

Guidelines for the Routine Application of the Peptide Hits Technique

Ji Gao, Mark S. Friedrichs, and Ashok R. Dongre

Pharmaceutical Research Institute, Bristol-Myers Squibb Company, Princeton, New Jersey, USA

Gregory J. Opiteck

Caprion Pharmaceuticals, Montreal, Quebec, Canada

A set of guidelines has been developed for using the peptide hits technique (PHT) as a semi-quantitative screening tool for the identification of proteins that change in abundance in a complex mixture. The dataset that formed the basis for these experiments was created using a cell lysate derived from the yeast *Saccharomyces cerevisiae*, spiked at various levels with serum albumin (BSA), and analyzed by LC/MS/MS and SEQUEST. Knowing that the level of only one protein (BSA) actually changed in the mixture allowed for the development and refinement of the necessary bioinformatics and statistical analyses, e.g., principal component analysis (PCA), normalization, and analysis of variation (ANOVA). As expected, the number of BSA peptide hits changed in proportion to the amount of BSA added to the sample. PCA was able to clearly distinguish between the spiked samples and the untreated sample, indicating that PCA may be able to classify samples, e.g., healthy versus diseased, in future experiments. The use of an endogenous "housekeeping" protein was found to be superior to the use of total hits for data normalization prior to analysis. An ANOVA based model readily identified BSA as a protein of interest, that is, one likely to be changing from amongst the background proteins, indicating that an ANOVA model may be able to identify individual proteins in target or biomarker discovery experiments. General guidelines based on these combined observations are set forth for future analyses and the rapid screening for candidate proteins of interest. (J Am Soc Mass Spectrom 2005, 16, 1231–1238) © 2005 American Society for Mass Spectrometry

The use of protein biomarkers to detect diseases in their earliest stages holds the potential to save thousands of lives [1]. The advancement of proteomic techniques has recently led to the discovery of a series of biomarkers for different disease states [2–6]. While the traditional proteomic method (two-dimensional gel electrophoresis (2D-PAGE) followed by mass spectrometry) to elucidate differences between normal and disease-state samples still remains in use, this technique is not adequately reproducible, labor intensive, and hard to automate [7, 8]. Recent advances in mass spectrometers and their related informatics platforms, in particular the ion trap mass spectrometer and the SEQUEST algorithm, were fueled in part by the development of multidimensional high performance liquid chromatography (HPLC) techniques [9–12]. However, the lack of quantification in these methods led to the later development of the isotope coded affinity tag (ICAT) [13, 14] method and other labeling techniques (SILAC, GIST, AQUA) [15–17].

While these labeling techniques are being developed and refined, our laboratory has focused on performing semi-quantitative proteomics using the peptide hits technique (PHT). PHT relies on collating and summing up total number of peptide hits generated from tandem mass spectrometry data for a particular protein. Earlier studies in our laboratory indicate that the total number of peptide hits for a specific protein semi-quantitatively reflect its abundance in a given protein mixture [5, 18]. Although much less precise than the labeling methods, this PHT is much easier to implement and therefore less prone to errors or experimental artifacts. The PHT has been proven successful in several biomarker and target discovery programs [2, 3, 5, 19]. Liu et al. and Blondeau et al. have used a similar concept named spectral sampling and successfully estimated relative protein abundance in complex protein mixtures [20–22].

Previous projects using the PHT demonstrated that statistical tools commonly used for gene array experiments were also applicable to proteomic data [3, 18]. The purpose of the project described here was to refine, optimize, and guide the application of these bioinformatic techniques by analyzing a well-defined sample. Therefore, a set of recommendations or guidelines would be established for performing data analysis on

Published online June 23, 2005

Address reprint requests to J. Gao, Clinical Discovery Technologies, Bristol-Myers Squibb Company, P.O. Box 5400, Princeton, NJ 08534-5400, USA. E-mail: ji.gao@bms.com

Table 1. Samples information

Treatments	BSA loading (μg)	Yeast loading (μg)
A	0.00	50.00
B	0.20	50.00
C	0.39	50.00
D	0.78	50.00
E	1.56	50.00
F	3.13	50.00
G	6.25	50.00
H	12.50	50.00

future biomarker and/or target discovery projects where the compositions of the samples would be unknown or less well-defined. In the present case, different known amounts of serum albumin (BSA) were spiked into a yeast whole cell lysate, digested, run through LC/MS/MS system, and then searched against SEQUEST. The data was subsequently analyzed using PCA to broadly distinguish the spiked samples from the untreated samples. Various normalization methods were explored using total peptide hits and “housekeeping” proteins. Finally, an ANOVA based method was applied to detect the changes in BSA from amongst the background yeast proteins.

Materials and Methods

Protein Samples

A protease deficient strain of the baker’s yeast *Saccharomyces cerevisiae* termed SGY 1276, a Squibb Genetic Yeast which was derived from the publicly available strain Y197, was grown in 10 mL volumes of SD minimal media (Clontech, Palo Alto, CA) at 30 °C. The culture media was supplemented with uracil, tryptophan, and leucine. The cells were harvested in log phase (o.d. <1; $\sim 1 \times 10^7$ cells per mL) by centrifugation at 1000 g at 4 °C for 30 min. Each pellet was washed with phosphate buffered saline solution (Sigma, St. Louis, MO) and lysed in 500 μL Y-Per (Pierce, Rockford, IL) according to the manufacturer’s instructions. The lysates were pooled and protein concentration was determined by BCA assay [23]. Bovine serum albumin was obtained (Sigma) and used without further purification.

