triplicate LC-MS/MS runs were used in the analysis.

note that the distribution of iTRAQ values is not symmetrical around the mean). These observations emphasize once again the importance of making a sufficient number of iTRAQ measurements for every single protein to obtain reproducible and accurate quantitative results.

Protein Differential Expression Analysis

In an attempt to verify the applicability of these findings to differential expression analysis and biomarker discovery in complex cellular extracts, three independent iTRAQ experiments with E2 and Tam treated MCF-7 cells were performed. One 4-plex (involving four cell culturing conditions) and one double 2-plex experiment (involving two experimental replicates of two cell culturing conditions) were conducted for optimization work, and a second double 2-plex experiment was conducted for the final differential protein expression analysis. Global correction factors and standard protein spike iTRAQ ratios, as discussed earlier, are provided in Table 3/rows 4 to 7, and Table 4/columns 3 and 4, respectively. The double 2-plex experiment was performed with MCF-7 cells grown in the presence of E2 (at physiological levels of 1 nM) as a control, and E2 (10 pM) + Tam (at lethal levels of 1 μ M) as a treatment. The concentration of E2 in the Tam treated sample was maintained below levels that can counteract the Tam effect, i.e., at 10 pM. Two experimental replicates of each cell state were processed and labeled separately [E2 (1 nM) treated cells were labeled with reagents 114 and 115, and E2 (10 pM)/Tam (1 μ M) treated cells were labeled with reagents 116 and 117], mixed, cleaned-up, and analyzed by five consecutive LC-MS/MS runs. A 36% increment in the number of identified proteins was observed when combining the results from the first two consecutive runs, and a 13% increment was observed after adding results from a third and fourth injection. After the fifth injection, the increment in the number of protein I.D.s was only ~8%, thus, further injections were not performed. A total of 407 proteins ($P < 0.001$) were identified in the five combined LC-MS/MS runs when the peptides were filtered with the Xcorr versus charge state filter set at 1.9, 2.2, and 3.8 for $z = 1, 2,$ and 3, respectively. The global correction factors ranged from 1.44 to 2.23 (Table 3/row 6), and the normalized iTRAQ values for the 9 proteins spiked into each batch of cells ranged from 0.52 to 2.16 (Table 4/column 4).

Initially, for all datasets, all proteins with $P < 0.001$ that were matched by peptides that passed the Xcorr versus charge state filter (1.9, 2.2, 3.8) were considered for analysis. Using such conditions, the false positive protein identification rate was ~4%. However, when only proteins matched by two peptides were considered, as typically required for confident protein identification, the false positive rate dropped to zero. As a result, for the E2/Tam treated samples, somewhat less stringent data filtering parameters were chosen, i.e., the Xcorr versus charge state peptide filter was set to 1.5,

2.0, and 3.0 for singly, doubly, and triply charged ions, respectively. A total of 530 proteins ($P < 0.001$) were identified, of which 302 proteins were matched by two peptides (spectral counts) with a false positive identification rate similar to the previous case, i.e., ~4%. The list of 302 proteins was identified by a total of 4074 tandem mass spectra (2648 with $P < 0.001$) corresponding to 1388 unique peptides (732 with $P < 0.001$). To improve overall quantitation accuracy, the list of 302 proteins was further refined according to the following criteria (Figure 4): (1) As not all peptides generated iTRAQ ratios, only the proteins that were quantified by at least two complete sets of peptide iTRAQ measurements were considered for analysis (i.e., two sets of 116/114, 117/114, 116/115, and 117/115 ratios). Multiple iTRAQ measurements on the same peptide were allowed to improve quantitation accuracy; (2) a preliminary global normalization was performed to obtain an estimate of peptide/protein iTRAQ ratios; (3) proteins that were quantified by only two sets of peptide iTRAQ measurements that were clearly contradictory after preliminary normalization (e.g., one showing up-regulation and the other down-regulation at a larger than 2-fold level), were eliminated from the list; (4) for proteins quantified by multiple sets of iTRAQ measurements on the same peptide, tandem mass spectra that

