Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Microbiology

Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles

Paul Wilmes[1] and Philip L Bond[2]

Molecular techniques are providing unprecedented insights into the organismal and functional make-up of natural microbial consortia. Apart from nucleic acid based approaches, community proteomics has the potential to provide a high-resolution representation of genotypic and phenotypic traits of distinct community members. With the recent availability of extensive genomic sequences from different microbial ecosystems, community proteomics has thus far been applied to activated sludge, acid mine drainage biofilms, freshwater and seawater, soil, symbiotic communities, and gut microbiota. Although these studies differ considerably in the depth of coverage of their respective protein complements, they highlight the power of community proteomics in providing a conclusive link between community composition, physiology, function, interaction, ecology, and evolution.

Addresses
[1] Department of Earth and Planetary Science, University of California at Berkeley, Berkeley, CA 94720, USA
[2] Advanced Water Management Centre, The University of Queensland, Brisbane, Queensland 4072, Australia

Corresponding author: Bond, Philip L (phil.bond@uq.edu.au)

Current Opinion in Microbiology 2009, **12**:310–317

This review comes from a themed issue on
Techniques
Edited by Jerome Garin and Grant Jensen

Available online 4th May 2009

1369-5274/$ – see front matter

DOI 10.1016/j.mib.2009.03.004

Introduction

Microorganisms represent the predominant mode of life on Earth and, consequently, microbial proteins constitute the primary catalytic entities that underlie the major biochemical transformations on our planet. Recent technological developments are facilitating the comprehensive extraction, separation, and identification of proteins from natural microbial communities. This approach, termed community-proteomics or meta-proteomics (see Box 1), has the potential for solving one of the major challenges facing microbial ecologists in providing a high-resolution representation of community structure and function.

Community genomic sequencing projects, that analyze genomic DNA directly from environmental samples, are providing opportunities to genetically characterize microbial ecosystems [1]. These studies vastly expand our knowledge of the genetic diversity and the physiological potential within a rapidly increasing set of selected environments that include activated sludge [2], acid mine biofilms [3••], seawater [4], human guts [5], and termite hindguts [6••]. The exponentially growing DNA sequence data (genomic and metagenomic) form the foundation for postgenomic approaches that provide functional insight into microbial ecosystems. Hence, the field of microbial ecology is currently entering the era of Systems Biology with community proteomics occupying a keystone role.

The study of protein expression within environmental samples is not an entirely new concept [7,8]. However, this review discusses recent community proteomic studies (from 2004 onwards), that have developed on the back of vast accumulation of DNA sequences from a range of microbial habitats, improved protein separation techniques, and high-throughput protein identification by mass spectrometry.

How does community proteomics work?

The procedure for a community proteomic experiment is analogous to that utilized for a proteomic study of an axenic culture. The procedural steps are briefly summarized as: firstly, sample preparation, secondly, protein extraction, thirdly, separation of the proteins or peptides (usually in two dimensions), and lastly, mass spectrometry (MS) analysis followed by *in silico* spectral matching for identification of proteins (see Figure 1 for a detailed workflow).

Protein or peptide separations are typically performed either by classical one-dimensional or two-dimensional polyacrylamide gel electrophoresis (2DE), or by liquid chromatography (LC), respectively. 2DE provides a comprehensive visual representation of the protein complement after staining and comparison of protein spot densities is convenient for the detection of differential protein expression in response to changes in condition or community structure and function [9]. Chosen spots are excised and digested with a protease (typically trypsin is used) before mass spectrometry analysis. Major limitations associated with the 2DE approach are comigration of proteins into discrete spots hampering accurate quantification and identification [9]. Also, large and hydrophobic

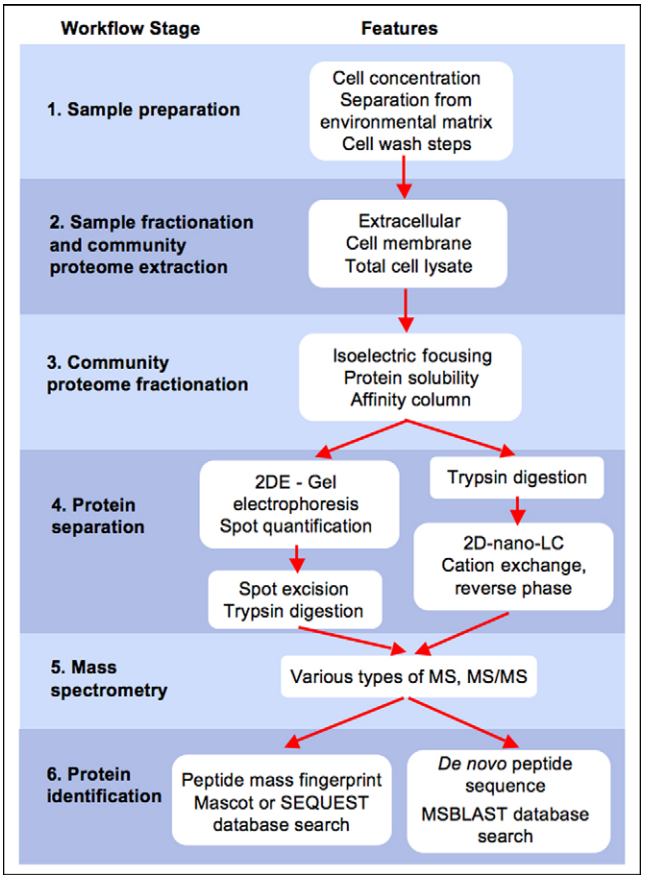
Box 1 Meta-proteomics, community-proteomics environmental-proteomics, or meigma-proteomics? Different prefixes have thus far been used to describe the bulk analysis of genomic, transcriptomic, proteomic, and metabolomic complements of mixed microbial communities. The prolific prefix 'meta' may not accurately reflect the fact that the analyses are of mixtures of distinct genomes, transcriptomes, proteomes, or metabolomes. A more accurate prefix may be the Greek for mixture, that is, meigma. However, for ease and clarity we employ the term 'community proteomics' throughout the present review since it reflects the idea of distinct community members each with their own distinct 'omes'.