Proteolysis

The yeast cell lysate was subjected to chloroform methanol protein precipitation and the protein pellet was brought back into solution with 8 M urea/100 mM ammonium bicarbonate. To gain statistical power in future data analysis, triplicate sets of samples were prepared. For each set of samples, yeast proteins were split into eight aliquots and different amounts of BSA were spiked into the various yeast solutions to make eight protein mixtures with final BSA concentrations of 0, 2.0, 3.9, 7.8, 15.6, 31.3, 62.5, and 125 ng/ μL , respectively (These eight conditions would be referred to as treatment A, B, C, D, E, F, G, and H). Each treatment

contained the same concentration of total yeast protein (1000 ng/ μL). All the samples were digested as previously described [18]. Samples were subsequently dialyzed (Dispo-Microdialyzer, 500 Dalton, The Nest Group, Southborough, MA) against 0.2 % acetic acid for 3 h at room temperature. Samples were stored at 4 °C before loading onto the LC/MS/MS system.

Automated LC/MS/MS System

Reversed-phase high performance liquid chromatography coupled to tandem mass spectrometry was conducted using a microbore column (Zorbax 300 SB-C18, 3 μm , 1 \times 150 mm, Agilent, Wilmington, DE). The autosampler (Famos, Dionex/LC Packings, San Francisco, CA) was supplied with mobile phase from a binary pump (Agilent 1100). Each injection made use of a 50 μL loop (The amount of BSA and yeast on the column for each of the eight treatments was described in Table 1). The flow rate through the microbore column was held constant at 50 $\mu\text{L}/\text{min}$. The mobile phases (MP) used were as follows: MPA = 0.001% trifluoroacetic acid (TFA) + 0.1% acetic acid + 0.2% 2-propanol in water, MPB = 0.001% TFA + 0.1% acetic acid + 0.2% 2-propanol + 95% acetonitrile in water. The gradient was 0–2 min, 0–10% MPB; 2–62 min, 10–40% MPB; 62–67 min, 40–100% MPB; 67–70 min, 100% MPB; 70–80 min, 0% MPB.

The effluent from the RPLC column was connected to an ion trap mass spectrometer (LCQDeca, Thermo-Electron, San Jose, CA) equipped with a standard electrospray source. The capillary temperature was set to 250 °C, and the electrospray voltage was set to 5 kV. Automatic gain control was switched on with a target value of 5×10^7 for each survey scan which maximum injection time was 50 ms. For each MS/MS scan, the

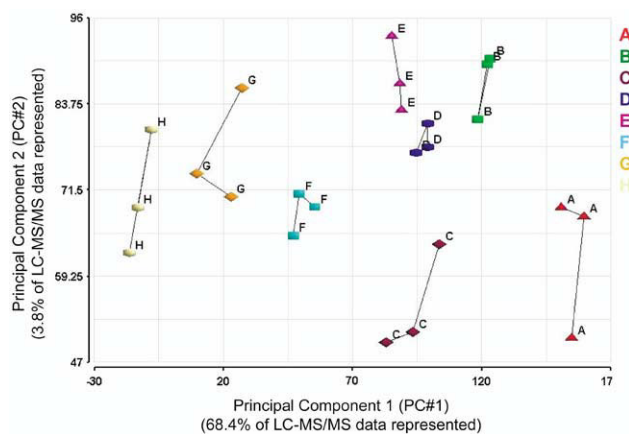


Figure 1. Two dimensional representation of the principal component analysis (PCA) of all 24 datasets (eight treatments in triplicate) and 700 protein identifications across 24 datasets. The X-axis represents PC#1 and the Y-axis represents PC#2. The first two principal components represent 72.3% of the LC-MS/MS data. The data clearly shows that replicates of each treatment (A to H) are readily clustered. There is an overall trend from treatment A to treatment H (from low BSA spike to high BSA spike) as well.

