High Performance Mass Spectrometry Facility

The High Performance Mass Spectrometry Facility (HPMSF) provides state-of-the-art mass spectrometry and separations instrumentation that has been refined for leading-edge analysis of biological problems with a primary emphasis on proteomics. Challenging research in proteomics, cell signaling, cellular molecular machines, and high-molecular-weight systems receive the highest priority for access to the HPMSF. Current research activities within the facility include proteomic analyses of whole cell lysates, analyses of organic macromolecules and protein complexes, quantification using isotopically labeled growth media, targeted proteomics analyses of subcellular fractions, and nucleic acid analysis of RNA and DNA oligomers. More than a dozen microbial systems are currently being studied within the HPMSF by researchers from throughout the country. In addition, several ongoing projects in higher order systems, including mammalian systems, are investigating a broad range of biological questions from cancer screening to infectious diseases to fundamental questions of post-translational modifications and protein-protein interactions.

The facility provides a complete suite of mass spectrometers, ranging from a conventional triple-quadrupole instrument (Figure 1) to state-of-the-art Fourier transform ion cyclotron resonance (FTICR) spectrometers (Figures 2 and 3) that enable very high-sensitivity and high-resolution mass spectrometry at 1 ppm mass measurement accuracy. Proteomics analysis on these FTICR spectrometers is supported by a group of five ion trap spectrometers (Figure 4), and a quadrupole time-of-flight spectrometer (Figure 5). These spectrometers are coupled with very high-resolution separations (500-1000 peak capacity) that greatly benefit these areas of research (Figure 6). Focused research projects on biomolecular complexes and macromolecules are supported on the 7-tesla (T) FTICR spectrometer (Figure 7), which is a flexible instrument that can be configured in many different ways.

The HPMSF is committed to maintaining its state-of-the-art mass spectrometry and separations capabilities. To this end, the facility's staff work closely with researchers at Pacific Northwest National Laboratory (PNNL) to develop new capabilities such as the ion funnel, Dynamic Range Enhancement Applied to Mass Spectrometry (DREAMS) and data analysis tools, which are incorporated into the capabilities of the facility as they become available. This work is supported by a separate 3.5-T FTICR spectrometer (Figure 8) so sample analysis is not interrupted by this developmental work. Facility staff are highly skilled in all of the areas required for proteomics research, from sample preparation to analysis and data interpretation, and they are available to help develop methodologies to tackle these challenging problems. As needed,

Instrumentation & Capabilities

Mass Spectrometers

- Quadrupole time-of-flight spectrometer
- Five ion trap spectrometers

FTICR Spectrometers

- 11.5-T FTICR spectrometer
- 7-T FTICR spectrometer
- 9.4-T FTICR spectrometer

Additional Capabilities

- 5 custom high-performance liquid chromatography (HPLC) systems
- 1 Agilent capillary HPLC system

PNNL staff outside the facility can be accessed as matrixed members of the facility. Since its inception, more than 100 separate user projects have been undertaken in the facility with some of them extending to over a year in duration.

The HPMSF has developed state-of-the-art software for the acquisition and analysis of FTICR mass spectra. This software package is called ICR-2LS. It is a Windows-based application that enables many of the unique instrument control functions developed in our laboratory. This same software allows automated spectral interpretation of raw FTICR data. The spectral interpretation features are integrated into the application and, thus, are not easily transferred. As part of our efforts to disseminate our developments in mass spectrometry, researchers outside EMSL can obtain this code from our website at http://www.emsl.pnl.gov/docs/msd/mass_spec/home/software.html.

Mass Spectrometry Capabilities

The ThermoQuest TSQ 7000 (Figure 1) is a triple-quadrupole mass spectrometer designed for use with an electrospray ionization (ESI) source. The TSQ 7000 has a mass range of 50 to 4000 m/z with peak widths as low as 0.7 Da and can be operated in normal or tandem mass spectrometry mode using collision-induced dissociation (CID), which takes place in the second quadrupole at a nominal but variable gas pressure of 23 mTorr. The system can be operated in either positive or negative ion mode with the addition of SF_6 as a sheath gas.

Figure 1. Triple-quadrupole mass spectrometer.

The ultra-high-performance FTICR mass spectrometer pictured in Figure 2 uses a wide-bore (205 mm), passively shielded 11.5-T superconducting magnet. The spectrometer is equipped with an ESI source and an ion funnel. The 11.5-T FTICR instrument has a resolution in excess of 150,000 at m/z 1,000 and a mass accuracy of 1 ppm for peptide samples with molecular weights ranging from 500 to 2000 Da. Ions are collected external to the magnet in a series of quadrupoles that allow the researcher to eliminate uninteresting ions before analysis in the FTICR instrument. The DREAMS process is a unique capability of this facility. The 11.5-T FTICR instrument can be fitted with an HPLC system and is equipped with an infrared (IR) laser for multi-photon dissociation of samples for tandem mass spectrometry in the ICR cell.

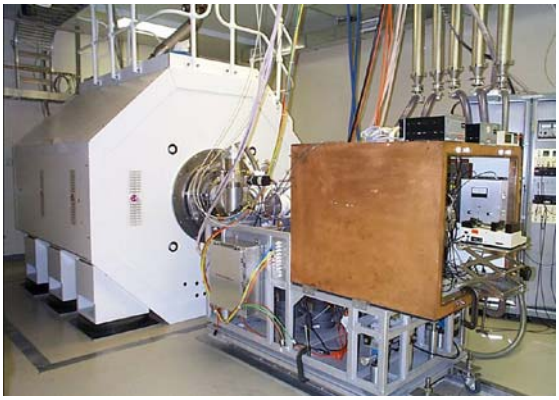

Figure 2. 11.5-T FTICR mass spectrometer.

The newest FTICR spectrometer (Figure 3) in the HPMSF is a 9.4-T, 150-mm bore, actively shielded Bruker Daltonics APEX III instrument. The original ion source has been replaced with a custom PNNL-built source that incorporates a dual-channel ion funnel for simultaneous introduction of calibrant ions, DREAMS technology, and automatic gain control. These enhancements permit maximum use of the spectrometer's capabilities by maintaining the optimum number of ions in the ICR cell throughout a liquid chromatographic (LC) separation. The outstanding resolution of over 60,000 is maintained throughout the separation with sensitivity comparable to the other FTICR instruments in the facility. This system is seamlessly integrated with PNNL's automated high-pressure (5000 psi) LC system for unattended operation 24 hours a day, seven days a week.

Figure 3. 9.4T FTICR mass spectrometer.

Five ion trap mass spectrometers (Figure 4) from ThermoQuest are available in the facility: two LCQ Classics (liquid chromatograph interfaced with an ion trap mass analyzer), two LCQ Duos providing improved sensitivity, and one LCQ DECAXP providing the latest in sensitivity improvements from Finnigan. The ion trap is a three-dimensional quadrupole ion-trap-based instrument designed for use with ESI sources. These instruments are well-suited to tandem mass spectrometry experiments due to their very high collection efficiency for product ions. The mass range of this instrument is 150 to 2000 m/z, but can be extended to 4000 m/z for some applications. The LCQ has a maximum resolution of 10,000 in the zoom-scan mode and 4000 in full-scan mode. In addition, the system is easily operated in either positive or negative ion mode with the addition of SF_6 as a sheath gas.

Figure 4. Ion trap mass spectrometers.

The API QSTARTM Pulsar spectrometer (Figure 5) is a state-of-the-art instrument that combines the robust operation of a quadrupole instrument with the speed and resolution of a time-of-flight spectrometer. The new LINAC Pulsar high-pressure collision cell, which pulses ions into the time-of-flight analyzer, offers superior mass accuracy and sensitivity. Innovative software applications include Information-Dependent Acquisition that enables automated MS to MS/MS acquisition for maximum extraction of information from a single LC/MS run and the powerful data acquisition and processing capabilities of Analyst and BioAnalyst.

Figure 5. Quadrupole time-of-flight mass spectrometer.

A signature capability of the facility is the efficient coupling of capillary separations to our mass spectrometers. Instruments for both LC and capillary electrophoretic separations are available (Figure 6). Two ABI 270A electrophoresis systems are available along with one Agilent capillary LC and one Shimadzu capillary LC system. Unique to this facility are the PNNL-developed LC systems that deliver constant-pressure gradient separations at up to 5000 psi. PNNL has recently finished the development of an integrated automated HPLC system. It features a PAL model autosampler from LEAP Technologies with a cooled sample holder, VALCO high-pressure valves, and ISCO syringe pumps. Computer software has been developed that allows the system to be configured with any of our spectrometers through Microsoft distributed communication (DCOM) protocols.

Figure 6. Capillary separations capability.

The FTICR mass spectrometer shown in Figure 7 is based on a 7-T, 160-mm bore, superconducting magnet and is equipped with a custom ESI source (see Figure 1). The 7-T FTICR has high mass-resolving power (e.g., greater than 2,000,000 has been obtained for insulin), while unit resolution is routinely achievable during online capillary isoelectric focusing (CIEF) experiments for proteins with relative molecular mass (Mr) less than 30,000u. Mass accuracy less than 5 ppm is typical for peptide/protein samples with molecular masses ranging from 500 to 30,000u. A detection limit of approximately 10 attomoles has been obtained with online LC and CIEF separations.

Figure 7. 7-T FTICR mass spectrometer.

The high-performance FTICR mass spectrometer pictured in Figure 8 is coupled to a 3.5-T, 330-mm bore magnet and is equipped with an ESI source. The 3.5-T FTICR has a mass resolution of 50,000 to 150,000 and a mass accuracy of 3 to 7 ppm for protein samples with molecular weights ranging from 5000 to 20,000 Da. The ion optics on this system are the most highly developed in the facility. The spectrometer incorporates DREAMS technology and a dual channel ion funnel for simultaneous introduction of a calibrant. The detection limit of the instrument has recently been improved to $3x10^{-20}$ moles or about 18,000 molecules (based on sample consumption).

Figure 8. 3.5-T FTICR mass spectometer

Upgrades

This year saw several important enhancements of the capabilities of the HPMSF:

- A new Bruker Daltonics APEX III data station was acquired to replace the ODYSSEY data station on the 11.5-T FTICR instrument.

- Five of our custom HPLC systems were fully automated, enabling 24-hour operation of all ion trap mass spectrometers and the 9.4-T and 11.5-T FTICR spectrometers. This upgrade has greatly improved throughput in the facility, enabling more complex user projects to be undertaken.

- PNNL's DREAMS and ion funnel technologies were added to the Bruker Daltonics 9.4-T FTICR spectrometer. The combined system was successfully interfaced with our automated HPLC systems to complete the first automated LC-FTICR spectrometer capable of unattended, continuous operation for 24 hours.

- Twenty-eight additional CPU nodes were acquired to improve our data analysis capabilities. When added to the 20 existing CPU nodes, this upgrade will help keep our analysis capabilities abreast of our increased throughput for the mass spectrometers.

Functional Genomics and Proteomics of Mitochondria

W Xiao,[(a)] CM Scharfe,[(a)] and RW Davis[(a)]
(a) Stanford University, Palo Alto, California

About half of the expected mitochondrial proteins in humans are known at this time, and already a quarter of these known proteins are associated with human Mendelian disorders (OMIM). Mitochondrial core functions such as oxidative phosphorylation, amino acid metabolism, fatty acid oxidation, and iron-sulfur cluster assembly have been highly conserved during evolution, suggesting that a systematic identification of mitochondrial proteins in model organisms will accelerate the search for new human mitochondrial disease genes. In yeast, 477 proteins (469 encoded by the nuclear genome) show conclusive evidence of mitochondrial localization (MitoP2 database, this study). About 30% of these proteins have evidence of orthologs in humans (MitoP2 database). Identification of the yeast mitochondrial proteome is far from complete. Of the predicted complement of proteins that make up the organelle, 40% are still considered unknown (Westermann and Neupert 2003), although many genome-wide and functional systematic studies have been applied. In this project, we generated a component list of the mitochondrial proteome by first identifying mitochondrial proteins using mass spectrometry, and then by integrating the structural data with 19 published datasets relevant to the study of mitochondria.

By combining the purification of whole mitochondrial organelles from yeast cell cultures with the direct measurement of the proteins present in these fractions using mass spectrometry, we developed a new approach for identifying mitochondrial proteins. Yeast cells from four different growth conditions including fermentable (glucose) and non-fermentable (lactate) substrates for both natural and synthetic culture media were purified by both density-gradient and free-flow electrophoresis. Preparations were separated into mitochondrial membrane and matrix fractions and then analyzed separately for protein content. In total, 20 fractions were analyzed by high-resolution liquid chromatography tandem mass spectrometry (LC MS/MS) (Washburn et al. 2003; Ferguson and Smith 2003). In addition, to increase sensitivity, eight of the fractions were further analyzed by high sensitivity, high-dynamic-range LC/Fourier transform ion cyclotron resonance MS (Lipton et al. 2002; Smith et al. 2002). Therefore, 28 experimental datasets were generated that yielded mass and time tags for 2086 proteins through a yeast proteome database search (Washburn et al. 2003). For evaluation of the identified proteins, we annotated a mitochondrial reference set of 477 proteins based on experimental evidence from single gene studies (MitoP2 database, this study).

This list was used to evaluate the performance of the proteomics dataset. Among 2086 different proteins 17% were known mitochondrial, covering 75% of the reference set (361 of 477).

To define a more confident candidate list, we increased the stringency of our identification criteria and considered only proteins identified with at least two peptide tags. This more conservative measure yielded 546 proteins, of which 256 (46.9%) were known to be mitochondrial. Sorting of the 546 candidates by the number of experiments in which they were found demonstrated that the probability of identifying a mitochondrial protein correlated with the number of tags and the number of experiments [Figure 1(A)].

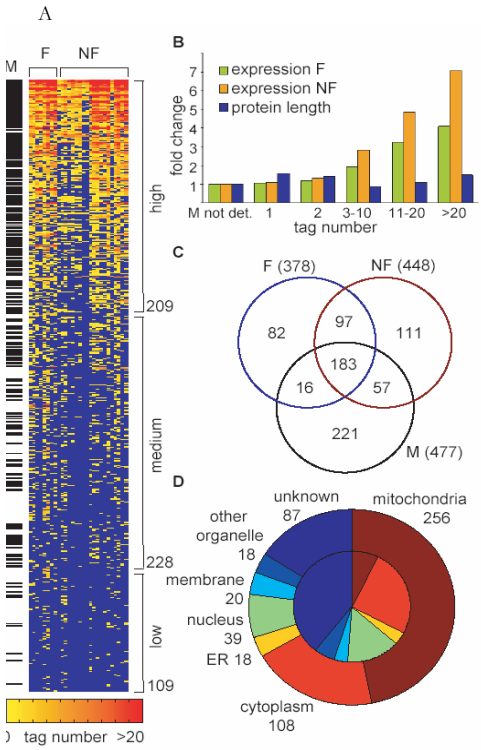

Figure 1. Distribution of identified proteins.

A comparison of the 28 datasets revealed that mass spectrometric analysis of mitochondria prepared under the same conditions added, on average, 16 known mitochondrial proteins to the dataset, while on average, 40 additional proteins were identified when growth conditions or preparation methods were changed. As expected, the proteins identified from membrane fractions were enriched for known membrane proteins and matrix fractions for known soluble proteins (belonging to matrix or inter-membrane space). However, overall a similar percentage of proteins were identified from membrane and matrix fractions, indicating that, compared to 477 mitochondrial reference proteins, no significant bias in the representation of proteins associated with membrane and soluble fractions was observed. To assess the correlation between protein detection and expression level systematically, we performed genome-wide mRNA expression analysis by means of high-density oligonucleotide arrays under the same fermentable and non-fermentable growth conditions. This analysis showed that absolute mRNA expression levels increased with higher peptide tag numbers [Figure 1(B)]. While genes identified by proteomics had median expression levels of 1.2- to 7.1-fold higher than their unidentified mitochondrial counterparts, the number of tag identifications did not correlate with protein length, supporting a bias of the proteomics analysis primarily towards higher abundant proteins.