proteins are not separated well by 2DE, for example, membrane proteins [10], and the method suffers from poor automation. Otherwise 2DE has good protein separation power especially when combined with prefractionation procedures, and chosen proteins can be readily identified after proteolytic digestion, MS analysis that produces specific peptide mass fingerprints (PMF) and database searching.

In the LC-based approach, complex protein samples are digested and the resulting peptides are separated, subjected to MS analysis, and PMFs generated (Figure 1). The approach circumvents many limitations of the 2DE approach and allows for high-throughput analysis identifying some thousands of proteins within one to two days [11••]. In particular, it allows the high-throughput analysis of insoluble membrane proteins [12] and, hence, membrane fractions are typically generated in LC-based community proteomics [13••]. The proteomic approach to separate and analyze at the peptide level, as opposed to protein separation, is referred to as *shotgun* or *bottom up* proteomics. Various LC and MS systems are used, in particular, 2D nano-LC (strong cation exchange followed by reverse phase LC) coupled to an LTQ-Orbitrap mass spectrometer has proven well-suited for the analysis of mixed microbial proteomes [14••,15•,16••].

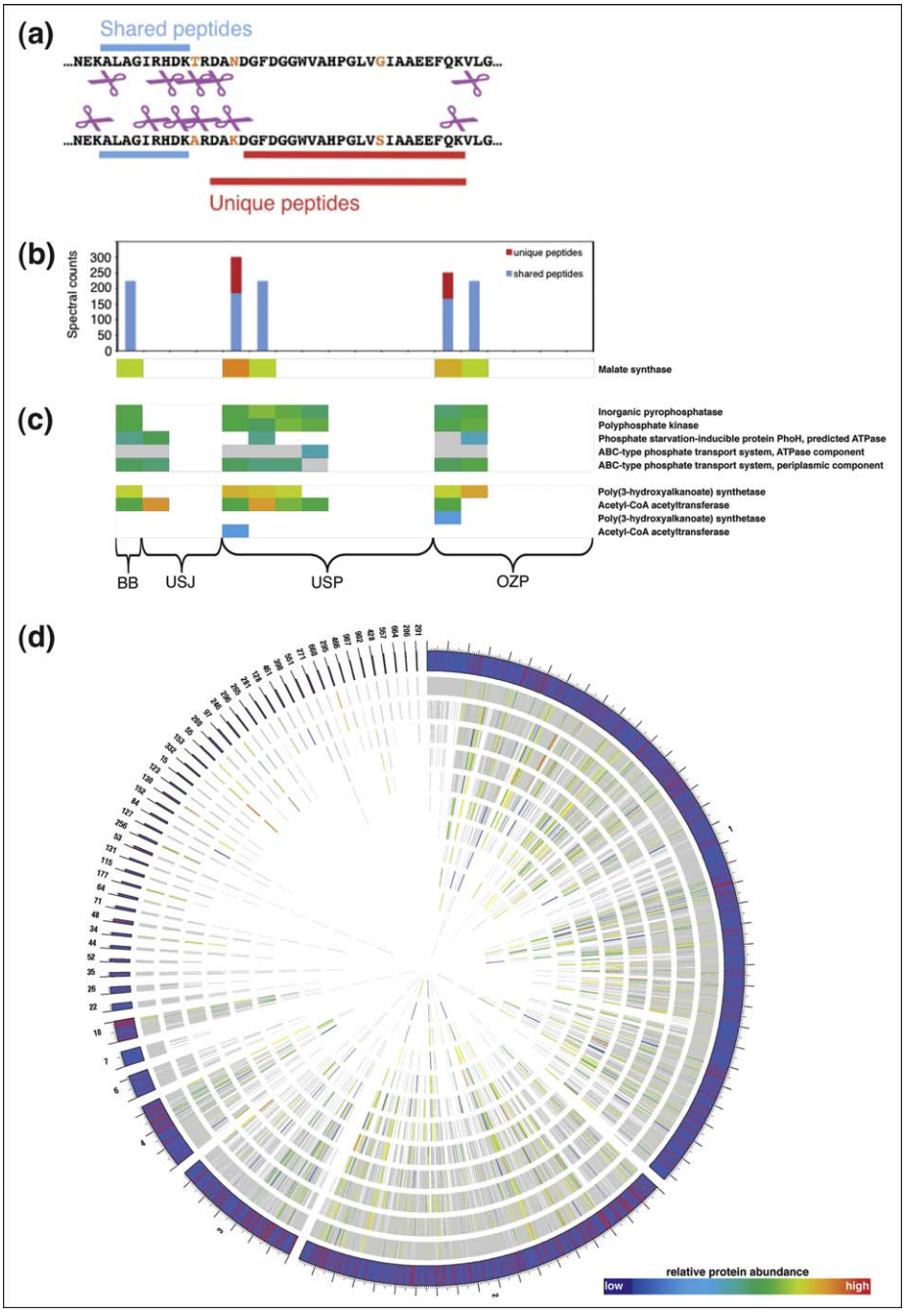
Parent proteins can be identified by sequence database matching of the PMFs using search algorithms such as Mascot [17] and SEQUEST [18], and these are most successful if representative genomic sequences are available. However, community proteomics is complicated by strain variation of protein species (Figure 2) and precise identification requires high-mass accuracy peptide measurements and tandem MS (MS/MS) involving fragmentation of the peptides [19]. The generation of characteristic MS/MS fragmentation patterns allows more precise spectral matching than from MS alone (for a detailed description on strain-resolved proteomics using LC–MS/MS, see VerBerkmoes *et al.* [11••]). Additionally, the fragmentation pattern can be employed to generate the *de novo* protein sequence and search for homologous sequences using MSBLAST [20]. *De novo* peptide sequencing is especially useful for protein identification when corresponding sequence data are unavailable [21•,22•]. High-mass accuracy spectral matching can be