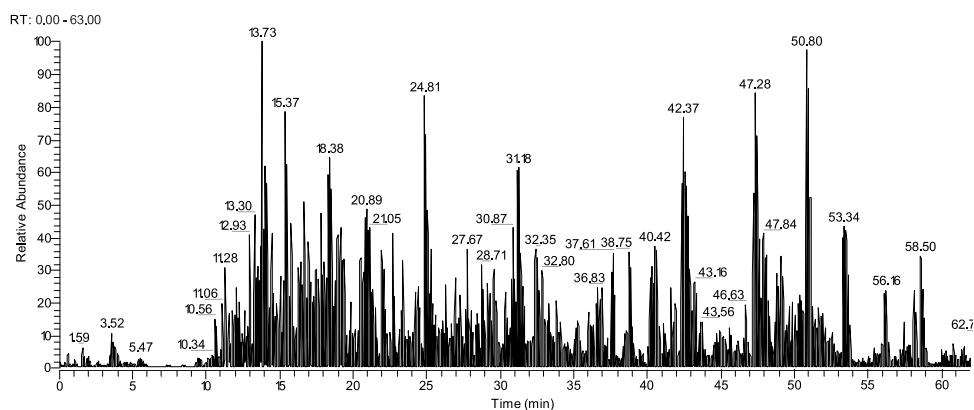
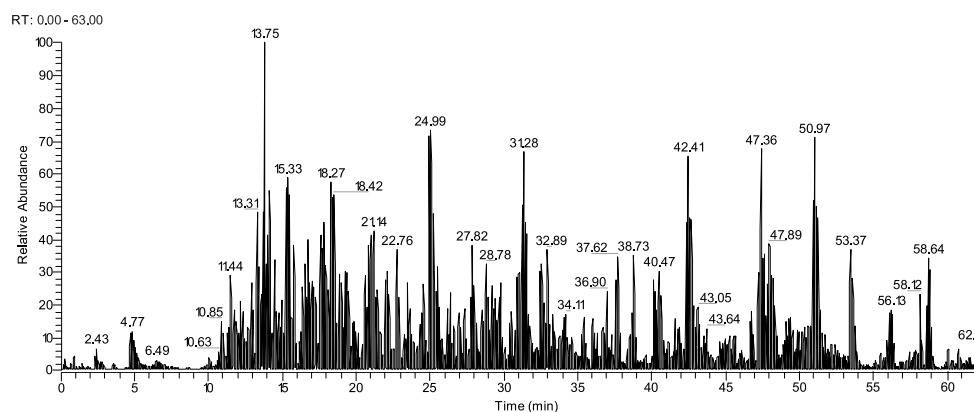
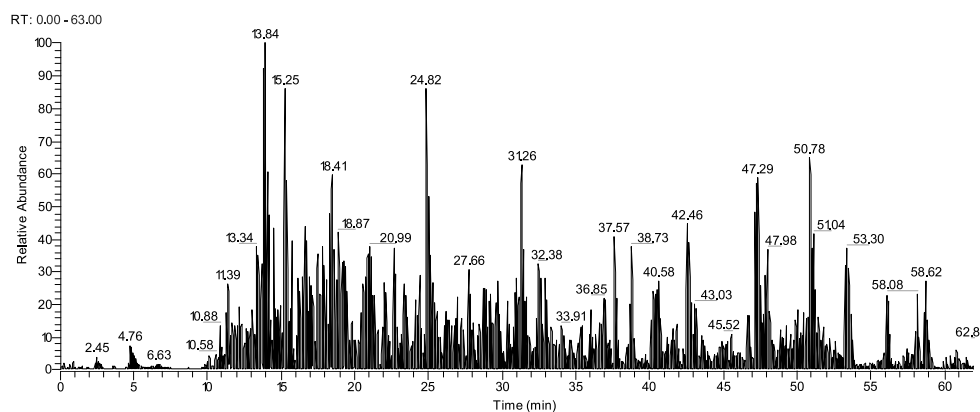
(a) Replicate 1**(b) Replicate 2****(c) Replicate 3**

Figure 2. Representative total ion chromatograms for replicate samples. The ion chromatograms shown are replicate samples for treatment D.

target value was 2×10^7 ions with maximum injection time of 400 ms. The instrument was set to trigger data dependent fragmentation of the three most intense ions during the MS survey scan. A total of 24 data sets were collected.

Data Analysis

The SEQUEST algorithm was applied to each of the data file sets using the yeast proteome database (YPD, Incyte, Beverly, MA) plus BSA [24]. No SEQUEST score

Table 2. Pearson's correlation coefficient among triplicate data sets across all eight treatments

Treatments	Replicates 1 and 2	Replicates 1 and 3	Replicates 2 and 3
A	0.97	0.96	0.96
B	0.96	0.96	0.96
C	0.96	0.96	0.95
D	0.96	0.96	0.96
E	0.94	0.95	0.95
F	0.95	0.95	0.95
G	0.94	0.93	0.94
H	0.95	0.94	0.94

filters were applied to the data set. For a typical LC/MS/MS data set, approximately 4500 MSMS spectra were acquired, about 2800 of those were assigned to peptides. These numbers were reproducible across all datasets. The total number of protein identifications was greater than 5700 from the 24 data files which would include both false positives and false negatives. The highest scoring SEQUEST peptide hit was used for each spectrum searched. The 24 data files were compiled in a spreadsheet program (Excel, Microsoft, Redmond, WA). In order to make the dataset more manageable, proteins were deleted if they were not observed in at least 6 out of the 24 data files. This reduced the number of total protein identifications to 700. Data analyses were performed in Partek for PCA, Student's *t*-test and an ANOVA model with Bonferroni correction (Partek Inc., St. Charles, MO).

