research articles Journal of proteome research

iTRAQ Experimental Design for Plasma Biomarker Discovery

Xiaomin Song,[†] Julia Bandow,[‡] Jamie Sherman,[†] J. David Baker,[‡] Paul W. Brown,[§] Michael T. McDowell,[‡] and Mark P. Molloy*,[†,||]

Australian Proteome Analysis Facility Ltd., Macquarie University, Sydney, Australia, Pfizer Global Research and Development, Ann Arbor, Michigan, 48105, Pfizer Global Research and Development, St. Louis, Missouri, 63006, and Department of Chemistry & Biomolecular Sciences, Macquarie University, Sydney, Australia

Received January 30, 2008

There is considerable interest in using mass spectrometry for biomarker discovery in human blood plasma. We investigated aspects of experimental design for large studies that require analysis of multiple sample sets using iTRAQ reagents for sample multiplexing and quantitation. Immunodepleted plasma samples from healthy volunteers were compared to immunodepleted plasma from patients with osteoarthritis in eight separate iTRAQ experiments. Our analyses utilizing ProteinPilot software for peptide identification and quantitation showed that the methodology afforded excellent reproducibility from run to run for determining protein level ratios (coefficient of variation 11.7%), in spite of considerable quantitative variances observed between different peptides for a given protein. Peptides with high variances were associated with lower intensity iTRAQ reporter ions, while immunodepletion prior to sample analysis had a negligible affect on quantitative variance. We examined the influence of different reference samples, such as pooled samples or individual samples on calculating quantitative ratios. Our findings are discussed in the context of optimizing iTRAQ experimental design for robust plasma-based biomarker discovery.

Keywords: mass spectrometry • plasma • biomarker • quantitation • experimental design

Introduction

Several different mass spectrometry (MS) approaches have been described for conducting comparative quantitative proteomics. Commonly, these approaches employ protein/peptide labeling strategies to enable relative sample quantitation, although some nonlabeling approaches have also been described.[1–4] The three most widespread approaches for protein/peptide labeling are chemical derivatization (e.g., Isotope coded affinity tags (ICAT) and Isobaric Tags for Relative and Absolute Quantitation (iTRAQ)[5,6]), metabolic labeling (e.g., Stable isotope labeling of amino acids in culture (SILAC)[7,8]) and enzymatic labeling during protein digestion using deuterated water.[9,10] iTRAQ is a chemical labeling approach that incorporates stable isotopes into an NHS-ester derivative amine tagging reagent that, when combined with mass spectrometry, allows comparative, quantitative multiplexing analysis. In this study, we used the 4-plex iTRAQ reagent kit. Recently, an 8-plexed version of the iTRAQ reagents was introduced. Each protein sample is proteolytically digested with trypsin, labeled with one of the isobaric tags, mixed with the other different iTRAQ labeled samples, and analyzed by tandem mass spectrometry (MS/MS). Proteins are identified using the peptide product ion spectra and are quantified by the relative intensities of the 4-plex iTRAQ reporter ions detected in the 114−117 m/z region of the product ion spectra. At least one of the labeled samples is a reference sample allowing the relative quantities of each peptide in comparative samples to be determined as a ratio of the quantity of the same peptide in the reference sample. A scoring function is then used to determine the weighting contribution of each peptide ratio toward the final protein ratio.[11] iTRAQ is ideally suited for biomarker applications as it provides both quantitation and multiplexing in a single reagent and has been applied to the analysis of clinical samples such as human blood serum or plasma, cerebrospinal fluid, disease tissues, or for *in vitro* profiling of cells to identify differentially expressed proteins.[12–14]

The experimental design of a large-scale biomarker discovery experiment weighs heavily toward its success, requiring a great deal of strategic consideration and planning. For iTRAQ experiments, some key variables that require consideration include sample size based on the degree of technical and biological variation, the type and composition of the reference sample, and whether sample pooling is a viable option. Currently, there is a lack of comprehensive information in the literature regarding these issues. Of the few available technical reports, Wu et al.[15] compared DIGE, ICAT, and iTRAQ to show that iTRAQ is the most sensitive proteomics quantitation method among the three techniques evaluated, based on the number of detected peptides. They used six standard proteins to determine quantitative reproducibility for iTRAQ experiments and reported standard deviations (SD) in the range of 0.088−0.145. Gan et al.[16] recently used cell lysate samples and

* To whom correspondence should be addressed. Mark P. Molloy, Ph.D., Tel, +612 9850 6218; fax, +612 9850 6200; e-mail, mmolloy@proteome.org.au.
[†] Australian Proteome Analysis Facility Ltd., Macquarie University.
[‡] Pfizer Global Research and Development, Ann Arbor, MI.
[§] Pfizer Global Research and Development, St. Louis, MO.
[||] Department of Chemistry & Biomolecular Sciences, Macquarie University.