Figure 1



Community proteomics sample preparation, extraction, separation and identification routes. The workflow for a community proteomic analysis may consist of six stages. Sample preparation may be required in stage 1. Cells may need to be concentrated or purified away from interfering substances, for example humic acids in soil. Protein extraction is performed next (stage 2) and fractions of interest may be targeted, for example, extracellular, membrane, soluble, and whole cell fractions [14••]. Cell lysis may involve French Press lysis, sonication, chemical lysis, or bead beating [21•]. The procedures in these stages must have minimal effect on the protein expression itself and sufficiently preserve the extracted proteins. To assist latter separations, the extracted protein complement may be fractionated (stage 3), for example, divided into groups by preparative liquid isoelectric focusing or based on protein solubility. Protein or peptide separations may be performed by two-dimensional polyacrylamide gel electrophoresis (2DE) or by liquid chromatographic (LC) methods, respectively (stage 4). Following 2DE, gel images are analyzed and spots quantified. Chosen spots are then excised and digested with a protease (trypsin) for mass spectrometry (MS) analysis [e.g. matrix assisted laser desorption ionization time-of-flight (MALDI-ToF)]. For LC, the protein mixture is digested with trypsin before separation. Following LC (2D-nano-LC), the separated peptides are directly introduced into the mass spectrometer (e.g. by electrospray ionization) and mass spectra acquired e.g. using a hybrid LTQ-Orbitrap mass spectrometer]. MS analysis may also involve the fragmentation of the peptides and recording of the MS/MS spectra. If required, *de novo* protein sequence data can be determined from the MS/MS data. Algorithms such as Mascot [17] and SEQUEST [18] enable the MS or MS/MS peptide mass fingerprint data to be searched against sequence databases for protein identification. Acronyms: 2DE: two-dimensional polyacrylamide gel electrophoresis; 2D-nano-LC: two-dimensional nano liquid chromatography.

Figure 2



Strain-resolving protein expression within the '*Candidatus Accumulibacter phosphatis*' (*A. phosphatis*) population [16••]. **(a)** Region of amino acid alignment of orthologous *A. phosphatis* encoded malate synthase genes. Amino acid substitutions are highlighted in orange and scissors indicate trypsin cleavage sites. Generated peptides that are detected by high-mass accuracy mass spectrometry may be unique (red) or shared (blue) between protein variants. **(b)** Summing of detected peptides (spectral counts) allows the estimation of relative abundance of *A. phosphatis* protein variants identified using distinct genomic assemblies (USJ: US sludge Jazz assembly; USP: US sludge Phrap assembly; OZP: Australian sludge Phrap assembly [2]); gray color indicates no peptides identified from an aligned protein variant; white color indicates absence of aligned variant that were aligned against the *A. phosphatis* composite genome that acts as the backbone (BB; based on a 90% amino acid identity cut-off [16••]). **(c)** Examples of heterogeneous expression among enzyme variants of paramount importance to enhanced biological phosphorus removal, including enzymes involved in phosphate and polyhydroxyalkanoate transformations. **(d)** Alignment of variant genomic fragments against the *A. phosphatis* composite genome (outer concentric ring in blue, USJ scaffold numbers indicated, and locations of inserted genes in the aligned genomic fragments highlighted in red). Relative *A. phosphatis* protein abundances highlighted according to each locus on the first inner concentric ring. Aligned variant genomic fragments with corresponding protein abundance in the following concentric rings. Gray color indicates no protein identification; white color indicates absence of aligned protein variant. The image was generated with Circos (M Krzywinski, http://1477mkweb.bcgsc.ca/circos/).

used to positively identify proteins from closely related organisms, but the ability to do so decreases rapidly with amino acid sequence divergence [23•]. Consequently, *de novo* peptide sequencing may prove indispensable in diverse samples in which the contribution of distinct microbial populations has to be assessed but for which full DNA sequence coverage is unavailable. Thus, with the advent of powerful *de novo* peptide sequencing algorithms, the implementation of the approach may become routine. For example, the *de novo* peptide sequencing approach was used to identify more than 100 proteins that were differentially expressed following exposure of bacterial communities to cadmium [22•].