Results and Discussion

The data was first subjected to principle component analysis (PCA). PCA is a method commonly used in microarray research to draw cursory, broad correlation between mRNA expression profiles by reducing the dimensionality of the datasets [25–27]. PCA is designed to capture the variance in a dataset in terms of its principle components. In effect, one is trying to reduce the dimensionality of the data to summarize the most important (i.e., defining) parts whilst simultaneously filtering out noise. As the LC/MS/MS dataset represented here was multi-dimensional, it was beneficial to perform PCA in order to draw broad correlation among protein profiles. The PCA matrix consisted of twenty four datasets (eight treatments consisting of three replicates per treatment) and 700 protein identifications across the twenty four datasets. After PCA was performed on this 24×700 matrix, the first principal component represented 68.4% of the data and the second principal component represented 3.8% (see Figure 1). The remaining 27.7% of data were distributed among remaining 21 principal components with no single principal component representing more than 3.8% of the data. In essence, ~70% of the LC-MS/MS data represented by the first two principal components of the PCA clearly distinguished the spiked samples from the untreated samples. Furthermore, the overall

trend A–H (low BSA spike to high BSA spike) was captured by PCA as well. This indicates that PCA could be a rapid data analysis tool to identify broad trends in protein profiling experiments. It could also be used to triage samples taken from a broad population in order to find protein profiles that correlate with treatment, response, or adverse event. Such analyses could also be useful to analyze data in order to ensure that the data is clustered by biological rather than artificial (time of analysis, date of analysis, analyst, etc.) effects.

Figure 2 represents typical total ion chromatograms generated during LC/MS/MS analysis from replicate samples for a specific treatment. It demonstrates qualitative reproducibility for replicate analysis. Additionally, Pearson correlation coefficients were generated among the triplicate datasets to study the quantitative reproducibility of replicate analyses for each treatment. A Pearson correlation coefficient is a dimensionless index that ranges from –1.0 to 1.0 and reflects the extent to which a linear relationship exists between two data sets. If all the proteins had exactly the same number of peptide hits in each replicate experiment, the correlation coefficient would be 1.00. For the eight treatments, A, B, C, D, E, F, G, and H, the correlation coefficients of all replicate analyses were above 0.90 (Table 2) indicating high overall quantitative reproducibility.

The following set of analyses focused on the change in BSA peptide hits as a function of loading. As ex-

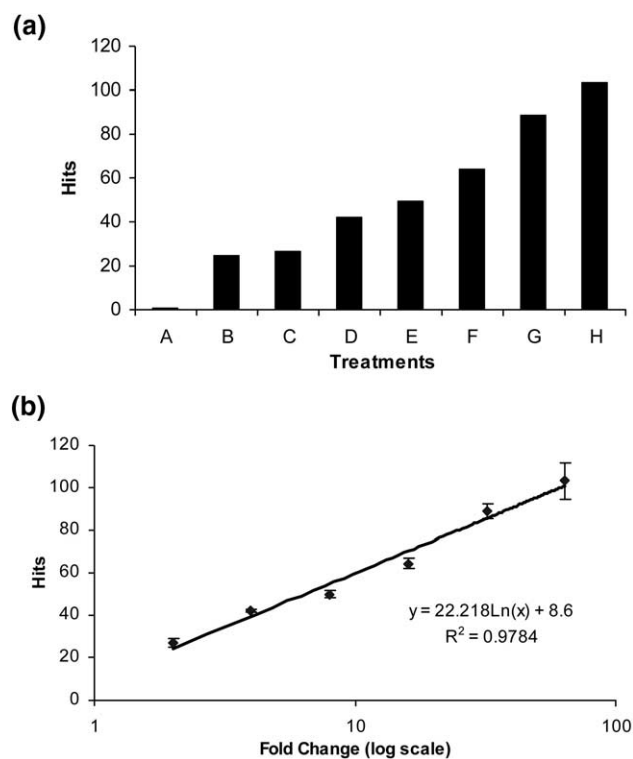


Figure 3. The change of peptide hits as a function of loading. (a) shows as the abundance of BSA increases, the peptide hits increase. (b) demonstrates peptides hits correlate with BSA level in a logarithmic fashion.

Table 3. (a) BSA hits change with the abundance of BSA

Treatments	BSA loading (μg)	BSA fold change	Peptide hits (sample 1)	Peptide hits (sample 2)	Peptide hits (sample 3)	Average	%CV
A	0.00	0	2	1	0	1	100.0
B	0.20	0	24	27	23	25	8.4
C	0.39	2	27	27	26	27	2.2
D	0.78	4	43	40	43	42	4.1
E	1.56	8	47	51	51	50	4.6
F	3.13	16	68	63	62	64	5.0
G	6.25	32	80	97	90	89	9.6
H	12.50	64	101	104	105	103	2.0

pected, the peptide hits for BSA increased as the amount of BSA in the samples increased, though not in a 1:1 ratio, but instead in a clear logarithmic relationship (Figure 3). With the standard curve of $R^2 = 0.9784$, experimental peptide hits value can be easily back-calculated to fold change of the protein. A Student's *t*-test was used to determine if the changes of BSA hits observed between treatments were significant ($P < .05$). Among all treatments, hit changes between every two-fold change in BSA spiking (except treatments B versus C) were statistically significant (see Table 3). Furthermore, hit changes as small as 16% (treatment D versus E) could be detected with good probability (P value = .032). The comparison between B and C did not pass the Student's *t*-test (P value = .092), this could be attributed to the non-linearity of the PHT technique at the lower protein concentration. However, a fourfold change (between B and D) in BSA was found to be statistically significant (P value = .008). The results indicated that if the change of peptide hits of a protein between two treatment groups was statistically significant, most likely the biological change of abundance of the underlying protein was at least twofold. This is fortuitous, as this magnitude of change is that which is typically discussed when searching for biomarkers or targets [16, 28, 29].

