MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Mar. 2009, p. 62–70 Vol. 73, No. 1
1092-2172/09/$08.00+0 doi:10.1128/MMBR.00028-08

Environmental Proteomics: a Paradigm Shift in Characterizing Microbial Activities at the Molecular Level

Martin Keller[1]* and Robert Hettich[2]

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831,[1] and Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831[2]

INTRODUCTION

Bacteria represent a major portion of Earth's biota, play essential roles in the nutrient cycle, interact with animals and plants, and directly impact climate. The total number of bacteria on Earth is estimated to lie between 4×10^{30} and 6×10^{30} cells, representing 350 to 550 Pg of total cellular carbon (1 Pg = 10^9 tons), or 60 to 100% of the estimated total carbon in plants (61). Nonetheless, the role and individual contributions of the various microbial species within the environment remain largely unknown.

The application of molecular tools to analyze microbial diversity resulted in a tremendous increase of new phylotypes. When Woese and colleagues defined the major bacterial taxa almost 20 years ago, they recognized only 11 taxa (66). After more than two decades of environmental 16S rRNA gene sequencing from almost every habitat on Earth, at least 41 additional taxa have been described, bringing the total number of bacterial taxa to more than 53 (46). All current environmental genomics studies originated, for the most part, from the cultivation-independent survey approach to studying natural microbial populations described by Pace and colleagues (40, 41). Lately, we have seen an explosion in total-genome and metagenomic sequencing projects based on a shotgun sequencing approach (52, 53). In addition, the repertoire of tools for "postenvironmental genomics" is expanding, with microarray, proteomic, and metabolomic experiments greatly expanding the already remarkable discoveries provided by environmental genomics (5). At the same time, the use of these molecular techniques has led to the realization that the microbial diversity found in almost all investigated environments is much larger than ever anticipated (14, 24, 46). To further evaluate species diversity, Schloss and Handelsman constructed rarefaction curves for each phylum to compare the extent of sampling and relative richness of each phylum at various taxonomic levels using 56,215 partial 16S rRNA gene sequences in a single analysis (46). The unexpected, relatively flat slope of the bacterial rarefaction curves suggested that either current sampling methods are not adequate to identify 10^7 to 10^9 different species of bacteria or these estimates are high. Nevertheless, only an extremely small fraction of these microbially diverse populations can be cultured in the laboratory, as evident by the relatively small number of bacterial species that have been cultivated and validly published thus far (4,800 as of July 2003) (24).

Does Microbial Composition Affect Ecosystem Processes?

The presence of this tremendous diversity in combination with the finding of significant lateral gene transfer within these environments challenge the conventional understanding and definition of a microbial species and the evolution of microbes in general. Although it has been suggested that "there is a continuity of energy flux and informational transfer from the genome up through cells, community, virosphere, and environment," it is not clear whether genomes are discrete or if they change and adapt to the needs and pressures imparted by a specific environment (12). As with macroorganisms, a growing body of evidence indicates that microbial composition also affects ecosystem processes, including CO_2 respiration and de-

* Corresponding author. Mailing address: Biosciences Division, Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008, MS-6026, Oak Ridge, TN 37831. Phone: (865) 576-0567. Fax: (865) 574-9223. E-mail: kellerm@ornl.gov.

composition (2), autotrophic and heterotrophic production (36, 37), and nitrogen cycling (19). Baas-Becking and Beijerink postulated decades ago that microbial taxa have preferred environments (see reference 35): "Everything is everywhere, but the environment selects." This hypothesis implies that microorganisms are frequently dispersed globally and that they are subsequently selected by the environments in which they reside on the basis of their functional capacities. Existing communities would thus constantly be challenged by intruders from nonspecialist taxa that may occasionally survive simply by chance, acquiring the necessary functionality through horizontal gene transfer (58). The paper by von Mering et al. supports this hypothesis, showing strong environmental preference along lineages but with a time-dependent decay. They observed a remarkable time-dependent stability of habitats and showed that for any two microbial isolates, the similarity of their annotated habitat is strongly correlated to their evolutionary relatedness (58). Even strains related only at the level of taxonomic order are still significantly more frequently found in the same environment than a random pair of isolates. Thus, most microbial lineages remain associated with a certain environment for extended time periods, and successful competition in a new environment seems to be a rare event, requiring more than just the acquisition of a few essential functions (58). Therefore, a question for future research is as follows. What are the traits that lead to the wide variety of colonization, diversification, and extinction rates in microorganisms (35)?

Proteomics in the Postgenomic Era

The tremendous increase in genome sequencing capacity, coupled with significant cost reductions, has led to a new wave of metagenomics in which whole-community DNA shotgun sequencing can be conducted to characterize at least the dominant members of microbial consortia, thereby bypassing the need to isolate and culture individual microbial species. However, current shotgun sequencing and in silico assembly algorithms are challenged in the assembly and assignment of sequences to specific species, strains, or ecotypes by an inherent intraspecies genetic complexity (11). The assignment of function and the metabolic contribution of specific microbial ecotypes or species to the investigated environment make the binning of sequences an important and challenging task. The genome coverage needed to allow sequence assembly and binning is hard to predict, especially when underrepresented species are of interest. Overall, despite all the tremendous benefits derived from shotgun sequencing approaches, the postgenomic era also demonstrated the limitations of nucleic acid-based methods for providing extensive information regarding the functional interplay between members of a microbial community in situ (34). This is the impetus for the field of proteomics, in that this experimental approach is designed to provide comprehensive *qualitative* and *quantitative* measurements of the final gene products (i.e., proteins) as biomarkers of the metabolic activity occurring in microbial communities. Of course, it is critical to realize that the monumental 16S rRNA work and deep metagenome sequencing/annotation are key elements for the success of the proteomic measurements. The 16S rRNA data provide vital information about the species membership of a sample, which is a required input for meaningful proteome evaluations of environmental microbial communities. The metagenome sequence provides an inventory of all possible gene products and thus provides the overall catalog from which proteome identifications are derived.

WHAT IS PROTEOMICS?

