Differential Label-free Quantitative Proteomic Analysis of *Shewanella oneidensis* Cultured under Aerobic and Suboxic Conditions by Accurate Mass and Time Tag Approach*§

Ruihua Fang‡, Dwayne A. Elias§, Matthew E. Monroe§, Yufeng Shen§, Martin Mcintosh‡, Pei Wang‡, Carrie D. Goddard§, Stephen J. Callister§, Ronald J. Moore§, Yuri A. Gorby§, Joshua N. Adkins§, Jim K. Fredrickson§, Mary S. Lipton§, and Richard D. Smith§¶

We describe the application of LC-MS without the use of stable isotope labeling for differential quantitative proteomic analysis of whole cell lysates of *Shewanella oneidensis* MR-1 cultured under aerobic and suboxic conditions. LC-MS/MS was used to initially identify peptide sequences, and LC-FTICR was used to confirm these identifications as well as measure relative peptide abundances. 2343 peptides covering 668 proteins were identified with high confidence and quantified. Among these proteins, a subset of 56 changed significantly using statistical approaches such as statistical analysis of microarrays, whereas another subset of 56 that were annotated as performing housekeeping functions remained essentially unchanged in relative abundance. Numerous proteins involved in anaerobic energy metabolism exhibited up to a 10-fold increase in relative abundance when *S. oneidensis* was transitioned from aerobic to suboxic conditions. Molecular & Cellular Proteomics 5:714–725, 2006.

Two of the most commonly used methods for quantitative proteomics are two-dimensional electrophoresis (2DE)¹ coupled to either MS or MS/MS and LC-MS/MS (1). In the 2DE-based approach, intact proteins are separated by 2DE, and the abundance of a protein is determined based on the stain intensity of the protein spot on the gel. The identity of the protein is now generally determined by MS analysis peptides after proteolysis of the protein spot. Since its inception in the mid-1970s, the 2DE-based approach has been routinely used for large scale quantitative proteomic analysis. Some of the disadvantages of this approach are that it is difficult to automate and has a limited detection capacity for proteins with extreme ranges in pI values, for hydrophobic proteins, and for low abundance proteins (1). It is generally believed that the intensity of an MS signal for a particular peptide does not always reflect its abundance due to ion suppression effects among co-eluting species. Thus, the LC-MS/MS-based approach often uses stable isotope labeling techniques, e.g. with ¹⁵N, ¹⁸O, stable isotope labeling by amino acids in cell culture (SILAC), and ICAT, to provide relative quantification (2–4). Although potentially providing the greatest accuracy, isotopic labeling has some disadvantages. Labeling with stable isotopes is expensive, and some labeling procedures involve complex processes and yield artifacts. It also can be computationally difficult to reliably define the isotopic “pairs” for relative quantification due to possible differences in the LC elution time of the labeled forms, incomplete mass spectrometric resolution of the isotopic pairs (5, 6), or the presence of other unresolved components.

An alternative approach involves “label-free” methods that use relative peptide peak intensities that are generally performed in conjunction with a data normalization procedure. A linear correlation between the amount of an analyte and its peak area can be obtained by using a sufficiently low LC flow rate and a small amount of sample because ESI approaches optimum (∼100%) efficiency under such conditions (7–9). This correlation has been demonstrated with simple mixtures of a few analytes (7, 8) and with peptides from mixtures of several proteins spiked into serum (10, 11). However, in the case of complex biological samples in which thousands of peptides are measured in one LC-MS analysis, many peptides co-elute (i.e. ion suppression effects could cause non-linear responses). These complex samples also require sufficient

From the §Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland Washington 99352 and ‡Computational Proteomics Laboratory, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024

Received, September 13, 2005, and in revised form, January 5, 2006

Published, MCP Papers in Press, January 9, 2006, DOI 10.1074/mcp.M500301-MCP200

¹ The abbreviations used are: 2DE, two-dimensional electrophoresis; AMT, accurate mass and time; UMC, unique mass class; AR, abundance ratio; SAM, significant analysis of microarray; FDR, false discovery rate; CV, coefficient of variance; PIPES, 1,4-piperazinediethanesulfonic acid; DO, dissolved oxygen; PMT, potential mass and time; Xcorr, cross-correlation; NET, normalized elution time; DOT, dissolved O₂ tension; BPI, base peak intensity.
reproducibility across multiple LC-MS analyses so that real biological differences can be distinguished from experimental variability.

Several research groups have undertaken broad studies to address these issues. Wang et al. (11) examined the reproducibility of LC-MS analyses of 105 serum samples and found the median coefficient of variation (CV) was 25.7%. They spiked various amounts of a mixture of two non-human proteins into a human serum sample and found that the signal of the peptides from the spiked proteins correlated linearly with the amounts of those proteins (11). Radulovic et al. (12) compared serum replicate analyses and obtained a good correlation ($R^2 = 0.84$) in scatter log$_{10}$ plots of peak intensities. They added troponin to the serum and were able to reproducibly detect the mass spectral peaks corresponding to the troponin and therefore reliably estimate the concentration of that protein in a complex matrix (12). Using very small diameter capillary LC with FTICR with an LC flow rate of ~20 nL/min and sample sizes of 50 ng and less, Smith et al. (13) recently showed that the peak area of highly abundant peptides in yeast proteome samples correlated linearly with the amount of sample used. Here we describe a label-free differential quantitative proteomic study using both peptides present in both samples types and unique to samples types with the accurate mass and time (AMT) tag approach. This approach was used to analyze the dissimilatory metal-reducing bacterium *Shewanella oneidensis* strain MR-1 cultured in bioreactors under aerobic (20% dissolved O$_2$ tension) and suboxic continuous culture conditions (<0.1% O$_2$).