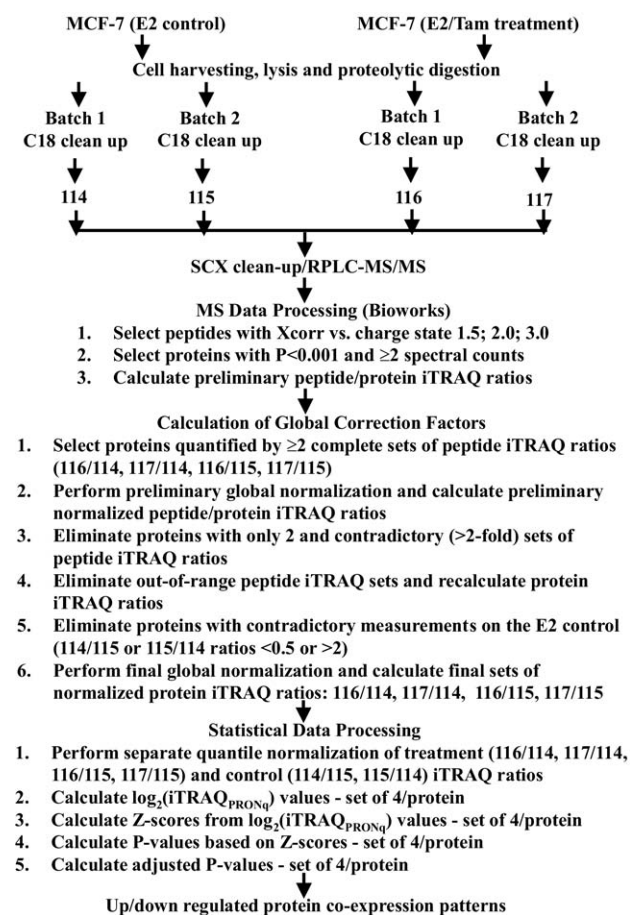


Figure 4. iTRAQ experimental and data processing outline.

displayed out-of-range iTRAQ ratios were eliminated, and the protein average iTRAQ values were manually re-calculated (e.g., when a >5-fold difference between the value of the iTRAQ ratio for one peptide versus other peptides with the same amino acid sequence was observed); (5) proteins for which the two control samples (E2 treated cells) displayed after global normalization a greater than 2-fold change one versus the other (i.e., 115/114 or 114/115 ratios were <0.5 or >2) were eliminated from the list; and (6) the final global correction factors for normalization were recalculated. These additional data filtering criteria resulted, ultimately, in a list of 255 proteins (>93% being quantified by ≥ 3 sets of iTRAQ measurements). We note that, overall, there was no significant change in the value of the global correction factors when less stringent criteria were used for peptide filtering according to Xcorr versus charge state, and when the proteins were selected according to the above described strategy (Table 3, row 7 versus row 6).

After global normalization, the global iTRAQ ratios matched the theoretical values, i.e., (1:1):(1:1). However, individual protein iTRAQ ratios continued to display, occasionally, either a relatively broad range of values, or conflicting results (e.g., some proteins displayed up/down-regulation when ion 114, but not ion 115, was used as a reference, or vice versa). As a result, the set of 255 proteins was subjected to a more advanced statistical evaluation. The two treatment datasets corresponding to the E2/Tam treated cells (labeled with reagents 116 and 117) were quantified relative to both control datasets corresponding to the E2 treated cells (labeled with reagents 114 and 115). Thus, 4 sets of iTRAQ ratios were generated for each protein: 116/114, 116/115, 117/114 and 117/115. The standard deviations of each iTRAQ dataset were as follows: (a) for ion 114 considered as a reference, $SD_{114/114} = 0$, $SD_{115/114} = 0.31$, $SD_{116/114} = 0.38$ and $SD_{117/114} = 0.43$; and (b) for ion 115 considered as a reference, $SD_{114/115} = 0.29$, $SD_{115/115} = 0$, $SD_{116/115} = 0.47$ and $SD_{117/115} = 0.53$. The standard deviations of the two control sets ($SD_{115/114} = 0.31$ and $SD_{114/115} = 0.29$) were obviously of smaller value and defined the intrinsic spread of the iTRAQ measurements characteristic to this sample processing protocol and ion trap PQD technology. We note that the iTRAQ ratios 114/115 and 115/114—as calculated by the BioWorks software—did not always represent the reciprocal of one another. The standard deviations of the four treatment sets were of larger value ($0.38 \leq SD \leq 0.53$), as a result of protein expression ratio alterations due to the Tam treatment. To produce consistent threshold values for outlier identification, the datasets were subjected to quantile normalization [49], performed separately for the control and the treatment data, to generate sets not only having the same mean, but having the same distribution. The two controls (114/115 and 115/114) generated an $SD = 0.3$, and the four treatments (116/114, 117/114, 116/115, and 117/115) an $SD = 0.45$. Based on the SD of the control, roughly, proteins with a ~ 2 -fold change in expression level (beyond $1 + 3 \times SD$;

$SD = 0.3$) would qualify as outliers. As a result, the data were further evaluated on a \log_2 scale. Overall, for each protein, four sets of the following quantities were calculated: quantile normalized iTRAQ ratio ($iTRAQ_{PRONq}$), $\log_2(iTRAQ_{PRONq})$, Z-score of $\log_2(iTRAQ_{PRONq})$, P value and adjusted P value. Within each dataset (116/114, 117/114, 116/115, and 117/115), individual Z scores for each protein were calculated according to eq 5:

$$Z = [\log_2(iTRAQ_{PRONq}) - MEAN_{\log_2 iTRAQ(C)}] / SD_{\log_2 iTRAQ(C)} \quad (5)$$

where $MEAN_{\log_2 iTRAQ(C)}$ and $SD_{\log_2 iTRAQ(C)}$ are the mean and SD of the \log_2 iTRAQ ratios for the quantile normalized control datasets ($SD_{\log_2 iTRAQ(C)} = 0.43$). The goal was to identify those proteins whose relative abundances (i.e., individual iTRAQ ratios) change in response to Tam treatment. The value of $MEAN_{\log_2 iTRAQ(C)}$ was very close to zero, which is the expected log-ratio value under the null hypothesis of no change. Corresponding to each Z-score in (5), a P value was computed (for a two-sided test) as:

$$P\text{-value} = 2 * [1 - CDF (ABS(Z))] \quad (6)$$

where CDF denotes the cumulative distribution function of the standard normal distribution and ABS denotes the absolute value (we note that these computed P -values refer to protein quantitation, and are different from the MS^2 -related P -values generated by the BioWorks software for protein identification).