A comparison between fermentable and non-fermentable growth conditions revealed that more proteins were detected under respiration (448) than fermentation (378) [Figure 1(C)], a finding that is consistent with the activation of oxidative phosphorylation during aerobic growth. To our surprise, however, of the 477 known mitochondrial proteins, 183 were identified under both growth conditions, suggesting that at least 38% of the mitochondrial machinery is present at moderate-to-high abundance even under fermentable growth conditions. This finding indicates the presence of a core mitochondrial protein set that could exist under multiple growth conditions. Mitochondria of yeast cells grown under fermentable conditions are more fragile and, therefore, more difficult to purify. This is reflected in a greater number of proteins with known non-mitochondrial localization in the fermentation dataset. Of the 546 candidates, 203 proteins are known to localize outside

mitochondria, mainly to the nucleus, plasma membrane, cytoplasm, and endoplasmic reticulum [Figure 1(D)] (Mewes et al. 2002). In addition to contaminants co-purified with the fractions, identification of these proteins lends further support to the physical interaction of mitochondria with other cellular compartments and the existence of proteins with multiple localizations (Achleitner et al. 1999).

Analysis of functional categories of the detected proteome permits an evaluation of the proteomics approach in comparison to functional approaches of gene expression analysis and quantitative deletion phenotype screening (Steinmetz et al. 2002). In the proteomics dataset, proteins annotated as localized outside mitochondria were not significantly enriched for any of the known functional classes (Mewes et al. 2002). In contrast, known mitochondrial proteins were primarily enriched for known mitochondrial functions such as energy production, transport and sensing, protein fate, and amino acid metabolism. In comparison, quantitative deletion phenotype screening enriched mainly for proteins involved in genome maintenance, transcription, and translation. Measurements of mRNA expression enriched predominantly for proteins involved in energy production, the majority of which however seem to be localized outside mitochondria.

Overall, the proteomics approach achieved a 5.9-fold enrichment for mitochondrial proteins over genome levels, which is comparable to the 6.1-fold enrichment achieved by the deletion approach. Both methods identified about equal numbers of previously known mitochondrial proteins. However, the overlap between proteomics and deletion phenotypes was less than 30%, supporting the hypothesis that the most abundant proteins are not necessarily most important for fitness (survival and reproduction), and vice versa. The results also demonstrate that proteomics and systematic deletion screening are complementary approaches because they identify different structural and functional subsets of mitochondrial proteins.

Three lines of evidence further support the success of this integrative analysis for defining the yeast mitochondrial proteome. First, the enrichment level for known mitochondrial proteins correlated with the level of the MitoP2 score and the number of experiments in which candidates were identified by proteomics: for the high, medium and low classes [Figure 1(A)], the median predictive scores were 97, 91, and 76, respectively. Second, MitoP2 prediction was confirmed by import experiments. Ten out of 14 candidates tested with MitoP2 scores greater than 90 were imported into isolated yeast mitochondria, and seven of these were supported with signal sequence cleavage. This ratio (10/14) predicts that 71% of the new candidates could be imported into mitochondria indicating that 519 of the 575 proteins (MitoP2 greater than 90) may thus be true positives (380 + 139). Third, an investigation of known subunits in mitochondria revealed that most of the components of known complexes were assigned a high MitoP2 score. Comparison with the proteomics dataset showed that, while some of the assembly factors of respiratory chain complexes IV and V and subunits of TIM22 were not detected by proteomics, the integrative analysis defined them correctly as mitochondrial proteins. This observation provides further support to the advantage gained by an integrative approach that combines various datasets over any single dataset.

The advantage of the integrative analysis that combines structural and functional approaches is the high coverage of various mitochondrial components. With this approach, we were able to detect proteins that had dual localization with high confidence. By including functional datasets in the calculations and using a localization list as a reference, our candidate list

is strongly enriched for mitochondrial localized proteins, but not limited to those proteins. Consequently, because the MitoP2 calculation is based on both structural and functional datasets, the score gives not only a prediction of mitochondrial localized proteins but also contains proteins that may localize outside mitochondria but affect mitochondrial function and biogenesis from there.

Finally, our study has implications for human diseases. To date, 129 mitochondrial proteins have been implicated in human Mendelian disorders (MitoP2 database) (DiMauro and Schon 1998; Wallace 1999). The integration in yeast identified 213 new human orthologs to the 195 new yeast mitochondrial candidates (MitoP2 greater than 90) (MitoP2 database). This set of 213 proteins provides new candidates for putative human mitochondrial disorders, where intervals have been mapped but no responsible gene has been identified to date (Steinmetz et al. 2002).

References

Achleitner G, B Gaigg, A Krasser, E Kainersdorfer, and SD Kohlwein. 1999. "Association Between the Endoplasmic Reticulum and Mitochondria of Yeast Facilitates Interorganelle Transport of Phospholipids Through Membrane Contact." *European Journal of Biochemistry* 264:545-553.

Ferguson PL and RD Smith. 2003. "Proteome Analysis by Mass Spectrometry." *Annual Review of Biophysics and Biomolecular Structure* 32:399-424.

DiMauro S and EA Schon. 1998. "Nuclear Power and Mitochondrial Disease." *Nature Genetics* 19:214-215.

Lipton MS, L Pasa-Tolic, GA Anderson, DJ Anderson, and DL Auberry. 2002. "Global Analysis of the *Deinococcus radiodurans* Proteome by Using Accurate Mass Tags." *Proceedings of the National Academy of Science* 99:11049-11054.

Mewes HW, D Frishman, U Guldener, G Mannhaupt, and K Mayer. 2002. "MIPS: A Database for Genomes and Protein Sequences." *Nucleic Acids Research* 30:31-34.

Smith RD, GA Anderson, MS Lipton, L Pasa-Tolic, and Y Shen. 2002. "An Accurate Mass Tag Strategy for Quantitative and High-Throughput Proteome Measurements." *Proteomics* 2:513-523.

Steinmetz LM, C Scharfe, AM Deutschbauer, D Mokranjac, and ZS Herman. 2002. "Systematic Screen for Human Disease Genes in Yeast." *Nature Genetics* 31:400-404.

Wallace DC. 1999. "Mitochondrial Diseases in Man and Mouse." *Science* 283:1482-1488.

Washburn MP, A Koller, G Oshiro, RR Ulaszek, and D Plouffe. 2003. "Protein Pathway and Complex Clustering of Correlated mRNA and Protein Expression Analyses in *Saccharomyces cerevisiae*." *Proceedings of the National of Academy Science* 100:3107-3112.

Westermann B and W Neupert. 2003. "'Omics' of the Mitochondrion." *Nature Biotechnology* 21:239-240.

Secretory and Membrane Proteins of *Pseudomonas aeruginosa*

W Xiao,[(a)] RD Davis,[(a)] and M Mindrinos[(a)]
(a) Stanford University, Palo Alto, California

Pseudomonas aeruginosa is a gram-negative bacterium of increasing importance because it is one of the top three opportunistic pathogens in humans, causing premature death in patients with cystic fibrosis, HIV infection, organ transplants, or cancer (Quinn et al. 1998). It is the most common cause of various nosocomial infections including sepsis in burned or extensively injured patients and is resistant to antibiotics. Consequently, there is significant interest in the proteins responsible for this virulence, particularly those of the cellular membrane. The phenotypic resistance of gram-negative bacteria is a direct consequence of the complex structure of their cell envelope, which acts as a barrier and prevents drug molecules from reaching their target sites or increases their active efflux, primarily mediated by various classes of integral membrane proteins (Nikaido 1998). It has been indicated that antibiotic resistance is due to integral outer membrane protein channels (porins) and membrane protein complexes known as multi-drug efflux pumps that transport antibiotic and biocide molecules out of the cell (Nikaido 1998).

Large-scale analysis of hydrophobic integral membrane proteins from complex protein mixtures is an important and challenging aspect of mass spectrometry (MS)-based membrane proteomics. Although large-scale proteomic studies based on in-gel digestion of proteins separated using two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) have recently shown significant improvements in protein coverage, very hydrophobic integral membrane proteins are generally not amenable to this approach, primarily because of issues related to the insolubility of these proteins and precipitation occurring at their isoelectric points (Nouwens et al. 2000; Santoni et al. 2002).

Membrane Protein Identification

The membrane subproteome of *P. aeruginosa* was probed using two techniques we developed specifically for MS identification of highly hydrophobic integral membrane proteins (Blonder et al. 2002; Goshe et al. 2003). The protocol for this investigation is outlined in Figure 1. Both techniques were used concomitantly in this investigation to achieve a more comprehensive analysis of membrane proteins than could be obtained by using only one approach. When analyzed using the SEQUEST protein database search algorithm, all the collected tandem mass spectrometry (MS/MS) spectra of both experiments produced a total of 9951 fully tryptic peptides of which 2727 peptides were unique. From these data, a total of 786 protein identifications were observed: 623 proteins from the unlabeled sample and 163 proteins from the labeled sample. The

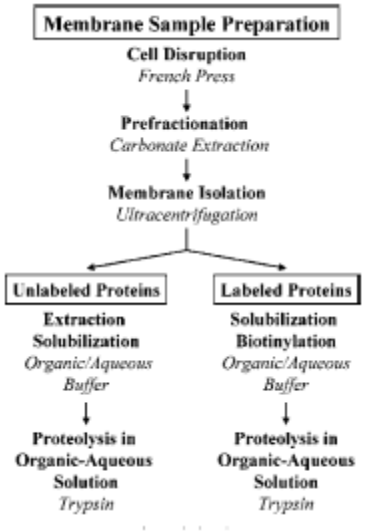

Figure 1. Experimental protocols used for the analysis of the membrane proteome.

set of 786 identified proteins contains 544 identified from the unlabeled sample and 84 from the labeled sample, corresponding to 707 unique proteins. Using results presented in a previous report describing a global analysis of yeast proteins using multidimensional chromatography and database searching of MS/MS spectra using SEQUEST (Peng et al. 2003), the percentage of false positives associated with our study was determined to be no greater than 10%. To increase the confidence of positive identification of cysteinyl-labeled peptides to greater than 90%, we added the requirement of the presence of at least two label-specific ions in the MS/MS spectrum (Borisov et al. 2002). On the basis of these numbers, the overall enrichment of proteins containing at least one transmembrane domain (TMD) relative to all those identified was 42% and 44% for the labeled and unlabeled samples, respectively.

Importantly, all four subunits of succinate dehydrogenase [*sdhA* (PA1583), *sdhB* (PA1584), *sdhC* (PA1581), and *sdhD* (PA1582)] predicted by the genome sequence were confidently identified in the unlabeled sample out of which *sdhC* and *sdhD* are very hydrophobic. Subunits *sdhA*, *sdhB*, and *sdhC* were also identified in the labeled aliquot. Together these results reflect the efficiency of the techniques used in extracting, solubilizing, and labeling highly hydrophobic transmembrane peptides, which lead to unambiguous identification of highly hydrophobic integral membrane proteins from complex mixtures using conventional one-dimensional micro-liquid chromatography tandem mass spectrometry (μLC MS/MS) analysis.

Transferases and Cytochromes

Prolipoprotein diacylglyceryl transferase (PA0341) is a hydrophobic integral membrane protein with four mapped TMDs. This protein, not previously identified in *P. aeruginosa,* was characterized by identifying four peptides. SFFQLMDFIAPLVPIGLGAGR, which is the first peptide, is very hydrophobic and completely covers the second mapped TMD. PA0341 is an essential factor for the growth, division, and viability of bacterial cells at nonpermissive temperatures. Homologues were identified in *Escherichia coli*, *Salmonella typhymurium*, and *Haemophilus influenzae* (Gupta et al. 1993). PA0341 is encoded by the *lgt* gene and is involved in fatty acid and phospholipid metabolism, translation, and post-translational modifications (Qi et al. 1995). As a cytoplasmic membrane enzyme, it is a part of the sequential catalysis that involves prolipoprotein signal peptidase and the *lnt* gene, which encodes for apolipoprotein N-acyl transferase (PA3984) (Gupta et al. 1993). PA3984 has six mapped TMDs and is involved in the formation of lipid-modified proteins. The protein modification is achieved by transferring the diacylglycerol group from phosphatidyl-glycerol to the sulfhydryl group of a cysteinyl residue present at the C-terminal portion of the signal sequence of the protein (Gupta et al. 1993).

MdoH and *mdoG* genes are necessary for the synthesis of membrane-derived oligosaccharides that occupy the periplasmic space of gram-negative bacteria and for glucosyl transferase activity. *MdoH* codes for an integral membrane protein expressing 76% homology with the *mdoH* gene of *E. coli* whose protein spans the cytoplasmic membrane and is required for the expression of disease symptoms and development of hypersensitive reaction on non-host plants (Loubens et al. 1993). The *mdoH* protein of *P. aeruginosa* (PA5077) was detected by four peptides from both the unlabeled and labeled samples. Three proteins (PA1554, PA1557, and PA4430) from the cytochrome super-family were also identified by

hydrophobic peptides. These proteins were not previously identified in *P. aeruginosa* or other prokaryotes according to the current *P. aeruginosa* Community Annotation Project Database. In addition, 14 other proteins from this super-family were identified. Additional proteins, not previously identified in *P. aeruginosa* until applying our MS-based approach, include NADH dehydrogenase I chain H (PA2643) and heat shock protein PA2643 whose homologues were characterized previously in *E. coli* exhibiting 84% and 63% sequence similarity, respectively (Weidner et al. 1993; Kornitzer et al. 1991).

Antibiotic Resistance, Adaptation, and Protection Proteins

Since the most important targets of antibacterial agents are located within the cytoplasmic membrane or cytoplasm, the outer membrane presents a natural barrier to antibiotics and biocides preventing their access to target sites (Denyer et al. 2002). Porins or channel-type proteins are important factors of antibiotic resistance of *P. aeruginosa* because they prevent the passage of hydrophilic antibiotics via a size-exclusion mechanism, whereas the dense lipopolysaccharide bilayer of the outer membrane slows down the penetration of hydrophobic macromolecules inside the cell (Poole 2002). However, the lower permeability of the outer membrane alone cannot completely protect the microbe, so additional cellular mechanisms are necessary to create the high level of antibiotic resistance exhibited by *P. aeruginosa* (Poole 2002). Hydrolytic enzymes possessing lactamase activity are omnipresent within the periplasmic space. We confidently identified a probable penicillin amidase (PA1032) by detecting multiple peptides. Notably, this enzyme was not previously characterized in *P. aeruginosa* and could serve as a viable pharmacological target.

Conclusion

This work presents a global qualitative analysis of the membrane subproteome of *P. aeruginosa* strain PAO1 in which hydrophobic membrane proteins are confidently identified on a comprehensive, large-scale basis using μLC MS/MS. Several biologically and medically significant classes of integral membrane proteins involving antibiotic resistance, adaptation, and susceptibility were identified in this investigation, regardless of their hydrophobicity and membrane location.

References

Blonder J, MB Goshe, RJ Moore, L Pasa-Tolic, CD Masselon, MS Lipton, and RD Smith. 2002. "Enrichment of Integral Membrane Proteins for Proteomic Analysis Using Liquid Chromatography-Tandem Mass Spectrometry." *Journal of Proteome Research* 1:351-360.

Borisov OV, MB Goshe, TP Conrads, VS Rakov, TD Veenstra, and RD Smith. 2002. "Low-Energy, Collision-Induced Dissociation Fragmentation Analysis of Cysteinyl-Modified Peptides." *Analytical Chemistry* 74:2284-2292.

Denyer SP and JY Maillard. 2002. "Cellular Impermeability and Uptake of Biocides and Antibiotics in Gram-Negative Bacteria." *Journal of Applied Microbiology* 92:35S-45S.

Goshe MB, J Blonder, and RD Smith. 2003. "Affinity Labeling of Highly Hydrophobic Integral Membrane Proteins for Proteome-Wide Analysis." *Journal of Proteome Research* 2:153-161.

Gupta SD, K Gan, MB Schmid, and HC Wu. 1993. "Characterization of a Temperature-Sensitive Mutant of *Salmonella typhimurium* Defective in Apolipoprotein N-Acyltransferase." *The Journal of Biological Chemistry* 268(22):16551-16556.

Kornitzer D, D Teff, S Altuvia, and AB Oppenheim. 1991. "Isolation, Characterization, and Sequence of an *Escherichia coli* Heat Shock Gene, Htpx." *Journal of Bacteriology* 173:2944-2953.

Loubens IL, A Debarbieux, JM Bohin, JP Lacroix, and JP Bohin. 1993. "Homology Between a Genetic Locus (Mdoa) Involved in the Osmoregulated Biosynthesis of Periplasmic Glucans in *Escherichia coli* and a Genetic Locus (Hrpm) Controlling Pathogenicity of *Pseudomonas syringae*." *Molecular Microbiology* 10:329-340.

Nikaido H. 1998. "Multiple Antibiotic Resistance and Efflux." *Current Opinion in Microbiology* 1:516-523.

Nouwens AS, SJ Cordwell, MR Larsen, MP Molloy, M Gillings, MD Willcox, and BJ Walsh. 2000. "Complementing Genomics with Proteomics: The Membrane Subproteome of *Pseudomonas aeruginosa* Pao1." *Electrophoresis* 21:3797-3809.