All community proteomic methods face challenges because of the large complexity of protein species and the large dynamic range of protein levels. This complexity increases for the LC approach as tryptic digestion produces some dozens of peptides per protein. A technique gaining popularity uses 'off-gel' separation by isoelectric focusing (IEF) to divide the peptide mix into many fractions before LC–MS/MS [24]. This approach has additional benefit if the IEF is performed on whole proteins instead of peptides. Thus, the protein mixture is simplified in each fraction and this increases the chance of protein identification by subsequent LC–MS/MS analysis.

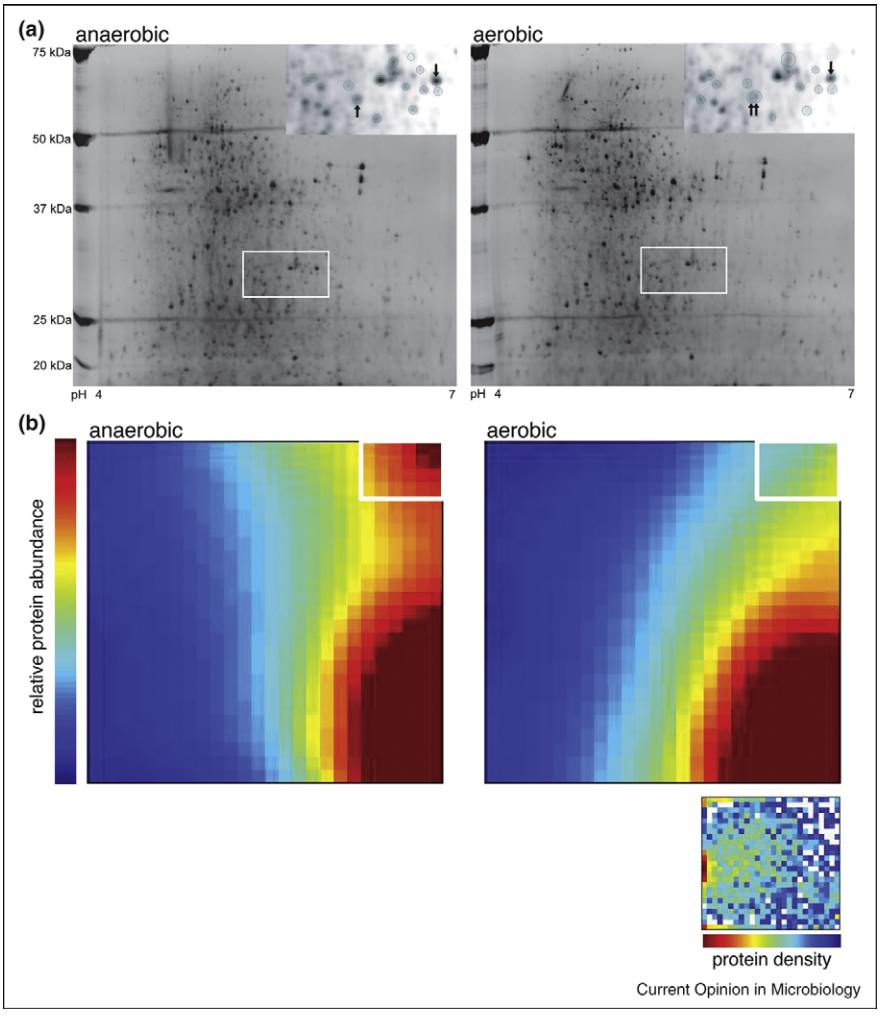
The story so far

The application of community proteomics is expanding rapidly, as recent studies directly examine community functional information in a phylogenetic context. As discussed, protein expression profiles based on gel separation provide overviews for the targeted identification of interesting proteins. Apart from activated sludge [9,21•,25], this approach has been applied to an estuary transect [26], contaminated soil and groundwater [27•], *Riftia pachyptila* endosymbionts [28], infant fecal samples [29], freshwater samples following exposure to heavy metals [30], lake water [31], proteins associated with exoploysaccharides in activated sludge [32,33], and sheep rumen [34]. Other studies have employed LC or combinations of gel and LC separation. Proteins identified within dissolved organic matter (DOM) in soil and water have been examined to determine the presence of broad taxonomic groups of microorganisms and highlight interesting functional differences between the microbial communities in forest soil (high abundance of peroxidases) and a peat bog (proteins involved in methanogenesis) [35•]. Similarly, proteins that form major constituents of DOM in seawater were identifiable because of the homology of identified peptides [36]. Because of the lack of comprehensive genomic sequences in the majority of these studies, the number of proteins identified was rather limited but it needs to be stressed that all studies did provide interesting functional insights.

