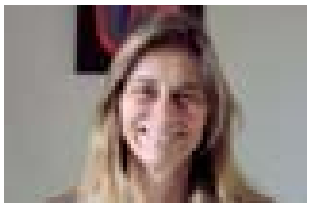
Metaproteomics for analysis of microbial function in the environment



Philip L Bond[1]* *& Margaret Wexler*[2]

[1]Advanced Wastewater Management Centre, The University of Queensland, Brisbane QLD 4072 Australia, [2]School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK
Tel: (+61) 7 3346 7841
Fax: (+61) 7 3365 4726
Email * phil.bond@uq.edu.au

This report briefly describes the use of proteomic analyses to examine protein expression directly from environmental samples (termed metaproteomics). This approach has potential for solving one of the major challenges facing microbial ecologists, by providing insight into microbial function directly within samples.

The emerging opportunity

It has become increasingly important to study microbial communities directly in their environments. Recent molecular analysis of environmental samples has vastly increased understanding of microbial diversity, but the next big challenge is to understand details of function in these environments and particularly to link this to phylogenetic information.

Recent metagenomic sequencing projects that analyse genomic DNA directly from environmental samples are providing opportunities to make the above link. These studies vastly expand our knowledge of the genetic diversity and the physiological and metabolic potential within selected environments that include seawater samples [1,2], an acid mine biofilm [3] and activated sludge [4]. The exponentially escalating DNA sequence information (genomic and metagenomic) provides tremendous potential for application of high throughput functional approaches. Transcriptomics and proteomics can be applied directly within mixed culture to detect expression profiles and provide functional insight of microbial environments.

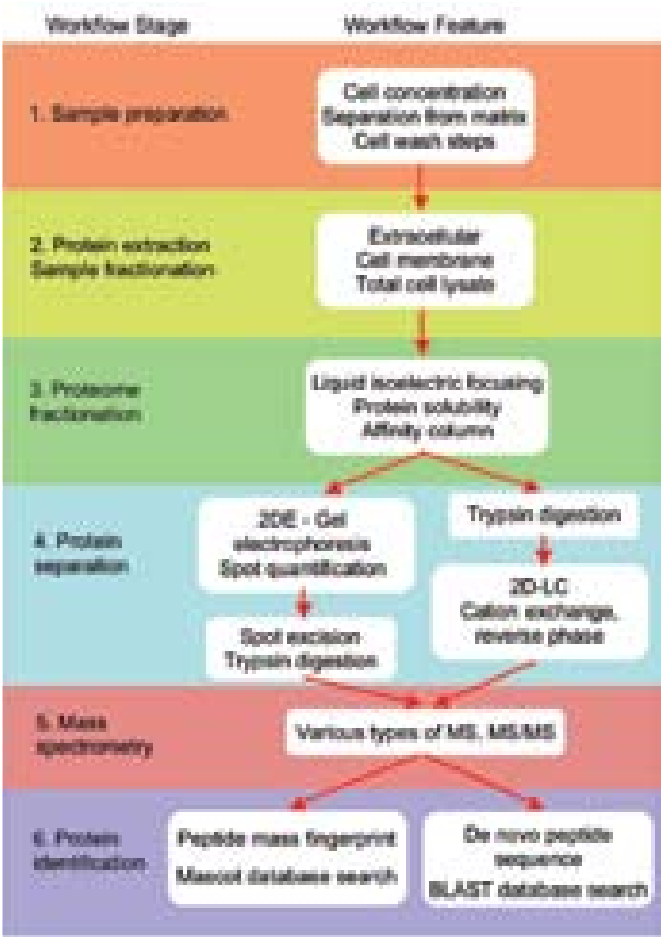
MICRO-FACT

Metagenomics has been used to study a number of diverse environmental samples including the human gut, symbionts of a marine worm, wastewater sludge, soil, water from an iron mine, sunken whale skeletons and ocean waters.

The procedure for metaproteomic analysis is basically that utilised for proteomic study of pure culture. This involves: i) sample preparation; ii) protein extraction; iii) separation of the proteins or peptides, usually in two dimensions; iv) mass spectrometry (MS) analysis for identification of the proteins (Figure 1). Detecting protein expression from environmental samples is not new [5,6], but interest has recently grown. Sequence data, improved protein separation techniques and the rapidly improving protein identification by mass spectrometry, provide new opportunities to apply large-scale proteomics and protein identification to environmental samples.

Major challenges associated with metaproteomics include adequate protein or peptide separation and identification of the proteins. The separations are typically performed by either two-dimensional polyacrylamide gel electrophoresis (2DE), or

Figure 1 Workflow for a metaproteomic analysis consisting of six stages – each provides some challenges and potential problems. See text for acronym descriptions.



liquid chromatography (LC). For 2DE, gels provide a visual representation of the proteome and are convenient for comparison of protein spot density between gels (Figure 2). Gel images are analysed to detect differential protein expression in response to a change in condition or community function. Chosen spots are excised and trypsin digested for mass spectrometry analysis. For LC, the protein mixture is trypsin digested prior to separation and the separated peptides then flow directly into the mass spectrometer for identification (Figure 1).