**EXPERIMENTAL PROCEDURES**

**Cell Culture**—Steady state fermentor grown cultures were used in this study. The cultures (3 liters) were grown in fed-batch mode using a Bioflow 3000 model fermentor (New Brunswick, Inc.) and allowed to achieve steady state before sampling and harvesting. The medium was similar to that described previously (14) but was carbon-limited and contained 0.907 g/liter PIPES, 0.3 g/liter NaOH, 1.5 g/liter NH$_4$Cl, 0.1 g/liter KCl, 0.6 g/liter Na$_2$PO$_4$, 0.213 g/liter Na$_2$SO$_4$, and 16.85 g/liter oxalate as well as vitamins, minerals, and amino acids (glutamate, arginine, and serine). The cultures used O$_2$ as the terminal electron acceptor from a house air source (4 liters/min flow rate) and were maintained at 50% dissolved oxygen (DO) (controlled by a DO probe) during logarithmic growth. Once the culture attained an A$_{600}$ = 0.6 (final A$_{600}$ ~ 0.9), the DO was set to 20% for aerobic cultures and 0.1% O$_2$ for suboxic cultures while fresh medium was infused into the reactor at a dilution rate of 0.1/h to achieve steady state. Other monitored parameters included maintaining a pH of 7.00 (±0.03), a constant agitation of the culture at 450 rpm, and temperature of 30 °C.

Once steady state was attained for three fermentor volume changes, the culture was harvested for LC-MS proteomic analysis and sample archives. These samples were pelleted (11,900 × g for 8 min at 4 °C) and frozen at −80 °C until processing for proteomic analysis. In each case, harvesting took ~20 min from emptying the fermentor to placing the cell pellets in storage at −80 °C and resulted in ~1.2 × 10$^8$ cells/ml (a total of ~3.6 × 10$^{12}$ cells) and translated to ~400 mg of protein.

**Sample Preparation**—Cell pellets were washed and suspended in 100 mM NH$_4$HCO$_3$ buffer. Cells were lysed by bead beating, and cell debris were removed by centrifugation at 14,000 rpm at 4 °C for 5 min. Proteins in the supernatant were denatured with 8 M urea and reduced with DTT for 30 min at 60 °C. The sample was diluted 10-fold with 100 mM NH$_4$HCO$_3$ and digested with sequencing grade trypsin (Promega, Madison, WI) at 37 °C for 5 h with a trypsin:protein ration of 1:50 (w/w). The sample was cleaned using an appropriately sized C$_{18}$ solid-phase extraction column and concentrated in a SpeedVac to a volume of ~50–100 μL. An aliquot was used for a BCA assay, and the rest was frozen in liquid nitrogen and stored at −80 °C (15).

**Generating a Potential Mass and Time (PMT) Tag Database for *S. oneidensis*—**Trypsin-digested *S. oneidensis* samples grown under various culture conditions were fractionated by gradient strong cation exchange chromatography as described previously (13) into as many as 100 and as few as 25 fractions. Fractions were concentrations to about 1 μg/μl for analysis by LC-MS/MS. Whether fractionated or unfractonated the normalized peptide elution times were not found to be significantly different. The fractionated samples were subjected to LC-MS/MS analysis with a conventional ion trap mass spectrometer (LCQ, ThermoFinnigan, San Jose, CA) operating in a data-dependent MS/MS mode over a series of segmented m/z ranges. The MS/MS spectra were analyzed by the search/identification SEQUEST program against an *S. oneidensis* fasta database. Fully and partially tryptic peptides with a minimum cross-correlation (Xcorr) score (i.e. Xcorr = 2 for peptides with molecular mass >1000 daltons and 1.5 for peptides with molecular mass ≤1000 daltons) were termed PMT tags and stored in a Microsoft Structured Query Language relational database (referred to as the PMT tag database). The elution times of the peptides were normalized by regressing the peptide elution times observed against the elution time predicted for each peptide that was determined using a neural network-based model that was developed by training data from 20 species and over 140,000 unique filtered peptide identifications (16).