Another set of statistical analysis was implemented to differentiate BSA from the yeast proteins among the eight treatments, knowing that only BSA was truly changing across the treatments. An ANOVA based method with Bonferroni correction was selected as a method of choice to analyze the

data. ANOVA is a powerful statistical method that can be used to separate and estimate the different causes of variation while Bonferroni correction applies a very conservative and rigorous statistical adjustment for multiple comparisons [30–32]. The results showed that 37 out of the 700 proteins passed the ANOVA with Bonferroni correction test (Table 4). In other words, 37 proteins showed a statistically

Table 4. Proteins that passed ANOVA with Bonferroni correction (Bonferroni = P value \times 700)

	Protein ID	Bonferroni	Pearson
1	P1_228883	1.22E-12	1.00
2	TDH3	3.36E-06	-0.94
3	TDH2	6.18E-06	-0.94
4	ENO1	8.16E-06	-0.90
5	ADH1	9.15E-06	-0.92
6	TDH1	1.04E-05	-0.94
7	PDC1	1.28E-05	-0.92
8	ENO2	1.82E-05	-0.92
9	CDC19	3.13E-05	-0.93
10	EFT1	6.59E-05	-0.89
11	SSB2	1.13E-04	-0.90
12	EFT2	2.11E-04	-0.88
13	SSB1	2.49E-04	-0.90
14	IPP1	2.61E-04	-0.85
15	YIL152W	4.97E-04	0.88
16	PGK1	5.46E-04	-0.92
17	YNL134C	6.29E-04	-0.92
18	CIT1	1.00E-03	-0.88
19	MET6	1.29E-03	-0.92
20	ARA1	1.40E-03	-0.78
21	PGI1	1.51E-03	-0.94
22	ACO1	2.20E-03	-0.93
23	SSA1	2.29E-03	-0.88
24	YHB1	2.73E-03	-0.90
25	HXK2	5.58E-03	-0.87
26	CYS3	6.32E-03	-0.88
27	GND1	7.35E-03	-0.92
28	SHM2	1.08E-02	-0.72
29	MET17	1.33E-02	-0.81
30	PDC6	1.41E-02	-0.85
31	HXK1	1.53E-02	-0.86
32	TEF1	1.54E-02	-0.91
33	TEF2	2.92E-02	-0.90
34	SSA2	3.40E-02	-0.84
35	ILV5	3.42E-02	-0.88
36	SSC1	3.51E-02	-0.85
37	UBI4	3.86E-02	-0.87

Table 3. (b) Percent change and Student *t*-test P value of BSA hits observed between neighboring treatments

Treatments	% Hit change	Student's <i>t</i> -test P value
B vs. C	7.4%	0.092
C vs. D	35.7%	0.003
D vs. E	16.0%	0.032
E vs. F	21.8%	0.022
F vs. G	28.0%	0.032
G vs. H	13.5%	0.036
*B vs. D	40.50%	0.008

*Comparison between B and D was performed because changes between B and C did not pass the Student's *t*-test (P value = .092).

significant difference among the experiments according to the model, even though only BSA was known to change. However, among the 37 proteins, BSA had the lowest P value (Bonferroni = $1.22E-012$), implying that it was most likely changing. The other 36 yeast proteins were false positives. In order to find out if there was any correlation between the changes of the 36 proteins and BSA loading across all eight treatments. Pearson correlation coefficient was calculated by comparing peptide hits of each of 36 proteins with those of BSA. Interestingly, out of the 36 false positives, 35 yeast proteins were inversely correlated with changes of BSA. This is probably the result of large amount of BSA peptides ions competing for fragmentation selection during the survey scans, thus increasing the likelihood of other peptides not being selected, fragmented, and identified. Figure 4 demonstrated a typical example: the change of average hits of Eno1 across all eight treatments compared with that of BSA. However, even though the peptide hits of Eno1 decreased while hits of BSA increased from treatment A to treatment H, the change of BSA (Bonferroni = $1.22E-12$) was much more significant than that of Eno1 (Bonferroni = $8.16E-06$).

To adjust for the “competition effect” and thereby reduce the false positives, two normalization (or scaling) methods that are commonly used in mRNA expression (transcriptional profiling) data analysis were explored. The first method scales the mean of the peptide hit values within a given run by multiplying all values by a constant. In this case, the peptide hits were summed for each individual run, and these sums were used to determine the run with the median total peptide hits. This value was then used as the basis for calculating the peptide hits multiplier (PHM) for each run (see eq 1).

$$\text{PHM}_n = T_{\text{median}}/T_n \quad (1)$$

Where PHM_n is PHM at treatment n , $n \in \{A, \dots, H\}$; T_{min} is the median total peptide hits across all the treatment, and T_n is the total peptide hits at treatment n , $n \in \{A, \dots, H\}$. The PHM was then used to scale the data in each

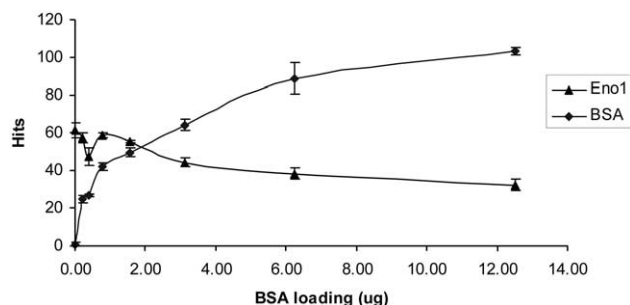


Figure 4. Numbers of peptide hits for Eno1 are shown to inversely correlate with those of BSA showing there is a competing effect between the two proteins. However, the change of BSA is more significant.