In comparison to lipids and nucleic acids, proteins are promising alternative markers of biological function, since they reflect the actual activity with respect to metabolic reactions and regulatory cascades and provide more direct information about microbial activity than functional genes and even their corresponding messenger RNAs (63). One might ask the following question. Why not simply use the well-developed approach of transcriptomics to profile gene expression and thus avoid the need for proteomic measurements? To address this issue, it is important to recognize that proteomics not only characterizes the final gene products but also provides detailed information about protein abundances, stabilities, turnover rates, posttranslational modifications, and protein-protein interactions, all of which provide critical metabolic activity information well beyond the genome and transcriptome levels. The use of proteins as alternative markers led to the establishment of *proteomics* as a research area. Proteomics is defined as the complete protein complement of the expressed genome and involves diverse techniques that provide a macroscopic view of what is expressed and present under different growth conditions, thereby enabling more constructive targeted experimentation (13). An early criticism of the emerging field of proteomics was the notion that that this measurement would provide information on only the most abundant housekeeping proteins and thus be of very limited value for microbiology research. However, the advent of more sophisticated and higher-throughput chromatographic-mass spectrometric instrumentation has greatly advanced the depth of proteome coverage, at least for microbial species. At present, it is readily possible to use proteomic approaches to identify at least 50 to 70% of the predicted proteome for most bacteria grown under a single growth condition. For example, ~2,000 to 2,500 proteins can be identified for bacteria with genomes of ~4,000 open reading frames. While there is considerable speculation about what fraction of a bacterial genome is actually expressed under a single growth condition, estimates suggest that most bacteria may employ only 50 to 80% of their predicted genes under a single growth condition. Thus, the current level of proteome measurement is already fairly deep into the dynamic range of proteins expressed. Clearly, low-copy-number proteins (such as transcription factors) are still difficult to identify, but this experimental approach has demonstrated remarkable success at probing well beneath "only the most abundant housekeeping proteins."

The logical expansion of single-organism proteomics is its application to whole microbial consortia and environments. Rodriguez-Valera (44) proposed the term "metaproteome" to describe the genes and/or proteins most abundantly expressed in environmental samples. The term was derived from "metagenome," which reflects the compound genome of the whole microbiota found in nature (17, 45). Wilmes and Bond proposed the term "metaproteomics" for the large-scale charac-

terization of the entire protein complement of environmental microbiota at a given point in time (64).

Among various proteomic techniques, mass spectrometry (MS) has emerged as the primary method for characterizing the presence and identities of proteins in native biological systems (4). One of the main drivers for this development is the unparalleled ability to acquire *nontargeted*, high-content, qualitative, and quantitative protein information about biological samples of enormous complexity. MS-based proteomics has several biological applications. In many pioneering studies, it was used to make an inventory of the content of subcellular structures and organelles, creating valuable repositories of information about the localization of proteins in cells and tissues (2). MS-based proteomic experiments involve several steps and can be separated into two major approaches: (i) two-dimensional (2D) gel electrophoresis (2D-GE) MS employing either peptide mass fingerprinting or tandem MS (MS/MS) or (ii) liquid chromatography (LC)-MS/MS (Fig. 1).

2D-GE Combined with MALDI-TOF MS

The first proteomics approach developed (and, in fact, still widely employed today) relies on the separation of complex protein samples by 2D-GE, which combines isoelectric focusing in the first dimension and denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (PAGE) in the second dimension (3, 16). The traditional approach to gel analysis involves detecting and differentiating the spot patterns from related but distant samples by quantifying and normalizing the spot volumes and then matching the spot patterns of each gel or gel set to an image selected as the master. Comparative proteomics using images from different gels has been impaired by gel-to-gel variation, hindering both the detection and quantification of proteins (23). To overcome these issues, differential in-gel electrophoresis was developed, which involves the use of multiple dyes for multiplex samples (54). In differential in-gel electrophoresis, samples are labeled with spectrally resolvable fluorescent CyDyes (GE Healthcare) prior to electrophoresis. These samples are then mixed before isoelectric focusing and resolved on the same 2D gel. This separation is followed by the in-gel digestion of selected spots and protein identification by MS. In the case of proteins from fully sequenced organisms, the identification is achieved by computational scoring of correlations between predicted peptides (obtained from the genome sequence) and measured peptides, as determined from peptide mass fingerprinting (matrix-assisted laser desorption–ionization [MALDI] time-of-flight [MALDI-TOF] MS) or fragmentation data of the peptides (electrospray ionization [ESI] source MS/MS). Identification of peptides from unsequenced species can be achieved in two ways: either the peptides and their MS/MS fragments resemble those of known proteins (in the case of moderate to high homology) or the peptides are sequenced by MS and de novo computational methods (Fig. 1). The de novo method focuses on analyzing the peptide fragmentation information directly to decipher amino acid sequences and is particularly attractive for cases in which the incomplete genome sequence is available. Owing to the ever-increasing number of fully sequences genomes, there is a fairly strong chance that homologies will be found. Challenges for the use of 2D-GE–MS for proteomic applications include limitations in the molecular sizes, pI ranges, and hydrophobicities of the proteins that can be analyzed. Furthermore, even with automated techniques, the ability to identify more than a few hundred proteins from a gel plate containing a thousand spots is still very difficult and generally low throughput.