Following protein or peptide separation, MS most commonly achieves identifications if metagenomic sequence is available. For example, MS or tandem MS (MS/MS) can be used to generate peptide mass fingerprints (PMF) for protein identification [7] (Figure 2). Another approach for identification is to estimate the *de novo* protein sequence from the MS/MS data, then search for homologous sequences. The latter requires extra effort compared to the PMF approach; however, *de novo* sequencing is especially useful for protein identification when corresponding metagenomic sequence data are unavailable. As pure culture microbial genome sequencing ascends, this increases the opportunity to find sufficient protein sequence homology across species and thus attain identification of proteins extracted and analysed from organisms whose genome has not been sequenced. This approach was recently used to identify more than 100 proteins that were differentially expressed following exposure of bacterial communities to cadmium [8].

Examples of metaproteomic studies

So far only a handful of studies in the literature examine the proteome of mixed culture samples. These studies include detection of proteins in high abundance during biological phosphorus removal in activated sludge wastewater treatment [9,10]. Proteins associated with dissolved organic matter in soil and water have been analysed to detect the presence of broad taxonomic groups of microorganisms [11]. Expression profiles that have been examined include: an estuary transect [12], infant faecal samples [13] and freshwater samples following exposure to heavy metals [14]. In a landmark study, high-throughput proteomic analyses have recently been performed on acid mine biofilms [15] (see more below).

Metaproteomics was first applied to a laboratory-scale activated sludge reactor [9]. In that study, comparisons of proteome profiles are made to determine metabolic details of a wastewater treatment process known as enhanced biological phosphorus removal (EBPR). Large-scale protein separation was performed by 2DE. Protein expression was compared between the two operational stages (anaerobic and aerobic) of EBPR. However, only minor differences in protein levels were detected (Figure 2). Subsequently, these studies have focused on identification of highly expressed proteins. Initially these were identified by MS/MS *de novo* sequencing of peptides [9]. However, when metagenome sequences of EBPR sludges became available [4], this facilitated protein identification by analysis of PMF patterns. By this means, over 30% of highly expressed proteins chosen from 2DE gels, could be matched to the dominant organism on the metagenome database. These identified proteins are potentially involved in major EBPR carbon transformations (polyhydroxyalkanoate synthesis, glycogen metabolism, tricarboxylic acid cycle reactions) and phosphate transport (unpublished data). These results of comparative expression offer insight into EBPR biochemistry and enable refinement of the EBPR metabolic model.

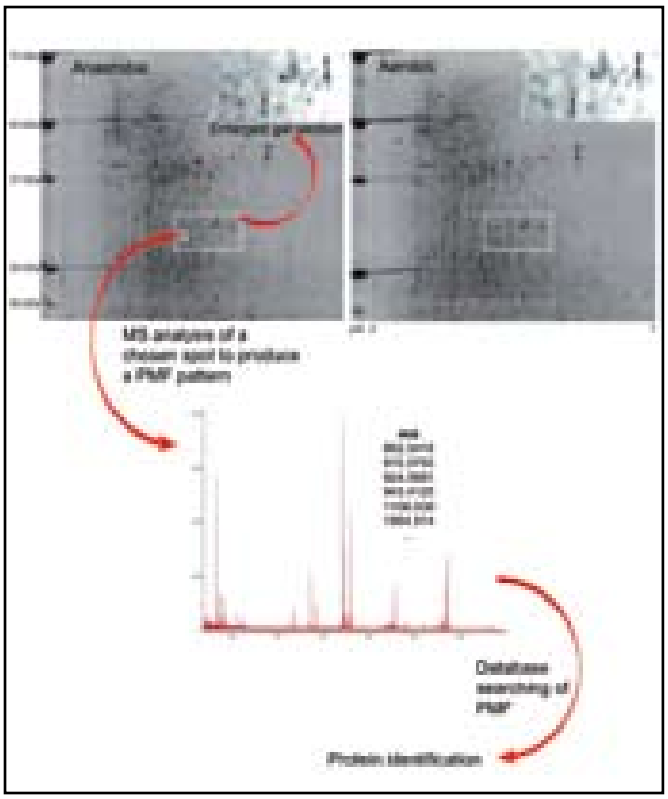


Figure 2. 2DE gels and MS of proteins expressed in the activated sludge phosphorus removal process (images provided by Paul Wilmes). The 2DE first dimension separation (horizontal) was based on protein charge and the second on molecular weight. Protein expression is compared here between the anaerobic and aerobic stages of the phosphorus process in wastewater treatment plants. Differential protein expression is highlighted in enlargements. Spots of interest are excised and analysed by MS to produce PMF patterns for protein identification. (System is described in the text).