Table 5. Proteins that passed ANOVA with Bonferroni correction after brightness adjustment

	Protein ID	Bonferroni
1	BSA	5.07E-12
2	YLR322W	2.90E-06
3	YIL152W	1.64E-05
4	TDH1	3.60E-03
5	ARA1	7.17E-03
6	YNL134C	7.70E-03
7	CIT1	9.43E-03
8	TDH2	9.93E-03
9	YHB1	1.10E-02
10	PGI1	1.39E-02
11	YDL072C	1.84E-02
12	IPP1	1.85E-02
13	CDC19	1.90E-02
14	EFT1	3.96E-02
15	CYS3	4.03E-02
16	BNI5	4.47E-02

run, and at the end of this calculation, the summed values of peptide hits for each run were all identical. This PHM correlates to the Brightness Adjustment which is typically used for transcriptional profiling data analysis [33–35]. The same ANOVA model with Bonferroni correction was applied to the normalized data. The results showed, 16 proteins out of the 700 total proteins passed the test, and therefore false positives were reduced by over 50% (from 36 to 15) through the scaling process. As in the first analysis, BSA had the lowest Bonferroni value (Bonferroni = $5.07E-012$, Table 5).

Another popular normalization method applied in mRNA expression data analyses uses the intensity of a “housekeeping” protein (which is supposed to stay constant throughout treatment regimes) to scale the intensities of all proteins [36]. In order to reduce the number of false positives from the current 15 to zero, Fructose-Biphosphate Aldolase 1 (FBA1) was used to scale the original data set. This was based on the assumption that the abundance of FBA1 was, or should have been, constant across all samples. FBA1 is a fairly high abundant yeast protein with codon bias value of 0.868 [37], and it did not pass the ANOVA/Bonferroni test on original dataset. The reasons stated above made FBA1 a good candidate to perform normalization across datasets. In order to use FBA1 as a normalization protein, all the intensity values with each run would have to be normalized by a constant, so the peptide hits of FBA1 would be the same cross all the treatments after normalization (Figure 5a versus Figure 5b). After the normalization, the same ANOVA/Bonferroni model was applied and only four proteins showed a statistically significant difference. BSA once again came out on top of the list (Bonferroni = $8.28E-07$, Table 6). Even though there were still three false positives in the final data set, the P value with Bonferroni correction was still more than four orders of magnitude lower than the next closest protein (YIL152W) which averaged three hits

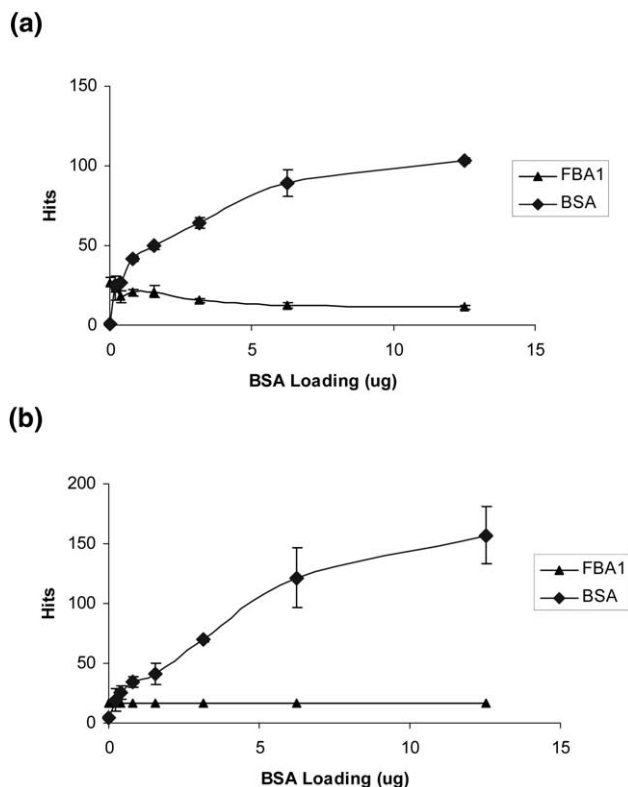


Figure 5. Numbers of peptide hits for FBA1 and BSA are shown before and after FBA1 normalization.

across the samples (Figure 6). Furthermore, error rate of three false positives out of 700 proteins was only 0.4%, and the false negative rate was effectively zero. Considering that the peptide hits technique is both intended to be, and used as, a simple and rapid whole proteome screening technique, and that any candidates so discovered will always be validated through at least one subsequent orthogonal method, such as immunoassay or AQUA, the error rates associated with this approach would not be of significant concern.