Peng JM, JE Elias, CC Thoreen, LJ Licklider, and SP Gygi. 2003. "Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC-MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome." *Journal of Proteome Research* 2:43-50.

Poole K. 2002. "Mechanisms of Bacterial Biocide and Antibiotic Resistance." *Journal of Applied Microbiology* 92:55S-64S.

Qi HY, K Sankaran, K Gan, and HC Wu. 1995. "Structure-Function Relationship of Bacterial Prolipoprotein Diacylglyceryl Transferase: Functionally Significant Conserved Regions." *Journal of Bacteriology* 177:6820-6824.

Quinn JP. 1998. "Clinical Problems Posed by Multiresistant Nonfermenting Gram-Negative Pathogens." *Clinical Infectious Diseases* 27:S117-S124.

Santoni V, M Molloy, and T Rabilloud. 2000. "Membrane Proteins and Proteomics: Un Amour Impossible?" *Electrophoresis* 21:1054-1070.

Weidner U, S Geier, A Ptock, T Friedrich, H Leif, and H Weiss. 1993. "The Gene Locus of the Proton-Translocating NADH: Ubiquinone Oxidoreductase in *Escherichia coli*. Organization of the 14 Genes and Relationship Between the Derived Proteins and Subunits of Mitochondrial Complex I." *Journal of Molecular Biology* 233:109-122.

Lyme Disease Spirochete Causative Agent (*Borrelia burgdorferi*) and Closely Related Strains

X Yang[(a)] and DG Camp II[(b)]
(a) State University of New York at Stony Brook, Stony Brook, New York
(b) Pacific Northwest National Laboratory, Richland, Washington

The genome sequencing of the B31 strain of the bacterium *Borrelia burgdorferi* exposed the blueprint of the causative agent of Lyme disease and opened the door for whole genome and proteome investigations into the pathogenic factors of infection (Fraser et al. 1997). Numerous studies have centered on the exposed surface proteins of the bacterium, which elicit strong humor responses during infection of the host (Craft et al. 1986; Schwan et al. 1995). A subset of these proteins, outer surface proteins (Osps), has become the focus of vaccine development efforts, specifically OspA and OspC (Luft et al. 2002), but challenges remain in overcoming sequence heterogeneity found between strains (Hefty et al. 2002) as well as in elucidating the structural determinates of the proteins that will help confer protective immunity for the host (Kumaran et al. 2001; Wei et al. 2000).

A global snapshot of the proteome expressed in *B. burgdorferi* would significantly enhance the pathogenic analysis of the bacterium and demonstrate how expressed proteins differ (both quantitatively and in sequence/structure) between varying strains of *B. burgdorferi*, with known variabilities in pathogenicity within a host. Recent advances in proteomic analysis now make it possible to identify large numbers of proteins with significant increases in throughput and dynamic range of detection (Washburn et al. 2001; Smith et al. 2002). We have achieved the first proteomic analysis of three strains of the bacterium *B. burgdorferi*, with comparison of results obtained from the B31, N40, and JD-1strains. Strain B31 corresponds to the majority of infections of Lyme disease while infections with strains N40 and JD-1 do not cause the systemic effects of the disease. Of interest is the molecular reasoning behind the differences in pathology observed between the strains, and detection and identification of those proteins that participate in the infectivity pathways. Considering that only the B31 strain has been completely sequenced, and that numerous proteins remain unknown or hypothetical with regard to function, there is an enormous amount of potential information that can be obtained from such a proteome analysis.

Results

Analysis of the three strains of *B. burgdorferi* was accomplished using multidimensional separation techniques applied at the peptide level and incorporating strong cation exchange (SCX) chromatography with reversed-phase liquid chromatography coupled to tandem mass spectrometry (Washburn 2001). In this study, we are using an off-line approach in which we tryptically digest proteins from a cellular lysate from each strain and, then, collect separate peptide fractions from a limited SCX analysis. A total of greater than 44,000 peptides have been confidently identified, corresponding to a total of 7007 unique peptides and 665 unique proteins representing ~38% of the entire *B. burgdorferi* strain B31 genome.

A large protein overlap is observed (i.e., 72%) between the comparison of any two strains of *B. burgdorferi* with greater than 50% of the confidently identified proteins observed in all three

strains. Of potential interest are those proteins that do not overlap and were detected in only a single strain. There are several possible reasons the observed protein overlap exists, including the differences in relative protein abundance and the specific differences in protein sequence not reflected in the database. However, the overlapping subset of proteins will assist in characterizing the differences between each of the strains.

Not only of interest are the proteins found to be present in only a single strain, but also proteins whose amino acid sequence obviously diverges from that of strain B31, thus precluding identification of all the peptides for each protein from each strain due to a breakdown in sequence homology. Table 1 demonstrates this using the percent peptide coverage of a list of detected proteins to determine the sequence homology with strain B31 and to potentially identify proteins of interest in either strain JD-1 or strain ND40 that differ in sequence comparison to B31. A limited list of Osps and hypothetical proteins were selected as an example. Peptide coverage for a number of B31 proteins approaches 100% with large differences observed between the B31 peptide coverage and the other two strains. Specifically, the peptide coverage of OspC, OspD, and the outer membrane porin demonstrate clearly the differences that can be observed between the three *B. burgdorferi* strains. Also charted in Table 1 is a subset of hypothetical proteins, about which little is known, that display a similar divergence in peptide identification as observed for the characterized outer membrane proteins. Based upon this proteomic coverage, it would appear that each identified hypothetical protein is present in strain B31 and that this information leads to the possibility of targeting those hypothetical proteins in the other two strains, based upon the variations of peptide coverage per strain. Decreases in peptide coverage for each protein could stem from both sequence divergence from the B31 database or from a specific reduction in the relative abundance of the protein in that strain. In either case, such coverage analysis helps to target possible proteins of interest without the need for sequencing the entire genome of a strain.

Table 1. Percent peptide coverage of selected outer membrane and hypothetical proteins.

Protein	B31	N40	JD-1
OspA (BBA15)	93[(a)]	79	81
OspB (BBA16)	87[(a)]	83	87[(a)]
OspC (BBB19)	83[(a)]	**13**	**13**
OspD (BBJ09)	88[(a)]	77	**28**
Outer Membrane Porin (BBA74)	84	**18**	89[(a)]
Hypothetical Protein BBI39	78[(a)]	**42**	**50**
Hypothetical Protein BBG01	74	**42**	**10**
Hypothetical Protein BB0713	69	80	64
Hypothetical Protein BBH37	47	**0**	39
Hypothetical Protein BBJ34	33	**15**	**0**
Hypothetical Protein BBJ08	40	**0**	**3**

All sequences are based upon Institute for Genomic Research *B. burgdorferi* B31 database and contain the complete N-terminal leader sequence, which was not detected by peptides for any of the above proteins.

(a) These proteins have a peptide coverage >98% if the leader sequence is removed for the percent coverage calculation.

The data can also be examined by functional category. Observed is a high overall percent coverage (>75%) for those categories involved in critical cellular pathways, (protein synthesis, energy metabolism, and transcription) and a low overall unique percentage for these categories (8 to 19%) as well. In contrast, the four categories that were found to have the highest unique percentage (unknown function, conserved hypothetical, cell envelope, and hypothetical) also have relatively low percent coverage. This correlates with the understanding that differences (protein sequence and abundance) between strains will most likely be found in the four categories just mentioned and generally not within proteins involved in central cellular functions.

One can also analyze the distribution of the detected proteins based on chromosome/plasmid location. *B. burgdorferi* has extensive genomic information located on 21 distinct plasmids that compose >50% of its genomic material. Proteins were detected for each plasmid with a 60% protein coverage observed for the chromosome, but the percent coverage of the plasmids was on average significantly less. This is in agreement with the low percent coverage observed for hypothetical proteins that are predominately located on plasmids. There is a distinct contrast in the protein coverage between plasmids, 5 to 30%, which implies that there are differences in the number and/or abundance of proteins originating from these plasmids. Similar to the analysis by functional category, this analysis shows a much higher percent of unique proteins between the three strains for those proteins specifically transcribed and translated from plasmids (mostly hypothetical proteins) versus the chromosome. These trends suggest that differences in expressed proteins, possibly either by sequence or abundance, are generally linked to cellular envelope or unknown/hypothetical proteins that are plasmid-associated.

Discussion

The strains examined here were selected on the basis of their divergent pathogenic nature within a host, and the parallel analysis of such strains will help elucidate novel factors involved in infectivity and Lyme disease progression. A limited amount of protein sequence information is known about the two comparative strains, N40 and JD-1, which correlates well with the peptide coverage results reported here. For OspA, there is >99% homology between strain B31 and both strain N40 and strain JD-1. This is reflected in the high percent peptide coverage seen for each protein with the <1% of mutations in the N40 and JD-1 sequences directly correlating with areas of the protein sequence not detected by peptides generated with the B31 database (data not shown). For OspC there is an approximate 74% and 80% homology for strains N40 and JD-1, respectively, with strain B31. Such relative low homology greatly affects the obtainable peptide coverage per strain, as observed with OspC in Table 1.

On average, more proteins were found to be unique to only one strain in either the hypothetical, conserved hypothetical, unknown function, or cell envelope categories. This finding demonstrates the large reservoir of undiscovered proteins that are potentially involved in cell pathogenesis.

References

Craft JE, DK Fischer, GT Shiamamoto, and AR Steere. 1986. "Antigens of *Borrelia burgdorferi* Recognized during Lyme Disease." *Journal of Clinical Investigation* 78:934-939.

Fraser CM, S Casjens, WM Huang, GG Sutton, R Clayton, R Lathigra, O White, KA Ketchum, R Dodson, EK Hickey, M Gwinn, B Dougherty, JF Tomb, RD Fleischmann, D Richardson, J Peterson, AP Kerlavage, J Quackenbush, S Salzberg, M Hanson, R van Vugt, N Palmer, MD Adams, J Gocayne, J Weidman, T Utterback, L Watthey, L McDonald, P Artiach, C Bowman, S Garland, C Fujii, MD Cotton, K Horst, B Hatch, HO Smith, and JC Venter. 1997. "Genomic Sequence of a Lyme Disease Spirochete." *Nature* 390:580-586.

Hefty PS, EJ Sarah, MJ Caimano, SK Wikel, and DR Akins. 2002. "Changes in Temporal and Spatial Patterns of Outer Surface Lipoprotein Expression Generate Population Heterogeneity and Antigenic Diversity in the Lyme Disease Spirochete." *Infect Immunogenetics* 70:3648-3478.

Kumaran D, S Eswaramoorthy, B Luft, S Koide, JJ Dunn, CL Lawson, and S Swaminathan. 2001. "Crystal Structure of Outer Surface Protein C (Ospc) from the Lyme Disease Spirochete." *The EMBO Journal* 20:971-978.

Luft BJ, JJ Dunn, and CL Lawson. 2002. "Approaches Toward the Directed Design of a Vaccine Against *Borrelia burgdorferi*." *The Journal of Infectious Diseases* 185:S46-S51.

Schwan TG, J Piesman, WT Golde, MC Dolan, and PA Rosa. 1995. "Induction of an Outer Surface Protein on *Borrelia burgdorferi* During Tick Feeding." *Proceedings of the National Academy of Sciences* 92:2909-2913.

Smith RD, GA Anderson, MS Lipton, L Pasa-Tolic, Y Shen, TP Conrads, TD Veenstra and HR Udseth. 2002. "An Accurate Mass Tag Strategy for Quantitative and High-Throughput Proteome Measurements." *Proteomics* 2:513-523.

Washburn MP, D Wolters, and JR Yates III. 2001. "Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology." *Nature Biotechnology* 19:42-247.

Wei D, X Huang, X Yang, JJ Dunn, BJ Luft, S Koide, and CL Lawson. 2000. "Structural Identification of a Key Protective B-Cell Epitope in Lyme Disease Antigen Ospa." *Journal of Molecular Biology* 302:1153-1164.

Extension of Proteomic Technology to Study Brain-Related Diseases and to Produce a Three-Dimensional Image Map of Protein Expression Throughout the Brain

RC Barry,[(a)] HM Mottaz,[(b)] EA Livesay,[(a)] DG Camp II,[(a)] RD Smith,[(a)] DM Sforza,[(c)] and DJ Smith[(c)]

(a) Pacific Northwest National Laboratory, Richland, Washington
(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(c) University of California Los Angeles, Los Angeles, California

Researchers from the University of California Los Angeles and Pacific Northwest National Laboratory (PNNL) are collaborating to develop the first integrated three-dimensional model of the transcriptome and proteome constructs of the mouse brain. This model will be used to reveal the expression patterns of mRNA and proteins throughout the normal brain and to ultimately compare the expression patterns of normal regions to similar regions apparently damaged from disease, drug abuse, or other trauma. A twofold strategy is currently being used to 1) systematically map the protein and mRNA expression levels in spatially registered 1-mm^3 volume elements (known as voxels, see Figure 1), and 2) investigate the regulation of mRNA and proteins in brain regions known to be specifically affected by certain diseases. Our initial pilot study for brain disease research will be on methamphetamine-induced brain damage, which produces physiological effects similar to Parkinson's Disease in mice and primarily damages the striatum of the brain. Researchers at PNNL will seek to identify and quantify the differential expression of proteins within the damaged striatum, as compared to a healthy striatum, and establish a list of proteins that are most intimately associated with Parkinson's-type symptoms.

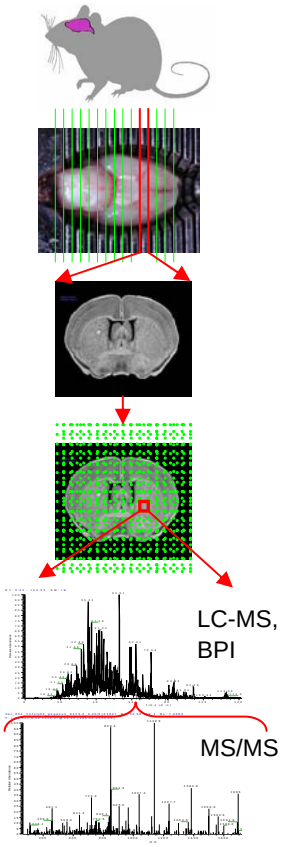

Figure 1. Mechanical preparation of voxels and mass spectrometric analysis.

To quantitatively map protein expression levels within voxels (i.e., samples containing <50 µg of tissue), efficient protein extraction procedures from lipid-laden tissues are being developed. The high-resolution chromatographic separations are being refined to analyze <5 µg of extracted protein by using smaller-diameter columns and higher pressures (15-30 µm, 10-20,000 psi), optimizing the Fourier transform infrared cyclotron resonance mass spectrometers for maximum sensitivity, and developing an automated high-throughput nanoscale sample-handling platform to process the ~600 voxels obtained from a single mouse brain.

Recently, several different protein extraction protocols for brain tissue were evaluated, and we have confidently identified 906 proteins, using automated mass spectral analysis software and very strict assignment rules (Table 1). This is the most extensive list of isolated and

Table 1. Comparison of methods for protein extraction and identifications for mouse brain tissue.

Overlap					
WS/WI	DC/WS	RG/WS	DC/WI	RG/WI	DC/RG
36%	46%	57%	57%	56%	44%
Comparison to Literature					
Method	Ultra-High Resolution RPLC and Ion Trap MS	High-Resolution 2-DE	High-Resolution 2-DE	High-Resolution 2-DE	High-Resolution 2-DE
Unique Proteins	906	437	210	180	90

identified brain proteins that we are aware of and contains twice as many proteins as previously reported in the literature by researchers using two-dimensional gel electrophoresis (2-DE) approaches.

Two detergent-based protocols [one an in-house mixture (DC) and the other a commercially available reagent (RG)] were tested for their ability to extract proteins from lipid-laden tissue and because of their potential suitability for automated sample handling. These two methods were compared with a more laborious technique that requires homogenizing the sample using glass beads, separating the water-soluble (WS) fraction from the water-insoluble fraction (WI) in an ultracentrifuge, dissolving the WI pellet, and then subjecting each fraction to trypsin proteolysis.

The advantage of the subfractionation scheme is the ability to reduce sample complexity and gain information about the subcellular location of proteins. The WS fraction (442 unique proteins) typically contains cytosolic and loosely associated peripheral membrane proteins, while the WI fraction (283 unique proteins) contains mostly membrane-associated proteins. The DC and RG methods resulted in 265 and 264 unique proteins, respectively. Thus, by comparing the overlap of proteins identified from using our detergent-based protocols with those from the WS and WI fractions, we are able to evaluate the types of proteins extracted using DC or RG reagents (Table 1). Both detergents performed equally well, although the brain material was much more completely dissolved when using the DC solution as compared to the RG solution. For automation purposes, both samples would require treatment in an ultracentrifuge (14,000 × g) before liquid chromatography mass spectrometry analysis.