**LC-FTICR Analysis**—*S. oneidensis* samples from cells cultured under aerobic and suboxic conditions were digested with trypsin and analyzed in triplicate by LC-FTICR using a fully automated, custom built capillary LC system coupled on line to a 9.4-tesla (Bruker Daltonics) FTICR mass spectrometer without any fractionation. 5 μg of a trypsin-digested peptide sample were used for each analysis. Separations on the LC system were achieved using 5,000 p.s.i. reversed-phase capillary columns (60 cm × 150-μm inner diameter, Polymicro Technologies) packed with 5-μm Jupiter C$_{18}$ particles (Phenomenex, Torrance, CA) and two mobile phase solvents that consisted of 0.2% acetic acid and 0.05% TFA in water (A) and 0.1% TFA in 90% acetonitrile, 10% water (B). The LC flow rate was ~1.8 μL/min when equilibrated to 100% mobile phase A. Eluent from the HPLC was infused into the mass spectrometer by an ESI interface to an electrodynamic ion funnel assembly coupled to a radio frequency quadrupole for collisional ion focusing and highly efficient ion accumulation before transport to the cylindrical ICR cell for analysis. Mass spectra were acquired with ~10$^3$ resolution. Deconvolution and deisotoping of the mass spectral peaks were performed using ICR-2LS software developed in house implementing the THRASH algorithm (17). Using the THRASH algorithm, FTICR-MS signals are processed in each mass spectrum based on local signal to noise level criteria. With an applied signal to noise threshold of 5 there is no consistent bias with intensity on the confidence of identification.

**Peptide Identification of LC-FTICR Analysis**—Peptide identification and abundance determination were performed using software developed in house. Briefly those spectral peaks having a mass value initially within a tolerance of ±12 ppm and eluting in a minimum of three consecutive scans were defined as a unique mass class (UMC), i.e. a distinct putative peptide. The elution time of the UMC was defined as the peak apex and was normalized by linear regression
against the normalized elution time (NET) values of peptides in the PMT tag database. The UMCs were then compared with a subset of peptides in the PMT tag database that were filtered by the criteria described in Supplemental Table 1. This step involved comparing the mass and NET value of each UMC to the mass and NET value of the PMT tags using a mass tolerance of 5 ppm and a NET tolerance of 5%. This relatively large variation is, in part, due to the use of linear regression models for mass and the normalized elution times between LC analyses. The linear regression model used for NET regression (16) works well overall, but when local variations occur during the chromatography process they are compensated in part through wider tolerances (i.e., a 1% variation using non-linear means may be masked by a 5% variation with linear means).

Annotation of Hypothetical Proteins—BLASTP and PSI-BLAST were used to compare proteins for homology (http://www.ncbi.nlm.nih.gov/BLAST). Pfam (www.sanger.ac.uk/Software/Pfam) was used to predict the domain, and SignalP (www.cbs.dtu.dk/services/SignalP) was used to predict the signal sequence. PSORT (psort.nibb.ac.jp) was used to predict membrane sequence localization. The threading program mGenTHREADER (bioinf.cs.ucl.ac.uk/psipred/psiform.html) and PROSPECT (www.bioinformaticssolutions.com/products/prospect.php) were used to predict the three-dimensional structures of proteins.

RESULTS AND DISCUSSION

Differential Quantitative Analysis Using the AMT Tag Approach—The three steps used in the present differential quantitative analysis of S. oneidensis MR-1 using the AMT tag approach are summarized in Fig. 1 (also see “Experimental Procedures”). First, peptide samples obtained from tryptic digests of proteins from S. oneidensis cultured under various conditions were analyzed by two-dimensional LC-FTICR, respectively. Peptide identities were determined by matching against the peptides in the PMT tag database (a). Abundance ratios of peptides were calculated, and the abundance ratios of proteins were determined based on the abundance ratio of their peptides (b).

**Fig. 1.** Overview of the experimental flow. Trypsin-digested samples from different biological states were analyzed by LC-FTICR, respectively. Peptide identities were determined by matching against the peptides in the PMT tag database (a). Abundance ratios of peptides were calculated, and the abundance ratios of proteins were determined based on the abundance ratio of their peptides (b).
LC-FTICR. Mass spectral peaks with monoisotopic masses were grouped by mass values that fell within a specified mass range and by detection within a range of consecutive scans and designated as UMCs, each having a single representative mass, LC NET, and abundance value. The peptide identities for a UMC were determined by searching against the mass and LC elution times of peptides in the PMT tag database. Finally abundance ratios (ARs) of peptides and proteins were determined using peptides that were detected in at least four of the six replicates, and average values of the peptides were assigned to the missing values in the other replicates. The ARs of peptides and their corresponding proteins were calculated to compare the S. oneidensis proteomes under aerobic and suboxic conditions.

Experimental Design—Two biological replicates (CR23 and CR25) of S. oneidensis cells were cultured under steady state growth conditions in separate bioreactors with 20% dissolved O2 tension (DOT) and then switched to steady state growth under suboxic (<0.1% DOT) conditions. At both aerobic and suboxic growth conditions, samples were taken from each reactor, and each sample was digested with trypsin and analyzed three times by LC-FTICR. Each culture condition thus had a total of six replicates: two biological replicates, each with three MS replicates.

Confidence of Peptide Identifications—A total of 3919 fully or partially tryptic peptides, corresponding to 971 distinct proteins, were identified after matching the UMCs detected in LC-FTICR analyses using the PMT tag database. The measured mass and normalized elution time errors from matching UMCs were within ±5 ppm and ±5%, respectively (data not shown). Among these 3919 peptides, 2343 peptides (corresponding to 668 proteins) were identified in at least four of six replicates and were subjected to quantitative analysis.