Conclusions

The PHT continues to mature as a useful technique for semi-quantitative proteome screening. Peptide hits were shown to correlate with protein abundance in a logarithmic fashion while the PCA method was shown to distinguish patterns of interest in large data sets. Furthermore, the ANOVA statistical model with Bonferroni correction was well suited to proteomic

Table 6. Proteins that passed ANOVA/Bonferroni after data was normalized according to FBA1

	Protein ID	Bonferroni
1	BSA	8.28E-07
2	YIL152W	1.12E-03
3	YDL072C	6.61E-03
4	HAP1	1.01E-02

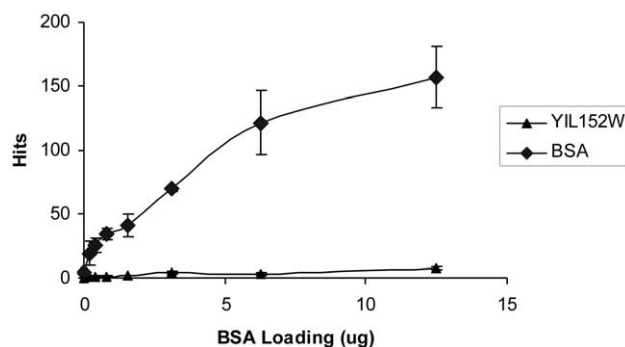


Figure 6. Numbers of peptide hits for YIL152W and BSA are shown after FBA1 normalization. It shows that the change of BSA is much more significant than that of YIL152W.

data analysis. Two different scaling techniques (total peptide hits adjustment and an abundant protein adjustment) were explored and of these, adjustment based on the use of an abundant protein was found to generate the lowest rate of false positive identifications (0.4%) while introducing no false negatives.

On the basis of the analyses shown here, we propose some general guidelines for future routine analyses using the peptide hits technique: design experiments to incorporate duplicate or even triplicate measurements (depending on sample source); prepare all samples in exactly the same way, at exactly the same time, and preferably by the same person in order to maximize reproducibility and minimize artifacts; screen initial data set by PCA to ensure that there are no underlying issues with the data set, e.g., time of day, day of week, analyst, instrument effects, etc. and that the treatment effect is dominant. Based on the data presented in this manuscript, PHT does well across a dynamic range of about two orders of magnitude. Therefore, a twofold change in protein abundance can be readily detectable using the PHT. Many of the current biostatistical tools such as PCA, ANOVA, *t*-test, as well as other vigorous statistical analyses are amenable to protein profiling and can serve to reduce the number of false positives. Furthermore, as in any profiling technique, once a list of protein candidates is generated, follow-on experiments must be performed to confirm the results and eliminate false positives.

Acknowledgments

D. J. Meyer (Partek, Inc.) was very helpful in helping the authors to analyze the data. Rameh Hafezi was instrumental in growing the yeast cultures. Douglas M. Robinson, Li-An Xu, and Keith Ho all provided critical reading and helpful comments for the manuscript.

References

- Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **2002**, *359*, 572–577.