10.1021/pr800072x CCC: $40.75
Published on Web 06/13/2008

Table 1. Preparation of Sample Pools[a]

samples	Pool 1 (vol in μL)	Pool 2 (vol in μL)	Pool 3 (vol in μL)
C1	50	50	x
C2	50	50	x
C3	50	50	x
C4	50	x	50
C5	50	x	50
C6	50	x	50
C7	x	x	x
D1	50	50	x
D2	50	50	x
D3	50	50	x
D4	50	x	50
D5	50	x	50
D6	50	x	50
D7	x	x	x
D8	x	x	x

[a] "C" refers to control samples, "D" refers to patient samples. "x" indicates sample was not added.

Table 2. Frequency of Immunodepletion Runs for Each Sample To Remove Top 12 Abundant Proteins

sample name	no. depletions conducted
Pool 1	4[a]
Pool 2	2[b]
Pool 3	2[b]
C1	5[a]
C2	2[b]
C3	2[b]
C7	2[b]
D1	2[b]
D2	2[b]
D3	2[b]
D7	2[b]
D8	1[b]

[a] Flow through from individual depletions were pooled, then samples were aliquoted for subsequent experiments. [b] Flow through fractions from individual depletions were not pooled but used independently following immunodepletion.

reported an average SD of 0.11 for identical samples analyzed in a single iTRAQ run, a SD of 0.23 for the same sample analyzed in repeated iTRAQ runs, and a SD of 0.25 for biological replicates analyzed in a single iTRAQ run.

Given the considerable worldwide efforts focused on plasma biomarker discovery,[17] and the requirement that these studies are large in scale, there is a clear need to evaluate more thoroughly iTRAQ experimental design. Furthermore, the limited technical data available for iTRAQ experimental design is not drawn from plasma samples, which pose special analytical challenges for mass spectrometry analysis due to the large dynamic range in protein concentrations.[18] In this paper, we evaluate several key aspects of experimental design described above to develop an optimized experimental design strategy for robust, multiple-sample, plasma-based iTRAQ biomarker studies.

Experimental Procedures

Plasma Samples. Potassium-EDTA plasma samples were collected by venipuncture from 15 human females. The plasma was collected, aliquoted, and stored frozen at −70 °C. All study volunteers were >40 years of age and in generally good health. All sample collections and experiments were approved by a local ethics board and informed consent was obtained from all donors. The study subjects were in two groups; eight presented as osteoarthritis (OA) patients and seven were age-matched controls.

OA patients were defined as having frequent knee symptoms in the past year, a Kellgren and Lawrence grades 2 or 3 on one or both knees, and a body mass index (BMI) ≥30. Age-matched control volunteers had no evidence of OA in the knee and a BMI <28.

Preparation of Reference Sample Pools. Pooled plasma reference samples were prepared prior to immunodepletion by combining 50 μL of plasma from individual samples (Table 1). Pool 1 was prepared from 6 controls (C) and 6 patient samples (D). Pool 2 was a subset of Pool 1, containing 3 controls and 3 patient samples. Pool 3 was formed from a different subset of Pool 1 using 3 controls and 3 patient samples that were different from those used for Pool 2. Samples C7, D7, and D8 were not added to any pool.

Immunodepletion of High-Abundance Proteins. High-abundance proteins in plasma were depleted using ProteomeLab IgY-12 chromatography column kit (Beckman Coulter). The column depleted serum albumin, IgG, α1-antitrypsin, IgA, IgM, transferrin, beta-haptoglobin, α1-acid glycoprotein (Orosomucoid), α2-macroglobulin, fibrinogen, apolipoprotein AI and apolipoprotein AII. For each plasma sample, 120 μL was loaded per injection. After depletion, pools were created by combining equal amounts of protein from several samples as indicated in Table 2. Some samples were split into two aliquots prior to depletion so that the depletion was performed twice independently on the same sample. For example, samples C2_1 and C2_2 originate from the same control plasma sample C2 but were depleted separately and used in different iTRAQ runs. Table 2 displays the sample for iTRAQ runs before and after the immunodepletion.

Protein Reduction, Alkylation, and Protein Assay. The proteins from the immunodepletion column flow through were precipitated overnight in acetone at −20 °C. After acetone precipitation, the protein pellets were resuspended in 0.5 M triethylammonium bicarbonate (TEAB), pH 8.5, and 0.1% (w/v) sodium dodecyl sulfate (SDS), reduced with 5 mM Tris(2-carboxyethyl)phosphine (TCEP) for 1 h at 60 °C, and alkylated with 10 mM *s*-methylmethanethiosulfonate (MMTS) at room temperature for 10 min. After reduction and alkylation, the proteins in each depleted sample were measured by Bradford Protein Assay and stored in 100 μg aliquots.