The confidence levels of peptide identifications from the LC-FTICR analyses were also evaluated by comparing the distribution of SEQUEST Xcorr scores of the much larger set of peptides resulting from the larger set of LC-MS/MS analyses (Fig. 2a) with the distribution authenticated by LC-FTICR analyses (Fig. 2b). The Xcorr score measures the quality of the matched MS/MS spectra against a theoretical spectrum for a peptide. Hence a larger Xcorr value generally indicates a better spectral match and a greater degree of quality for the identified peptide. Fig. 2a displays only peptides with a charge state of +2, which represents the largest fraction of peptides (~80%) identified by SEQUEST analysis of MS/MS generated spectra. Of the peptides with an Xcorr >2, ~62% of the peptide identifications had Xcorr scores within the range of 2.0–2.5. The large percentage of peptide identifications with scores in this range skewed the distribution in Fig. 2a to the left. In contrast, the distribution of peptides with a charge state of +2 that were authenticated by LC-FTICR analyses resulted in a bell-shaped histogram with only 4.5% of these peptides having Xcorr scores in the 2.0–2.5 range (Fig. 2b). The majority of the peptides (>95%) confirmed by FTICR measurement had Xcorr scores above 3.

It is a common practice to define an arbitrary Xcorr score as a “threshold of quality” to determine whether a peptide is correctly identified by SEQUEST. For peptides with a charge state of +2, this threshold has been reported to vary from 1.5 to 2.5 (19, 20). Although the vast majority of peptides confirmed by LC-FTICR MS measurement are of higher confidence, with ~95% of them having an Xcorr score above 3, a majority of the peptides identified by LC-MS/MS alone using a simple Xcorr >2 filter are in the lower confidence range with an Xcorr score between 2.0 and 2.5. These scores show that many of the peptides identified by MS/MS alone with Xcorr values at the lower end (Xcorr of 2–2.5) include many false positive identifications and can be “filtered out” using accurate mass measurements as shown previously (18). Both the abundance spectrum of the identified peptides and the overall population of UMCs detected by LC-FTICR covered the entire abundance range. However, the abundance distribution of the high confidence peptide population sequenced is skewed toward peptide ions of higher intensity. This is consistent with the expectation that low abundance species will result in lower quality MS/MS spectra that will more often result in incorrect identifications (Supplemental Fig. 1).

Preprocessing and Normalization of LC-FTICR Data—There are many sources of systematic variation in LC-MS-based proteomic experiments that affect protein abundance measurements. These sources include variations in sample preparation, LC separation, electrospray ionization efficiency, and mass spectrometer performance. Normalization is the term used to describe the process of removing or correcting for such variations.

Two global normalization schemes, in which the same normalization factor is used for all the spectral peaks, were
evaluated in this study. In both cases, one file is used as a reference, and the rest of the files are normalized against this reference. One scheme, first applied by Wang et al. (11), used the median value for the intensity ratio of all the spectral peaks between the files as the normalization factor for the file. The second scheme used 302 peptides from 56 putative "housekeeping" proteins (e.g. ribosomal proteins, translation elongation factors, and chaperone proteins) for normalization. The normalization factor was taken as the ratio of the intensities of these housekeeping peptides after linear regression of the nonreference file over the reference. This approach is based on the assumption that the abundance of proteins in this functional category should remain relatively constant in cells growing at the same dilution rate (0.1 h⁻¹) in defined medium regardless of the dissolved O₂ status of the culture. This anticipated behavior of housekeeping proteins is supported by the following two observations. 1) The linear regression R² values for the scatter plots of all LC-FTICR analysis pairwise comparisons between different culture conditions of different biological samples (R² = 0.88 ± 0.04, n = 18) are very similar to those between the same culture conditions of different biological sample (R² = 0.93 ± 0.02, n = 18). 2) The ARs of these proteins between the two culture conditions are in the range of 0.67–1.3, suggesting that the abundances of these proteins generally remain relatively constant.

Reproducibility—Upon normalization of the peptide abundances across all the LC-FTICR analyses to correct for possible systematic bias, both experimental and instrument reproducibility were evaluated using base peak intensity (BPI) chromatograms, scatter plot analyses, and CVs. Fig. 3 (a–f) shows the BPI chromatograms of six aerobic replicates, whereas Fig. 3 (g–l) shows the BPI chromatograms of six suboxic replicates. The BPI profiles from the same culture conditions are very similar in appearance, indicating good instrument and biological reproducibility, whereas prominent features that differ between the chromatograms of samples from aerobic and suboxic cultures are readily apparent by visual inspection.