Currently, we are investigating ways to extract more biomolecules from a given voxel while developing a more automation-friendly protein extraction technique. In particular, we are testing methods to subfractionate the brain material into an aqueous layer (containing peptides, RNA, and aqueous soluble small molecule metabolites) and an organic layer (containing lipids and hydrophobic metabolites), with the interphase containing DNA, thereby establishing an information-rich technique for a more complete analysis of each voxel.

High-Throughput FTICR Strategies for the Molecular Analysis of Cancer

DG Camp II[(a)]
(a) Pacific Northwest National Laboratory, Richland, Washington

The cell membrane serves as the primary interface between the cell and its surrounding environment. Events such as cell-cell communication, binding of chemical messengers, interaction with pathogens, environmental perturbations, and transport functions are often mediated by specific interactions with cell surface-exposed membrane proteins. Cellular differentiation and other highly regulated parallel biochemical processes determine the protein composition of the cell membrane, and the dynamic turnover of membrane proteins provides a unique opportunity for the application of new proteomic approaches.

The proteomic labeling method used here significantly advances the amount of coverage, quality, and throughput for investigations of the surface-exposed membrane subproteome of human mammary epithelial cells (HMEC). This method circumvents the need to specifically isolate the plasma membrane and focuses on surface-exposed regions of membrane proteins including secreted and shed proteins while simultaneously avoiding the challenge of working with hydrophobic polypeptides imbedded in the phospholipid bilayer. Specific labeling of cell surface-exposed proteins was achieved by using the water-soluble labeling reagent sulfosuccinimidyl-6-(biotinamido) hexanoate (sulfo-NHS-LC-biotin), where the primary amines of lysine side chains are modified by the introduction of a covalently attached biotin affinity label. This specific labeling of the cell surface-exposed proteins was facilitated by the presence of the negatively charged sulfo group, $(R\text{-}SO_3)^{1-}$, on the labeling reagent, which retards passage of the reagent through the cell membrane because of electrostatic forces, thereby effectively localizing the labeling of primary amines on surface-exposed proteins. While labeling membrane proteins with sulfo-NHS-LC-biotin has been previously reported (Sabarth et al. 2002; Shetty et al. 2001; Ghebrehiwet and Peerschke 1998; Schuberth et al. 1996; Bradburne et al 1993), this project is the first application of affinity labeling with sulfo-NHS-LC-biotin to HMEC cells followed by subsequent characterization of avidin-enriched, biotin-labeled tryptic peptides using high-resolution reversed-phase capillary liquid chromatography coupled to electrospray ionization ion trap tandem mass spectrometry (LC MS/MS).

Using this labeling and enrichment method resulted in the identification of 363 biotinylated peptides using the SEQUEST protein database search algorithm where the output was filtered using criteria published by Washburn et al. 2001. The Human database available through the National Cancer Institute was used in the analysis of the MS/MS data by SEQUEST. Figure 1(A) shows the LC total ion chromatogram (TIG) for avidin-enriched, biotinylated peptides; Figure 1(B) shows the mass spectrum for a representative tryptic peptide, WLK*GGVVLKEDALPGQK (where the symbol "*" represents the site of lysine biotinylation in the peptide); and Figure 1(C) shows the MS/MS spectrum for the tryptic peptide and includes both the ion series and SEQUEST scoring parameters. The identity of the labeled peptide is further strengthened by the observation of the biotin label fragment (34 0.468 Da) in the MS/MS spectrum.

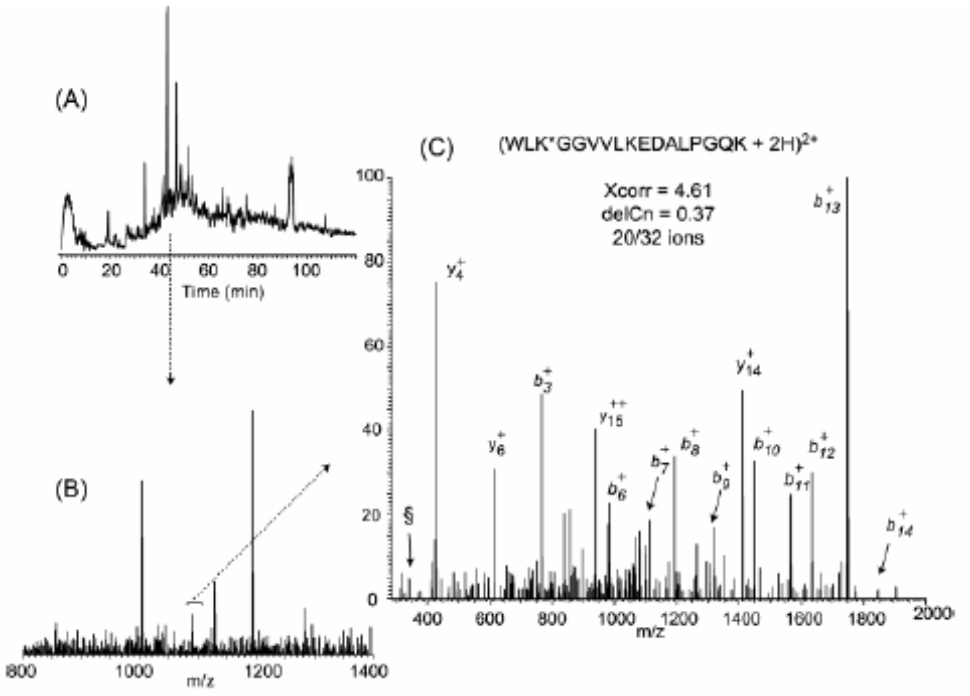

Figure 1. TIC for an avidin-enriched, biotinylated, and tryptically digested HMEC (strain 184 A1L5) protein sample is shown in (A). In the parent mass spectrum (B) acquired on the liquid chromatograph interfaced with an ion trap mass analyzer, a peptide of m/z =1090 is selected for dissociation. The resulting MS/MS spectrum for this ion and the b/y fragmentation series (C) were used to identify the sequence of the peptide, using SEQUEST, as WLK*GGVVLKEDALPGQK (where "*" is a site of biotinylation). Further verification that the peptide contains the biotinylation is seen from the peak (340.5 Da) in the MS/MS spectrum, representing the dissociated label indicated with the symbol "§".

In total, 45 different proteins could be identified using the biotinylated peptides from two independent experiments (different batches of cell culture) that were searched against the National Cancer Institute's Human database. These proteins are primarily either membrane proteins or membrane-associated proteins located on the surface of HMEC cells. Of these identified proteins, 34 are integral membrane proteins containing one or more transmembrane domains, two are anchored proteins, three are identified as secreted or extracellular proteins, and six are annotated simply as membrane or glycoproteins. Each of these proteins represents a potential biomarker for studying cellular response to chemical treatment or environmental perturbation, or changes in cellular physiology due to infection or cellular disease states. Quantitation of specific changes in cell surface-exposed proteins in the membrane subproteome could be achieved by employing a combination of stable isotope labeling (Goodlett et al. 2001) with the biotinylation approach reported here.

In summary, we have shown that this method can greatly enrich the affinity-labeled peptides in the samples that are used for LC MS/MS analysis. We have shown that this labeling provides a mechanism for reducing the overall complexity of the sample and significantly enhances the identification by LC MS/MS of the labeled peptides. We have also demonstrated that the labeling is inclusive enough to simultaneously provide a "snapshot" of the cell surface-exposed proteins in the membrane subproteome.

References

Bradburne JA, P Godfrey, JH Choi, and JN Mathis. 1993. "*In Vivo* Labeling of *Escherichia coli* Cell Envelope Proteins with N-Hydroxysuccinimide Esters of Biotin." *Applied and Environmental Microbiology* 59:663-668.

Ghebrehiwet B and EI Peerschke. 1998. "Platelet Receptors for the Complement Component C1q: Implications for Hemostasis and Thrombosis." *Immunobiology* 199:225-238.

Goodlett DR, A Keller, JD Watts, and R Newitt. 2001. "Differential Stable Isotope Labeling of Peptides for Quantitation and *De Novo* Sequence Derivation." *Rapid Communications in Mass Spectrometry* 15:1214-1221.

Sabarth N, S Lamer, U Zimny-Arndt, and PR Jungblut. 2002. "Identification of Surface Proteins of *Helicobacter pylori* by Selective Biotinylation, Affinity Purification, and Two-Dimensional Gel Electrophoresis." *Journal of Biological Chemistry* 277:27896-27902.

Schuberth HJ, A Kroell, and WJ Leibold. 1996. "Biotinylation of Cell Surface MHC Molecules: A Complementary Tool for the Study of MHC Class II Polymorphism in Cattle." *Immunology Methods* 189:89-98.

Shetty J, AB Diekman, FC Jayes, and NE Sherman. 2001. "Differential Extraction and Enrichment of Human Sperm Surface Proteins in a Proteome: Identification of Immunocontraceptive Candidates." *Electrophoresis* 22:3053-3066.

Washburn MP, D Wolters, and JR Yates III. 2001. "Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology." *Nature Biotechnology* 19:242-247.

Technology for Global and Quantitative Proteome Analysis of Human Tumors

DG Camp II[(a)]
(a) Pacific Northwest National Laboratory, Richland, Washington

Reversible phosphorylation of proteins on serine, threonine, and tyrosine residues is one of the most prevalent post-translational modifications, and nearly 30% of all proteins expressed in eukaryotic cells are phosphorylated to some extent (Hubbard et al. 1993). The general significance of phosphorylation has been well appreciated as a central mechanism that regulates nearly every aspect of cellular life and touches almost every known signaling pathway. Despite its importance, the global analysis of protein phosphorylation remains a major analytical challenge for several reasons including the low abundance of many signaling phosphoproteins within cells, the low stoichiometry of phosphorylation at any specific site for a given protein, and the limited dynamic range for current analytical techniques employed for the study of protein phosphorylation.

Recently, several novel methodologies have been developed for selective enrichment of phosphopeptides from complex mixtures using chemical modification strategies or immobilized metal affinity chromatography (IMAC). Researchers at Pacific Northwest National Laboratory (Goshe et al. 2001; Oda et al. 2001) concurrently and independently, developed two similar methods for modifying and enriching phosphoseryl (pSer) and phosphothreonyl (pThr) peptides. Both methods involve hydroxide ion-mediated, β-elimination of the O-phosphate moiety and the addition of 2-ethanedithiol (EDT) followed by biotinylation and isolation using avidin affinity chromatography (Goshe et al. 2001). Our method, termed PhIAT (phosphoprotein isotope-coded affinity tag), incorporates stable isotope labeling by using either light ($HSCH_2CH_2$-SH, EDT-d0) or heavy ($HSCD_2CD_2SH$, EDT-d4) isotopic versions of EDT, allowing relative quantitation of protein phosphorylation states. One disadvantage of the β-elimination modification is that it is not applicable to phosphorylated tyrosine (pTyr) residues.

Although some promising results have been demonstrated with the PhIAT approach, limitations associated with avidin affinity chromatography including nonspecifically bound peptides as well as reduced sample recovery of a subpopulation of the biotinylated peptides have been observed. To address these issues, a novel isotope-labeling strategy, termed the phosphoprotein isotope-coded solid-phase tag (PhIST) approach, has been developed using an IST in place of the biotin affinity tag implemented in our original PhIAT approach. An IST similar to that used by Zhou and co-workers (Zhou et al. 2002) was coupled with the PhIAT approach to more effectively isolate phosphopeptides from complex peptide mixtures.

To evaluate the efficiency of the PhIST labeling method for identification and quantitation of phosphopeptides, two equivalent samples of β-casein were used for labeling, according to the procedures outlined in Figure 1. When compared to the earlier PhIAT approach, the results of these analyses suggest at least a 25-fold improvement in overall sensitivity. The

utility of this approach was explored, applying it to the human breast cancer cell line MCF-7. Initially, 200 μg of soluble protein extracts obtained from the MCF-7 cell lysate was used for each PhIST labeling. The MCF-7 proteins were denatured, reduced, and alkylated before being subjected to β-elimination and EDT addition. The enriched PhIST-labeled peptides were analyzed by liquid chromatography tandem mass spectrometry (LC MS/MS), and the peptides were identified using the SEQUEST protein database search algorithm.

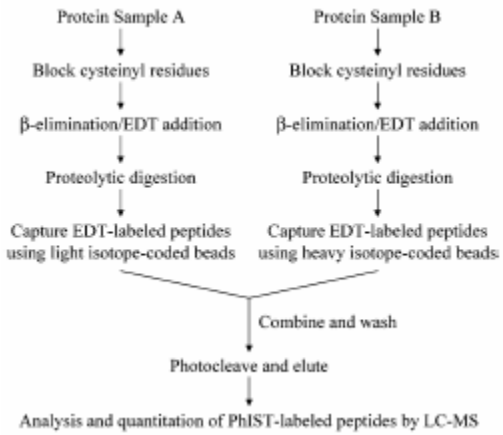

Figure 1. Strategy for quantitative analysis of phosphopeptides from two different samples using the PhIST approach.

The LC MS/MS analysis of the MCF-7 sample resulted in identification of a total of 86 PhIST-labeled peptides and 13 unlabeled peptides using the SEQUEST filtering criteria of Peng et al. (2003). Twenty-eight of these peptides were further confirmed by manually inspecting the LC MS/MS spectra, and all of these peptides were confidently identified as either the PhIST-L- or PhIST-H-labeled versions, but not both, because of software constraints in the data-dependent, MS-acquisition method used in the LC MS/MS analysis. By manually examining the precursor MS spectra, nearly 70% of the identified peptides were observed as light- and heavy-labeled peptide pairs. The average abundance ratio was 1.02 + 0.24, which is in agreement with measured protein abundance ratios using other stable isotope-labeling approaches (Han et al. 2001). Although the abundance ratios of the remaining pairs were difficult to measure because of low signal-to-noise ratios of the centroid MS spectra, these data suggest the merit of quantifying relative phosphorylation using the PhIST approach with higher resolution and sensitivity mass spectrometers (e.g., time of flight MS or Fourier transform ion cyclotron resonance MS). In addition to quantitation, the exact pSer or pThr residue can be elucidated by the presence of the PhIST label. For example, residue Ser254 of heat shock protein HSP 90-β was determined to be PhIST-labeled and, thus phosphorylated, which agrees with a previous report using other biochemical methods to verify this site as phosphorylated (Lees-Miller and Anderson 1989). Unfortunately, most of the identified phosphorylation sites of the 28 peptides are undocumented in the current literature. To probe the potential of the PhIST approach to identifying phosphorylation sites, the proteins identified by the 28 peptides were analyzed using a phosphorylation motif-based tool, Scansite (Yaffe et al. 2001). The results indicated that nearly 65% of the PhIST-labeled sites are predicted to be phosphorylation sites and suggest that these sites are potentially involved in biological pathways. Those sites not predicted by Scansite could be either novel motifs not contained within the Scansite database or potential sites of O-linked glycosylation. It is known that O-linked carbohydrates are also susceptible to β-elimination via this approach, although the kinetics may be significantly slower. Nevertheless, these data, particularly those identified peptides that are multiply labeled, clearly demonstrate the broad utility of the PhIST approach for identifying phosphopeptides from complex protein mixtures.

Better understanding of biological signaling networks requires global identification and quantitation of transient protein phosphorylation events. Although the analysis of protein phosphorylation has been a rapidly evolving field, the goal of phosphoproteome analysis for cells and tissues remains an analytical challenge. The PhIST-labeling approach developed represents another step forward toward achieving the proteome-wide analysis of phosphoproteins. The results obtained from the casein protein mixtures and the MCF-7 cell lysate demonstrate the efficacy and sensitivity of the PhIST method. Overall, the method is significantly simpler and provides improved sample recovery and increased sensitivity, relative to the original PhIAT approach.

References

Goshe MB, TP Conrads, EA Panisko, NH Angell, TD Veenstra, and RD Smith. 2001. "Phosphoprotein Isotope-Coded Affinity Tag Approach for Isolating and Quantitating Phosphopeptides in Proteome-Wide Analyses." *Analytical Chemistry* 73:2578-2586.

Han DK, J Eng, H Zhou, and R Aebersold. 2001. "Quantitative Profiling of Differentiation-Induced Microsomal Proteins Using Isotope-Coded Affinity Tags and Mass Spectrometry." *Nature Biotechnology* 19:946-951.