LC-MS/MS

Newer MS-based proteomics approaches rely on gel-less technologies, most of which integrate online high-performance LC-MS/MS. In this case, LC-MS/MS combines the separation power of LC with both molecular mass (MS1) and fragmentation (MS2) information for the proteolytic peptides in a complex mixture. A peptide mixture from the sample of interest can be obtained by the proteolytic digestion of a solution-phase protein mixture or a gel band/spot resulting from electrophoretic separation. These peptides are then introduced into a one-dimensional or multidimensional LC system, which can be operated either offline or online with MS detection and characterization (4) (Fig. 1). Link and colleagues and Washburn et al. have pioneered an online, multidimensional LC-MS/MS approach termed MudPIT (multidimensional protein identification technology) in which the peptide mixture is separated by microcapillary 2D LC followed by MS/MS (29, 59). In the first stage, the acidified peptide mixture is separated by a multiple-step-gradient elution from a strong cation-exchange chromatographic phase directly onto a high-performance reversed-phase column, which is then flushed with a linear organic gradient eluent directly into the nanoelectrospray source of the mass spectrometer. For tandem MS measurements, the mass-to-charge ratios of the peptide ions are first measured by MS to determine the molecular mass of each peptide. Each peptide ion is then isolated in a high-throughput manner and collisionally dissociated by controllable impact with a neutral target gas. The mass-to-charge ratios of the resultant fragments are then measured, producing a tandem mass spectrum. Computational matching algorithms are then used to compare the experimental fragmentation spectrum to the predicted fragmentation spectrum of the putative peptide sequence suggested by the genome information. For these LC-MS/MS approaches, it is now fairly routine to identify several thousand nonredundant proteins from individual bacterial samples. Challenges for this approach include difficulty in accurately quantifying the peptides and/or proteins and the need for fairly complex and expensive instrumentation. Wei et al. optimized peptide separation through an online three-cycle liquid chromatography system (three-dimensional LC-MS/MS). The complex peptide mixture was loaded directly onto the microcapillary column filled with C_{18} reversed-phase (RP1) material as the first phase of separation. Peptides were fractionated through RP1, subfractionated in the following strong-cation-exchange phase (SCX), and separated further on a second reverse-phase segment (RP2) with a combination of reverse-phase and salt gradients through the three phases in an iterative process (60). The three-dimensional LC-MS/MS system demonstrated high separation power and tolerance to detergent. MS/MS data are acquired from the separated peptides and used for database searches to identify peptides and, thus, the proteins to which they belong.

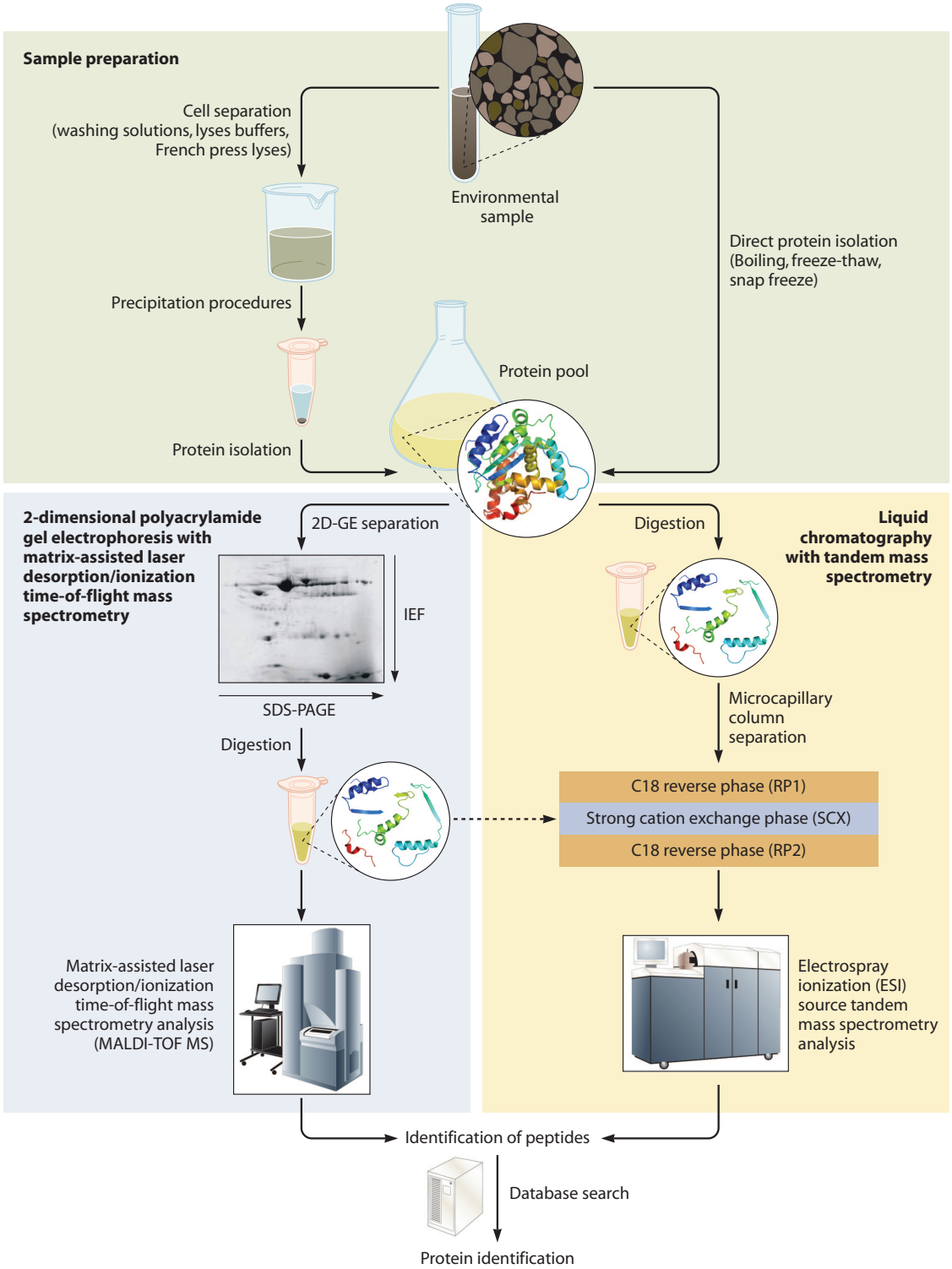

FIG. 1. Overview of two major environmental proteomic strategies. In one strategy (top left), proteins are separated by 2D-GE followed by MS analysis with either peptide mass mapping or LC-MS approaches. An alternate strategy (top right) involves protein digestion followed by online liquid LC separation and MS characterization. There are dramatic differences in measurement throughput, automation, and depth of protein identification for these two related but distinct approaches, as discussed in the text. IEF, isoelectric focusing; SDS, sodium dodecyl sulfate.