To evaluate instrument and biological reproducibility, we examined scatter plots (Fig. 4) of peptide abundances for S. oneidensis from (a) MS replicates of duplicate samples collected from the same bioreactor at the same time (i.e. aerobic phase), (b) different samples collected from the same bioreactor at different time points (i.e. under suboxic and aerobic growth conditions), (c) samples from different bioreactors operating under identical aerobic continuous culture conditions, and (d) different bioreactors and different culture conditions (i.e. suboxic and aerobic). We observed good reproducibility, indicated by root sum square (R²) values, for replicates of samples from the same bioreactors cultivated.
under identical conditions (Fig. 4a, $R^2 = 0.95–0.96$) and for samples from different bioreactors cultivated under identical conditions (Fig. 4c, $R^2 = 0.83–0.87$). The lower $R^2$ value associated with Fig. 4c relative to Fig. 4a is consistent with expected experimental variation, a factor that is often underestimated. On the other hand, the $R^2$ value associated with Fig. 4b, which compares the peptide abundances of proteins from MR-1 cells cultivated under aerobic and suboxic continuous culture from the same bioreactor, was significantly lower at 0.75–0.77. Furthermore the $R^2$ value of Fig. 4d, which compares the aerobic and suboxic cultures from different bioreactors, was even lower at 0.65–0.67. These results indicate that the abundance of certain peptides, and therefore the proteins they represent, changed significantly in suboxic cultures compared with aerobic cultures.

The distribution of CVs from all 12 LC-FTICR analyses can be seen in Supplemental Fig. 2. The median and mean CVs of the normalized peptide abundances are 16.0 and 18.2%, respectively.

**Confidence of “One-state” Peptides**—Peptides that were identified in one culture condition but not the other in at least four of six replicates were defined as one-state peptides, whereas peptides identified in both aerobic and suboxic samples were referred to as “two-state” peptides. It is unlikely that the present set of one-state peptides includes significant false positive identifications due to the strict filtering criteria used. As an example, fumarate reductase flavoprotein subunit precursor (SO0970) showed a total of 28 peptides identified, among which 17 were two-state and 11 were one-state peptides. Importantly all 11 one-state peptides were identified in at least four of the six suboxic replicates, but none were identified in the aerobic replicates. If these peptides were false positives caused by random errors, then they would likely have been distributed in a random fashion, which some would have been identified in the aerobic and others in the suboxic samples. The abundances of the 11 one-state peptides in the suboxic samples were generally much lower than the 17 two-state peptides (Fig. 5a), suggesting that the SO0970 one-state peptides...
in the aerobic sample were probably below the detection limit due to the lower protein abundance in this sample.

171 one-state peptides, corresponding to 53 proteins, were identified and quantified, among which 169 were fully tryptic and two were partially tryptic. The high Xcorr values for these one-state peptides are indicative of highly confident identifications (see Supplemental Table 2). For example, 85% of these peptides had a +2 charge, and of those, 94% had Xcorr values ≥2.5, and 80% had Xcorr values ≥3. Their abundance measurements also exhibit good reproducibility as indicated by the distribution of their CVs (see Supplemental Fig. 3) with median and mean values of 22.0 and 23.3%, respectively.

Determination of ARs for Two-state and One-state Peptides—For two-state peptides, the AR between the aerobic and suboxic conditions was calculated as the ratio of arbitrary peptide abundances measured by LC-FTICR analysis. To calculate the AR of one-state peptides, the smallest magnitude abundance value of the entire LC-FTICR analysis was used as the reference value. The actual abundance should be equal to or lower than the reference value, and the actual AR should therefore be equal to or higher than the calculated AR.

On closer examination of proteins containing both one-state and two-state peptides, the average AR in the majority of the one-state peptides was observed to be greater than that in the two-state peptides (Fig. 5b). For example, the average AR of the 11 one-state peptides of SO0970 was 7.1 ± 0.8, whereas the average AR of the 17 two-state peptides of SO0970 was 3.9 ± 0.4. This trend was also observed in all other proteins that contained both one-state and two-state peptides (Fig. 5b), suggesting that two-state peptides may have suffered from a reduction in MS response compared with one-state peptides or that the use of the lowest abundance value may under estimate the lowest detectable species for the overall population.

Reduction in MS response can be caused by several reasons, and e.g. for ESI it is anticipated that significant effects can arise due to ionization suppression and space charge effects, particularly for highly abundant species. Although it is...
difficult to estimate the degree of reduction in MS response of a peptide in the mixture because it depends on complex factors, including concentrations, proton affinities, and surface activities of various co-eluting analytes (21) as well as separation efficiencies, the above present results suggest that low abundance peptides in general suffer less reduction in their MS response than do high abundance peptides.

Identification and Confirmation of Significant Changes in Relative Protein Abundance—Distinguishing real biological change from experimental noise with proteomics and transcriptomics is a challenge because thousands of peptides or genes are measured in the same experiment, meaning that even when no biological differences exist, some large number of individual features may still display changes that meet ordinary levels of significance (i.e. a p value of 0.05, or a 2-fold change). One approach to accommodate multiple comparisons is to identify the false discovery rate (FDR), which has been commonly used in transcriptomics, for example using statistical analysis of microarrays (SAM) (22). The false discovery rate estimates the fraction of all significant features that are due to experimental chance and due to systematic differences.