Tryptic Digestion and iTRAQ Labeling. A total of 100 μg of protein was digested overnight with trypsin at 37 °C at a ratio of 1:20, trypsin to protein. Digested samples were labeled with the iTRAQ reagents following the protocol provided by the vendor (Applied Biosystems, Foster City, CA). Briefly, one vial of iTRAQ labeling reagent was used for every 100 μg of protein. Ethanol was used to solubilize the iTRAQ reagent then added to the peptide sample ensuring a final organic concentration of at least 60% (v/v). After 1 h of iTRAQ labeling, the reaction was quenched by adding 50 μL of water. The samples were then mixed at equal ratios and dried by centrifugal evaporation. Table 3 shows the sample labeling design for each experiment.

Strong Cation Exchange Chromatography. iTRAQ labeled peptides were fractionated by strong cation exchange liquid chromatography (SCX) using a PolySulfethyl A 200 mm × 2.1 mm, 5 μm, 200 Å column (PolyLC, Columbia, MD). Buffer A was 5 mM phosphate 25% (v/v) acetonitrile, pH 2.7, and buffer B was 5 mM phosphate, 350 mM potassium chloride, 25% (v/v) acetonitrile, pH 2.7. The dried iTRAQ labeled sample was

Table 3. Overview of Experimental Design[a]

	iTRAQ 114	iTRAQ 115	iTRAQ 116	iTRAQ 117
Run1	Pool 1	C1	Pool 2_1	Pool 3_1
Run2	C1	Pool 1	Pool 3_2	D1_1
Run3	Pool 2_2	C2_1	Pool 1	C1
Run4	D3_1	D7_1	C1	Pool 1
Run5	D2_1	Pool 1	C1	C7_2
Run6	Pool 1	D3_2	C3_2	C1
Run7	D1_2	C1	D2_2	Pool 1
Run8	C2_2	Pool 1	C1	C3_2
Run9	C1	D8_1	D7_2	C7_2

[a] Sample names refer to Table 1. Sample names appended with an underscore and a number indicating immunodepletion run number.

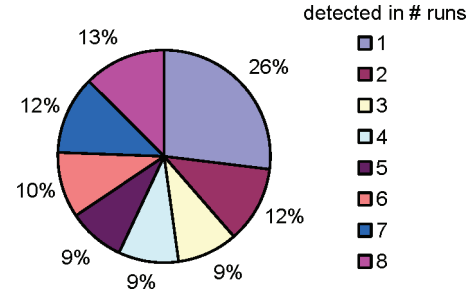

Figure 1. The percentage of nonredundant peptide observations per replicate MS runs for C1/Pool 1 peptides. Thirteen percent (13%) of peptides were detected in all 8 runs, while 26% of all peptides were detected in only a single run.

resuspended in buffer A and applied to the SCX column. After sample loading and washing with buffer A for 37.5 min, buffer B concentration increased from 10% to 60% in 46 min and then ramped to 100% at a flow rate of 200 μL/min. Fifteen (15) fractions were collected from the SCX separation then dried by centrifugal evaporation.

Reverse Phase NanoLC ESI MS/MS. Dried SCX fractions were solubilized in 100 μL of 0.1% (v/v) formic acid, 2% (v/v) acetonitrile. Thirty microliters of sample was loaded onto a reverse phase peptide Captrap (2 μg capacity) (Michrom Bioresources, Auburn, CA) and desalted for 10 min with buffer A (0.1% (v/v) formic acid) at 10 μL/min. After desalting, the trap was switched in-line with a 150 μm $\times$ 10 cm C18, 3 μm, 300 Å ProteCol column (SGE). The buffer B (90% (v/v) acetonitrile, 0.1% (v/v) formic acid) gradient started from 5% to 10% in 2 min and then to 50% in 80 min to elute peptides.

The LC eluent was subject to positive ion nano flow electrospray analysis using a QStar XL MS/MS system (Applied Biosystems) in an information dependent acquisition mode (IDA). In IDA mode, a TOFMS survey scan was acquired (m/z 350−1600, 0.5 s), with the three most intense multiply charged ions (counts >25) in the survey scan sequentially subjected to product ion analysis. Product ion spectra were accumulated for 2 s in the mass range m/z 100−2000 with a modified Enhance All mode Q2 transition setting favoring low mass ions so that the reporting iTRAQ ion (114, 115, 116 and 117 m/z) intensities were enhanced for quantitation. Dynamic exclusion was used with a 2 min and 150 ppm window.

Data Analysis. MS/MS data were analyzed using ProteinPilot v1.0 (Applied Biosystems) which uses the Paragon algorithm[11] to perform database matching for protein identification, protein grouping to remove redundant hits and comparative quantitation. The Swiss-Prot *Homo sapiens* protein database was used for all searches.

Data was normalized for loading error by bias corrections calculated using ProteinPilot. All reported data were based on 95% confidence for protein identification as determined by ProteinPilot (ProtScore $\geq$ 1.3). A further requirement was a protein p-value, which ensured protein identification and quantitation was not based on a single peptide hit. The default ion intensity threshold in ProteinPilot for calculating peptide ratios was 40 counts.

All protein iTRAQ ratios were transformed to base 2 logarithm values. In base 2 logarithm space, a 2-fold change in levels is reported as −1 or 1 for down-regulated and up-regulated changes, respectively.