To control for multiple testing within each of the four treatment experiments (116/114, 117/114, 116/115, and 117/115), we used the Benjamini-Hochberg method [50] for controlling the (expected) false discovery rate (FDR). From each of the four lists of P values we computed adjusted P values. The adjusted P value of a protein is the level at which the FDR would be controlled if that protein and all other proteins with smaller P values would be considered as having significant changes in relative abundance. For example, if the P value corresponding to the Z-score of a protein was 0.001 and the adjusted P value was 0.018, then calling significant the relative abundance change for this protein (and any other protein with P value < 0.001) means that the FDR is controlled at level 0.018 (i.e., 1.8%). Note that the adjusted P value is always equal to or larger than the (original) P value because it accounts for multiple testing. Providing adjusted P values, as in Table 5, is more informative than only providing the information whether a certain protein is significant at a pre-chosen FDR level (say 0.05 or 5%) or not significant. For example, if the FDR level was prespecified at the 0.05 level, then a protein with adjusted P value of 0.066 and a protein with adjusted P value of 0.812 would both be considered as having no significant change in relative abundance, although the first protein clearly has some evidence supporting a change in abundance.

Table 5. Differentially expressed proteins in the E2/Tam double 2-plex experiment (ordered in decreasing value of $-10 \times \lg P$, P being the MS²-related P values). Conditions: MCF-7 cells were grown in the presence of E2/Tam using the following conditions: (A) E2 (1 nM) and (C) E2 (10 pM)/Tam (1 μ M). Cell condition A was labeled with reagents 114 and 115, and cell condition C was labeled with reagents 116 and 117. The samples were mixed in a ratio (A:A):(C:C) of (1:1):(1:1). Peptides were filtered with the Xcorr versus charge state filter set at 1.5, 2.0, and 3.0, and only proteins with $P < 0.001$ and matched by two peptide iTRAQ measurements were counted (see additional selection criteria in the text).

SwissProt i.d.	Protein	$-10 \times \lg P$ (MS ²)	MW	Peptide Hits/ Unique	iTRAQ sets	$\text{Log}_2(\text{iTRAQ}_{\text{PRONG}})/\text{adjusted } P \text{ value}$				#
						116/114	117/114	116/115	117/115	
Q15365 UP	Poly(rC)-binding protein 1; nucleic acid binding, accelerator of mRNA metabolic processes*	115	37474	8/2	5	1.18	1.18	1.08	0.91	0
P30041 DOWN	Peroxiredoxin-6 (antioxidant protein2); response to oxidative stress	113	24888	10/5	5	0.066212	0.066212	0.106215	0.199833	4
P31930 DOWN	Ubiquinol-cytochrome-c reductase complex core protein I; respiration, oxidative, reduction, and phosphorylation, electron transport	111	52612	12/4	8	6.70E-05	0.00063	0.000153	0.003526	3
P08195 DOWN	4F2hc cell-surface antigen heavy chain (CD98 antigen); amino acid transporter, neoplastic cell growth*	103	57909	9/3	2	-1.08	-1.28	-1.26	-1.43	4
Q5RI18 DOWN	Heterogeneous nuclear ribonucleoprotein U; nucleic acid binding, RNA metabolism	93	27387	5/2	3	0.106215	0.041074	0.045644	0.017745	2
Q15181 DOWN	Inorganic pyrophosphatase; phosphate metabolic processes	91	32639	5/2	5	-1.31	-1.56	-1.90	-1.74	4
Q6ZUD8 DOWN	CDNA FLJ43793 fis, clone TESTI4000014 (LRP130); RNA metabolism	86	157862	5/2	2	0.035208	0.00753	0.00063	0.00196	2
P50454 DOWN	Serpin H1 precursor (collagen-binding protein) (47 kDa HSP); proliferation-inducing gene 14	80	46411	17/7	12	-0.99	-2.16	-1.11	-2.16	2
P13693 DOWN	Translationally-controlled tumor protein (p23) (Fortilin); Ca binding/transport, anti-apoptotic	80	19583	4/2	2	0.14474	6.70E-05	0.094267	6.70E-05	4
P05455 DOWN	Lupus La protein (La autoantigen); mRNA/tRNA binding, metabolic processes	69	46808	4/2	3	-1.90	-2.05	-1.74	-1.90	4
O95817 UP	BAG-3 molecular chaperone regulator (BCL-2-binding); Anti-apoptotic	67	61557	8/4	4	0.00063	0.000153	0.00196	0.00063	3
Q96QV6 DOWN	Histone H2A type 1-A; transcription, DNA replication, repair, chromosomal stability*	59	14094	6/2	3	-1.74	-1.00	-1.79	-1.51	4
P49588 DOWN	Alanyl-tRNA synthetase, (renal carcinoma antigen NY-REN-42); tRNA binding/processing	52	106743	5/3	4	0.084941	0.142992	0.001306	0.010281	3
Q99729 DOWN	Heterogeneous nuclear ribonucleoprotein A/B(ABBP1); RNA metabolism/processing	49	36590	8/4	4	-1.34	-1.24	-1.16	-1.18	4
Q16695 DOWN	Histone H3.1t; transcription regulation, DNA replication and repair, chromosomal stability*	42	15367	12/4	10	0.030601	0.049065	0.073528	0.066212	1
Q96Q06 DOWN	KIAA1881 protein (fragment); triacylglycerol packaging into adipocytes	30	132978	3/3	3	-1.51	-1.46	-1.86	-1.86	3
P01966	Hemoglobin α -bovine (standard)	163	15044	68/8	40	0.010281	0.014054	0.000773	0.000773	0
P12763	α -2HS-glycoprotein-bovine (standard)	110	38394	62/7	37	-1.79	-1.74	-1.56	-1.56	0
P00921	Carbonic anhydrase 2-bovine (standard)	110	28965	64/13	28	0.001306	0.00196	0.00753	0.00753	0
P02070	Hemoglobin β -bovine (standard)	95	15944	65/9	23	1.69	1.21	1.69	1.69	0
						0.002665	0.057909	0.002665	0.002665	0
						-1.74	-1.43	-1.34	-1.24	0
						0.00196	0.017745	0.030601	0.049065	0
						-2.05	-1.86	-1.24	-1.06	0
						0.000153	0.000773	0.049065	0.119007	0
						-1.46	-1.51	-1.51	-1.37	0
						0.014054	0.010281	0.010281	0.025312	0
						-1.24	-0.99	-1.00	-0.79	0
						0.049065	0.14474	0.142992	0.298926	0
						-2.45	-1.34	-1.65	-0.86	0
						3.26E-06	0.030601	0.003526	0.232871	0
						0.73	0.85	0.23	0.24	0
						0.350996	0.243858	0.829882	0.822678	0
						0.93	0.91	0.62	0.76	0
						0.185702	0.199833	0.459976	0.318928	0
						0.32	0.01	0.16	-0.18	0
						0.755072	0.988855	0.877253	0.863449	0
						0.14	0.21	-0.11	0.01	0
						0.891243	0.849519	0.909453	0.988855	0