Hubbard MJ and P Cohen. 1993. "On Target with a New Mechanism for the Regulation of Protein Phosphorylation." *Trends in Biochemical Sciences* 18:172-177.

Hunter T. 2000. "Signaling-2000 and Beyond." *Cell* 100:113-127.

Lees-Miller SP and CW Anderson. 1989. "Two Human 90-kDa Heat Shock Proteins are Phosphorylated *In vivo* at Conserved Serines that are Phosphorylated *In vitro* by Casein Kinase II." *Chemico-Biological Interactions* 264:2431-2437.

Oda Y, T Nagasu, and BT Chait. 2001. "Enrichment Analysis of Phosphorylated Proteins as a Tool for Probing the Phosphoproteome." *Nature Biotechnology* 19:379-382.

Peng J, JE Elias, CC Thoreen, LJ Licklider, and SP Gygi. 2003. "Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC-MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome." *Journal of Proteome Research* 2:43-50.

Yaffe MB, GG Leparc, J Lai, T Obata, S Volinia, and LC Cantley. 2001. "A Motif-Based Profile Scanning Approach for Genome-Wide Prediction of Signaling Pathways." *Nature Biotechnology* 19:348-353.

Zhou HL, JA Ranish, JD Watts, and R Aebersold. 2002. "Quantitative Proteome Analysis by Solid-Phase Isotope Tagging and Mass Spectrometry." *Nature Biotechnology* 19:512-515.

Proteomic Profiling of *Yersinia pestis*

MS Lipton,[(a)] KK Hixson,[(b)] and S McCutchen-Maloney[(c)]

(a) Pacific Northwest National Laboratory, Richland, Washington
(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(c) Lawrence Livermore National Laboratory, Livermore, California

The genome of the *Yersinia pestis* strain KIM, a gram-negative, pathogenic bacterium, consists of one chromosome and 3 plasmids (pCD1, pMT1, and pPCP1). *Y. pestis* normally cycles between a flea vector and a mammalian host, causing disease and probable death in humans if not treated. The goal of this work is to create as complete a profile of the proteins expressed by *Y. pestis* as possible. To this end, we used multiple growth conditions and liquid chromatography tandem mass spectroscopy (LC MS/MS) for the global protein analyses. Our interpretation of the spectra used the 4221 protein-encoding open reading frames (ORFs) predicted by the primary annotation from the Sanger Institute website (http://www.sanger.ac.uk). We excluded all genes with frame shifts from this analysis.

Four growth conditions—37°C without Ca^{2+}, 37°C with Ca^{2+}, 26°C without Ca^{2+}, and 26°C with Ca^{2+}—were used to stimulate the expression of as wide a range of proteins as practical. These conditions were chosen because both temperature and calcium stress were known to dramatically affect several biochemical pathways in *Y. pestis*. The cells from each growth condition were lysed, and the proteins were extracted either globally or as membrane and cytosolic fractions. The collected protein samples were then digested with trypsin and multiple analyses by LC MS/MS were performed on an ion trap mass spectrometer. Peptide identifications from the LC MS/MS analysis were made using the peptide identification algorithm SEQUEST. For each growth condition the following analyses were performed: full mass range analyses of the global digest in triplicate, seven small mass range analyses of the global digest in triplicate, 90 full mass range analyses of each fraction from two-dimensional separation (i.e., strong cation exchange followed by reverse-phase LC analysis) of the global digest, full mass range analyses of both the cytosol and membrane fractions in duplicate, and seven small mass range analyses of the cytosol and membrane fractions in duplicate. This approach required a total of 520 LC MS/MS analyses, all of which were needed because of the complexity of proteomic samples and the limited charge capacity of ion trap mass spectrometers. Because the full mass range of interest is large (200 to 2000 Da), the limited charge capacity requires that only a small part (1/7) of the mass range be used if a complete analysis of a complex sample is required. This extensive analysis of these four growth conditions yielded a large data set of putative mass tag (PMT) identifications. A subset of these PMTs was subsequently validated with high mass measurement accuracy Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) analysis. The greater charge capacity, resolution, and mass measurement accuracy of the FTICR MS coupled with the LC elution time permitted the analysis of each protein fraction from each growth condition in a single LC MS analysis.

The final results of these analyses are displayed in Table 1 where the distribution of observed ORFs in each genomic element is listed. The table illustrates that the method was effective in observing at least 60% of the predicted proteomes of *Y. pestis*, and there was no obvious bias for or against any proteins translated from any one genetic element.

Table 1. Proteomic observations of ORFs from *Y. pestis*.

Genetic Element	Predicted ORFs	Observed ORFs	Percent Coverage
Total	4223	3125	74
Chromosome	4016	2994	74.6
pPMT1 plasmid	103	65	63.1
pCD1 plasmid	94	60	63.8
PPCP1 plasmid	10	6	60

A further analysis of the observed proteins can be done by distributing them into their functional categories. Figure 1 is a representation of such a distribution. The size of a slice indicates the number of proteins from the annotated genome that fall into the category indicated. Each slice has been shaded to indicate the percentage of proteins from that category that has been identified by the proteomic profiling.

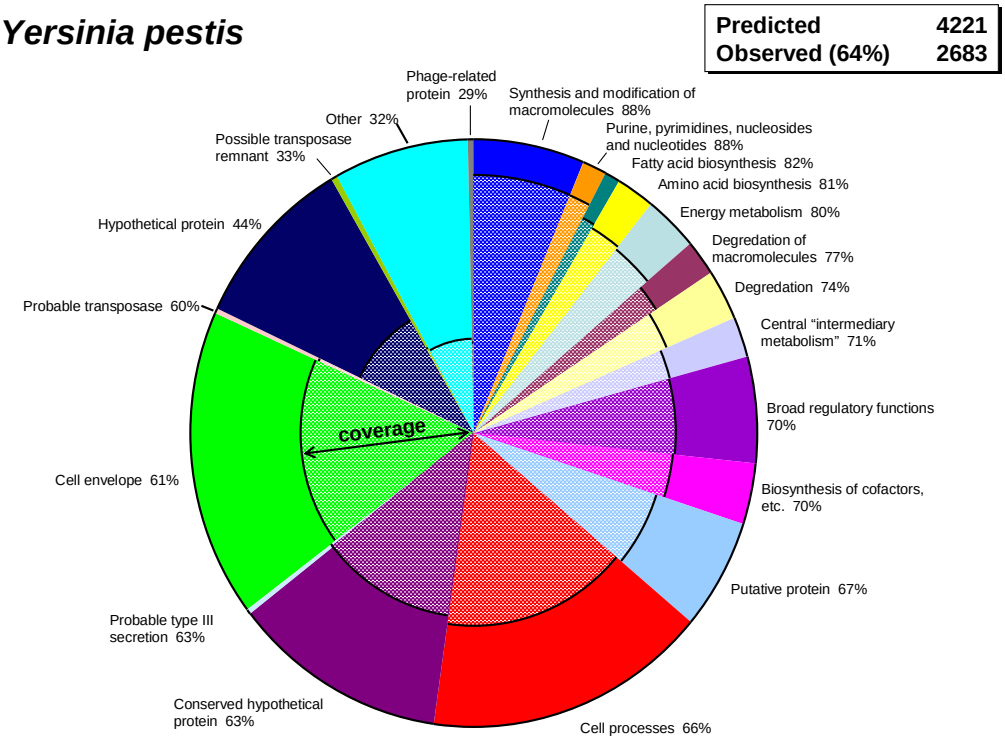

Figure 1. Distribution and coverage of proteins identified using accurate mass and time tags from studies of *Y. pestis*.

A general evaluation of the reliability of the identifications can be had by determining if our expectations for the expression of critical proteins have been met. The categories associated with housekeeping functions, protein synthesis, amino acid biosynthesis, cell envelope membrane, etc., had representation in general of over 80% (Figure 1). Most of these proteins are required in a living cell under any conditions, so an analysis that purported to represent the proteins in a cell would have to detect these in high abundance. Because we do this analysis, we can examine the other proteins detected with reasonable confidence in their importance. This is of immediate interest for hypothetical and conserved hypothetical proteins. These are proteins whose expression has not been previously confirmed by

physical methods. This proteomic profiling of *Y. pestis* has confirmed the expression of 63% and 44% of the conserved hypothetical and hypothetical proteins, respectively. By confirming the expression of these proteins, the corresponding ORFs have been identified as possible targets for further study. Continuing work on *Y. pestis* is focused on identifying the pathogenic proteins expressed under the four different growth conditions, some of which will be found among these hypothetical proteins.

Biomolecular Mechanisms for Microbe-Fe (III) Oxide Interactions in *Geobacter* Species

KK Hixson,[(a)] DR Lovely,[(b)] K Nevins,[(b)] and MS Lipton[(c)]

(a) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

(b) University of Massachusetts, Boston, Massachusetts

(c) Pacific Northwest National Laboratory, Richland, Washington

The *Geobacter* family of bacteria is known for its ability to reduce groundwater contaminants such as uranium, technetium, and other heavy metals into insoluble solid forms. This ability makes the *Geobacter* family an attractive candidate for bioremediation of contaminated sites by reducing contaminant mobility through soil and into groundwater. The mechanism for this reduction in contaminant mobility is the transfer of an electron to the heavy metal through cell surface proteins of the bacteria. To gain a detailed understanding of the intermediate steps in this mechanism and to identify which proteins may be involved in the bioremediation aspects of this organism, investigations of surface-exposed proteins are underway using the unique sample preparation techniques and instrumentation of the W.R. Wiley Environmental Molecular Sciences Laboratory. Cells of *Geobacter sulfurreducens* were grown under two environmental conditions relevant to heavy metal reduction: 1) where fumarate was used as the end electron acceptor to mimic conditions without metal and 2) where Fe(III) citrate was used to mimic the presence of a heavy metal. After growth was stopped, the surface-exposed proteins of the cells were labeled and extracted. The proteins were then digested with trypsin and passed over a substrate that would bind strongly to the label. The labeled tryptic peptides are retained on the substrate and eluted separately. The labeled peptide mixture was then analyzed via high-pressure liquid chromatography and high-performance mass spectrometry.

The bacteria were labeled by two different biotinylation reagents: sulfo-NHS-biotin, which labels primary amine groups, and biotin-PEO-amine, which labels carboxyl groups. Because the reagents bind to two different functional groups, the raw numbers of proteins identified will indicate which of these functional groups is exposed in the greatest number and, therefore, is probably more important in metal reduction. The numbers of unique membrane proteins that have been identified are listed in Table 1. The data show that the primary amine labeling technique was most comprehensive.

Table 1. Unique outer membrane proteins identified from *G. sulfurreducens* grown with either fumarate or Fe(III) citrate as the electron donor.

	Primary Amine Labeled	Carboxyl Labeled
Fumarate	73	8
Fe(III)	69	15

Table 1 gives no indication of how many of the proteins identified in the two growth conditions are the same. A detailed look at the proteins identified shows that there is very little overlap between the two groups. Of primary importance in this study are the unique proteins that are exposed on the cell surface in response to the presence of iron in the extracellular environment. Table 2 is a comprehensive list of not just the proteins uniquely

expressed but of all the labeled surface-exposed peptides identified exclusive only to the metal reduction from the Fe(III) citrate growth. Peptides identified in both Fe(III) citrate and fumarate (non-metal-containing growth) or peptides unique only to the fumarate growth were excluded. The proteins detected are listed in numerical order with multiple peptide hits for the same protein indicated by redundancy in the open reading frame number. The listed peptides identify potential protein targets involved in reducing harmful metals into insoluble forms so as to prevent contaminant mobility into clean groundwater. In addition, the label itself provides information as to what domain of each identified protein may be directly involved in the electron transfer.

Table 2. Identified surface exposed peptides expressed only with Fe(III) citrate reduction.

Carboxyl Labeled Peptides

Reference	Annotation	Sequence
ORF00042	OmpA domain protein	K.AQTITE*APVAVASQASDTAGR.A
ORF00042	OmpA domain protein	R.GSDE*YNLALGEK.R
ORF00042	OmpA domain protein	R.GSD*EYNLALGEK.R
ORF00042	OmpA domain protein	R.AE*AAETSLER.I
ORF00197	ATP synthase F1, epsilon subunit (atpC)	K.ILSE*EVDEITATGALGEFSVLPGHAPFLTSLK.I
ORF00309	lipoprotein, putative	K.IE*QALQEAQAAK.A
ORF00555	serine protease, HtrA/DegQ/DegS family (htrA)	K.SGSLAEE*AGILPGDIVR.E
ORF00595	OmpA domain protein	K.AADFLKEYPDATAVIEGHTDSVGSD*AYNQK.L
ORF00595	OmpA domain protein	K.EYPDATAVIEGHTDSVGSD*AYNQK.L
ORF00595	OmpA domain protein	K.DTDGDGVIDD*LDK.C
ORF00917	conserved hypothetical protein	R.D*AVREILCRCRQLGVTTV.A
ORF01712	GTP-binding protein (hflX)	R.E*REVIITDTVGFIR.S
ORF02546	hypothetical protein	K.D*FTFKPSTNVSVSYF.T
ORF02900	branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein, putative	K.FIE*LAGADAAEGIMLPAGK.L
ORF02967	type II secretion system protein, putative	K.TVGSGD*AISSVVTIGTR.N
ORF05218	penicillin-binding protein, 1A family	R.YGE*EQLYK.E
ORF05512	hypothetical protein	K.NCATE*VDSLQK.K

Primary Amine Labeled Peptides

Reference	Reference	Sequence
ORF00077	hypothetical protein	I.AHGVDPDHDKTPLK*KAK*K*.L
ORF00306	lipoprotein, putative	K.ILEQK*VFAIATR.K
ORF00574	NADH dehydrogenase I, J subunit (nouJ)	R.SSLTGNK*GEISPELIHR.I
ORF00578	NADH dehydrogenase I, M subunit (nuoM)	K.MSPSLEK*LIVHVK.T
ORF00578	NADH dehydrogenase I, M subunit (nuoM)	K.TK*QQGAVAQVETVPAAPQVMPAAAVVPAEHK.-
ORF00578	NADH dehydrogenase I, M subunit (nuoM)	K.QQGAVAQVETVPAAPQVMPAAAVVPAEHK*.-
ORF00991	high-molecular-weight cytochrome c family protein	R.NTK*VVTMHEMEKGK*.S
ORF00991	high-molecular-weight cytochrome c family protein	N.K*PVGMAAMEK.G
ORF00991	high-molecular-weight cytochrome c family protein	K.VVTMHEMEK*GK.S
ORF00991	high-molecular-weight cytochrome c family protein	R.VVFSHK*AHIGK.K
ORF00991	high-molecular-weight cytochrome c family protein	K.EVLLSSPGTGK*IIFSHK.L
ORF00991	high-molecular-weight cytochrome c family protein	R.NK*PVGMAAMEK.G
ORF00991	high-molecular-weight cytochrome c family protein	K.IIFSHK*LHAGK.M
ORF01113	cytochrome c family protein	K.K*AVDLLSASTAK.H
ORF01307	nickel-dependent hydrogenase, small subunit	K.K*PIAEK.L
ORF01560	mce related protein family	R.FASDPK*LYDNLVR.L
ORF01564	peptidase, M16 family, putative	K.TFDK*PLSTFGTVQEIK.L
ORF01826	conserved hypothetical protein	R.IEK*DTYIIFPVIYVK.G
ORF02178	cytochrome c, putative	R.FVVVQHTIK*DVGGKPTK.G
ORF02202	methyl-accepting chemotaxis protein	Q.K*LADLK.I
ORF02258	hypothetical protein	K.TGK*AITQGEVNLK*VQGPDNK.A
ORF02258	hypothetical protein	K.GMK*ETHHIAVEFK.D
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	R.AVIDGLEADVVTLALAYDIDEIAGK*AK.L
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	K.GK*GGGPVTVK.Q
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	R.GQGDVLLAWENEAFLAVNELGK*DK.F
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	K.VK*LFTIDDTFGGWQAAQK.K
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	R.ELYQDYNAAFAAHWK*GK*GGGPVTVK.Q
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	K.HYYRPIDK*NVAAK.Y
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	K.DK*FQIVVPSVSILAEPPVTVVDK.V
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	K.GK*GGGPVTVK*QSHGGSGK.Q
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	R.AVIDGLEADVVTLALAYDIDEIAGKAK*.L
ORF02287	sulfate ABC transporter, periplasmic sulfate-binding protein (cysP)	R.AVAEAYLK*YLYTPAGQEIAAK.H
ORF02388	SCO1/SenC family protein	R.IYGLISNVELMNELTK*AGLK.-
ORF02432	hypothetical protein	R.IEEK*LAGLR.E
ORF02455	Peptidase family M48 family	K.VVK*DDSVNAFAIPGGR.V
ORF02462	carbonic anhydrase family protein	R.HLVK*EGTLK.I
ORF02462	carbonic anhydrase family protein	G.SGGVGVTADEALQQLMDGNK*R.Y
ORF02462	carbonic anhydrase family protein	K.AKYDLDDGTVVLMDGKK*.-
ORF02516	lemA protein (lemA)	R.RADLIPNLVEVVK*GYAK*HEAETLTAVTEAR.A