Results

Assessment of Quantitative Reproducibility of iTRAQ Experiments. We constructed a series of experiments using iTRAQ reagents with the aim of determining an optimized experimental design for plasma biomarker discovery (Table 3). Using data from 8 independent iTRAQ runs of samples C1 + Pool 1, we determined the run-to-run technical variation that is attributed to the many experimental steps of enzymatic digestion, chemical derivatization, SCX fractionation, and LC-ESI mass spectrometry data acquisition. These samples were randomized with regard to the four iTRAQ labels used for derivatization in order to account for any systematic bias that might be associated with utilizing the same iTRAQ reporter label. Furthermore, it should be noted that, for each of these 8 runs, in addition to C1 + Pool 1 samples, two other samples that differed in each run were also analyzed (Table 3). On average, 105 nonredundant proteins were detected per run with 95% confidence using the Paragon algorithm in the ProteinPilot software. Seventy-three proteins represented by at least two peptides each were quantitated in all 8 iTRAQ runs, while 111 proteins were quantitated in at least 3 out of the 8 runs. The frequency of repeat detection of nonredundant peptides as a function of MS run is shown in Figure 1. Approximately 13% of the total number of nonredundant peptides were detected in all eight runs, while 26% of all peptides were observed in only a single run.

Figure 2A displays the histogram of standard deviations (SD) for the protein ratios detected in at least three of the eight iTRAQ runs. Overall, the mean SD was 0.14, the median was 0.09, and 90% of the quantified proteins had a SD smaller than 0.3. This equates to a mean coefficient of variation (CV) of 11.7%. The 73 proteins detected in all eight runs displayed even lower variance (CV = 7.7%) as these are among the most abundant proteins producing high iTRAQ reporter ion counts.We next evaluated the additional variance contributed by genetic differences between individual samples. First, we examined variances from healthy volunteers using four samples (C1, C2, C3, C7) compared to the Pool 1 sample (as the reference denominator). A total of 106 proteins were quantified across all four samples. Figure 2B displays the histogram of the SD for the protein ratios. The mean of the SD was 0.29 (28% CV), the median was 0.22, with 90% of the quantified proteins having an SD of ≤0.60. It should be remembered that this variance also includes underlying variation due to all experimental steps. The variance for OA samples (D1, D2, D3, D7) referenced to Pool 1 was slightly greater, with mean SD 0.33, median SD 0.25, with 90% of the data having SD ≤ 0.60 (Figure 2C). Comparison