Table 5. Continued

SwissProt i.d.	Protein	-10xlgP (MS ²)	MW	Peptide Hits/ Unique	iTRAQ sets	Log ₂ (iTRAQ _{PRONq})/adjusted P value				#
						116/114	117/114	116/115	117/115	
P02662	α-S1-casein-bovine (standard)	85	24513	42/7	23	0.63	0.40	0.36	0.37	0
P00711	α-Lactalbumin-bovine (standard)	85	16236	29/4	16	0.448557	0.680047	0.726054	0.715218	0
P62894	Cytochrome c -bovine (standard)	73	11565	32/8	19	-0.62	0.298926	-1.00	-1.13	0
P02663	α-S2-casein-bovine (standard)	48	26002	28/5	15	0.459976	0.797563	0.142992	0.084941	0
P02666	β-Casein-bovine (standard)	33	25091	15/3	10	0.05	0.797563	0.28	0.63	0
						0.955714	0.448557	0.788099	0.448557	0
						0.27	0.891243	-0.14	0.12	0
						0.797563	0.448557	0.891243	0.909453	0
						0.15	0.40	0.82	0.82	0
						0.891243	0.680047	0.271358	0.271358	0

* Confirmed by experimental duplicate; # the number of datasets (out of a maximum of 4) in which the protein had an adjusted P-value of 0.05 or less. Protein function was obtained from the EXPASY Proteomics Server website [36].

Proteins that were selected as biomarker candidates had to qualify as outliers in all four datasets according to the following two criteria: (1) at least three out of four iTRAQ ratios, and (2) the average of all 4 iTRAQ ratios, had to display a minimum of 2-fold change in relative expression levels on the log₂ scale, i.e., log₂(iTRAQ_{PRONq}) ≥ |±1| (positive values for the log₂ ratios corresponding to up-regulated proteins, negative values for down-regulated proteins, and zero for no change in relative expression level). This corresponded to average Z-scores ≥ |±2.5|. A scatter plot of the average Z-scores for the 255 protein set, identified in the E2/Tam treated cells, is provided in the Supplemental material. Table 5 provides a list of the selected putative biomarkers with corresponding MS²-related P-values, molecular weight (MW), total peptide hits/unique peptides, number of quantifying iTRAQ sets/protein, set of four log₂(iTRAQ_{PRONq}) and corresponding adjusted P values for each protein, and the number of datasets (out of a maximum of four) in which the protein had an adjusted P value of 0.05 or less. A total of two proteins were found to be up-regulated and 14 down-regulated. For example, BAG-3 molecular chaperone (O95817) was identified by eight peptide hits (tandem mass spectra), four unique peptides, and quantified by four peptide iTRAQ sets that were averaged into one final protein iTRAQ set (116/114, 117/114, 116/115, and 117/115). The log₂ values of this final set and the corresponding adjusted P values are listed in the next columns. All four log₂ values were >1, showing a larger than 2-fold up-regulation of this protein, and three out of the four adjusted P values were <0.005, controlling the FDR at <0.5%. One of the adjusted P values was 0.057909, equivalent to an FDR ~5.8%, thus exceeding only marginally the typical threshold of 5%. For the other up-regulated protein, Poly(rC)-binding protein (Q15365), the adjusted P values indicated control of the FDR at ~6.6%, ~6.6%, ~10.6%, and 19.9%, respectively. A tandem mass spectrum of a relevant peptide belonging to Poly(rC)-binding protein 1 is provided in the Supplemental material. Again, if a 5% FDR threshold would have been a prespecified requirement, this protein would not have qualified as showing a significant change in abundance. In addition, we note that for the standard bovine protein spikes shown in Table 5, none of the log₂(iTRAQ_{PRONq}) and adjusted P values would have justified the selection of these standards as up/down regulated candidates, confirming, thus, the utility of our statistical approach.