Ubiquitous proteins of biogeochemical importance (proteorhodopsins) were readily identifiable in environmental samples because of the availability of extensive gene sequences [37]. A comprehensive community proteogenomic approach involving deep genomic and proteomic sampling was pioneered on acid mine biofilms [13••,14••,38••] (see below). This approach has since been applied to the enhanced biological phosphorus removal (EBPR) activated sludge system that is typically dominated by as yet uncultured organisms putatively named '*Candidatus Accumulibacter phosphatis*' [16••] and to human fecal samples [15•]. An analogous approach was applied to microbial communities inhabiting the euphotic zone of the Sargasso Sea and led to the identification of several proteins linked to the dominant organisms SAR11, *Prochlorococcus* and *Synechococcus* that are reflective of their lifestyle in this nutrient-limited environment [39••]. The power of combining genomics and proteomics on communities of immediate biotechnological interest (bioenergy) was highlighted by an integrative study of termite hindgut symbiotic bacteria involved in lignocellulose degradation [6••].

A full-cycle community proteomic investigation including comprehensive extraction, purification, separation, and identification following MS was first applied to a laboratory-scale activated sludge reactor operated in the UK [21•]. In this study, comparisons of proteome profiles, generated by 2DE, were made to determine metabolic details of the EBPR wastewater treatment process. Protein expression was compared between the two operational stages (anaerobic and aerobic; Figure 3a) of EBPR. However, only minor differences in protein levels were detected (Figure 3a) [21•,40]. Consequently, the study focused on the identification of highly expressed proteins by *de novo* peptide sequencing. When metagenomic sequences of EBPR sludges became available [2], these facilitated protein identification by analysis of PMF patterns with over 30% of highly expressed proteins chosen from 2DE gels being matched to the dominant organism [9]. Importantly, the identified proteins are involved in major EBPR carbon transformations. In a further study, using reference metagenomic sequences from the EBPR sludges cultured in the United States and Australia [2], the application of shotgun proteomics provided much deeper insight into the metabolic transformations of *A. phosphatis* in the UK sludge (10% proteome coverage). Interesting findings from placing identified proteins into metabolic context include the importance of denitrification, fatty acid cycling, and the glyoxylate shunt [16••]. This study also used strain-resolved proteomics to differentiate the expression of co-occurring protein variants within the EBPR sludge dominated by the *A. phosphatis* population [16••] (Figure 2). This revealed that 59% of the most abundant protein variants derived from flanking *A. phosphatis* populations and not from the dominant *A. phosphatis* strain in

Figure 3



Time-resolved community proteomics highlighting subtle differences in protein abundance at the end of the anaerobic (120 min) and aerobic (330 min) phases within activated sludge performing enhanced biological phosphorus removal and dominated by *A. phosphatis*. **(a)** Two-dimensional polyacrylamide gel electrophoresis protein expression profiles. **(b)** Gene Expression Dynamics Inspector [53] profiles of LC–MS/MS normalized spectral abundance factor data. Interesting proteins that are slightly more abundant in the anaerobic phase are highlighted. Proteins in the highlighted clusters include a phosphate transport regulator (distant homolog of PhoU), an inorganic pyrophosphatase, an acyl-CoA dehydrogenases (fatty acid beta oxidation) and a nucleoside diphosphate kinase (governs the relative levels of GTP and ATP in cells).

the sequenced sludges. A significant subset of these was involved in core-metabolism and EBPR-specific pathways, suggesting an essential role for genetic diversity in maintaining the stable performance of microbial-mediated wastewater treatment (Figure 2).

The most extensive community proteomic analyses to date have been performed on acid mine biofilms that exhibit comparatively low diversity [13••,14••,38••]. The mine is characterized by low pH (~0.8) and microbially mediated iron oxidation that contributes to the acid mine drainage production. Here, an initial shotgun proteomic approach allowed the identification of more than 2000 proteins [13••]. High protein coverage (48%) was obtained for the dominant microorganism (*Leptospirillum* group II). One highly abundant protein, annotated as a hypothetical, was further investigated and found to be an iron oxidizing cytochrome (Cyt_{579}), a key component of energy generation in the biofilms [13••]. Thus, the proteomic results were instrumental in guiding the ensuing biochemical investigations [41]. Lo *et al.* demonstrated that the high-resolution tandem mass spectrometry could differentiate between peptides originating from discrete populations within the mixed microbial community [38••]. By assigning peptides to two different sequenced *Leptospirillum* group II populations they inferred the genome architecture of a third unsequenced *Leptospirillum* group II population and demonstrated extensive

interpopulation recombination [38••]. This approach was expanded to conduct an extensive *Leptospirillum* group II proteomic genotype survey from 27 distinct biofilm samples [14••]. The protein expression patterns suggest selection for particular recombinant types and, thus, revealed that recombination is a mechanism for fine-scale adaptation [14••], demonstrating the power of integrated genomics and proteomics to contribute extensively to our understanding of microbial ecology and evolution.