We used both a SAM-based approach and an AR approach to determine the FDR of significantly changed proteins. The SAM procedure was applied to peptides, and a modified t statistics, a false discovery rate (q value), and AR values were calculated for each peptide (22). The AR of the protein was then calculated as the average AR of those peptides with q < 0.001. In the AR approach, we approximate a null distribution by taking the pool of AR of proteins of the replicates of the same culture conditions (aerobic versus aerobic and suboxic versus suboxic) and then counting the number of proteins having >2-fold change; this approximates the number of false positives we expect to have a significant change on their own (proteins above the line in Supplemental Fig. 4a). We then

![Fig. 6. Comparison of housekeeping and differentially expressed proteins.](image)

**a**, abundance ratio of proteins with peptides found in both aerobic and suboxic states (squares). **b**, abundance ratio of proteins with peptides identified in only the aerobic or suboxic samples (triangles). For these proteins, the lowest abundance (signal level) value of the entire LC-FTICR analysis was used as the reference value. **a** and **b**, abundance ratio of housekeeping proteins (diamonds).
determined the number of protein ARs of suboxic versus aerobic conditions having the same change or greater; these included both false positives and true positives (proteins above the line in Supplemental Fig. 4b). The FDR associated with this change is computed as the ratio of the false positives to the total positives. For those proteins that contained both one-state and two-state peptides, the average AR for one-state peptides was calculated separately from those of two-state peptides. The ARs for proteins with both one-state and two-state peptides were calculated in the same manner as in the SAM-based approach.

We found the SAM-based approach to be more sensitive, indicating 56 proteins changed by more than 2- and 4-fold based on two-state and one-state peptides, respectively, where the FDR at the peptide level is <0.1% (see Supplemental Table 3). Based on the AR approach, 46 proteins, all of which had been revealed by the SAM-based approach, were found to have changed by more than 2- and 4-fold based on two-state and one-state peptides, respectively (see Supplemental Table 2), where the FDR at the protein level is <3.0%. It should be noted that all of those 46 proteins judged by the AR approach to have changed significantly were also identified using the SAM-based approach, and further, of those that were common, the $R^2$ for a linear regression for the abundance of the two approaches was greater than 0.98. Fig. 6 shows that housekeeping proteins remain constant in comparison with those proteins that changed significantly. No major differences were observed between the results of the two normalization schemes (data not shown).

**Biological Validation**—It is well known in microarray data analysis that different data analysis procedures may give different results due to the complexity of the experimental process and the noise level of the data. Very often biological information is needed to validate the results. The same problems exist in proteomic data analysis; therefore, we examined the biological relevance of the proteins that exhibited significant changes in AR (see Supplemental Table 2). We observed broad agreement between the quantitative proteomic results and the functionalities of the proteins in the system and previous published results. For example, fumarate reductase (SO0970), the sole physiological fumarate reductase in MR-1 (23), previously shown to be more highly expressed under anaerobic and low oxygen growth conditions (24, 25), increased by as much as 7-fold. Several decaheme cytochrome c proteins (SO1777, SO1778, and SO1779), which are essential for anaerobic electron transport in the reduction of iron citrate and MnO$_2$ (26, 27), were 4–7-fold more abundant in cells from the suboxic cultures relative to the 20% DOT cultures. Formate dehydrogenase (SO4509 and SO4513), which serves as a major electron donor to a variety of inducible respiratory pathways that use terminal acceptors other than molecular oxygen (28), increased by as much as 10-fold. The tricarboxylic acid cycle is a principle route of assimilation and catabolism during aerobic growth, and many of its enzymes have repressed activities under anaerobic condition (24, 28, 29). In agreement with these observations, we found that a majority of enzymes involved in the tricarboxylic acid cycle (or in facilitating it), such as malate dehydrogenase (SO0770), isocitrate lyase (SO1484), isocitrate dehydrogenase (SO2629), citrate synthase (SO1926), and acetyl-CoA synthase (SO2743), decreased by 2–7-fold. These results suggest that the large scale relative quantitation of complex proteomic samples can be obtained using an LC-MS platform without the use of stable isotope labeling.

Our analyses also revealed six unidentified proteins that exhibited significant changes in relative abundance under the two cultivation conditions (see Supplemental Table 2). Fig. 7 shows the peptide abundances of these six proteins. In this figure, the peptide abundances in the six suboxic sample replicates are distinctly different from those in the six aerobic sample replicates. Protein SO0409, which increased 7-fold under suboxic conditions, was originally annotated by...
The qualitative proteomic results presented here agree well with the gene profile studies reported previously in the following aspects (24). Both our study and the earlier studies showed that under anaerobic conditions various proteins involved in aerobic respiration are repressed, and various proteins involved in anaerobic respiration are induced. For example, proteins involved in tricarboxylic acid cycles were found to be decreased in anaerobic conditions by both the gene profile studies and the proteomic studies presented here. Proteins involved in anaerobic respiration are found to be induced by both platforms. More specifically, fumarate reductase was observed to be increased by the proteomic study presented here and by gene profile studies and proteomic studies by two-dimensional gel electrophoresis reported previously (24).

However, many other gene expression results were not correlated with the proteomic data. For example, many genes with regulatory functions, such as sensor histidine kinase, two-component response regulators, etc., were quantified by microarray experiments but not quantified by the proteomic experiments presented here. This may arise from many regulatory proteins being present at very low abundance and therefore not readily detected with confidence by the proteomic platform without enrichment. The observation that the gene expression data generally do not correlate well with proteomic results has been made by others (30). The reasons for these discrepancies are complex and have been discussed previously (30). Additionally we have observed much greater biological variability in experiments grown in shaker flasks as used previously (24) versus those grown in a bioreactor as used in our study.