Necessary Improvements for MS-Based Proteomics

Although the MS-based proteomic approach is already very powerful, improvements of several technical aspects will help to better meet the demands for higher throughput and proteome coverage without sacrificing information content. First, faster and more efficient separations of complex protein and peptide mixtures through various chromatographic methods need to be further explored and optimized. Newer separation approaches, such as gas-phase ion mobility fractionation (55, 56), suggest promise to alleviate some of the limitations of the current chromatographic technologies. Conventional chromatographic separations are notoriously time-consuming, so recently, attention has been given to the exploration of newer methodologies, such as ultra-high-pressure LC (31, 42), to expedite the separation process and thus increase the duty cycle of the experiment. Second, advances in MS scan speed would allow more frequent sampling of ions from these very complex mixtures. Higher rates of sampling would translate into more tandem mass spectra acquired per unit of time, which would in turn enable higher-resolution chromatography methods to be used. Increased sampling rates should also improve the dynamic range, because lower-abundance ions are more likely to be detected. By coupling increased scanning speed to continued improvements in sensitivity and mass accuracy measurements, a large gain in the dynamic range of proteome measurement could be realized, thereby leading to the ability to perform proteomics on an extremely small amount of cells, enabling single-cell proteomics. Increased resolution and mass accuracy should also strengthen confidence in peptide identifications and facilitate the discovery of protein modifications. Third, advances in top-down MS for sequence-based characterization of intact proteins can allow patterns of modifications on a protein to be correlated with specific activities or functions (67). At present, top-down MS is most efficient for small proteins (<25 kDa) and presents difficulties for analyzing larger proteins (67). Key areas for the improvement of top-down MS are the development of more general fragmentation methods for large proteins and of higher-throughput and more robust methods to introduce intact proteins into the mass spectrometer (4).

ENVIRONMENTAL PROTEOMICS

Optimization of Sample Preparation Protocols

Sample preparation is one of the critical elements in environmental proteomics and, while seemingly somewhat mundane, has been an Achilles' heel that has impeded progress in this field. As shown for environmental genomic library construction, significant method development was necessary to develop protocols leading to representative environmental genomic libraries from soil and other environments (14). Developing methods to obtain an environmental representative protein extract will be challenging and will differ for the two major proteomics methods (LC versus 2D gel separation). Ogunseitan developed and evaluated two methods for extracting proteins from water, sediments, and soil samples (38, 39). In the first method, microbial proteins were extracted from 1 g soil or sediment by boiling the samples in a solution, while in the second method, the same quantities of environmental sample were incubated for 1 h at 0°C in a solution followed by four 10-min freeze-thaw cycles. The boiling method recovered high concentrations of proteins from wastewater but not from soils and sediments. The freeze-thaw method performed better for soils and sediments (38, 39). Singleton et al. examined a variety of methods to extract total soil protein. A simple snap-freeze protein extraction technique using liquid nitrogen was found to extract the most protein from soil samples compared to a bead-beating method used commonly for DNA extraction from soil (48). Schulze et al. developed a protocol to analyze soil proteins isolated from dissolved organic matter, targeting extracellular microbial enzymes that are possibly important in the carbon cycle. Water-soluble proteins were extracted from soil by dissolving soil minerals with hydrofluoric acid, and the extracted proteins were identified by MS coupled to LC (47). Wilmes and Bond reported a method that enabled the successful extraction and purification of the entire proteome from a laboratory-scale activated sludge system optimized for enhanced biological phosphorus removal (EBPR) followed by separation by 2D-GE. The protocol involves various washing solutions, lysis buffers, French press lyses, and precipitation procedures (64). Benndorf et al. developed an extraction protocol to use proteins from soil or groundwater as functional biomarkers based on the separation of proteins from the inorganic and organic constituents of the soil matrix by a combination of 0.1 M NaOH treatment and phenol extraction. The incubation of the soil with NaOH released humic acids and proteins from soil minerals and simultaneously disrupted microorganisms. The subsequent phenol extraction separated the proteins from the humic organic matter (3). These samples could be separated on one-dimensional or 2D gels. However, high humic acid content within soil samples is still a limiting factor, and further-improved extraction methods will be necessary to obtain samples with sufficient purity for downstream analyses.

2D Electrophoresis-Based Environmental Proteomics

Early environmental proteomics research focused almost exclusively on the use of 2D-GE–MS technology. In this case, environmental samples were collected and evaluated using a 2D gel display followed by downstream MS analysis. The main focus for most of the described papers was not to extensively identify each and every protein spot in the complete gels but rather to focus on interesting spots that changed location or intensity as a function of sample or growth condition.

Wilmes and Bond used 2D-PAGE coupled with MALDI-TOF MS on a laboratory-scale activated sludge system optimized for EBPR. This method enabled the successful extraction and purification of the entire proteome, its separation by 2D-PAGE, and the mapping of this metaproteome. Highly expressed protein spots were excised and identified using quadrupole time-of-flight MS with de novo peptide sequencing. The isolated proteins were putatively identified as an outer membrane protein (porin), an acetyl coenzyme A acetyltransferase, and a protein component of an ABC-type branched-chain amino acid transport system. These proteins were postulated to stem from the dominant and uncultured *Rhodocyclus*-type polyphosphate-accumulating organism in the activated sludge (64).

Lacerda et al. used 2D-PAGE coupled with MALDI-TOF/TOF MS and de novo peptide sequencing for the separation and identification of proteins differentially expressed over time within a bacterial community following exposure to an inhibitory level of cadmium (27). This was the first report detailing the dynamic metaproteomic response of a unsequenced microbial community to an environmental insult using de novo sequencing (27). More than 100 unique, differentially expressed proteins were identified through database searching and de novo sequencing. Proteins of importance in cadmium shock included ATPases, oxidoreductases, and transport proteins (22).

Markert et al. performed a 2D gel-based proteomics approach on the bacterial endosymbiont of the deep-sea tube worm *Riftia pachyptila*, revealing that three major sulfide oxidation proteins constitute about 12% of the total cytosolic proteome. This is consistent with the conventional understanding that sulfur is the terminal electron acceptor in deep-sea species where oxygen is limiting. In addition, the proteome work provided strong evidence that the *Riftia* symbionts, which have been considered a prime example for chemolithoautotrophic carbon fixation via the Calvin cycle, partly use the reductive tricarboxylic acid cycle for autotrophic carbon fixation (33).