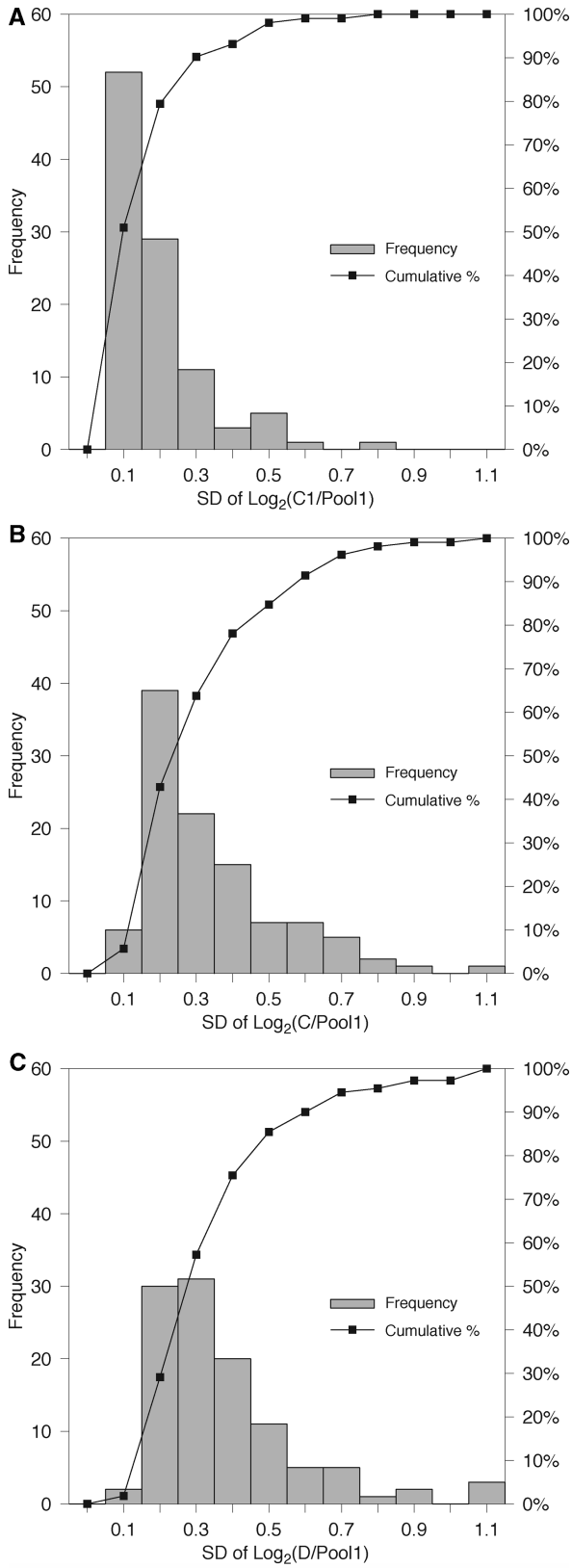

Figure 2. Distribution and frequency of iTRAQ protein ratio standard deviations (SD) for replicate samples. (A) Histogram of the SD (log_2) from technical replicates from eight independent C1/Pool 1 runs (111 proteins). (B) Histogram of SD (log_2) from biological replicates of control samples using C1/Pool 1, C2/Pool 1, C3/Pool 1 and C7/Pool 1 runs (106 proteins). (C) Histogram of SD (log_2) from biological replicates of OA samples using D1/Pool 1, D2/Pool 1, D3/Pool 1 and D7/Pool 1 runs (102 proteins).

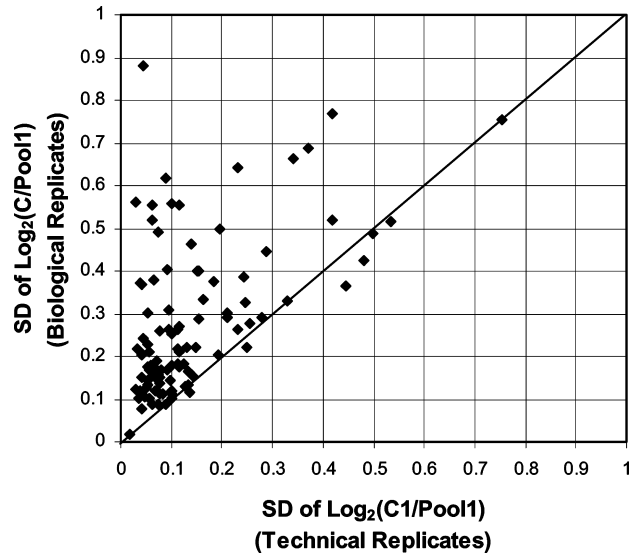

Figure 3. Comparison of SD (log_2) using 103 proteins from eight technical replicates (C1/Pool 1) to four control biological replicates (C/Pool 1).

of the technical variation alone (Figure 2A) and the combined technical and biological variation (Figure 2B) clearly illustrates that there is significant biological variation that needs to be considered when determining appropriate experimental sample sizes. The degree of biological variation should be determined experimentally. While the above result is unsurprising, Figure 3 shows that variance does not increase consistently for all proteins, rather it increases significantly for some proteins while others show little additional biological variation.

Quantitative Reproducibility of Immunodepletion. As plasma contains many highly abundant proteins, immunodepletion is required to access lower abundant proteins for biomarker discovery. We used iTRAQ to determine the degree of variation introduced by immunodepletion. The iTRAQ ratios were compared for four different samples that were depleted on two occasions using antibody depletion columns of the same lot number. The variance attributed to the immunodepletion process was almost negligible, being only slightly greater than technical variation, with a mean SD of 0.15 and median SD of 0.1 (data not shown). The high reproducibility of this data supports the use of immunodepletion as a standard strategy for plasma biomarker studies. As a note, we did not investigate variation that might result from using different antibody lots in the depletion columns.

Assessment of Peptide Level Quantitative Reproducibility. We noted that there was considerable variance in the iTRAQ ratios of individual peptide ions for a given protein within a MS run. In many instances, we observed peptides that differed by as much as 2-fold from the final protein ratio calculated by ProteinPilot. To examine whether variance was associated with iTRAQ reporter ion intensity, we plotted the intensity of ions used to calculate Pool 1 + C1 ratios in the eight runs against the peptide ratio variance (Figure 4). This data showed that approximately 75% of the peptides have a variance below 0.5. The remaining 25% have a higher variance and are predominantly, but not exclusively, ions of lower intensity. It would be unwise to simply discard the data based on low ion intensity as a considerable portion of reliable data is contained here. Nonetheless, these observations suggest caution should be taken in relying on iTRAQ quantitation that is obtained from

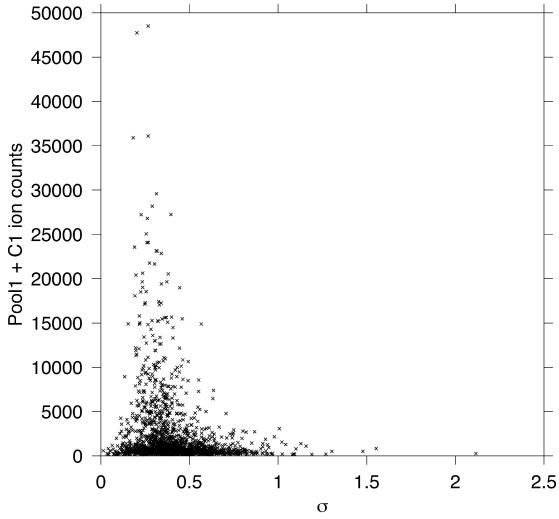

Figure 4. Peptide standard deviation for C1/Pool 1 in 8 runs as a function of iTRAQ reporter ion intensity. Seventy-five percent (75%) of the data has SD < 0.5.

low intensity reporter ions, particularly when single occurrences or few replicate measurements are obtained.