*label attachment site

Primary Amine Labeled Peptides (continued)

Reference	Annotation	Sequence
ORF02516	lemA protein (lemA)	R.RADLIPNLVEVVK*GYAK*HEAETLTAVTEAR.A
ORF02528	soluble lytic murein transglycosylase, putative	R.ALAVFNTIPLK*EQLAGFNDR.V
ORF02545	pilin domain protein	R.VK*AYNSAASSDLR.N
ORF02546	hypothetical protein	A.AGK*IPTTTMGGK.D
ORF02699	fatty acid/phospholipid synthesis protein PlsX (plsX)	N.ELTRETNAILK*NISFDYEG.Y
ORF02713	outer membrane efflux protein	R.DQNDVLHVLLQLPGK*EEIIPVDIPTR.D
ORF02713	outer membrane efflux protein	R.VK*AGVLPAMEILNAEFGVASRIE
ORF02741	glyceraldehyde 3-phosphate dehydrogenase 1 (gap)	R.VPTPNVSVVDLVVTLKK*.K
ORF02914	lipoprotein, putative	R.SGSSK*SVAPDFLFPFAVGK.-
ORF02936	phosphoribosylglycinamide formyltransferase (purN)	R.KVSILQGA.-
ORF02985	ATP-dependent protease La (lon)	R.AQILGIDK*EFHK*KAD.I
ORF03082	capsule polysaccharide export protein, putative	K.YK*VVSVADQAQAAISNIAQLNAQLIAQEIQLSNLR.K
ORF03082	capsule polysaccharide export protein, putative	R.EFK*VQEAIVDLLTK.Q
ORF03104	hypothetical protein	R.GTMTLENFADLFPHQAK*SLGENIVK.F
ORF03176	membrane-associated zinc metalloprotease, putative	K.RGEVIK*TFR.V
ORF03300	high-molecular-weight cytochrome c family protein	K.NAAPVIFSHDIHLKK*.Y
ORF03341	acetyl-CoA carboxylase, biotin carboxyl carrier protein (accB)	K.ENAQPVEFGEPLFIVEPM.-
ORF03341	acetyl-CoA carboxylase, biotin carboxyl carrier protein (accB)	M.K*LMNEIEAEFR.C
ORF03341	acetyl-CoA carboxylase, biotin carboxyl carrier protein (accB)	K.GQVLCIVEAMK*LMNEIEAEFR.C
ORF03353	type IV pilus biogenesis protein PilQ	R.SDVK*IQVQLR.Q
ORF03353	type IV pilus biogenesis protein PilQ	K.AIEAQNGYLVIETSGEIKDFK*FFR.L
ORF03353	type IV pilus biogenesis protein PilQ	R.EIK*VAETGEGAR.V
ORF03438	hypothetical protein	K.ANK*VSNLAK.L
ORF03635	hypothetical protein	K.TVTVTVKDELTLDKFK*.D
ORF03641	hypothetical protein	P.YGK*VDIQIDKGGVELDLNSR.Y
ORF03775	HDIG domain, putative	R.LGSSSELIEQVSQLADHLYSHK*IIADK.G
ORF03846	membrane protein, putative	R.EEYILLLPFTPSK*R.D
ORF04011	pyruvate carboxylase (pyc)	K.AGDVLMVTEAMK*METNIK.A
ORF04344	protein-export membrane protein SecD (secD)	R.EGLTNIVVQLPGISDPK*R.A
ORF04344	protein-export membrane protein SecD (secD)	K.AVEGTLDLIATDLEDALSAK*TLR.Y
ORF04371	lipoprotein, putative	R.YGGDLVGNPEATVAHTGGTYTSTIPIILPADAK*K.G
ORF04397	amino acid ABC transporter, amino acid-binding protein	K.KIVIQNMAFDGLIPALKTGK*.I
ORF04397	amino acid ABC transporter, amino acid-binding protein	R.K*K*VNAFLADFR.K
ORF04423	lipoprotein, putative	R.LAK*GLVVSPLDGVVALR.D
ORF04476	multidrug resistance protein, HlyD family	R.AVFPNPK*QTLLPGMFVR.A
ORF04499	hypothetical protein	K.IESAMNNQFGDSLSNLGVQPGK*R.I
ORF04536	polyheme membrane-associated cytochrome c	R.TIK*ATNGAAGPVVNWDPNNNR.T
ORF04536	polyheme membrane-associated cytochrome c	K.AITDADGILGFVNSHYLAAGGQLFGK*TGYEYATR.S
ORF04536	polyheme membrane-associated cytochrome c	R.EAGENIK*AITDADGILGFVNSHYLAAGGQLFGK.T
ORF04751	transcription antitermination protein NusG (nusG)	I.PSETVVELKKGEKK*.T
ORF04975	sensor histidine kinase	R.SFQELTGLMHQAVALK*R.A
ORF05392	hypothetical protein	K.KEEAAAPAAEKK*.-
ORF05392	hypothetical protein	K.K*PVK*K*AK.K
ORF05392	hypothetical protein	K.K*EEAAAPAAEKK.-
ORF05392	hypothetical protein	K.K*EEAAAPAAEKK*.-
ORF05392	hypothetical protein	K.K*EEKKEEAAAPAAEKK*.-
ORF05392	hypothetical protein	K.K*EEK*KPVKK.A
ORF05392	hypothetical protein	K.K*EEKKPVK*KAK*.K
ORF05392	hypothetical protein	K.K*EEKKPVK*K.A
ORF05392	hypothetical protein	K.K*EEK*KPVK.K
ORF05392	hypothetical protein	K.KEEK*KPVK.K
ORF05392	hypothetical protein	K.K*EEKK*PVKK*.A
ORF05426	hypothetical protein	K.AAAINQAVAHVK*DIYSR.N
ORF05426	hypothetical protein	K.VK*AAAINQAVAHVK.D
ORF05459	conserved hypothetical protein	R.PSGPDEGVLVK*SITVQK.G
ORF05459	conserved hypothetical protein	K.ADDNEHSLYEK*GVSAYK.S
ORF05459	conserved hypothetical protein	K.AVFSQQPSAVPAAAAK*R.E
ORF05536	peptidyl-prolyl cis-trans isomerase A (ppiA)	K.NVLDYVK*SGFYDGTIFHR.V
ORF05793	membrane protein, putative	R.K*TASGYLENTISSPELSLNPGEGR.A
ORF05793	membrane protein, putative	R.APGFGLDQNAVFVPSADALTVGPGEKK*.Q

Global Analysis of *Shewanella oneidensis* Strain MR-1 Proteome Using Accurate Mass Tags

MS Lipton,[(a)] DA Elias,[(a)] MF Romine,[(a)] YA Gorby,[(a)] J Fredrickson,[(a)] and RD Smith[(a)]
(a) Pacific Northwest National Laboratory, Richland, Washington

Shewanella oneidensis strain MR-1 has potential applications in the bioremediation of metals due to its ability to enzymatically reduce and precipitate a diverse range of heavy metals and radionuclides. A thorough understanding of how MR-1 responds to changes in electron acceptor type and concentration and the enzymatic pathways involved in these reactions is critical for effectively using metal-reducing bacteria for remediation. A critical step in this process is being able to characterize the proteome of MR-1, or the entire protein complement of the cell, expressed under a given set of conditions. The ability to conduct these analyses in a high-throughput, comprehensive, and quantitative mode provides a powerful means for defining regulatory and metabolic networks in bacteria. The protein complement from a single microbial genome can vary significantly as a function of cell-cycle stage, cell differentiation, environmental conditions (e.g., nutrients, temperature, stress), and even association with other organisms.

We have used a new technology based on the combination of global tryptic digestion of cellular proteins, high-resolution liquid chromatography and tandem mass spectrometry, and high-field Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) to define a proteome of an organism. One approach for protein identification based upon global approaches for protein digestion and accurate mass analysis has resulted in the generation of an accurate-mass-and-time (AMT) tag database for *S. oneidensis*. Using this approach, we have identified over 60% of the proteome for *S. oneidensis*. Although we have not detected a large percentage of the proteins coded on the plasmid, the majority of the proteins are coded from the chromosome, and we were able to detect a high percentage of these proteins.

We detected proteins spanning all functional categories, as shown in Figure 1. The majority of these proteins, not surprisingly, were housekeeping proteins including functional categories such as the synthesis of nucleotides and nucleosides, protein synthesis, transcription, and metabolism. Much less represented are the conserved hypothetical and the hypothetical proteins, although this finding is much less surprising because these proteins are only predicted to be expressed with no homology to any other known proteins. In an effort to analyze post-translational modifications, the analysis of tryptic peptides was combined with analysis of intact proteins.

Using highly stringent criteria for peptide identification, we analyzed proteome data generated from mass spectrometry analysis of 172 tryptic digests of *S. oneidensis* MR-1 proteins for the occurrence of peptides that were associated with proteins less than 101 amino acids in length or that were added to the genome annotation after its initial deposit in Genbank. Using stringent criteria for acceptance, 51 proteins less than 101 amino acids in length were identified. Three were associated with recent additions to the Genbank annotation and two were among those removed from the original annotation. In addition, 27 peptides were detected from proteins added in the most recent annotation that were greater than

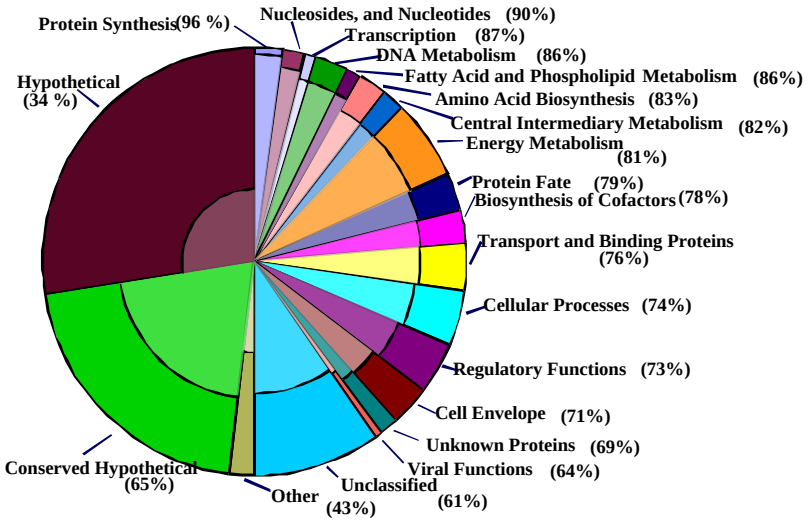

Figure 1. Detected proteins sorted by functional category.

101 amino acids in length. The number of newly predicted proteins that were validated was 31 out of a total 109. We also identified peptides that mapped to positions downstream to proposed degeneracies in SO0101, SO0419, SO0561, SO0590, SO0738, SO1113, SO1211, SO3079, SO3130, SO3240, SO4231, SO4328, SO4422, and SO4657, genes that were identified by The Institute for Genetic Research (TIGR) but not deposited in Genbank. These results demonstrate the utility of AMT tag methodology for validation of protein-encoding genes predicted by *in silico* analysis of genome sequences, including those recoded during translation to enable full-length proteins to be expressed.

For example, SO4494.1 overlaps with SO4493, one of the genes not deposited in Genbank by TIGR, but described as a conserved hypothetical with an authentic frameshift at colossal magnetorresistance (CMR). BlastP analysis of our repaired version of SO4493 suggests that this protein has greater than 60% identity with the MR-1 ISSod10 transposase OrfB protein (185 amino acids in length). The absence of any detectable unique peptides for SO4493 and the presence instead of two peptides unique to SO4494.1 (Figure 2) suggests that the truncated form of this transposase-like protein was expressed in the samples we processed.

```
SO4532    AYLFGALCPATGATEAILAPFANSDYMHEHLKLISAATEFGRHALVIMDG
            |||| ||||| ||||  || | | || || ||   |  |||| | ||
SO4493    MHLF GAVCPATGETEAIITPFVNK DIMHQHLELIAKRTKPGRHAVVVMDV
SO4494.1 MHLF GAVCPATGETEAIITPFVNK DIMHQHLELIAK RTKPGRHAVVVMDV

SO4532    AGWHQRDLADDFDNLSIVKIPPYSPELNPMEQVWQWLRQHVLANRSFKGY
          ||||  | | |  |||| | ||||||||| |||| |||||||||| |||
SO4493    AGWHTNDIAADIPNLSILKLPPYSPELNPIEQVWSWLRQHYLANRCFKGD
SO4494.1  RVGIPMTLLQISPTCRYSNCR HTPQN

SO4532    DDIVEQCSITWNTFIKEPLRVMQLCSRQWAKMNS
            |||  |    | ||    ||   ||| |||
SO4493    EDIVDACNQALNSFISCTKRVIKMCSREWAKVTEL
```

Figure 2. Alignment of SO4532 and SO4494.1 to SO4493. SO4532 encodes ISSod10 transposase OrfB (the N-terminal 51 residues are not shown). The protein sequence of SO4493 was derived by frameshifting at the A residue shown in bold. The regions encoding peptides identified by mass spectrometry are GAVCPATGETEAIITPFVNK, PGRHAVVVMDVRVGIPMTL, and RTKPGRHAVVVMDVRVGIPMTLLQISPTCRYSNCR.

Duplex Foldamers from Assembly-Induced Folding

B Gong,[(a)] X Yang,[(a)] H Zeng,[(a)] S Ahn,[(b)] and S Martinovic[(c)]
(a) State University of New York, Buffalo, New York
(b) Pacific Northwest National Laboratory, Richland, Washington
(c) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

Molecular duplexes formed by the combination of sequence-specific self assembly and folding of the component strands were designed, synthesized, and verified.

Evidence from our experiments indicates that an initial self-assembly process is followed by the folding of the component strands, which leads to the final folded, dimeric species (Figure 1) (Yang et al. 2003; Zeng et al. 2003). Such assembly-induced folding systems are reminiscent of biomacromolecules such as duplex DNA and DNA-recognizing proteins. The assembled and folded structures are stabilized by both H-bonding and aromatic stacking interactions. The modular nature of this system enables convenient design of longer oligomers that should assemble and fold into duplex foldamers with multiple stacks of H-bonded units. Similar to the stacking of Watson-Crick base pairs in duplex DNA, the stacking of the H-bonded units in our duplex foldamers not only provides additional stabilization to the assembled structure but also shields the H-bonds from solvent molecules (in our case chloroform). On the basis of this novel assembling and folding motif, a wide variety of sequence-specific, self-assembled duplex foldamers that are stable in competitive media, such as aqueous solution, should become available in the near future.

Figure 1. Illustration of molecular duplex formation including a combination of sequence-specific self assembly and folding of the component strands.

This project involves the characterization of a new class of sequence-specific, H-bonded organic molecular duplexes. Two oligoamide molecular strands with complementary hydrogen-bonding sequences were found to form stable molecular duplexes in chloroform. These H-bonded duplexes may eventually be developed into a whole set of programmable molecular recognition units. Our purpose in this project is to demonstrate specificity within the given system.

Figure 2 shows an example of duplex foldamers. Monomer A and monomer B form eight attractive hydrogen-bonding sites, thereby creating a stable duplex.

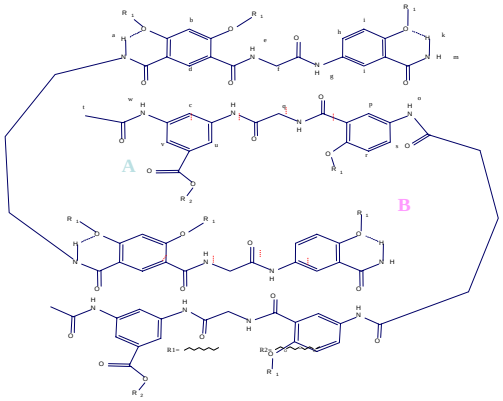

Figure 2. Example of the structure of the molecular duplex analyzed in this project.