Klaassens et al. showed for the first time the extraction of proteins, reproducible 2D gel electrophoresis, and tentative identification using MALDI-TOF MS of the metaproteome of a complex intestinal ecosystem of an uncultured infant fecal microbiota (25). Limitations in protein identification were based on the limited microbiome sequence information. The ongoing metagenomic library development will enable the meaningful identification of an extended list of proteins present in this complex environment (11, 28).

Wilmes and Bond reported the application of metaproteomics to compare two laboratory-scale sequencing batch reactors, one with EBPR performance and one which did not perform EBPR. The comparison of protein expression in the two sludge reactors demonstrated that their metaproteomes were substantially different, which was also reflected in their microbial community structures and metabolic transformations (65). Further studies of this system revealed that the microbial communities of the laboratory-scale EBPR batch reactor were dominated by the uncultured polyphosphate-accumulating organism "*Candidatus* Accumulibacter phosphatis." An extended metaproteomic study based on 2D gel separation matched 638 proteins across gels generated from the phosphate-removing sludge. They included enzymes involved in energy generation, polyhydroxyalcanoate synthesis, glycolysis, gluconeogenesis glycogen synthesis, the glyoxylate/tricarboxylic acid cycle, fatty acid beta oxidation, fatty acid synthesis, and phosphate transport and provided evidence to link the metabolic activities of "*Candidatus* Accumulibacter" to the chemical transformations observed in EBPR. An in-depth proteomic study and a comprehensive genome-wide alignment of more than 13,000 orthologous proteins uncovered substantial differences in protein abundance for enzyme variants involved in EBPR-specific pathways as well as the core metabolisms among the "*Candidatus* Accumulibacter" population. This suggests an essential role for genetic diversity in maintaining the stable performance of EBPR systems and, hence, demonstrates the power of integrated cultivation-independent genomics and proteomics for the analysis of complex biotechnical systems.

Finally, Benndorf et al. used a combination of sodium dodecyl sulfate-PAGE and LC-ESI-MS to analyze proteins extracted from contaminated soil and groundwater (3). Protein extracts were applied for sodium dodecyl sulfate-PAGE and 2D electrophoresis. Spots and bands were excised, and individual proteins were identified by online LC coupled to MS via an ESI source. To assess the suitability of this approach for the functional analysis of environmental metaproteomes, it was applied to soil that had been enriched in chlorophenoxy acid-degrading bacteria by incubation with 2,4-dichlorophenoxy acetic acid for 22 days. The identification of enzymes such as chlorocatechol dioxygenases was consistent with bacterial metabolic pathways expected to be expressed in these samples. (3).

LC-MS-Based Environmental Proteomics

The revolution in the use of gel-less online multidimensional LC-MS/MS technology for studying microbial isolates opened up a new regimen of comprehensive proteome characterization (4, 8), now enabling the identification of a few thousand proteins from an individual cultivated microorganism (15, 51, 57, 60). This permits a detailed, fairly deep glimpse into the molecular activities of the bacteria and now provides a robust technology which can be extended to environmental samples.

The recent application and demonstration of online multidimensional LC-MS/MS technology for a natural microbial community thriving in acid mine drainage demonstrated the first large-scale (i.e., greater than 2,000 proteins identified from a single sample) whole-community proteome characterization of a microbial consortium (43). Ram et al. combined shotgun MS-based proteomics with community genomic analysis to evaluate the in situ microbial activity of this low-complexity natural microbial biofilm. This community proteogenomics approach was successful in identifying more than 2,000 proteins from the five most abundant species in the biofilm (43). Subsequent work revealed how strain-resolved community proteomics could be used to unravel genome recombination in environmental samples (6, 32). These whole-community proteogenomics results revealed that genomes within this community were shaped by recombination involving large chromosomal regions that were derived from two closely related bacterial populations, suggesting that gene variation/rearrangement is crucial for adaptation to specific ecological niches (32).

This whole-community proteomics approach has been used to probe microbial strain-variant protein expression within activated sludge, verifying the importance of denitrification, fatty acid cycling, and glyoxylate bypass in EBRP (62).

More recent work has focused on the metaproteomics from the ocean environment, specifically, the Sargasso Sea. In this case, a total of 236 SAR11 proteins, 402 *Prochlorococcus* proteins, and 404 *Synechococcus* proteins were detected. Proteins implicated in the prevention of oxidative damage and protein refolding were abundant. The measurements support the view that competition for nutrients in oligotrophic systems is extreme but that nutrient flux is sufficient to sustain microbial community activity (49).

OUTLOOK: WHAT IS NEXT?

Improvements in Integrating Experimental MS Technology with Bioinformatics

The further development of environmental proteomics and its application to more complex microbial environments will require further development to push the current limits of this exciting and promising technology. As outlined above, improvements to current mass spectrometers with regard to both sensitivity and capacity will be required to further increase the dynamic range and speed of the analysis needed to profile the range of diversity expected in natural environmental microbial consortia. Despite the reported success in monitoring the complex proteomes of low-complexity microbial communities, the studies detect primarily proteins from the most abundant organisms but fail to dig deep enough into the more minor species (1). For more generic applications, it will be necessary to improve the ability of MS-based approaches to profile a much wider dynamic microbial range to include the simultaneous evaluation of not only bacterial species with large variations in abundance but also archaea and even viruses. While MS-based approaches are already very powerful for *qualitative* metaproteome investigations, there is a great need to develop and demonstrate improved approaches for *quantitative* measurements. Furthermore, the ability to characterize protein posttranslational modifications is essential for a more comprehensive understanding of how microbial species regulate their proteins for functionality. Since MS is exquisitely sensitive to mass shifts, the presence of modifications such as methylation, phosphorylation, and oxidation, etc., on peptides is readily resolvable, provided that adequate mass accuracy and appropriate bioinformatic techniques are used, and can be used to view a deeper level of posttranslational protein control.