Biological Relevance of Differentially Expressed Proteins

The estrogen positive MCF-7 cell line is commonly used in studies that target the elucidation of estrogen/anti-estrogen modulation of breast cancer development. The proliferation of estrogen receptor (ER)+ breast cancer cells is stimulated by the binding of estrogen/

estrogen-like hormones to ER proteins, such as ER α and ER β , that act as ligand-activated transcription factors [51–55]. The physiological level of 17 β -estradiol is \sim 1 nM, E2 being the most abundant circulating estrogen in humans [55]. Tamoxifen, a nonsteroidal compound, is commonly prescribed in hormonal breast cancer therapy. Competitive binding of Tam to ERs inhibits estrogen-mediated events (gene transcription, DNA synthesis, etc.) that lead to cell proliferation and tumor growth. Studies have shown that the effect of Tam on breast cancer cells is mediated through apoptosis [54]. The fine mechanisms that are at play, however, in E2 stimulation/Tam inhibition of cell proliferation are not known.

The entire list of 255 quantified proteins was subjected to biological categorization based on information extracted from the Gene Ontology (GO) database. A total of 149 proteins returned biological process related GO assignments, the major categories including processes related to mRNA/tRNA processing, transcription/translation, signaling, metabolic processes, nucleosome/spliceosome assembly, apoptosis, and protein binding/folding. Protein differential expression in this study is expected to occur for the following reasons: (a) as a direct result of Tam effect on the expression level of certain proteins; (b) as a result of cell response to stress inducing conditions; (c) as a result of cell accumulation in the G1 phase of the cell cycle (it has been shown that Tam treatment will arrest the cell cycle in the G1 phase) [56]; or, (d) as a result of changes in post-translational modifications (PTMs). The proteins that were found to be up- or down-regulated in the Tam treated cells are all involved in biological processes that are consistent with the sources that can induce changes in their expression level (Table 5). For example, the finding that BAG3 was up-regulated in Tam treated cells is very interesting, as it was recently shown that BAG3 has anti-apoptotic activity, and that expression of BAG3 in neoplastic cells sustains cell survival in response to stress inducing factors. It is also believed that BAG3 could be responsible for impairing response to therapy, and could represent a novel therapeutic target [57]. The other up-regulated protein, Poly(rC)-binding protein 1, is a nucleic acid binding protein that functions as an accelerator of mRNA metabolic processes and has roles in mRNA stabilization, translational activation and translational silencing [58, 59]. In addition, several proteins with role in cell proliferation and tumorigenesis that are up-regulated in cancerous cell states have been found to be down-regulated in the Tam treated cells. The preliminary double 2-plex experiment that returned 154 quantified proteins, and the 4-plex experiment that returned 145 proteins, enabled the confirmation of up/down regulation for four proteins from Table 5 according to most of the stringency parameters imposed in our study (see proteins marked with an asterisk). The other proteins in the list could not be confirmed, either as a result of the protein not being identified/quantified in the smaller datasets, or as a result of the protein not passing the quality criteria imposed for accurate quantitation.

While the E2/Tam cultured cells served only as a model system for evaluating the effectiveness of the PQD linear ion trap technology and of our newly developed method for identifying differentially expressed proteins, the present findings seem very promising, and additional studies involving the use of complementary techniques, such as Western blotting and MS multiple reaction monitoring, will be conducted on biological replicates to confirm the differential expression of these proteins.

Conclusions

In this work, a one-step LC-MS/MS approach, using iTRAQ/PQD ion trap MS technology, has been developed to perform differential expression profiling of complex proteomic cellular extracts. The effectiveness of the method was evaluated on MCF-7 breast cancer cells cultured in the presence of E2 and Tam. The detection of protein up/down regulation at a \sim 2-fold threshold change in expression level was enabled by the development of: (1) a data dependent MS acquisition strategy that involved the analysis of two experimental replicates of two cell states (double 2-plex), (2) an advance data filtering strategy that involved multiple decisions at the peptide and protein level, and (3) a statistical data validation strategy that involved global/quantile normalization, log₂ transformation, Z-score generation and adjusted *P* value calculation for controlling the FDR. Ultimately, the method enabled the detection/quantitation of 530/255 proteins (*P* < 0.001), and the selection of 16 differentially expressed proteins, demonstrating the potential of iTRAQ/PQD-MS for biomarker discovery applications. Most importantly, it was determined that these proteins are implicated in cancer-relevant biological processes such as cell proliferation, apoptosis, oncogenesis and metastasis. The corroborated change in expression level of these proteins in Tam treated cells has not been reported, so far, thus, the biological significance of this outcome requires further validation. It is anticipated that a proper assessment of protein up/down regulation in the larger context of biochemical signaling pathways will have a significant impact on elucidating the molecular mechanisms that control Tam inhibition of cell proliferation and cell response to stress inducing conditions. Ultimately, a better understanding of how breast tumor cells develop Tam resistance will lead to the identification of novel therapeutic targets and more effective treatment options for cancer patients.