Choice of iTRAQ Reference Sample. Biomarker discovery studies utilize iTRAQ reagents to enable sample multiplexing and relative protein quantitation. The number of samples that need to be analyzed to achieve statistical power normally will exceed the number of iTRAQ reporter labels available. Comparison between multiple iTRAQ runs can best be facilitated by analyzing a reference sample in each iTRAQ run against which the other samples are normalized. Because the quantitation is reported as a ratio (e.g., normal:disease), the choice of reference sample used as the denominator is an important one. We designed an experiment to examine the effect of subtly modifying the iTRAQ reference sample during iTRAQ comparisons. Pool 2 and Pool 3 were created as subsets of Pool 1 (Table 1), and used in an iTRAQ channel independently of the Pool 1 labeled channel (e.g., 114 m/z versus 116 m/z). This enabled us to calculate iTRAQ ratios using related but different reference samples.

In the majority of cases, the choice of reference sample did not negatively influence the calculated iTRAQ ratios. That is, the calculated ratios using Pool 2 or Pool 3 approximately equaled the ratio calculated if Pool 1 was used, even though the composition of Pool 2 and Pool 3 contained only 50% of the samples in Pool 1. This is illustrated in Figure 5 with a boxplot using the ratios for Vitamin D binding protein (VTDB_HUMAN). Nonetheless, we identified several cases where the use of Pool 2 or Pool 3 as the reference caused vastly altered ratios compared to those ratios calculated using Pool 1 as reference. In some cases, this produced fold-changes of more than three SD (see Figure 5).

Interestingly, in run 1 for fibronectin (FINC_HUMAN), the observed ratios calculated from C1 + Pool 2 are smaller than C1 + Pool 1, while the ratios of C1 + Pool 3 are larger than C1 + Pool 1 (Figure 5). This observation could be explained as differences in the abundance of fibronectin between samples that constitute Pool 2 and Pool 3. When these samples are combined to produce Pool 1, there is a "balancing-out" of protein quantities so that the mean quantity (i.e., Pool 1) falls between the ratios of Pool 2 and Pool 3. That is, the sum of Pool 2 and Pool 3 equals Pool 1.

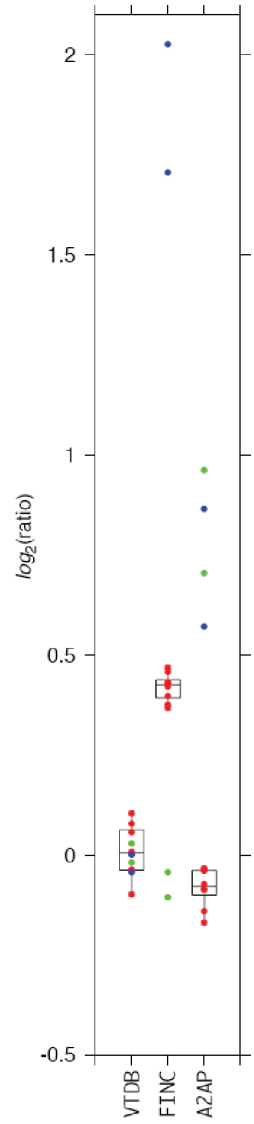

Figure 5. Box plots displaying $\log_2$ iTRAQ protein ratios for Vitamin D binding protein (VTDB), fibronectin (FINC) and α-2-antiplasmin (A2AP) from 8 technical replicate runs of C1/Pool 1 (red dots). $\log_2$ iTRAQ protein ratios are plotted as outliers dots using 2 technical replicates of C1/Pool 2 (green dots), and 2 technical replicates of C1/Pool 3 (blue dots).

While the above hypothesis can explain the observation for fibronectin, we observed other instances (approximately 10% of the data) that are not supported by this hypothesis. For example, for α-2-antiplasmin (A2AP_HUMAN), the ratios for C1 + Pool 2 and C1 + Pool 3 are both significantly greater than the C1 + Pool 1 ratio (Figure 5). These cases illustrate that consistency of the reference is necessary, otherwise even small changes to the reference sample is sufficient to alter which proteins are reported as differentially expressed. This argues strongly for experimental design to incorporate a single reference sample for use throughout the entire experiment, including future comparative studies.