As might be obvious, the ability to conduct metaproteome measurements is intimately tethered to metagenome sequencing success. Errors or omissions in experimental metagenome sequencing or annotation will be propagated directly into the metaproteome work. Thus, the depth and quality of the metagenome sequence directly control and somewhat limit the success of the metaproteome investigations. At present, there are at least two considerations that must be attended to: (i) the ability to get deeper metagenome sequence and more complete annotation will be critical to profile not only the species but also the strain composition of microbial consortia (because metaproteome data must be searched against metagenome results; missing genome segments will translate into unidentified peptides in the metaproteome data sets), and (ii) while exciting and new high-throughput DNA sequencing approaches such as 454 pyrosequencing are making their way into a variety of laboratories, the ability to assemble metagenomes from these data and thereby couple it to MS-based metaproteomic approaches are largely untested at this point and will require careful integration and evaluation.

Improved Sample Handling through Chip-Based Methods

The tremendous progress within the field of proteomics was achieved through significant instrumentation improvements. Sample preparations are still tedious and include preconcentration and digestion before MS analysis. Microfluidic systems are an important component of the ongoing push toward miniaturization and integration of analytical platforms for proteome characterization (7). Microchip techniques, which have been under intensive research over the last 20 years, have been developed to make the sample preparation and introduction to MS less problematic and harbor the opportunity to reduce the sample amount, analysis time, and costs (50). Srbek et al. previously reported the coupling of an integrated polymer microfluidic device with MS and analyzed proteins in complex biological samples isolated from plants and human tissue (50). This integrated polymer microfluidic device was built by laser ablation on polyimide film, and a noble metal electrode was deposited at the end of the channel to apply the voltage for ESI. A reverse-phase chromatography column was built inside the polymer-based chip to perform LC separation by coupling the nanofluidic pumping system. The chip is sandwiched between a stator and rotor valve (26). The overall performance of the microfluidic chip enabled a gradient liquid chromatographic separation and MS/MS identification of peptides with an estimated detection sensitivity of 5 to 7 fmol, demonstrated with both a plant and the human tissue sample (50). Horvatovich et al. compared label-free profiling of immunodepleted, trypsin-digested serum by a microfluidics-based LC-MS system with a conventional capillary LC-MS system (18). The chip-LC-MS system had a two-times-higher resolution on the LC dimension and resulted in a lower average charge state of the tryptic peptide ions generated in the ESI interface than the cap-LC-MS while requiring about 30-times-less sample (18). Ethier et al. developed a microfluidic proteomic reactor that greatly simplifies the processing of complex proteomic samples by combining multiple proteomic steps (9). The rapid extraction and enrichment of proteins from complex proteomic samples or directly from cells were performed in a 50-nl effective volume, resulting in an increased number of generated peptides. The sensitivity increased 10 times, and the reactor allowed the analysis of as little as 300 mouse testicular cancer cells (9). Hou et al. took the proteomic reactor further and developed a 96-well plate proteomic reactor that performs multiplexed trapping, enrichment, and biochemical processing of proteins and is coupled with protein fractionation using size exclusion chromatography for the large-scale identification of proteins (20). Other microfluidic-based devices include open channels, immobilized beads, and other solid-phase media (10). Huang et al. (21) and Liu et al. (30) used surface-adsorbed trypsin in channels to enable complete digestions in less than 5 s, the fastest tryptic digestions that have been reported.

There is widespread interest in the coupling of microfluidic devices to ESI or MALDI MS. Indeed, if a fully integrated method comprising chemical processing, sample preconcentration and cleanup, and 2D separations could be integrated with MS detection, the results could revolutionize the field of proteomics and would also have a major impact on environmental proteomics. However, the chip-to-MS interfaces are still not as solid and reliable and have therefore not been widely adapted so far (10).

CONCLUDING REMARKS

Environmental proteomics is gaining momentum and is already being used to study increasingly complex environments.

Key enabling breakthroughs include increased genome and metagenome sequencing capacities, faster and more sensitive mass spectrometers, and the ability to handle extremely large data sets. Although still in its infancy, environmental proteomics is expected to flourish and become more prevalent within the next few years. Improvements to sample preparation as well as devising a successful interface between mass spectrometers and sample preparation devices will be especially critical. Overall, however, environmental proteomics will provide enormous insights into microbial ecology that were not achievable only a couple of years ago. In addition, this burgeoning field will further enhance our knowledge regarding microbial cooperation and competition within complex natural ecosystems.

ACKNOWLEDGMENTS

M.K. and R.H. are partially supported by the U.S. Department of Energy through the BioEnergy Science Center. The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the U.S. Department of Energy Office of Science. Oak Ridge National Laboratory is managed by University of Tennessee–Battelle, LLC, for the Department of Energy under contract DOE-AC05-00OR22725.