Acknowledgments

The authors acknowledge partial support for this work by NSF grant Career BES-0448840. They thank Yang Xu (VBI/Virginia Tech) for providing help with cell culturing. JMA and IML acquired and processed the data, and wrote the manuscript; IH and IML performed the statistical evaluation of data.

Appendix A Supplementary Material

Supplementary material associated with this article may be found in the online version at doi:10.1016/j.jasms.2009.02.029.

References

- Choe, L. H.; Aggarwal, K.; Franck, Z.; Lee, K. H. A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia coli* cells. *Electrophoresis* **2005**, *26*, 2437–2449.
- Shi, Y.; Xiang, R.; Horváth, C.; Wilkins, J. A. Quantitative analysis of membrane proteins from breast cancer cell lines BT474 and MCF7 using multistep solid phase mass tagging and 2D LC/MS. *J. Proteome Res.* **2005**, *4*, 1427–1433.
- Xiang, R.; Shi, Y.; Dillon, D. A.; Negin, B.; Horváth, C.; Wilkins, J. A. Analysis of membrane proteins from breast cancer cell lines MCF7 and BT474. *J. Proteome Res.* **2004**, *3*, 1278–1283.
- Lill, J. Proteomic tools for quantitation by mass spectrometry. *Mass Spectrom. Rev.* **2003**, *22*, 182–194.
- Wu, W. W.; Wang, G.; Baek, S. J.; Shen, R. Comparative study of three proteomic quantitative methods, DIGE, cIAT, and iTRAQ, using 2D Gel- or LC-MALDI TOF/TOF. *J. Proteome Res.* **2006**, *5*, 651–658.
- Melanson, J. E.; Avery, S. L.; Pinto, D. M. High-coverage quantitative proteomics using amine-specific isotopic labeling. *Proteomics* **2006**, *6*, 4466–4474.
- Julka, S.; Regnier, F. Quantification in proteomics through stable isotope coding: A review. *J. Proteome Res.* **2004**, *3*, 350–363.
- Patwardham, A. J.; Strittmatter, E. F.; Camp, D. G.; Smith, R. D.; Pallavicini, M. G. Quantitative proteome analysis of breast cancer cell lines using 18 O-labeling and an accurate mass and time tag strategy. *Proteomics* **2006**, *6*, 2903–2915.
- Schneider, L. V.; Hall, M. P. Stable isotope methods for high-precision proteomics. *Drug Discov. Today Targets* **2005**, *5*(10), 353–363.
- Thiede, B.; Kretschmer, A.; Rudel, T. Quantitative proteome analysis of CD95 (Fas/Apo-1)-induced apoptosis by stable isotope labeling with amino acids in cell culture, 2-DE, and MALDI-MS. *Proteomics* **2006**, *6*, 614–622.
- Ong, S.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteom.* **2002**, *1*, 376–385.
- Everley, P. A.; Krijgsvelde, J.; Zetter, B. R.; Gygi, S. P. Quantitative cancer proteomics: Stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol. Cell. Proteom.* **2004**, *3*, 729–735.
- Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994–999.
- Wienkoop, S.; Weckewerth, W. Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in *Arabidopsis thaliana*. *J. Experimental Botany* **2006**, *57*(7), 1529–1535.
- Aggarwal, K.; Choe, L. H.; Lee, K. H. Shotgun proteomics using the iTRAQ isobaric tags. *Briefings Funct. Genom. Proteom.* **2006**, *5*(2), 112–120.
- Zieske, L. R. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J. Exp. Botany* **2006**, *57*(7), 1501–1508.
- Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007**, *7*, 340–350.
- Petti, F.; Thelemann, A.; Kahler, J.; McCormack, S.; Castaldo, L.; Hunt, T.; Nuwaysir, L.; Zieske, L.; Haack, H.; Sullivan, L.; Garton, A.; Haley, J. D. Temporal quantitation of mutant kit tyrosine kinase signaling attenuated by a novel thiophene kinase inhibitor OSI-930. *Mol. Cancer Ther.* **2005**, *4*, 1186–1197.
- Evans, F. F.; Raftery, M. J.; Egan, S.; Kjelleberg, S. Profiling the secretome of the marine bacterium *Pseudoalteromonas tunicata* using amine-specific isobaric tagging (iTRAQ). *J. Proteome Res.* **2007**, *6*, 967–975.
- Glückmann, M.; Fella, K.; Waidelich, D.; Merkel, D.; Kruff, V.; Kramer, P.; Walter, Y.; Hellmann, J.; Karas, M.; Kröger, M. Prevalidation of potential protein biomarkers in toxicology using iTRAQ reagent technology. *Proteomics* **2007**, *7*, 1564–1574.
- Ross, P. 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.; Pappin, D. J. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteom.* **2004**, *3*, 1154–1169.
- Dean, A. R.; Overall, C. M. Proteomics discovery of metalloproteinase substrates in the cellular context by iTRAQ labeling reveals a diverse MMP-2 substrate degradome. *Mol. Cell. Proteom.* **2007**, *6*, 611–623.
- Stensjö, K.; Ow, S. Y.; Barrios-Llerena, M. E.; Lindblad, P.; Wright, P. C. An iTRAQ-based quantitative analysis to elaborate the proteomic response of *Nostoc* sp. PCC 7120 under N₂ fixing conditions. *J. Proteome Res.* **2007**, *6*, 621–635.
- Cong, Y.; Fan, E.; Wang, E. Simultaneous proteomic profiling of four different growth states of human fibroblasts, using amine-reactive isobaric tagging reagents and tandem mass spectrometry. *Mech. Ageing Dev.* **2006**, *127*, 332–343.
- Siepen, J. A.; Swainston, N.; Jones, A. R.; Hart, S. R.; Hermjakob, H.; Jones, P.; Hubbard, S. J. An informatic pipeline for the data capture and submission of quantitative proteomic data using iTRAQ. *Proteome Sci.* **2007**, *5*(4), 1–9.