Electrospray ionization (ESI) is a soft technique that can ionize intact non-covalent complexes. Complexes of H-bonded duplexes were detected using electrospray ionization from chloroform. The intact complexes were observed in the negative-ion mode as a Cl^--bound species using Ph_4PCl as the negative charge donor. The soft

ionization capability of ESI allowed the use of Fourier transform infrared cyclotron resonance mass spectrometry (FTICR MS) to determine the molecular weight of the ionized complexes. The crucial question to be answered is whether the complex observed represents specific binding in solution prior to ionization or a non-specific aggregation of molecules that occurred after ionization in the gas phase. By working at low concentrations of the molecular species, we can eliminate the possibility of non-specific binding. In addition, the high resolution of the FTICR MS instrument enables us to determine the degree of completion of the reaction by separating the specific dimer from the non-specific tetramer. Since the m/z values for the (-1) charge state of the dimer is the same as the (-2) charge state of the tetramer, only the very high resolution of the FTICR spectrometer allows the isotopic peaks to be resolved in order to differentiate the specific from non-specific complexes as shown in Figure 3.

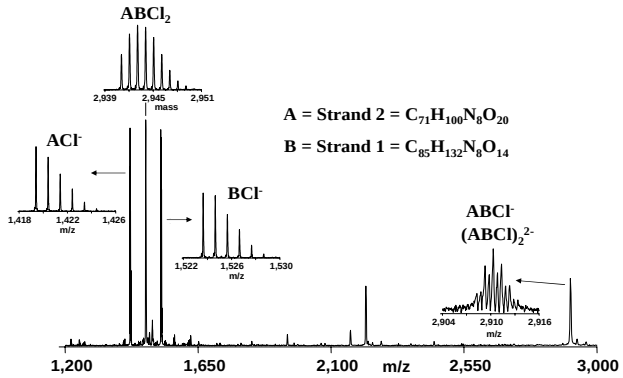

Figure 3. Negative ion mode ESI-FTICR mass spectrum of duplex AB (structure in Figure 2) sprayed from chloroform.

References

Yang X, S Martinovic, RD Smith, and B Gong. 2003. "Duplex Foldamers from Assembly-Induced Folding." *Journal of the American Chemical Society* 125:9932-9933.

Zeng H, X Yang, AL Brown, S Martinovic, RD Smith, and B Gong. 2003. "An Extremely Stable, Self-Complementary, Hydrogen-Bonded Duplex." *Chemical Communications* 13:1556-1557.

Identification of the Labeled Residue(s) of SecB Protein

LL Randall,[(a)] J Crane,[(a)] and S Martinovic[(b)]
(a) University of Missouri, Columbia, Missouri
(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

SecB, a small tetrameric cytosolic chaperone in *Escherichia coli*, facilitates the export of precursor polypeptides by maintaining them in a non-native conformation and passing them to SecA, which is a peripheral member of the membrane-bound translocation apparatus. To elucidate the interaction between SecB and the unfolded precursor, we are mapping the molecular path that the precursor takes when it binds SecB. The structure of protein SecB is shown in Figure 1. The use of site-directed mutagenesis to systematically replace single residues with cysteine coupled with electron paramagnetic resonance (EPR) spectroscopy allows scanning through SecB to identify sites of interaction. EPR spectroscopy requires an unpaired electron, which we provide by labeling the introduced cysteinyl residue in SecB with methanethiosulfonate (MTSL), a stable nitroxide spin label. Ideally, in our experimental approach, we would remove all native cysteines and introduce a single cysteine at the site of interest. However, we have been unable to replace two of the four native cysteines and still maintain the protein in an active form. Thus, each of our constructs has three cysteines. It is imperative that we know confidently that we quantitatively label the introduced cysteine and not the native cysteines.

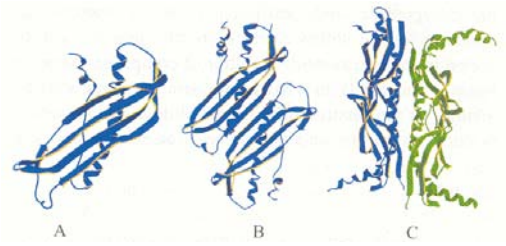

Figure 1. Structure of the *Hemophilus influenzae* SecB protein. Ribbon representations of (A) the monomer, (B) dimer, and (C) tetramer, where the view is orthogonal to the view of the dimer.

Mass spectrometry is recognized as the technique of choice for identifying and quantifying amino acid modifications in proteins. These are challenging experiments, but they are the only means of unambiguously determining if the targeted cysteine residue has been labeled correctly. To confidently identify the labeled residue position, a series of experiments are underway using the advanced instrumentation in the W.R. Wiley Environmental Molecular Sciences Laboratory. In the first experiment, the molecular mass of intact proteins was measured on a 7-tesla Fourier transform ion cyclotron resonance mass spectrometer (FTICR MS) in order to confirm the number of labels per protein. This approach is a quick means of determining that only one residue has been labeled and that the molecular weight of the label is correct. An example of this process is represented by the spectra in Figure 2. The mass spectrometer detects a series of multiply protonated versions of the

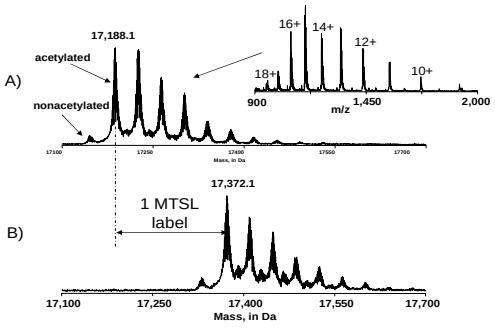

Figure 2. Zero charge ESI FTICR mass spectra of SecB protein: (A) wild type and (B) MTSL-labeled SecB. The inset (upper right) shows charge state distribution in the mass spectrum that was deconvoluted to (A).

protein as shown in the inset in the figure. To get a quantitative representation, each one of these charge-state spectra must be deconvoluted to the molecular weight of the uncharged protein (zero charge state). The resulting zero-charge spectra in the figure show the MTSL-labeled and non-MTSL versions of the SecB protein and easily allow the number of labels per protein to be determined.

In subsequent experiments, proteins were digested by trypsin to the peptide level. Liquid chromatography-mass spectroscopy (LC MS) and liquid chromatography-tandem mass spectroscopy (LC MS/MS) experiments on a liquid chromograph interfaced with an ion trap mass analyzer and a 9.4-tesla FTICR MS enabled identification of peptides and the labeled residue. Results covered 100% of the protein sequences in all analyzed proteins (that have replaced single residue) and showed which cysteine residue(s) were modified.

Determination of the Proteome of *Drosophila* Lipid Droplets

SP Gross,[(a)] S Cermelli,[(a)] and RJ Moore[(b)]

(a) University of California Irvine, Irvine, California

(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

In the eukaryotic cell, many cargos such as mitochondria, mRNA particles, endosomes, neuronal vesicles, pigment granules, virus particles, and lipid droplets move in a bi-directional manner along microtubules. It has been determined that the opposite-polarity molecular motors kinesin and dynein are responsible for this motion. It is known that the cell can control the average direction of cargo transport, and we assume that this is accomplished through a combination of regulatory and accessory proteins and post-translational protein modification. The regulation and details of such transport are not yet understood in any system. To gain some insights into how such transport is regulated at a molecular level, we are using a proteomic approach to investigate lipid droplet transport.

Using mass spectrometry, the complete proteome of the lipid droplets has been characterized. Initially, analysis of the first set of samples we provided revealed massive mitochondrial contamination of our lipid droplets preparation, so we developed a new purification protocol. The second set of samples substantially eliminated the contamination, and identified proteins associated with isolated lipid droplets. The proteins identified can be subdivided into the following general functional groups:

1. *Lipid Droplet Associated.* The presence of the proteins lipid storage droplets 1 and 2, well-known markers of *Drosophila* lipid droplets, validates the accuracy of the purification and proteomic techniques.

2. *Lipid Metabolism.* Several enzymes related to lipid and fatty acid metabolism have been found (long-fatty acyl-CoA ligase and dehydrogenase, acyl-CoA thioesterase, very-long-chain-acyl-CoA dehydrogenase, palmitoyl-protein hydrolase, alcohol dehydrogenase, fatty acid amide hydrolase, glycolipid mannosyltransferse, triacylglycerol lipase); other non-enzymatic proteins involved in fatty acids transport and metabolism are present [retinol and fatty acid binding proteins (RBP and FABP) and long-chain fatty acid transporter].

3. *Membrane Traffic.* Rho GTPase (Rab 1, 2, 5, 6, 7, 8, 9, 11, 35, snap receptor, synaptobrevin.

4. *Signaling Transduction.* Component of signaling cascade of protein kinase A and C, phospholipase A and C, protein serine/threonine kinase, 14-3-3 protein. Components of different signal-transduction pathways: Ras cascade, JNK cascade, protein kinase (PKA and PKC), Rho GTPase (Rab1, Rab5 and Rab 11) proteins.

5. *Motor Proteins and Related Proteins.* Microtubule and cytoskeletal binding proteins (dynein ATPase, myosin ATPase).

With this relatively complete proteome, we are trying to integrate this information with our other approaches and, in particular, to validate the functional relevance of a number of the more interesting proteins identified. Of particular interest is the lipid storage droplet 2 protein. We have made mutant flies that are lacking lipid storage droplet 2, which are surprisingly viable, though not healthy. Indeed, lipid droplet motion is severely impaired. We are currently doing a biophysical analysis to better understand the alteration of motion. In parallel, we have been trying to better understand the role of lipid storage droplet 2 by identifying interaction partners. We have done pull-downs using the lipid storage droplet 2 antibody, and in collaboration with researchers at the W.R. Wiley Environmental Molecular Sciences Laboratory, we are trying to identify the proteins that co-precipitate. These efforts are ongoing.

Recent papers on mammalian lipid droplets have highlighted the importance of a complete study of this peculiar organelle for lipid metabolism and related diseases. As a model organism, *Drosophila* has previously been an ideal system in which to identify molecules and define pathways involved in development, in part because of the powerful genetic approaches that are possible. Many of the molecules and pathways important in development in *Drosophila* are evolutionarily conserved between flies and humans. The same strengths that have made *Drosophila* such a useful model organism for increasing our understanding of mammalian developmental processes will also be relevant for unraveling the intricacies of molecular-motor-based cargo transport in general, and specifically lipid-droplet transport. The already-determined lipid droplet proteome and the ongoing efforts to define protein-protein interactions promise to tremendously increase our understanding of the transport of these vital cargos.

Investigation of Macrophage Activation in Response to Nanoscale Biomaterial Surface Features

A Golden,[a] HM Mottaz,[b] WJ Shaw,[c] MS Lipton,[c] and P Stayton[a]

(a) University of Washington, Seattle, Washington

(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

(c) Pacific Northwest National Laboratory, Richland, Washington

The control of the foreign body reaction to implanted biomaterials and tissue engineering scaffolds is a central aspect of biocompatibility. The design of biomaterial coatings that minimize chronic inflammation, yet encourage healing responses, is a key goal in the medical device field. Recent findings by researchers at the University of Washington and other laboratories have shown that materials with nanoscale fibrous features exhibit favorable *in vivo* biological responses. Underlying this response are differences in macrophage biology that are related to biomaterial architecture. Surface topography at nano- to meso-scales may be a major influence in the biocompatibility of a material. With advances in genomic and proteomic technologies, it is now possible to study in unprecedented detail the pro-inflammatory activation of macrophages as a function of the nanoscale features of biomaterials. This new mechanistic detail will identify common architectural properties that can be incorporated into biomaterial surfaces. The identification of the key signaling pathway components associated with the response to nanoscale features will also provide targets for the design of drug delivery coatings from biomaterials that can augment the surface architecture effect. This collaborative project combines the biomaterials expertise at the University of Washington with the proteomics expertise at Pacific Northwest National Laboratory (PNNL) to study the pro-inflammatory response of macrophages to materials with nanoscale surface features. The response of human primary macrophages on polytetrafluoroethylene (PTFE) materials that are topographically distinct, and well under 100 nm, will be studied using advanced proteomic techniques that employ mass spectrometry. Tandem mass spectrometry and Fourier transform ion cyclotron resonance (FTICR) mass spectrometry will be used to study signaling pathway activation as monitored by phosphorylation activation of pro-inflammatory pathway components. We aim to characterize dynamic signaling pathways in macrophage inflammatory responses to better understand and control the effect of material nanostructure.

Changes in chemical makeup or changes in the surface topography control the rapid responses of macrophages. In our test system, a well-characterized soluble inflammatory agent, lipopolysaccharide (LPS), is being used so that the method can be optimized around known signaling events. Following this, we will subject macrophages to a system where the surface chemistry is kept the same, while nanoscale features are varied. PTFE has been extensively used in implantable materials. A fibrous variety of PTFE studied at the University of Washington has been shown to elicit a more desirable encapsulation response, although the exact mechanism is yet to be understood. A comprehensive analysis of signaling pathway activation at the proteomics level will allow the critical responses to surface topography to be identified. Signaling proteome results will be compared with existing data of cytokine production, RNA expression, and *in vivo* response to all types of PTFE. Since signaling pathways are dynamic, multiple time points will be gathered to map the responses temporally. Through this approach, desirable outcomes, such as rapid healing and

progressive angiogenesis, could be separated at the level of signaling pathway activation from those pathway components connected with undesirable chronic inflammation—as a function of nanoscale topographical features in the PTFE materials.

High-resolution protein identification is achieved by combining information from capillary liquid chromatography and tandem mass spectrometry with electrospray ionizaton (ESI). In cases where sample quantity is limited, as with many human-derived samples, FTICR mass spectrometry is used so that detection of sub-picomolar quantities of a protein is possible. The ESI spectrum records phosphoric acid loss that occurs during fragmentation, providing an indication of phosphorylations from serine, threonine, and tyrosine residues. With the above methods, identification not only of the peptide's parent protein, but also the peptide's phosphorylation states is possible. Quantitative measurements of relative abundances and phosphorylation states require simultaneous analysis of the samples to be compared. Using primary cells from human donors requires that the proteins be labeled after isolation. A new method incorporates a stable oxygen isotope, ^{18}O, at the end of peptides of one sample during standard systematic protein cleavage, so that two samples can be run concurrently. Relative abundance calls can thus be achieved, similar in approach to previous comparative proteomic studies done with ^{14}N and ^{15}N by PNNL. Protein samples from exposed and control macrophages will be isolated and purified, and pairs will be distinguished by stable oxygen isotope labeling to obtain relative abundances and phosphorylation patterns present in exposed versus control macrophages. Control peptides with known phosphorylation sites will be included to validate phosphorylation calls made by software.

Currently, we have established protocols that allow consistent recovery of phosphoproteins from human macrophages isolated from donor blood samples. The primary macrophages can be reliably stimulated using LPS to initiate phosphorylation cascades. Western analysis is being used as a companion method for verification of the phosphorylation of selected signaling proteins, and enzyme-linked immunosorbent assay has been used to verify endpoints of signaling cascade activation. Since key phosphorylation events that can change the course of cell development may occur on only about 5% of a pool of a given signaling protein, it is crucial to enrich the samples for phosphorylations so that the high abundance of non-phosphorylated proteins does not mask the signal. While serine and threonine phosphorylations are more constitutive, the small percentage of key phosphorylation events on signaling proteins typically involve tyrosine residues that act as switches between active and inactive states as they are differentially phosphorylated. We are in the process of comparing the data from samples enriched using immuno-precipitation by an anti-phosphotyrosine antibody and from samples enriched by immobilized metal affinity columns that bind the negatively charged phosphate groups.

The results from this study will contribute to the future design of biomaterials and, in particular, will elucidate the cellular response to sub-micron topographies, an emerging field that will be critical in the design of small-sample laboratory units, single-cell analysis, and tissue engineering. In subsequent years, the project is designed to incorporate surface modification to combine the surface architecture aspect with rapid biological signal display.