Estimation of False Positive Identifications—For the present study we estimated a “false peptide discovery rate” for AMT tag proteomic measurements. This method uses the distribution of peak matched mass and elution times for this estimation. These distributions result in a peak, which is believed to represent mostly “true” positives, and a base line, which represents mostly “false” positives. The mass and elution time tolerances are simultaneously refined such that the base line component is reduced. The measurement of the base line and all the peptides observed below it represent the expected false positives for that analysis within the mass and elution tolerances used. These are then divided by the total number of identifications to give a measure of “relative risk” within each peak matching effort. We typically find this calculation for an experiment of this complexity to result in a relative risk of 5–10% for individual peptides; this can be lower for proteins when multiple peptides are used as a requirement (see Supplemental Fig. 5). Thus, most commonly we require two or more peptides to discern high and low confidence protein assessments. Although we will often maintain the “single peptide” protein observations, generally it is felt that manual analysis of the LC-FTICR MS and LC-MS/MS spectra that gave rise to an identification is required prior to drawing broad conclusions with regard to the single peptide proteins.

Finally it is important to recognize that this “false discovery rate” or relative risk differs in important parameters from the more typical calculations for LC-MS/MS results that are actually prerequisites for identifications using AMT tags. Instead this “rate” corresponds to the probability that a detected UMC may be misassigned in the context of the specific tolerances used to match the LC-MS/MS database (e.g. MS/MS filters, mass measurement accuracy, and NET cutoff) applied to a specific dataset.

Conclusions—The results presented here demonstrate that LC-MS label-free approaches using AMT tags can be successfully applied to differential quantitative proteomic analysis of complex biological samples. Our results agree well with findings from previous conventional biochemical and microbiological studies as well as with studies of the biological functions of the proteins in the system. Additionally the quantitative proteomic results presented here agree well with the gene profile studies reported previously (24). However, many gene expression results were not correlated with the proteomic data. The reasons for possible discrepancies in correlation are complex and have been discussed previously (30). Due to the observed greater biological variability in experiments grown in shaker flasks as used previously versus those grown in a bioreactor as used in our study we feel this study represents a better and more controlled “base line” for comparisons and analysis of “significant” changes.

The AMT tag approach is particularly suitable for high throughput and long term studies. The PMT tag database for a particular organism can also be used by different researchers for any quantitative proteomic analyses on that organism. Reproducibility is relatively easily accessed, quantification can then be conducted in a high throughput fashion, and data analysis is simplified. For the data presented here, we applied two global normalization schemes to normalize intensities across multiple LC-MS runs. One scheme uses the median ratios of peak intensities of one run over a reference run; this scheme has been successfully used by Wang et al. (11) for label-free LC-MS data analysis. The other scheme uses housekeeping proteins and has been used frequently in microarray data analysis (31, 32). We determined the dependence of peptide ratio on intensity, mass, and elution time domain by plotting the log ratio over the intensity, mass, and elution time, respectively (32), and we found that the log ratios are evenly distributed around zero across the range of these parameters indicating that the global normalization was quite effective.

Several factors are important for successful label-free LC-MS analysis (31). In our experience, it is important to have
highly reproducible processes for sample preparation, LC separation, electrospray, and MS performance. These can be achieved by following appropriate standard operating, system maintenance procedures and quality control. In our practice efforts, a mixture of tryptic peptides from 14 standard proteins was analyzed weekly, and the reproducibility was determined by scatter plot of peptide abundance. Normally an $R^2$ value >0.85 is considered to be sufficiently reproducible. Another important factor is the use of effective software tools for peak picking, deisotoping, deconvolution, and alignment/LC retention time normalization (31). It is emphasized that we believe that such approaches are best used to complement quantitation based upon the use of stable isotope labeling methods, which are capable of providing the best precision but are ultimately limited for reasons discussed in the Introduction. We emphasize that label-free methods can be less effective for highly abundant peptides when ionization suppression effects apply; however, the use of lower abundance peptides (as well as isotopic labeling approaches) can circumvent such difficulties. These observations suggest that “dual” approaches based upon use of both isotopic labeling and data from normalized peak intensity information should be particularly effective.

Acknowledgments—We acknowledge the contributions of the Shewanella Federation to the annotation of the MR-1 genome. We thank Penny Colton and Andrea Detter for critical manuscript reading.

* This work was supported by the Genomes to Life Program, United States Department of Energy (DOE) Office of Biological and Environmental Research. Pacific Northwest National Laboratory (PNNL) is operated for the DOE by Battelle under Contract DE-AC05-76RLO 1830. This research was performed in the Environmental Molecular Sciences Laboratory, a DOE national scientific user facility at PNNL. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “solely to indicate this fact.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “solely to indicate this fact.

The online version of this article (available at http://www.mcponline.org) contains supplemental material.

To whom correspondence should be addressed: Biological Sciences Division, Pacific Northwest National Laboratory, P. O. Box 999, MSIN: K8-98, Richland, WA 99352.