MICRO-FACT

'Bacterial/microbial flora' is a misnomer since these organisms are not plants – the correct term is 'microbiota'.

The most extensive metaproteomic analysis to date was performed on an acid mine biofilm of low diversity [15]. The mine is characterised by low pH (~0.8) and microbially mediated iron oxidation that contributes to the acid mine drainage production. Here, liquid chromatography was used to separate the protein mixture following protein extraction and trypsin digestion. This LC-MS/MS approach utilises 2D chromatographic separation (typically strong cation exchange with reversed phase), coupled with tandem MS for fragmentation and peptide mass pattern matching to a database [7,16]. Thus, corresponding DNA sequence is required for protein identification. In this case the metagenomic data set was of a similar biofilm from another part of the mine [3]. From the proteins identified (~2000) a high coverage (48%) of the predicted proteins for the dominating microorganism *(Leptospirillum* sp.) was obtained [15]. One highly abundant protein, annotated as a hypothetical, was further investigated and found to be an iron oxidising cytochrome, a key component of the energy generation in these biofilms. Here the proteomic results were instrumental in guiding the ensuing biochemical investigations.

Future directions

Proteomic analysis of mixed communities is challenging, especially in complex samples such as soil, as a typical analysis may only resolve <1% of the metaproteome [17]. Nevertheless, mixed community studies are exciting and timely given the improved techniques and capabilities of proteomics and environmental genome analyses. The approach holds great promise for comparative analysis to examine response to a range of environmental perturbations such as stress and redox, and for monitoring metabolic and physiological activities.

There are a number of potential metaproteomic applications suited to the different protein separation techniques (2DE and LC). Although 2DE is labour intensive, presently it is preferred for quantification of expression and comparative studies. There are also some useful in-gel aspects of 2DE. For example, the use of fluorescent labelling to run multiple samples on a common gel, a system known as DIGE, allows more accurate spot matching [18]. Another useful application is for incorporation of radiolabel to detect newly synthesised proteins [19]. The LC-MS/MS approach, together with advanced *de novo* sequencing [20], holds much promise for high throughput metaproteomics. Additionally, quantitative analysis of LC-MS data has recently been achieved in pure culture studies [21], and likely metaproteomic studies will follow.

References

1. DeLong EF, Preston CM, Mincer T *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 2006; 311:496-503.
2. Venter JC, Remington K, Heidelberg JF *et al*. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004; 304:66-74.
3. Tyson GW, Chapman J, Hugenholtz P *et al*. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 2004; 428:37-43.
4. Martin HG, Ivanova N, Kunin V *et al*. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnol* 2006; 24:1263-9.
5. Ehlers MM & Cloete TE. Protein profiles of phosphorus- and nitrate-removing activated sludge systems. *Water SA* 1999; 25:351-6.
6. Ogunseitan OA. Protein profile variation in cultivated and native freshwater microorganisms exposed to chemical environmental pollutants. *Microbial Ecol* 1996; 31:291-304.
7. Aebersold R & Mann M. Mass spectrometry-based proteomics. *Nature* 2003; 422:198-207.
8. Lacerda CMR, Choe LH & Reardon KF. Metaproteomic analysis of a bacterial community response to cadmium exposure. *J Proteome Res* 2007; 6:1145-52.
9. 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; 6:911-20.
10. Wilmes P & Bond PL. Towards exposure of elusive metabolic mixed-culture processes: the application of metaproteomic analyses to activated sludge. *Water Science Technol* 2006; 54:217-26.
11. Schulze WX, Gleixner G, Kaiser K *et al*. A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* 2005; 142:335-43.
12. Kan J, Hanson TE, Ginter JM *et al*. Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline Systems* 2005; 1:7-19.
13. 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; 73:1388-92.
14. Maron PA, Ranjard L, Mougel C *et al*. Metaproteomics: a new approach for studying functional microbial ecology. *Microbial Ecol* 2007; 53:486-93.
15. Ram RJ, VerBerkmoes NC, Thelen MP *et al*. Community proteomics of a natural microbial biofilm. *Science* 2005; 308:1915-20.
16. Link AJ, Eng J, Schieltz DM *et al*. Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol* 1999; 17:676-82.
17. Wilmes P & Bond PL. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* 2006; 14:92-7.
18. Lilley KS & Dupree P. Methods of quantitative proteomics and their application to plant organelle characterization. *J Exp Botany* 2006; 57:1493-9.
19. Savijoki K, Suokko A, Palva A *et al*. New convenient defined media for [^{35}S] methionine labelling and proteomic analyses of probiotic lactobacilli. *Letts App Microbiol* 2006; 42:202-9.
20. Ma B, Zhang KZ, Hendrie C *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrometry* 2003; 17:2337-42.
21. DeSouza L, Diehl G, Rodrigues MJ *et al.* Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J Proteome Res* 2005; 4:377-86.

MICRO-FACT

Metagenomics is the study of the collective genomes (DNA sequences) of microorganisms as they exist in complex microbial communities.