REFERENCES

1. **Banfield, J. F., N. C. VerBerkmoes, R. L. Hettich, and M. P. Thelen.** 2008. Proteogenomic approaches to natural microbial communities. Omics **9:**301–333.
2. **Bell, T., J. A. Newman, B. W. Silverman, S. L. Turner, and A. K. Lilley.** 2005. The contribution of species richness and composition to bacterial services. Nature **436:**1157–1160.
3. **Benndorf, D., G. U. Balcke, H. Harms, and M. von Bergen.** 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. ISME J. **1:**224–234.
4. **Cravatt, B. F., G. M. Simon, and J. R. Yates.** 2007. The biological impact of mass-spectrometry-based proteomics. Nature **450:**991–1000.
5. **DeLong, E. F.** 2004. Microbial population genomics and ecology: the road ahead. Environ. Microbiol. **6:**875–878.
6. **Denef, V. J., M. B. Shah, N. C. VerBerkmoes, R. L. Hettich, and J. F. Banfield.** 2007. Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis. J. Proteome Res. **6:**3152–3161.
7. **Devoe, D. L., and C. S. Lee.** 2006. Microfluidic technologies for MALDI-MS in proteomics. Electrophoresis **27:**3559–3568.
8. **Domon, B., and R. Aebersold.** 2006. Mass spectrometry and protein analysis. Science **312:**212–217.
9. **Ethier, M., W. M. Hou, H. S. Duewel, and D. Figeys.** 2006. The proteomic reactor: a microfluidic device for processing minute amounts of protein prior to mass spectrometry analysis. J. Proteome Res. **5:**2754–2759.
10. **Freire, S. L. S., and A. R. Wheeler.** 2006. Proteome-on-a-chip: mirage, or on the horizon? Lab Chip **6:**1415–1423.
11. **Gill, S. R., M. Pop, R. T. DeBoy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett, and K. E. Nelson.** 2006. Metagenomic analysis of the human distal gut microbiome. Science **312:**1355–1359.
12. **Goldenfeld, N., and C. Woese.** 2007. Biology's next revolution. Nature **445:** 369.
13. **Graves, P. R., and T. A. J. Haystead.** 2002. Molecular biologist's guide to proteomics. Microbiol. Mol. Biol. Rev. **66:**39–63.
14. **Green, B. D., and M. Keller.** 2006. Capturing the uncultivated majority. Curr. Opin. Biotechnol. **17:**236–240.
15. **Gupta, N., S. Tanner, N. Jaitly, J. N. Adkins, M. Lipton, R. Edwards, M. Romine, A. Osterman, V. Bafna, R. D. Smith, and P. A. Pevzner.** 2007. Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. Genome Res. **17:**1362–1377.
16. **Hanash, S. M.** 2000. Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. Electrophoresis **21:**1202–1209.
17. **Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman.** 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. **5:**R245–R249.
18. **Horvatovich, P., N. Govorukhina, T. H. Reijmers, A. G. J. van der Zee, F. Suits, and R. Bischoff.** 2007. Chip-LC-MS for label-free profiling of human serum. Electrophoresis **28:**4493–4505.
19. **Horz, H. P., A. Barbrook, C. B. Field, and B. J. M. Bohannan.** 2004. Ammonia-oxidizing bacteria respond to multifactorial global change. Proc. Natl. Acad. Sci. USA **101:**15136–15141.
20. **Hou, W. M., M. Ethier, J. C. Smith, Y. L. Sheng, and D. Figeys.** 2007. Multiplexed proteomic reactor for the processing of proteomic samples. Anal. Chem. **79:**39–44.
21. **Huang, Y., W. Shan, B. H. Liu, Y. Liu, Y. H. Zhang, Y. Zhao, H. J. Lu, Y. Tang, and P. Y. Yang.** 2006. Zeolite nanoparticle modified microchip reactor for efficient protein digestion. Lab Chip **6:**534–539.
22. **Kan, J., T. E. Hanson, J. M. Ginter, K. Wang, and F. Chen.** 2005. Metaproteomic analysis of Chesapeake Bay microbial communities. Saline Syst. **19:**7.
23. **Karp, N. A., R. Feret, D. V. Rubtsov, and K. S. Lilley.** 2008. Comparison of DIGE and post-stained gel electrophoresis with both traditional and SameSpots analysis for quantitative proteomics. Proteomics **8:**948–960.
24. **Keller, M., and K. Zengler.** 2004. Tapping into microbial diversity. Nat. Rev. Microbiol. **2:**141–150.
25. **Klaassens, E. S., W. M. de Vos, and E. E. Vaughan.** 2007. Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. Appl. Environ. Microbiol. **73**:1388–1392.
26. **Kutter, J. P., S. C. Jacobson, N. Matsubara, and J. M. Ramsey.** 1998. Solvent-programmed microchip open-channel electrochromatography. Anal. Chem. **70:**3291–3297.
27. **Lacerda, C. M. R., L. H. Choe, and K. F. Reardon.** 2007. Metaproteomic analysis of a bacterial community response to cadmium exposure. J. Proteome Res. **6:**1145–1152.
28. **Ley, R. E., D. A. Peterson, and J. I. Gordon.** 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell **124:**837–848.
29. **Link, A. J., J. Eng, D. M. Schieltz, E. Carmack, G. J. Mize, D. R. Morris, B. M. Garvik, and J. R. Yates.** 1999. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. **17:**676–682.
30. **Liu, Y., H. J. Lu, W. Zhong, P. Y. Song, J. L. Kong, P. Y. Yang, H. H. Girault, and B. H. Liu.** 2006. Multi layer-assembled microchip for enzyme immobilization as reactor toward low-level protein identification. Anal. Chem. **78:** 801–808.
31. **Livesay, E. A., K. Q. Tang, B. K. Taylor, M. A. Buschbach, D. F. Hopkins, B. L. LaMarche, R. Zhao, Y. F. Shen, D. J. Orton, R. J. Moore, R. T. Kelly, H. R. Udseth, and R. D. Smith.** 2008. Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. Anal. Chem. **80:**294–302.
32. **Lo, I., V. J. Denef, N. C. VerBerkmoes, M. B. Shah, D. Goltsman, G. DiBartolo, G. W. Tyson, E. E. Allen, R. J. Ram, J. C. Detter, P. Richardson, M. P. Thelen, R. L. Hettich, and J. F. Banfield.** 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature **446:**537–541.
33. **Markert, S., C. Arndt, H. Felbeck, D. Becher, S. M. Sievert, M. Hugler, D. Albrecht, J. Robidart, S. Bench, R. A. Feldman, M. Hecker, and T. Schweder.** 2007. Physiological proteomics of the uncultured endosymbiont of Riftia pachyptila. Science **315:**247–250.
34. **Maron, P. A., L. Ranjard, C. Mougel, and P. Lemanceau.** 2007. Metaproteomics: a new approach for studying functional microbial ecology. Microb. Ecol. **53:**486–493.
35. **Martiny, J. B. H., B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Ovreas, A. L. Reysenbach, V. H. Smith, and J. T. Staley.** 2006. Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. **4:**102–112.
36. **Naeem, S., D. R. Hahn, and G. Schuurman.** 2000. Producer-decomposer co-dependency influences biodiversity effects. Nature **403:**762–764.
37. **Naeem, S., and S. B. Li.** 1997. Biodiversity enhances ecosystem reliability. Nature **390:**507–509.
38. **Ogunseitan, O. A.** 1997. Direct extraction of catalytic proteins from natural microbial communities. J. Microbiol. Methods **28:**55–63.
39. **Ogunseitan, O. A.** 1993. Direct extraction of proteins from environmental-samples. J. Microbiol. Methods **17:**273–281.
40. **Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl.** 1986. Microbial ecology and evolution—a ribosomal-RNA approach. Annu. Rev. Microbiol. **40:**337–365.
41. **Pace, N. R., and T. L. Marsh.** 1985. RNA catalysis and the origin of life. Origins Life Evol. Biosphere **16:**97–116.
42. **Plumb, R. S., P. Rainville, B. W. Smith, K. A. Johnson, J. Castro-Perez, I. D. Wilson, and J. K. Nicholson.** 2006. Generation of ultrahigh peak capacity LC separations via elevated temperatures and high linear mobile-phase velocities. Anal. Chem. **78:**7278–7283.
43. **Ram, R. J., N. C. VerBerkmoes, M. P. Thelen, G. W. Tyson, B. J. Baker, R. C. Blake, M. Shah, R. L. Hettich, and J. F. Banfield.** 2005. Community proteomics of a natural microbial biofilm. Science **308:**1915–1920.
44. **Rodriguez-Valera, F.** 2004. Environmental genomics, the big picture? FEMS Microbiol. Lett. **231:**153–158.
45. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L.**

Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. **66:**2541–2547.
46. **Schloss, P. D., and J. Handelsman.** 2004. Status of the microbial census. Microbiol. Mol. Biol. Rev. **68:**686–691.
47. **Schulze, W. X., G. Gleixner, K. Kaiser, G. Guggenberger, M. Mann, and E. D. Schulze.** 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia **142:**335–343.
48. **Singleton, I., G. Merrington, S. Colvan, and J. S. Delahunty.** 2003. The potential of soil protein-based methods to indicate metal contamination. Appl. Soil Ecol. **23:**25–32.
49. **Sowell, S. M., L. J. Wilhelm, A. D. Norbeck, M. S. Lipton, C. D. Nicora, D. F. Barofsky, C. A. Carlson, R. D. Smith, and S. J. Giovanonni.** 4 September 2008. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J. doi:10.1038/ismej.2008.83.
50. **Srbek, J., J. Eickhoff, U. Effelsberg, K. Kraiczek, T. van de Goor, and P. Coufal.** 2007. Chip-based nano-LC-MS/MS identification of proteins in complex biological samples using a novel polymer microfluidic device. J. Sep. Sci. **30:**2046–2052.
51. **Thompson, M. R., N. C. VerBerkmoes, K. Chourey, M. Shah, D. K. Thompson, and R. L. Hettich.** 2007. Dosage-dependent proteome response of Shewanella oneidensis MR-1 to acute chromate challenge. J. Proteome Res. **6:**1745–1757.
52. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin.** 2005. Comparative metagenomics of microbial communities. Science **308:**554–557.
53. **Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield.** 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:**37–43.
54. **Unlu, M., M. E. Morgan, and J. S. Minden.** 1997. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis **18:**2071–2077.
55. **Valentine, S. J., M. Kulchania, C. A. S. Barnes, and D. E. Clemmer.** 2001. Multidimensional separations of complex peptide mixtures: a combined high-performance liquid chromatography/ion mobility/time-of-flight mass spectrometry approach. Int. J. Mass Spectrom. **212:**97–109.
56. **Valentine, S. J., X. Y. Liu, M. D. Plasencia, A. E. Hilderbrand, R. T. Kurulugama, S. L. Koeniger, and D. E. Clemmer.** 2005. Developing liquid chromatography ion mobility mass spectrometry techniques. Expert Rev. Proteomics **2:**553–565.
57. **VerBerkmoes, N. C., M. B. Shah, P. K. Lankford, D. A. Pelletier, M. B. Strader, D. L. Tabb, W. H. McDonald, J. W. Barton, G. B. Hurst, L. Hauser, B. H. Davison, J. T. Beatty, C. S. Harwood, F. R. Tabita, R. L. Hettich, and F. W. Larimer.** 2006. Determination and comparison of the baseline proteomes of the versatile microbe Rhodopseudomonas palustris under its major metabolic states. J. Proteome Res. **5:**287–298.
58. **von Mering, C., P. Hugenholtz, J. Raes, S. G. Tringe, T. Doerks, L. J. Jensen, N. Ward, and P. Bork.** 2007. Quantitative phylogenetic assessment of microbial communities in diverse environments. Science **315:**1126–1130.
59. **Washburn, M. P., D. Wolters, and J. R. Yates.** 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. **19:**242–247.
60. **Wei, J., J. Sun, W. Yu, A. Jones, P. Oeller, M. Keller, G. Woodnutt, and J. M. Short.** 2005. Global proteome discovery using an online three-dimensional LC-MS/MS. J. Proteome Res. **4:**801–808.
61. **Whitman, W. B., D. C. Coleman, and W. J. Wiebe.** 1998. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. USA **95:**6578–6583.
62. **Wilmes, P., A. F. Andersson, M. G. Lefsrud, M. Wexler, M. Shah, B. Zhang, R. L. Hettich, P. L. Bond, N. C. VerBerkmoes, and J. F. Banfield.** 2008. Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. ISME J. **2:**853–864.
63. **Wilmes, P., and P. L. Bond.** 2006. Metaproteomics: studying functional gene expression in microbial ecosystems. Trends Microbiol. **14:**92–97.
64. **Wilmes, P., and P. L. Bond.** 2004. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environ. Microbiol. **6:**911–920.
65. **Wilmes, P., and P. L. Bond.** 2006. Towards exposure of elusive metabolic mixed-culture processes: the application of metaproteomic analyses to activated sludge. Water Sci. Technol. **54:**217–226.
66. **Woese, C. R., E. Stackebrandt, T. J. Macke, and G. E. Fox.** 1985. A phylogenetic definition of the major eubacterial taxa. Syst. Appl. Microbiol. **6:**143–151.
67. **Zabrouskov, V., M. W. Senko, Y. Du, R. D. Leduc, and N. L. Kelleher.** 2005. New and automated MSn approaches for top-down identification of modified proteins. J. Am. Soc. Mass Spectrom. **16:**2027–2038.