- DeSouza, L.; Diehl, G.; Rodrigues, M. J.; Guo, J.; Romaschin, A. D.; Colgan, T. J.; Siu, K. W. M. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cIAT with multi-dimensional liquid chromatography and tandem mass spectrometry. *J. Proteome Res.* **2005**, *4*, 377–386.
- Lund, T. C.; Anderson, L. B.; McCullar, V.; Higgins, L.; Yun, G. H.; Grzywacz, B.; Verneris, M. R.; Miller, J. S. iTRAQ is a useful method to screen for membrane-bound proteins differentially expressed in human natural killer cell types. *J. Proteome Res.* **2007**, *6*, 644–653.
- Liu, T.; Donahue, K. C.; Hu, J.; Kurnellas, M. P.; Grant, J. E.; Li, H.; Elkabes, S. Identification of differentially expressed proteins in experimental autoimmune encephalomyelitis (EAE) by proteomic analysis of the spinal cord. *J. Proteome Res.* **2007**, *6*, 2565–2575.
- Keshamouni, V. G.; Michailidis, G.; Grasso, C. S.; Anthwal, S.; Strahler, J. R.; Walker, A.; Arenberg, D. A.; Reddy, R. C.; Akulapalli, S.; Thannickal, V. J.; Standiford, T. J.; Andrews, P. C.; Omenn, G. S. Differentially protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *J. Proteome Res.* **2006**, *5*, 1143–1154.
- Chaerkady, R.; Harsha, H. C.; Nalli, A.; Gucek, M.; Vivekanandan, P.; Akhtar, J.; Cole, R. N.; Simmers, J.; Schulick, R. D.; Singh, S.; Terubson, M.; Pandey, A.; Thuluvath, P. J. A quantitative proteomic approach for identification of potential biomarkers in hepatocellular carcinoma. *J. Proteome Res.* **2008**, *7*, 4289–4298.
- Meany, D. L.; Xie, H.; Thompson, L. V.; Arriaga, E. A.; Griffin, T. J. Identification of carbonylated proteins from enriched rat skeletal muscle mitochondria using affinity chromatography-stable isotope labeling and tandem mass spectrometry. *Proteomics* **2007**, *7*, 1150–1163.
- Griffin, T. J.; Xie, H.; Bandhakavi, S.; Popko, J.; Mohan, A.; Carlis, J. V.; Higgins, L. iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *J. Proteome Res.* **2007**, *6*, 4200–4209.
- Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B. Robust and sensitive iTRAQ quantitation on LTQ Orbitrap mass spectrometer. *Mol. Cell. Proteom.* **2008**, *7*(9), 1702–1713.
- Venable, J. D.; Wohlschlegel, McClatchy, D. B.; Park, S. K.; Yates, J. R. III. Relative quantification of stable isotope labeled peptides using linear ion trap-Orbitrap hybrid mass spectrometer. *Anal. Chem.* **2007**, *79*, 3056–3064.
- Sarvaiya, H. A.; Yoon, J. H.; Lazar, I. M. Proteome profile of the MCF7 cancer cell line: A mass spectrometric evaluation. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3039–3055.
- ExpASY/SwissProt (<http://www.expasy.org/>).
- Malorni, L.; Cacace, G.; Cuccurullo, M.; Pocsfalvi, G.; Chambery, A.; Farina, A.; Di Maro, A.; Parente, A.; Antonio Malorni, A. Proteomic analysis of MCF-7 breast cancer cell line exposed to mitogenic concentration of 17 β -estradiol. *Proteomics* **2006**, *6*, 5973–5982.
- Zhao, J.; Zhu, K.; Lubman, D. M.; Miller, F. R.; Shekhar, M. P. V.; Gerard, B.; Barder, T. J. Proteomic analysis of estrogen response of premalignant human breast cells using a 2-D liquid separation/mass mapping technique. *Proteomics* **2006**, *6*, 3847–3861.
- Huang, H.; Stasyk, T.; Morandell, S.; Dieplinger, H.; Falkensammer, G.; Griesmacher, A.; Mogg, M.; Schreiber, M.; Feuerstein, I.; Huck, C. W.; Stecher, G.; Bonn, G. K.; Huber, L. A. Biomarker discovery in breast cancer serum using 2-D differential gel electrophoresis/MALDI-TOF/TOF and data validation by routine clinical assays. *Electrophoresis* **2006**, *27*, 1741–1650.
- Brusic, V.; Marina, O.; Wu, C. J.; Reinherz, E. L. Proteome informatics for cancer research: From molecules to clinic. *Proteomics* **2007**, *7*, 976–991.
- Minafra, I. P.; Cancemi, P.; Fontana, S.; Minafra, L.; Feo, S.; Becchi, M.; Freyria, A. M.; Minafra, S. Expanding the protein catalogue in the proteome reference map of human breast cancer cells. *Proteomics* **2006**, *6*, 2609–2625.
- Kreunin, P.; Yoo, C.; Urquidí, V.; Lubman, D. M.; Goodison, S. Proteomic profiling identifies breast tumor metastasis-associated factors in an isogenic model. *Proteomics* **2007**, *7*, 299–312.
- Kirmiz, C.; Li, B.; An, H. J.; Clowers, B. H.; Chew, H. K.; Lam, K. S.; Ferrige, A.; Alecio, R.; Borowsky, A. D.; Sulaimon, S.; Lebrilla, C. B.; Miyamoto, S. A serum glycomics approach to breast cancer biomarkers. *Mol. Cell. Proteom.* **2007**, *6*, 43–55.
- Canelle, L.; Bousquet, J.; Pionneau, C.; Hardouin, J.; Kastylevsky, G. C.; Joubert-Caron, R.; Caron, M. A proteomic approach to investigate potential biomarkers directed against membrane-associated breast cancer proteins. *Electrophoresis* **2006**, *27*, 1609–1616.
- Harwood, M. M.; Bleecker, J. V.; Rabinovitch, P. S.; Dovichi, N. J. Cell cycle dependent characterization of single MCF-7 breast cancer cell by 2-D CE. *Electrophoresis* **2007**, *28*, 932–937.
- Chong, P. K.; Gan, C. S.; Pham, T. K.; Wright, P. C. Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J. Proteome Res.* **2006**, *5*, 1232–1240.

47. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6591–6596.
48. Gan, C. S.; Chong, P. K.; Pham, T. K.; Wright, P. C. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. Proteome Res.* **2007**, *6*, 821–827.
49. Bolstad, B. M.; Irizarry, R. A.; Åstrand, M.; Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **2003**, *19*(2), 185–193.
50. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300.
51. Furuya, Y.; Kohno, N.; Fujiwara, Y.; Saitoh, Y. Mechanisms of estrogen action on the proliferation of MCF-7 human breast cancer cells in an improved culture medium. *Cancer Res.* **1989**, *49*, 6670–6674.
52. Darbre, P.; Yates, J.; Curtis, S.; King, R. J. B. Effect of estradiol on human breast cancer cells in culture. *Cancer Res.* **1983**, *43*, 349–354.
53. Osborne, C. K.; Boldt, D. H.; Estrada, P. Human breast cancer cell cycle synchronization by estrogens and antiestrogens in culture. *Cancer Res.* **1984**, *44*, 1433–1439.
54. Stackiewicz, R.; Drucker L.; Radnay, J.; Beyth Y.; Yarkoni, S.; Cohen, I. Tamoxifen modulates apoptotic pathways in primary endometrial cell cultures. *Clin. Cancer Res.* **2001**, *7*, 415–420.
55. Schnaper, H. W. Estrogen—it's not just for reproduction anymore. *Kidney Int.* **1999**, *55*, 1577–1579.
56. Osborne, C. K.; Boldt, D. H.; Clark, G. M.; Trent, J. M. Effects of tamoxifen on human breast cancer cell cycle kinetics: Accumulation of cells in early G1 phase. *Cancer Res.* **1983**, *43*, 3583–3585.
57. Rosati, A.; Ammirante, M.; Gentilella, A.; Basile, A.; Festa, M.; Pascale, M.; Marzullo, L.; Belisario, M.A.; Tosco, A.; Franceschelli, S.; Moltedo, O.; Pagliuca, G.; Lerose, R.; Turco, M. C. Apoptosis inhibition in cancer cells: A novel molecular pathway that involves BAG3 protein. *Int. J. Biochem. Cell. Biol.* **2007**, *39*, 1337–1342.
58. Makeyev, A. V.; Liebhaber, S. A. The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. *RNA* **2002**, *8*, 265–278.
59. Leffers, H.; Dejgaard, K.; Celis, J. E. Characterisation of two major cellular poly(rC)-binding human proteins, each containing three K-homologous (KH) domains. *Eur. J. Biochem.* **1995**, *230*, 447–453.