REFERENCES

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422, 198–207
2. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999
3. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386
4. Washburn, M. P., Irlaszek, R., Deciu, C., Schizas, D. M., and Yates, J. R., III (2002) Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal. Chem.* 74, 1650–1657
5. Chelius, D., and Bondarenko, P. V. (2002) Quantitative profiling of protein-sin complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* 1, 317–323
6. Quadrini, M., Ducret, A., and Stocklin, R. (2004) Quantify this! Report on a round table discussion on quantitative mass spectrometry in proteomics.

**Molecular & Cellular Proteomics** 5.4

7. Cech, N. B., and Enke, C. G. (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* 20, 362–387
8. Tang, K., Page, J. S., and Smith, R. D. (2004) Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 15, 1416–1423
9. Voyksner, R. D., and Lee, H. (1999) Investigating the use of an octopole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. *Rapid Commun. Mass Spectrom.* 13, 1427–1437
10. Bondarenko, P. V., Chelius, D., and Shaler, T. A. (2002) Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* 74, 4741–4749
11. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* 75, 4818–4826
12. Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T. G., Foss, E., Mao, Y., and Emili, A. (2004) Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* 3, 984–997
13. Smith, R. D., Shen, Y., and Tang, K. (2004) Ultrarapid and quantitative analyses from combined separations-mass spectrometry for the characterization of proteomes. *Acc. Chem. Res.* 37, 269–278
14. Myers, C. R., and Nealon, K. H. (1999) Respiration-linked proton translocation coupled to anaerobic reduction of manganese(V) and iron(III) in Shewanella putrefaciens MR-1. *J. Bacteriol.* 172, 8232–8238
15. Lipton, M. S., Pasa-Tolica, L., Anderson, G. A., Anderson, D. J., Auberry, D. L., Battista, J. R., Daly, M. J., Fredrickson, J., Hixson, K. K., Kostantarhes, H., Masselon, C., Marklille, L. M., Moore, R. J., Romine, M. F., Shen, Y., Strittmatter, E., Tolic, N., Udseth, H. R., Venkateswaran, A., Wong, K. K., Zhao, R., and Smith, R. D. (2002) Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11049–11054
16. Petritis, K., Kangas, L. J., Ferguson, P. L., Anderson, G. A., Pasa-Tolica, L., Lipton, M. S., Auberry, K. J., Strittmatter, E. F., Shen, Y., Zhao, R., and Smith, R. D. (2003) Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. *Anal. Chem.* 75, 1039–1048
17. Horn, D. M., Zubarev, R. A., and McLafferty, F. W. (2000) Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* 11, 320–332
18. Smith, R. D., Anderson, G. A., Lipton, M. S., Pasa-Tolica, L., Shen, Y., Conrads, T. P., Veenstra, T. D., and Udseth, H. R. (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* 2, 513–523
19. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* 2, 43–50
20. Washburn, M. P., Irlaszek, R. K., and Yates, J. R., III (2003) Reproducibility of quantitative proteomic analyses of complex biological mixtures by multidimensional protein identification technology. *Anal. Chem.* 75, 5054–5061
21. Steen, H., and Pandey, A. (2002) Proteomics goes quantitative: measuring protein abundance. *Trends Biotechnol.* 20, 361–364
22. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5116–5121
23. Maier, T. M., Myers, J. M., and Myers, C. R. (2003) Identification of the gene encoding the sole physiological fumarate reductase in Shewanella oneidensis MR-1. *J. Basic Microbiol.* 43, 312–327
24. Beliaev, A. S., Thompson, C. K., Lin, H. H., Brandt, C. C., Li, G., Murray, A. A., Heidelberg, J. F., Giometti, C. S., Yates, J., III, Nealon, K. H., Tiedje, J. M., and Zhoui, J. (2002) Gene and protein expression profiles of Shewanella oneidensis during anaerobic growth with different electron acceptors. *Omics* 6, 39–60
25. Giometti, C. S., Khare, T., Tollaksen, S. L., Tsapin, A., Zhu, W., Yates, J. R., III, and Nealon, K. H. (2003) Analysis of the Shewanella oneidensis
proteome by two-dimensional gel electrophoresis under nondenaturing conditions. Proteomics 3, 777–785
26. Beliaev, A. S., Saffarini, D. A., McLaughlin, J. L., and Hunnicutt, D. (2001) MtrC, an outer membrane decahaem c cytochrome required for metal reduction in Shewanella putrefaciens MR-1. Mol. Microbiol. 39, 722–730
27. Myers, C. R., and Myers, J. M. (2002) MtrB is required for proper incorporation of the cytochromes OmcA and OmcB into the outer membrane of Shewanella putrefaciens MR-1. Appl. Environ. Microbiol. 68, 5585–5594
28. Jormakka, M., Byrne, B., and Iwata, S. (2003) Formate dehydrogenase—a versatile enzyme in changing environments. Curr. Opin. Struct. Biol. 13, 418–423
29. Scott, J. H., and Nealson, K. H. (1994) A biochemical study of the intermediary carbon metabolism of Shewanella putrefaciens. J. Bacteriol. 176, 3408–3411
30. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19, 1720–1730
31. Listgarten, J., and Emili, A. (2005) Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. Mol. Cell. Proteomics 4, 419–434
32. Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30, e15–e25