Challenges and future perspectives

So far, the application of community proteomics has provided unprecedented functional insight into microbial communities with limited diversity and/or that are dominated by a particular organismal group [13••,14••,16••,38••]. Detection limits for community proteomics suggest that each organism for which a protein is identified must be present at an abundance of at least a few percent of the total community. This represents a major limitation in ecosystems that harbor extensive species richness, for example, 10^6 taxa in a gram of soil [42]. Further complications for comprehensive proteomic coverage include the unevenness of species distribution within samples, the fact that protein expression levels within a cell may differ by six orders of magnitude [43], and the extensive fine-scale genetic heterogeneity within microbial populations [1]. Consequently, comprehensive proteomic analyses of mixed communities are challenging, and with current technology a typical analysis may only resolve $\ll$1% of the protein complement within diverse samples [25]. However, the range of detectable proteins will improve with future technical developments in proteomics especially advances in LC and MS. In addition, complex proteomes may be reduced in complexity before LC. Dividing protein complements into many fractions before LC, such as by IEF, holds promise to expand the dynamic range extensively [24]. Although, DNA sequencing technology is currently advancing at an astonishing rate, the implementation of MS-based *de novo* peptide sequencing will diminish the requirement for comprehensive genomic (transcriptomic) foundations and will allow the identifications of proteins from low abundance community members for which no genomic sequences are available.

Because of the ability of being able to infer taxonomic and functional information from protein expression data, proteomics lends itself ideally to the monitoring of community structure and function over space and time (Figure 3). For example changes in protein abundance may be monitored between different natural conditions, for example, diurnal or anaerobic/aerobic cycles. Rapid changes within the environmental conditions may be manifested at the post-translational level [44], and the MS and bioinformatic methodologies need to be refined to detect these. Another useful approach to detect rapid changes in protein expression is to observe incorporation of stable isotope-label or radio-label into newly synthesized proteins [45].

Although 2DE is labor intensive and has limitations regarding separation of proteins, it remains convenient for expression quantification and comparative studies. For example, multiple samples differentiated by fluorescent tags (known as DIGE) can be run on the same gels [46], or metabolically active portions of communities can be detected by incorporation of a labeled substrate [47•]. Given the potential superiority of LC–MS/MS approaches it is highly desirable to obtain quantitative information to detect systems-level responses to change. However, this is not so readily obtained from MS data, mainly because of the large variation in individual peptide chemistry. Furthermore, quantification may be confounded by the complexity of peptides such that only subsets of proteins may be identified from a sample [48•]. Nonetheless, emerging techniques for quantifying proteins from LC–MS/MS data include isotope-coded affinity tags (ICAT), metabolic labeling of proteins (using ^{13}C or ^{15}N) and isobaric tags for quantification (iTRAQ) [49]. These labeling techniques allow simultaneous analysis of multiple samples for comparison of the differentially tagged peptide abundances. Alternative methods to quantify MS data are commonly used, the so-called 'label-free' methods [48•]. One such method, spectral counting, relates the number of peptides detected, normalized to protein size to protein abundance. An alternative approach uses the peptide MS peak signal intensity, which is collected for each peptide during the chromatographic spread [49]. The subsequent chromatograph peak area is then proportional to the peptide's abundance. These 'label-free' methods are reportedly not as accurate as labeling approaches for quantification; however, there is much interest to use these simple approaches and application will increase as statistical treatment continues to improve [48•,50].

So far, community proteomic studies have been carried out on bulk samples. However, microbial communities exhibit distinct organismal and functional organization [51] and particular enzyme variants may be localized within distinct microniches [16••]. Hence, more fine-scale measurements will be necessary in future to resolve the functional significance of protein localization within microbial communities. In particular, mass spectrometry imaging techniques [52] show great promise for resolving fine-scale expression differences within microbial communities.

Conclusion

Community proteomics is providing unprecedented insight into genotypic and phenotypic traits within microbial consortia. In addition to other systems-level data that include genomics, transcriptomics, and metabolomics,

proteomics is providing high-resolution molecular data that is allowing us to glean a more complete picture of microbial community composition, function, physiology, interaction, ecology, and evolution. Such fundamental knowledge is essential for our understanding of the Earth's biogeochemical cycles, biotechnologies that rely on microbial communities as well as human health.

Acknowledgements

PW is supported by a United States Department of Energy Genomics: GTL Program (Office of Science) grant. PB acknowledges support by the Biotechnology and Biological Sciences Research Council, Industrial Partnership Grant BB/D013348/1, and by the Environmental Biotechnology Cooperative Research Centre (Australia).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wilmes P, Simmons SL, Denef VJ, Banfield JF: **The dynamic genetic repertoire of microbial communities**. *FEMS Microbiol Rev* 2009, **33**:109-132.

2. Martin HG, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He SM, Salamov AA, Szeto E *et al.*: **Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities**. *Nat Biotechnol* 2006, **24**:1263-1269.

3. •• Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF: **Community structure and metabolism through reconstruction of microbial genomes from the environment**. *Nature* 2004, **428**:37-43.

The first account of random shotgun sequencing of a microbial community to reconstruct genomes for uncultured microorganisms and provide the genomic platform for subsequent community proteomic studies.

4. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu DY, Paulsen I, Nelson KE, Nelson W *et al.*: **Environmental genome shotgun sequencing of the Sargasso Sea**. *Science* 2004, **304**:66-74.

5. Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE: **Metagenomic analysis of the human distal gut microbiome**. *Science* 2006, **312**:1355-1359.

6. •• Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N *et al.*: **Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite**. *Nature* 2007, **450**:560-565.

An intergrative metagenomic and proteomic study of the termite hindgut, providing insight into the microbial communities involved in lignocellulose degradation.

7. Ehlers MM, Cloete TE: **Protein profiles of phosphorus- and nitrate-removing activated sludge systems**. *Water SA* 1999, **25**:351-356.

8. Ogunseitan OA: **Protein profile variation in cultivated and native freshwater microorganisms exposed to chemical environmental pollutants**. *Microb Ecol* 1996, **31**:291-304.

9. Wilmes P, Wexler M, Bond PL: **Metaproteomics provides functional insight into activated sludge wastewater treatment**. *PLoS ONE* 2008, **3**:.

10. Santoni V, Molloy M, Rabilloud T: **Membrane proteins and proteomics: un amour impossible?** *Electrophoresis* 2000, **21**:1054-1070.

11. •• VerBerkmoes NC, Denef VJ, Hettich RL, Banfield JF: **Functional analysis of natural microbial consortia using community proteomics**. *Nat Rev Microbiol* 2009, **7**:196-205.

An excellent review on community proteomics, highlighting the significance of strain-resolved proteogenomics.

12. Wu CC, MacCoss MJ, Howell KE, Yates JR: **A method for the comprehensive proteomic analysis of membrane proteins**. *Nat Biotechnol* 2003, **21**:532-538.

13. •• Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, Shah M, Hettich RL, Banfield JF: **Community proteomics of a natural microbial biofilm**. *Science* 2005, **308**:1915-1920.

This comprehensive community proteomic investigation obtained high coverage of proteins from the dominant organism and led to a number of subsequent fruitful studies of AMD biofilms.

14. •• Denef VJ, VerBerkmoes NC, Shah MB, Abraham P, Lefsrud M, Hettich RL, Banfield JF: **Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation**. *Environ Microbiol* 2009, **11**:313-325.

Use of proteogenomics to reveal the significance of recombination for environmental adaptation in *Leptospirillum* group II strains.

15. • VerBerkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL *et al.*: **Shotgun metaproteomics of the human distal gut microbiota**. *ISME J* 2008, **3**:179-189.

An extensive community proteome examination of human fecal samples that demonstrates the use of the shotgun proteomic approach for the characterization of a highly complex sample.

16. •• Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B, Hettich RL, Bond PL, VerBerkmoes NC, Banfield JF: **Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal**. *ISME J* 2008, **2**:853-864.

An extensive community proteomic study that provides new insight into EBPR metabolism and highlights extensive strain variant protein expression among phosphorus removing organisms.

17. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data**. *Electrophoresis* 1999, **20**:3551-3567.

18. Eng JK, McCormack AL, Yates JR: **An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database**. *J Am Soc Mass Spectrom* 1994, **5**:976-989.

19. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR: **Protein sequencing by tandem mass-spectrometry**. *Proc Natl Acad Sci U S A* 1986, **83**:6233-6237.

20. Shevchenko A, Sunyaev S, Loboda A, Shevehenko A, Bork P, Ens W, Standing KG: **Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching**. *Anal Chem* 2001, **73**:1917-1926.

21. • Wilmes P, Bond PL: **The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms**. *Environ Microbiol* 2004, **6**:911-920.

The first report of a complete mixed community proteomic investigation involving protein isolation and identification following MS. This study was successful in obtaining protein identifications in the absence of corresponding metagenomic sequences.

22. • Lacerda CMR, Choe LH, Reardon KF: **Metaproteomic analysis of a bacterial community response to cadmium exposure**. *J Proteome Res* 2007, **6**:1145-1152.

Computer-based *de novo* peptide sequence analysis is used to identify proteins from a community sample in the absence of metagenomic sequences.

23. • Denef VJ, Shah MB, VerBerkmoes NC, Hettich RL, Banfield JF: **Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis**. *J Proteome Res* 2007, **6**:3152-3161.

This study demonstrates the effect of strain-level variation limiting protein identification by LC–MS/MS spectral matching.

24. Ye ML, Jiang XG, Feng S, Tian RJ, Zou HF: **Advances in chromatographic techniques and methods in shotgun proteome analysis**. *Trends Anal Chem* 2007, **26**:80-84.

25. Wilmes P, Bond PL: **Metaproteomics: studying functional gene expression in microbial ecosystems**. *Trends Microbiol* 2006, **14**:92-97.

26. Kan J, Hanson TE, Ginter JM, Wang K, Chen F: **Metaproteomic analysis of Chesapeake Bay microbial communities**. *Saline Syst* 2005, **1**:7-19.

27. • Benndorf D, Balcke GU, Harms H, von Bergen M: **Functional metaproteome analysis of protein extracts from contaminated soil and groundwater**. *ISME J* 2007, **1**:224-234.
Metaproteomics was used to detect proteins expressed in complex water and soil samples relevant to the degradation of contaminants.

28. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hugler M, Albrecht D, Robidart J, Bench S, Feldman RA *et al.*: **Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila***. *Science* 2007, **315**:247-250.