Characterization of the Stress Responses of *Deinococcus radiodurans*

MS Lipton,[(a)] HM Mottaz,[(b)] and AK Schmid[(c)]

(a) Pacific Northwest National Laboratory, Richland, Washington
(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(c) University of Washington, Seattle, Washington

Deinococcus radiodurans is a non-pathogenic bacterium that can withstand high levels of γ-irradiation (up to 5 Mrad), ultraviolet light, and desiccation (Battista 1997). The *D. radiodurans* strain R1 is a gram-positive non-pathogenic bacterium with a 3.1 Mbase genome that consists of two chromosomes, one megaplasmid, and one plasmid. The genome has been sequenced and annotated (White et al. 1999), and has sparked intensive investigation into the radioresistance mechanisms of the organism. This work used the 3116 protein-encoding open reading frames (ORFs) predicted by The Institute for Genomic Research (TIGR) annotation (ftp://ftp.tigr.org/pub/data/dradiodurans/GDR.pep). We excluded 71 ORFs predicted to contain frame shifts from this analysis. However, the annotation consists of the best prediction of ORFs encoded by an organism based on comparison to ORFs identified in other organisms and on the codon usage. The proteomic measurements provided a physical validation that the ORFs actually encode a protein. Despite this interest, little is known about how the organism responds to other environmental stresses.

To gain a better understanding of how *D. radiodurans* interacts with its environment, we undertook a study of stress responses and regulation of these responses in *D. radiodurans*. As an initial study, we focused on the response of *D. radiodurans* to heat shock, a stress/response system that has been well characterized in other bacteria. We conducted mutagenic analysis of two alternative sigma-factor genes of the organism, designated *sig1* and *sig2,* and found that both are important in protection against heat shock, although *sig1* appeared to be more important in protection than *sig2* (Schmid and Lidstrom 2002). In addition, *sig1* positively regulates the heat shock induction of the classical heat shock operons, *groESL* and *dnaKJ*. To investigate whether *sig1* also is involved in the regulation of other heat shock genes, we undertook a global proteomics investigation of wild type cells compared to *sig1* mutant cells under heat shock. Whole cell lysates of the proteins for both the wild type and *sig1* mutant were prepared for liquid chromatograph-mass spectrometer (LC MS) analysis on the Fourier transform infrared cyclotron resonance mass spectrometer (FTICR MS). Both the soluble fraction and the membrane fraction were collected and digested with trypsin. By using the previously created databases of peptides found in *D. radiodurans*, the peptides present in each sample can be determined in a single LC MS analysis on the FTICR MS. This permits replicate analysis to assess reproducibility and studies of different environmental conditions to be conducted efficiently. Preliminary analysis of the results of these global proteomics studies have indicated several novel proteins whose up or down regulation has been affected by the absence of the *sig1* factor. Continuing studies are underway to confirm the protein assignments and to quantify the changes in protein expression.

References

Battista J. 1997. "Against All Odds: The Survival Strategies of *Deinococcus radiodurans*." *Annual Review of Microbiology* 51:203-24.

Schmid A and ME Lidstrom. 2002. "Involvement of Two Putative Alternative Sigma Factors in the Stress Response of *Deinococcus radiodurans*." *Journal of Bacteriology* 184(22):6182-6189.

White O, JA Eisen, JF Heidelberg, EK Hickey, JD Peterson, RJ Dodson, DH Haft, ML Gwinn, WC Nelson, DL Richardson, KS Moffat, H Qin, L Jiang, W Pamphile, M Crosby, M Shen, JJ Vamathevan, P Lam, L McDonald, T Utterback, C Zalewski, KS Makarova, L Aravind, MJ Daly, and CM Fraser. 1999. "Genome Sequence of the Radioresistant Bacterium *Deinococcus radiodurans* R1." *Science* 286(5444):1571-1577.

Proteomic Analysis of Wild Type and Soil Colonization Deficient Strains of *Pseudomonas fluorescens*

SB Levy,[(a)] HM Mottaz,[(b)] MS Lipton,[(c)] I Lopez-Hernadez,[(a)] and EA Robleto[(a)]

(a) Tufts University School of Medicine, Boston, Massachusetts

(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

(c) Pacific Northwest National Laboratory, Richland, Washington

Understanding the ecology of release, growth, and establishment of populations in soil would be a significant step toward developing effective genetically modified microbes for bioremediation purposes. In *Pseudomonas fluorescens*, a microbe widely used for enhancing soil characteristics, we have identified a locus (AdnA) that affects several cell processes important for its behavior in soil. Mutants deficient in AdnA are defective in biofilm formation, motility, and attachment to sand and seeds. Of more importance, mutants affected in AdnA do not colonize and persist in soil as well as the wild type (Marshall et al. 2001). AdnA shows homology to FleQ, from *Pseudomonas aeruginosa*, a transcription factor of the NtrC/NifA family that is involved in flagella synthesis (Casaz et al. 2001). This type of regulator is Sigma 54-dependent and is known to activate transcription upon changing environmental conditions. Therefore, by identifying traits affected by AdnA, we expect to gain insights into how to improve soil colonization, persistence, and genetic determinants affecting soil colonization by comparing proteomes from the wild type and an AdnA-deficient mutant. Recently, we began to characterize the regulon controlled by AdnA using a genetic approach (Robleto et al. 2003).

To develop a more complete picture of the genes controlled by AdnA, we aim to compare the genes expressed in wild type *P. fluorescens* (Pf0-1) with those expressed in an AdnA-deficient mutant (Pf0-2x). Genetic evidence shows that AdnA-dependent genes are expressed at significantly lower levels in Pf0-2x than Pf0-1 (Robleto et al. 2003). Thus, protein extracts from Pf0-2x are likely to contain no activated peptides and/or peptides that are repressed by the AdnA transcription factor. Proteomic technologies offer promise in this kind of approach due to the broad coverage now being obtained, and the fact that translated products are compared, rather than just transcribed products revealed using microarray technology. In combination, microarray and proteomic approaches will allow a full description of genes under the control of AdnA.

Wild type (Pf0-1) and mutant (Pf0-2x) *P. fluorescens* were grown under conditions known to favor expression of AdnA-dependant genes in the wild type strain. Proteins have been extracted from cytoplasmic extracts, membrane fractions, and whole cells, and then subjected to trypsin digestion. Using Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS), the peptides are being individually identified, and by computational means, they have been compared to the draft genome sequence of *P. fluorescens*. This approach has enabled us to identify 763 and 1271 expressed genes in Pf0-1 and Pf0-2x, respectively, using peptides found in cytoplasmic extracts. Peptides found in membrane fractions revealed 437 expressed genes in Pf0-1 and 290 in Pf0-2x. Comparison of the data sets shows that, of the cytoplasmic proteins, 451 present in the Pf0-1 extract are absent from Pf0-2x. Among membrane proteins, 318 Pf0-1 proteins are absent from Pf0-2x. These data are highly likely to overestimate the number of proteins whose expression is governed by

AdnA, probably because of incomplete coverage of the proteome at this time. However, examination of the proteins absent from Pf0-2x will allow some candidate AdnA-regulated genes to be identified. It is known that AdnA regulates motility, and the proteomic data point to a number of flagella and chemotaxis proteins not previously known to be part of the AdnA regulon. For example, seven methyl-accepting chemotaxis proteins are unique to the membrane fraction of Pf0-1. Our previous genetic study only identified two proteins (Robleto et al. 2003).

While the AdnA regulon remains to be completely defined, an additional benefit from the proteomic data is the confirmation that a number of previously annotated "hypothetical" genes are indeed expressed (e.g., 55 hypotheticals from the membrane of Pf0-1). Moreover, proteins with no matches in databases have also been shown to be expressed (e.g., 84 from the Pf0-1 membrane fraction). These data will be used to aid in the thorough annotation of the genome sequence of Pf0-1 when it is completed.

References

Casaz P, A Happel, J Keithan, DL Read, SR Strain, and SB Levy. 2001. "The *Pseudomonas fluorescens* Transcription Activator *Adna* is Required for Adhesion and Motility." *Microbiology* 147:355-361.

Marshall B, EA Robleto, R Wetzler, P Kulle, P Casaz, and SB Levy. 2001. "The *Adna* Transcriptional Factor Affects Persistence and Spread of *Pseudomonas fluorescens* under Natural Field Conditions." *Applied Environmental Microbiology* 67:852-857.

Robleto EA, I Lopez-Hernandez, MW Silby, and SB Levy. 2003. "Genetic Analysis of the *Adna* Regulon in *Pseudomonas fluorescens*: Nonessential Role of Flagella in Adhesion to Sand and Biofilm Formation." *Journal of Bacteriology* 185:453-460.

User Projects

Determination of an Unknown Protein Implicated in Radiation Resistance in *Deinococcus radiodurans*

JR Battista, AM Earl
Louisiana State University, Baton Rouge, Louisiana
MS Lipton
Pacific Northwest National Laboratory, Richland, Washington

Molecular Analysis of Cancer by FTICR

BD Thrall
Pacific Northwest National Laboratory, Richland, Washington

Use of Mass Spectral Disassembly to Elucidate the Quaternary Structure of the Complex Heterotetrameric Sarcosine Oxidase

MS Jorns
MCP Hahnemann University, Philadelphia, Pennsylvania

Whole Genome Cloning and Expression of *Treponema pallidum* Open Reading Frames

TG Palzkill
Baylor College of Medicine, Houston, Texas

Isolation and Characterization of Novel Antimicrobial Peptides from the Zebrafish, *Danio rerio*

CH Kim, C Sullivan
University of Maine, Orono, Maine

Proteomic Analysis of Wild Type and Soil—Colonization Deficient Strains of *Pseudomonas fluorescents*

SB Levy, I Lopez-Hernadez, EA Robleto
Tufts University School of Medicine, Boston, Massachusetts
MS Lipton
Pacific Northwest National Laboratory, Richland, Washington

Comparative Display of *Deinococcus radiodurans* After Exposure to Ionizing Radiation

M Daly
Uniformed Services University of Health Sciences, Bethesda, Maryland

Proteomic Characterization of Subproteome of *Shewanella oneidensis*

MF Romine
Pacific Northwest National Laboratory, Richland, Washington

Functional Genomics and Proteomics of Mitochondria
W Xiao, CM Scharfe, RW Davis
Stanford University, Stanford, California

Study of Secretory and Membrane Proteins of *Pseudomonas aeruginosa*
W Xiao, M Mindrinos, RW Davis
Stanford University, Stanford, California

Accurate Mass Determination
ER Lacy
Medical University of South Carolina, Charleston, South Carolina

Hollow Helices as Folding Nanotubes with Tunable Cavity Size
B Gong
State University of New York at Buffalo (SUNY), Buffalo, New York

Determination of the Proteome of *Drosophila* Lipid Droplets
SP Gross, Je Martinez
University of California Irvine, Irvine, California

ICAT Labeling of *Paeruginosa* Membrane Subproteome
W Xiao, RW Davis
Stanford University, Stanford, California

Processing of the Cysteine Protease Aleurain
JC Rogers, CE Halls
Washington State University, Pullman, Washington

Mechanism of Action of G Protein-Coupled Receptors (GPCR) Studied by LC-FTICR Mass Spectrometry
EA Dratz, P Kraft
Montana State University, Bozeman, Montana

Identification of Covalent Modification of SecA
LL Randall
University of Missouri, Columbia, Missouri

Identification of Heat Shock Proteins in the Radioresistant Bacterium *Deinococcus radiodurans*
AK Schmid
University of Washington, Seattle, Washington

Identification of Proteins Differentially Expressed in Response IR
MA Coleman
Lawrence Livermore National Laboratory, Livermore, California

Proteome Analysis of a Breast Carcinoma Cell Line Overexpressing the HER2/neu Oncogene
K Williams, M Pallavicini
University of California San Francisco, San Francisco, California

Protein Signatures of Leukemic Cells
K Williams
University of California San Francisco, San Francisco, California

Identification and Relative Expression of Membrane Proteins in Breast Cancer Cell Lines
AJ Patwardhan, K Williams, M Pallavicini
University of California San Francisco, San Francisco, California
EF Strittmatter, R Smith, L Pasa-Tolic, DG Camp II
Pacific Northwest National Laboratory, Richland, Washington

Identification of a Novel Biomolecule in Deep-Sea Clams
PH Yancey
Whitman College, Walla Walla, Washington

Posttranslational Modifications of Chicken Brain Myelin Basic Protein Charge Isomers
R Zand
University of Michigan, Ann Arbor, Michigan

NCI-II (Development of a Global Accurate Mass and Time [AMT] Tag Database)
DG Camp II
Pacific Northwest National Laboratory, Richland, Washington

Determination of the C-Terminus of p110 sEGFR
RC Zangar
Pacific Northwest National Laboratory, Richland, Washington

Alpha Project (Functional Genomics)
O Resnekov
Molecular Sciences Institute, Berkeley, California
DG Camp II
Pacific Northwest National Laboratory, Richland, Washington

NCI-III (Process and Analyze Cys-containing Polypeptides and Phosphopeptides)
DG Camp II
Pacific Northwest National Laboratory, Richland, Washington

Proteomic Characterization of *Yersinia pestis*

MS Lipton
Pacific Northwest National Laboratory, Richland, Washington
SL McCutchen-Maloney
Lawrence Livermore National Laboratory, Livermore, California

Biomolecular Mechanisms for Microbe-Fe (III) Oxide Interactions in Geobacter Species

TJ Beveridge
University of Guelph, Guelph, Ontario, Canada
MS Lipton, DR Lovley, SE Childers
Pacific Northwest National Laboratory, Richland, Washington

Deinococcus radiodurans

MS Lipton
Pacific Northwest National Laboratory, Richland, Washington

Shewanella oneidensis

MS Lipton, JK Fredrickson
Pacific Northwest National Laboratory, Richland, Washington

Low Dose (Isolating Protein Complexes)

MS Lipton, C Bruckner-Lea
Pacific Northwest National Laboratory, Richland, Washington

Work with the Lyme Disease Spirochete Causative Agent *(Borrelia burgdorferi)* and Closely Related Strains

X Yang
State University of New York at Stony Brook, Stony Brook, New York
DG Camp II
Pacific Northwest National Laboratory, Richland, Washington

Rhodobacter sphaerodies

T Donohue
University of Wisconsin-Madison, Madison, Wisconsin
MS Lipton
Pacific Northwest National Laboratory, Richland, Washington

Proteomics of Morphology Determination in a Fungus

EA Panisko, JK Magnuson, LL Lasure
Pacific Northwest National Laboratory, Richland, Washington

Regulated Proteolysis and Bystander Effects in Radiation

DL Springer
Pacific Northwest National Laboratory, Richland, Washington

Global Measurements of Phosphopeptides in Human Mammary Epithelial Cells
RC Zangar
Pacific Northwest National Laboratory, Richland, Washington

Preliminary Work on the Proteomes of Brains and Dissected Brains Obtained from Control Mice and Treated Mice Simulating Parkinson's Disease
DJ Smith, DM Sforza
University of California, Los Angeles, California
DG Camp II
Pacific Northwest National Laboratory, Richland, Washington

Molecular Dysfunction in Alcoholic Liver Disease
A Banerjee
Wayne State University, Detroit, Michigan

Engineering Device Surfaces that Instruct Bell Behavior
P Stayton, A Golden
University of Washington, Seattle, Washington
WJ Shaw
Pacific Northwest National Laboratory, Richland, Washington

Abundance of Protein Components in Photosystem II Protein Complex Purified from Mutant Cyanobacterial Cells Lacking Individual Protein
HB Pakrasi
Washington University in St. Louis, St. Louis, Missouri

Using Mass Spectrometry to Assist in Identifying Genes for Enzymes of Known Activity, Unknown Open Reading Frame and Small Abundance
DH Kohl
Washington University in St. Louis, St. Louis, Missouri

Determination of the Peptide-Binding Sites on the Bacterial Chaperone Protein Secb
LL Randall
University of Missouri, Columbia, Missouri
VF Smith
U.S. Naval Academy, Annapolis, Maryland

Mass Spectrometric Analysis of Eukaryotic Proteasome
RA Maxwell
Pacific Northwest National Laboratory, Richland, Washington

Cellular Response to Hepatitis C Virus Infection: Global Quantitative Proteome AMT Tag Measurements of Cellular Protein Express
MG Katze, DL Diamond, EY Chan
University of Washington, Seattle, Washington

Staff

Harold R. Udseth, Staff Scientist, Technical Lead
(509) 376-3698, harold.udseth@pnl.gov

Kimberly S. Korenkiewicz, Administrative Secretary
(509) 373-0765, kimberly.korenkiewicz@pnl.gov

David J. Anderson, Scientist
(509) 376-7898, david.anderson@pnl.gov

Kim K. Hixson, Research Scientist
(509) 373-6170, kim.hixson@pnl.gov

Suzana Martinović, Senior Research Scientist
(509) 376-3709, suzana.martinovic@pnl.gov

Ronald J. Moore, Technologist
(509) 376-2721, ronald.moore@pnl.gov

Heather M. Mottaz, Scientist
(509) 376-4869, heather.mottaz@pnl.gov

Rhui Zhao, Research Scientist
(509) 376-7332, rui.zhao@pnl.gov