2. Fach, E. M.; Garulacan, L. A.; Gao, J.; Xiao, Q.; Storm, S. M.; Dubaquié, Y. P.; Hefta, S. A.; Opiteck, G. J. In vitro biomarker discovery for atherosclerosis by proteomics. *Mol. Cell. Proteom.* **2004**, *3*, 1200–1210.
3. Gao, J.; Garulacan, L.; Storm, S. M.; Hefta, S. A.; Opiteck, G. J.; Lin, J. H.; Moulin, F.; Dambach, D. M. Identification of in vitro protein biomarkers of idiosyncratic liver toxicity. *Toxicol. In Vitro* **2004**, *18*, 533–541.
4. Gao, J.; Garulacan, L.; Storm, S. M.; Opiteck, G. J.; Yves, D.; Hefta, S. A.; Dambach, D.; Dongre, A. R. Biomarker Discovery in Biological Fluids. *Methods Enzymol.* **2004**, in press.
5. Pang, J. X.; Ginanni, N.; Dongre, A. R.; Hefta, S. A.; Opiteck, G. J. Biomarker discovery in urine by proteomics. *J. Proteom. Res.* **2002**, *1*, 161–169.
6. Zheng, Y.; Xu, Y.; Ye, B.; Lei, J.; Weinstein, M. H.; O'Leary, M. P.; Richie, J. P.; Mok, S. C.; Liu, B. C. Prostate carcinoma tissue proteomics for biomarker discovery. *Cancer* **2003**, *98*, 2576–2582.
7. MacCoss, M. J.; Yates, J. R. III. Proteomics: Analytical tools and techniques. *Curr. Opin. Clin. Nutr. Metab. Care* **2001**, *4*, 369–375.
8. Haynes, P. A.; Yates, J. R. III. Proteome profiling—pitfalls and progress. *Yeast* **2000**, *17*, 81–87.
9. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. III. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17*, 676–682.
10. Washburn, M. P.; Wolters, D.; Yates, J. R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–247.
11. Opiteck, G. J.; Jorgenson, J. W. Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. *Anal. Chem.* **1997**, *69*, 2283–2291.
12. Opiteck, G. J.; Ramirez, S. M.; Jorgenson, J. W.; Moseley, M. A. III. Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal. Biochem.* **1998**, *258*, 349–361.
13. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **1999**, *19*, 1720–1730.
14. Griffin, T. J.; Gygi, S. P.; Rist, B.; Aebersold, R.; Loboda, A.; Jilkine, A.; Ens, W.; Standing, K. G. Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. *Anal. Chem.* **2001**, *73*, 978–986.
15. 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. Proteom.* **2002**, *1*, 376–386.
16. Verhoeckx, K. C.; Bijlsma, S.; de Groene, E. M.; Witkamp, R. F.; van der Greef, J.; Rodenburg, R. J. A combination of proteomics, principal component analysis and transcriptomics is a powerful tool for the identification of biomarkers for macrophage maturation in the U937 cell line. *Proteomics* **2004**, *4*, 1014–1028.
17. Stemmann, O.; Zou, H.; Gerber, S. A.; Gygi, S. P.; Kirschner, M. W. Dual inhibition of sister chromatid separation at metaphase. *Cell* **2001**, *107*, 715–726.
18. Gao, J.; Opiteck, G. J.; Friedrichs, M. S.; Dongre, A. R.; Hefta, S. A. Changes in the protein expression of yeast as a function of carbon source. *J. Proteom. Res.* **2003**, *2*, 643–649.
19. Whitney, G.; Longphre, M.; McKinnon, M.; Burke, J. R.; Garulacan, L.; Gao, J.; Hefta, S. A.; Opiteck, G. J. RAI3: A Novel Target for the Treatment of Chronic Obstructive Pulmonary Disease, unpublished.
20. Liu, H.; Bergman, N. H.; Thomason, B.; Shallom, S.; Hazen, A.; Crossno, J.; Rasko, D. A.; Ravel, J.; Read, T. D.; Peterson, S. N.; Yates, J., III; Hanna, P. C. Formation and composition of the *Bacillus anthracis* endospore. *J. Bacteriol.* **2004**, *186*, 164–178.
21. 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*, 4193–201.
22. Blondeau, F.; Ritter, B.; Allaire, P. D.; Wasiaik, S.; Girard, M.; Hussain, N. K.; Angers, A.; Legendre-Guillemain, V.; Roy, L.; Boismenu, D.; Kearney, R. E.; Bell, A. W.; Bergeron, J. J.; McPherson, P. S. Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3833–3838.
23. Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **1985**, *150*, 76–85.
24. Eng, J. K.; McCormack, A. L.; Yates, J. R. III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
25. Misra, J.; Schmitt, W.; Hwang, D.; Hsiao, L. L.; Gullans, S.; Stephanopoulos, G. Interactive exploration of microarray gene expression patterns in a reduced dimensional space. *Genome Res.* **2002**, *12*, 1112–1120.
26. Crescenzi, M.; Giuliani, A. The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. *FEBS Lett.* **2001**, *507*, 114–118.
27. Raychaudhuri, S.; Stuart, J. M.; Altman, R. B. Principal components analysis to summarize microarray experiments: Application to sporulation time series. *Pac. Symp. Biocomput.* **2000**, 455–466.
28. Borozdenkova, S.; Westbrook, J. A.; Patel, V.; Wait, R.; Bolad, I.; Burke, M. M.; Bell, A. D.; Banner, N. R.; Dunn, M. J.; Rose, M. L. Use of Proteomics to Discover Novel Markers of Cardiac Allograft Rejection. *J. Proteom. Res.* **2004**, *3*, 282–288.
29. Ren, D.; Penner, N. A.; Slentz, B. E.; Regnier, F. E. Histidine-rich peptide selection and quantification in targeted proteomics. *J. Proteom. Res.* **2004**, *3*, 37–45.
30. Miller, J. C.; Miller, J. N. *Statistics for Analytical Chemistry*, 2nd ed.; John Wiley and Sons: Chichester, England, 1988; pp. 65–70.
31. Bonferroni, C. E. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze **1936**, 3–62.
32. Miller, R. G. *Simultaneous Statistical Inference*, 2nd ed.; Springer Verlag: New York, NY, 1981.
33. Affymetrix, *Statistical Algorithms Reference Guide*; Part Number 701110, 2001.
34. Affymetrix, *Fine Tuning Your Data analysis*; 2001.
35. Ulrich, R. G.; Rockett, J. C.; Gibson, G. G.; Pettit, S. D. Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. *Environ. Health Perspect.* **2004**, *112*, 423–427.
36. Bas, A.; Forsberg, G.; Hammarstrom, S.; Hammarstrom, M.-L. Utility of the Housekeeping Genes 18S rRNA, β -Actin and Glyceraldehyde-3-Phosphate-Dehydrogenase for Normalization in Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction Analysis of Gene Expression in Human T Lymphocytes. *Scand. J. Immunol.* **2004**, *59*, 566–573.
37. Fitcher, B.; Latter, G. I.; Monardo, P.; McLaughlin, C. S.; Garrels, J. I. A sampling of the yeast proteome. *Mol. Cell. Biol.* **1999**, *19*, 7357–7368.