29. Klaassens ES, de Vos WM, Vaughan EE: **Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract**. *Appl Environ Microbiol* 2007, **73**:1388-1392.

30. Maron PA, Ranjard L, Mougel C, Lemanceau P: **Metaproteomics: a new approach for studying functional microbial ecology**. *Microb Ecol* 2007, **53**:486-493.

31. Pierre-Alain M, Christophe M, Severine S, Houria A, Philippe L, Lionel R: **Protein extraction and fingerprinting optimization of bacterial communities in natural environment**. *Microb Ecol* 2007, **53**:426-434.

32. Park C, Helm RF: **Application of metaproteomic analysis for studying extracellular polymeric substances (EPS) in activated sludge flocs and their fate in sludge digestion**. *Water Sci Technol* 2008, **57**:2009-2015.

33. Park C, Novak JT, Helm RF, Ahn YO, Esen A: **Evaluation of the extracellular proteins in full-scale activated sludges**. *Water Res* 2008, **42**:3879-3889.

34. Toyoda A, Iio W, Mitsumori M, Minato H: **Isolation and identification of cellulose-binding proteins from sheep rumen contents**. *Appl Environ Microbiol* 2009, **75**:1667-1673.

35. • Schulze WX, Gleixner G, Kaiser K, Guggenberger G, Mann M, Schulze ED: **A proteomic fingerprint of dissolved organic carbon and of soil particles**. *Oecologia* 2005, **142**:335-343.
This early study performed community profiling using shotgun proteomics to broadly categorize proteins within complex soil and water samples.

36. Powell MJ, Sutton JN, Del Castillo CE, Timperman AI: **Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry**. *Mar Chem* 2005, **95**:183-198.

37. Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL, Rappe MS, Laney S, Wilhelm LJ, Tripp HJ *et al.*: **Proteorhodopsin in the ubiquitous marine bacterium SAR11**. *Nature* 2005, **438**:82-85.

38. •• Lo I, Denef VJ, VerBerkmoes NC, Shah MB, Goltsman D, DiBartolo G, Tyson GW, Allen EE, Ram RJ, Detter JC *et al.*: **Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria**. *Nature* 2007, **446**:537-541.
This proteogenomic study detected functional strain-level genome rearrangements in the dominant Le AMD biofilm communities.

39. •• Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, Carlson CA, Smith RD, Giovanonni SJ: **Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea**. *ISME J* 2009, **3**:93-105.
This analysis of ocean communities indicates that a large proportion detected were periplasmic substrate-binding proteins, providing insight into marine ecosystem function.

40. Wilmes P, Bond PL: **Towards exposure of elusive metabolic mixed-culture processes: the application of metaproteomic analyses to activated sludge**. *Water Sci Technol* 2006, **54**:217-226.

41. Jeans C, Singer SW, Chan CS, VerBerkmoes NC, Shah M, Hettich RL, Banfield JF, Thelen MP: **Cytochrome 572 is a conspicuous membrane protein with iron oxidation activity purified directly from a natural acidophilic microbial community**. *ISME J* 2008, **2**:542-550.

42. Gans J, Wolinsky M, Dunbar J: **Computational improvements reveal great bacterial diversity and high metal toxicity in soil**. *Science* 2005, **309**:1387-1390.

43. Tyers M, Mann M: **From genomics to proteomics**. *Nature* 2003, **422**:193-197.

44. Mann M, Jensen ON: **Proteomic analysis of post-translational modifications**. *Nat Biotechnol* 2003, **21**:255-261.

45. Savijoki K, Suokko A, Palva A, Varmanen P: **New convenient defined media for [S-35]methionine labelling and proteomic analyses of probiotic lactobacilli**. *Lett Appl Microbiol* 2006, **42**:202-209.

46. Lilley KS, Dupree P: **Methods of quantitative proteomics and their application to plant organelle characterization**. *J Exp Bot* 2006, **57**:1493-1499.

47. • Jehmlich N, Schmidt F, von Bergen M, Richnow HH, Vogt C: **Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures**. *ISME J* 2008, **2**:1122-1133.
This study applies stable isotope probing with community proteomics, a useful approach for directed functional studies, not limited to 2DE.

48. • Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B: **Quantitative mass spectrometry in proteomics: a critical review**. *Anal Bioanal Chem* 2007, **389**:1017-1031.
An informative review of quantitative and statistical analyses of MS proteomic data.

49. Ong SE, Mann M: **Mass spectrometry-based proteomics turns quantitative**. *Nat Chem Biol* 2005, **1**:252-262.

50. Choi H, Fermin D, Nesvizhskii AI: **Significance analysis of spectral count data in label-free shotgun proteomics**. *Mol Cell Proteomics* 2008, **7**:2373-2385.

51. Wilmes P, Remis JP, Hwang M, Auer M, Thelen MP, Banfield JF: **Natural acidophilic biofilm communities reflect distinct organismal and functional organization**. *ISME J* 2009, **3**:266-270.

52. Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM: **Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues**. *Nat Med* 2001, **7**:493-496.

53. Eichler GS, Huang S, Ingber DE: **Gene expression dynamics inspector (GEDI): for integrative analysis of expression profiles**. *Bioinformatics* 2003, **19**:2321-2322.