We also tested whether a single individual sample or a pooled sample was the optimal choice as the reference sample. Our concern was that, in a pooled sample, rare proteins would be diluted to such an extent that they would not allow reliable ratios to be quantitated. iTRAQ runs 1−8 were investigated to determine the frequency of missing quantitation ratios de-

pending on the choice of reference sample (i.e., Pool 1 or C1). However, we found no statistical evidence that the composition of reference sample was directly linked to missing quantitation values. This suggests that pooling to make a reference sample does not negatively impact the ability to quantitate peptides from comparative individual samples.

Discussion

In this report, we investigated various parameters that are important to consider when designing large-scale iTRAQ experiments for plasma biomarker discovery. Our results show that even though there are several steps involved in the preparation of plasma samples for iTRAQ, the technique itself affords high quantitative reproducibility at the protein level which is key for detecting small changes in abundances with high statistical confidence. As an estimate of the variance from the technical process appears relatively minor (CV 11.7%), and with the knowledge that the time to conduct an iTRAQ experiment is relatively long, a challenge for biomarker discovery is to weigh the benefit of sampling large numbers of different individual samples, or focus on sampling replicates to overcome the problems of 'single-hit' peptides (Figure 1). The problem of single-hit peptides with large comparative studies is that much of the data is missing in each run which prevents comparative analysis. While replicate analysis of the same sample will detect some additional peptides, the MS duty cycle associated with information dependent data acquisition means that many single-hit peptides will still remain (and new ones obtained). In fact, the majority of data will simply be replicate analysis of abundant peptides, many of which already produced reliable data in the initial analysis. This leads us to conclude that the analysis of more individual samples should be prioritized above replicate analysis.

Representative biological samples were used to benchmark iTRAQ variance levels, allowing us to determine statistically sound sample sizes for future biomarker discovery experiments. We calculated sample size by using standard values of 0.05 and 0.2 for α and β error constants, respectively.[19] The experimentally determined SD value of 0.65 (which accounts for 90% data) was considered, meaning that a 2-fold or 1.5-fold change in expression would require profiling of a minimum of 8 or 20 samples, respectively.

As shown in Figure 4, for some of the iTRAQ reporter ions of lower intensity, we noted considerably large variances of peptide ratios. One explanation may be variable quantities of endogenously processed peptides present in the plasma samples. Such endogenous peptides may represent a different sample 'population' to the tryptic peptide 'population' produced by our sample processing of the intact proteins. It is well-known that plasma contains many endogenous protein fragments, and further, undergoes rapid proteolysis upon collection.[20–22] Even though we observed great diversity of quantitative data at the peptide level, run-to-run variation at the protein level was well-controlled due to the weighted scoring function of ProteinPilot that appears to favor iTRAQ reporter peptides of higher intensity. From a practical perspective, our observations suggest to err on the side of caution in relying on iTRAQ ratios obtained from single peptide measurements as, in our experience, are commonly observed with lower abundance plasma proteins. There is a strong likelihood that such peptide-level measurements inaccurately represent the protein level. Increasing peptide product ion spectra acquisition time will increase iTRAQ reporter ion signal intensity; however, this comes at the expense of decreased sampling, resulting in fewer protein identifications.

The use of ProteinPilot for peptide identification and quantitation resulted in fewer peptide identities compared to analysis using Mascot (data not shown). However, analyses using ProteinPilot are advantageous as there are no identity redundancies ensuring that quantitation is attributed to a peptide that is exclusively supported by MS/MS data. The limited numbers of individual proteins identified in this study, despite immunodepletion of 12 abundant proteins, highlights the challenges of plasma-based biomarker discovery. In our hands, similar iTRAQ analyses of tissues or cells identifies more than 1000 different proteins from a single run. Clearly, the presence of many highly abundant proteins in plasma limits the depth of analysis, warranting newer strategies to remove abundant plasma proteins or enrich for less abundant proteins to facilitate efficient biomarker discovery.

For plasma biomarker discovery experiments that require comparative analyses from multiple iTRAQ experimental runs, we have shown that it is imperative to use the same reference sample (the denominator) throughout these studies. In principle, our data supports the use of either a single individual sample or a pooled sample in the reference channel. Nonetheless, we favor the use of a pooled reference sample as material can be obtained by mixing small aliquots from multiple individual samples to provide sufficient reference material for use in all subsequent experiments. Further, the use of an individual sample as a reference may be practically limiting given the number of experiments that may be required in a biomarker study. As we have already indicated, it is not possible to achieve reliable quantitative results should the reference sample need to be changed midway through a comparative experiment due to reference sample exhaustion (Figure 5). A further disadvantage of the single individual reference is that it may not contain all of the proteins that are detected in comparative samples. In such a case, the quantitation software may produce an erroneous ratio. It is important that this type of error is considered prior to further statistical analysis.

A pooled reference sample consisting of aliquots from numerous samples (controls as well as patient samples) should ensure that the reference channel contains all proteins needed for quantitation. With the use of this approach, each iTRAQ run enables 3 individual samples (combinations of controls and disease samples) to be profiled and ratios determined relative to the pooled reference. Statistical tests can then be applied to examine differences between sample groups. Importantly, this approach maintains the analysis of individual samples which is useful for identifying interindividual variation and outliers. An alternate approach would be to simply construct two pools, one for each group (control and disease) and label with iTRAQ. This approach averages the protein profiles for each group and potentially offers a rapid approach to flag candidate biomarkers. However, this approach should be taken with caution as individual outlier samples may confusingly skew the averaged result. The clear advantage of this approach is the shortened acquisition time; however, only limited statistical analysis can be conducted, thereby requiring subsequent candidate confirmation strategies. A second generation commercially produced iTRAQ reagent offering 8 quantitation channels will be invaluable in reducing data acquisition time,[23] which would support analysis of individual samples in each iTRAQ channel.

Acknowledgment. We thank Chris Clarke and Rebecca Touma for their technical assistance and Dr. Ruth VanBogelen and Mr. Lindsay Woods for facilitating the research. This study was supported by Pfizer, Inc. and facilitated by access to the Australian Proteome Analysis Facility Ltd established under the Australian Commonwealth Government's Major National Research Facilities Scheme.

References

(1) Chelius, D.; Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **2002**, *1* (4), 317–323.
(2) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **2003**, *75* (18), 4818–4826.
(3) Liu, H.; Sadygov, R. G.; Yates, J. R., III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201.
(4) Wang, G.; Wu, W. W.; Zeng, W.; Chou, C. L.; Shen, R. F. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J. Proteome Res.* **2006**, *5* (5), 1214–1223.
(5) 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* (10), 994–999.
(6) 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.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–1169.
(7) Ong, S. E.; 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. Proteomics* **2002**, *1* (5), 376–386.
(8) Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell. Biol.* **2006**, *7* (12), 952–958.
(9) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **2001**, *73* (13), 2836–2842.
(10) Yao, X.; Afonso, C.; Fenselau, C. Dissection of proteolytic ^{18}O labeling: endoprotease-catalyzed ^{16}O-to-^{18}O exchange of truncated peptide substrates. *J. Proteome Res.* **2003**, *2* (2), 147–152.
(11) Shilov, I. V.; Seymour, S. L.; Patel, A. A.; Loboda, A.; Tang, W. H.; Keating, S. P.; Hunter, C. L.; Nuwaysir, L. M.; Schaeffer, D. A. The Paragon Algorithm: A next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* **2007**, *6* (9), 1638–1655.
(12) Abdi, F.; Quinn, J. F.; Jankovic, J.; McIntosh, M.; Leverenz, J. B.; Peskind, E.; Nixon, R.; Nutt, J.; Chung, K.; Zabetian, C.; Samii, A.; Lin, M.; Hattan, S.; Pan, C.; Wang, Y.; Jin, J.; Zhu, D.; Li, G. J.; Liu, Y.; Waichunas, D.; Montine, T. J.; Zhang, J. Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. *J. Alzheimers Dis.* **2006**, *9* (3), 293–348.
(13) 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. Differential 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* (5), 1143–1154.
(14) DeSouza, L.; Diehl, G.; Rodrigues, M. J.; Guo, J.; Romaschin, A. D.; Colgan, T. J.; Siu, K. W. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome Res.* **2005**, *4* (2), 377–386.
(15) Wu, W. W.; Wang, G.; Baek, S. J.; Shen, R. F. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J. Proteome Res.* **2006**, *5* (3), 651–658.
(16) 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* (2), 821–827.
(17) Maurya, P.; Meleady, P.; Dowling, P.; Clynes, M. Proteomic approaches for serum biomarker discovery in cancer. *Anticancer Res.* **2007**, *27* (3A), 1247–1255.
(18) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002**, *1* (11), 845–867.
(19) Motulsky, H. *Intuitive Biostatistics*; Oxford University Press: New York, 1995; pp 195−204.
(20) Richter, R.; Schulz-Knappe, P.; Schrader, M.; Standker, L.; Jurgens, M.; Tammen, H.; Forssmann, W. G. Composition of the peptide fraction in human blood plasma: database of circulating human peptides. *J. Chromatogr., B: Biomed. Sci. Appl.* **1999**, *726* (1−2), 25–35.
(21) Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2003**, *2* (10), 1096–1103.
(22) Yi, J.; Kim, C.; Gelfand, C. A. Inhibition of intrinsic proteolytic activities moderates preanalytical variability and instability of human plasma. *J. Proteome Res.* **2007**, *6* (5), 1768–1781.
(23) Pierce, A.; Unwin, R. D.; Evans, C. A.; Griffiths, S.; Carney, L.; Zhang, L.; Jaworska, E.; Lee, C. F.; Blinco, D.; Okoniewski, M. J.; Miller, C. J.; Bitton, D. A.; Spooncer, E.; Whetton, A. D. Eight-channel iTRAQ enables comparison of the activity of 6 leukaemogenic tyrosine kinases. *Mol. Cell. Proteomics* **2008**, *7* (5), 853–863.

PR800072X