Proteogenomics Reveals Key Insight Into The Microbial Activities Of Enhanced Biological Phosphorus Removal In Activated Sludge

M.G. Lefsrud[1], P. Wilmes[2], N.C. VerBerkmoes[1], M. Wexler[3], P.E. Abraham[1], M. Shah[1], P.L. Bond[4], J.F. Banfield[2], R.L. Hettich[1]

1) Oak Ridge National Laboratory, Oak Ridge, TN 2) University of California, Berkeley, CA 3) University of East Anglia, Norfolk, England 4) The University of Queensland, Brisbane, Queensland, Australia

OVERVIEW

- Biological wastewater treatment plants employing activated sludge rely on complex microbial communities to perform an array of different metabolic transformations (Figure 1).
- Excess inorganic phosphate in effluent water can have a negative effect on water quality and ecosystem nutrient balance.
- Present within this bacterial community is *Candidatus* "Accumulibacter Phosphatis", which is a polyphosphate accumulating organism with prime importance in enhanced biological phosphorus removal (EBPR) systems.
- The identities and metabolic transformations of these microorganisms remain largely unknown.
- With the advent of high throughput cultivation independent genomics (Tyson, *Nature*, 2004; Venter, *Science*, 2004) and cultivation independent proteogenomics techniques (Ram, *Science*, 2005), it is now possible to study these microbial assemblages and phosphorus removing microorganisms *in situ* to deduce future process optimization strategies.
- One of the most challenging aspects of these types of measurements is the dynamic range of the species and proteins they express in the sample.
- Microbial communities can vary from a few microbial species in extreme environments or hundreds to thousands in soil and water samples.
- This study explores the challenges of these communities and how new mass spectrometry technology can further enable deeper more accurate measurements.

EXPERIMENTAL

Samples and Sample Preparation

- The EBPR samples used in this study were grown within a laboratory-scale sequencing batch reactor performing EBPR at 55 mg/l phosphorus in the influent (Figure 2).
- The reactor was seeded with a microbial community from an English waste water treatment facility.
- Cellular material was harvested, washed, and the whole cells lysed by French Press.
- The proteins were denatured, reduced, and digested into peptides with trypsin.

LC/LC-MS/MS and Informatics

- All samples were analyzed via two-dimensional (2D) nano-LC MS/MS system with a split-phase column (RP-SCX-RP) on a linear ion trap (LTQ) and hybrid linear ion-trap LTQ-Orbitrap (Thermo Finnigan) (Figure 3) operated in data-dependent MS/MS mode with technical duplicates.
- All MS/MS spectra were searched with the SEQUEST algorithm and filtered at 2 different filter levels with DTASelect at the peptide level [Xcorrs of at least (Low) 1.5 (+1), 2.0 (+2) 3.0 (+3) or (High) 1.8 (+1), 2.5 (+2) 3.5 (+3)]. Only proteins identified with *two fully tryptic peptides* were kept.
- False positive levels were estimated by reverse database searching.
- Databases included the US and Australian metagenomes (Martin, *Nature Biotechnology*, 2006).
- High mass accuracy monoisotopic peaks were extracted from the Orbitrap full scans and compared to theoretical monoisotopic masses for the identified peptides.

RESULTS AND DISCUSSION

- Three different metagenome databases were assembled for these communities. These three databases were from similar EPBR microbial communities from the United States and Australia which were characterized by cultivation independent genomics techniques (Martin, *Nature Biotechnology*, 2006).
- The majority of identified proteins measured by MS/MS were from the uncultured polyphosphate accumulating bacterium, *Candidatus* "Accumulibacter Phosphatis" (Figure 4).
- The three databases were compared and revealed no significant statistical differences in the measurements (data not shown). However, all further analyses focused on only one US annotation (US Phrap) database.
- The MS/MS data resulted in over 5300 unique peptides from over 6900 total peptides identified (Table 1).
- Database searches resulted in over 1112 unique proteins from 1750 redundant identified proteins (Table 1, Figure 5). However, of the redundant proteins identified, over 300 belong to *Candidatus* (over 19%).
- The majority of conserved proteins, 188, were present during both growth states (anaerobic and aerobic), with 56 proteins only expressed during the aerobic growth state and 36 only during the anaerobic growth state (Figure 6).
- These conserved proteins are directly linked to the metabolic transformations observed in EBPR including enzymes involved in polyhydroxyalkanoate synthesis, glycolysis, gluconeogenesis, glycogen synthesis, glyoxylate/TCA cycle, fatty acid beta-oxidation, fatty acid synthesis, phosphate transport and polyphosphate metabolism. Similar to an earlier proteomic investigation of EBPR, implications include the involvement of fatty acid metabolism, aerobic activity of the glyoxylate shunt, and phosphate uptake and release mediated by the Pst system.
- The vast majority, greater than 75%, of identified MS/MS data with defined mass accuracy was within <± 10 ppm (Figure 7 and Figure 10).
- Less than 14% of of the identified MS/MS data was greater than ± 10 ppm for the low filter and less than 7% for the high filter (Figure 7).
- However, approximately 10.3% of identified MS/MS data collected in the Orbitrap did not provide accurate masses due to low signal to noise at the low filter levels (unresolved m/z). However when the filters were raised to the high levels, this number dropped to 8.5% (Figure 7).
- False positive rates (FPR) were highest with the LTQ low filter level at 17.8%, then decreased to 10.9% Orbitrap low filter, and decreased further to 0.79% for Orbitrap data between <± 10 ppm. At the high filter level, both the LTQ and Orbitrap were approximately 1% FPR, with a reduction to 0.04% for data <± 10 ppm (Figure 8).
- The total peptide count, highest with the LTQ and the low filter level, of 8959 decreased to 7951 when using the Orbitrap. When only Orbitrap data between <± 10 ppm was used, this dropped to 6061. At the high filter level, the LTQ yielded 5776 peptides and Orbitrap gave 5352, with a reduction to 4538 for data <± 10 ppm (Figure 9).
- The reduction in peptides from the largest count was 11% from the LTQ low filter to the Orbitrap low filter and 32%, 35%, 40%, and 49% from the LTQ low filter to the Orbitrap low filter <± 10 ppm, LTQ high filter, Orbitrap high filter and Orbitrap low filter <± 10 ppm, respectively.
- For the combined data, the average mass accuracy varied between 0.004 amu to 0.094 amu and from 3.0 ppm to 5.6 ppm over the 2 biological states and 2 technical replications. However, some fo the Orbitrap data was incorrect in assigning a charge states to the peptides MS/MS resulting in a significant number of peptides measured at a delta AMU whole integer (Figure 11).

MASS ACCURACY RESULTS

CONCLUSIONS

- The total peptide and proteins identified were greatest using the LTQ at low filter levels and decreased in sequence: Orbitrap low filter, LTQ high filter, and finally Orbitrap high filter.
- This suggests the LTQ actually has greater dynamic range in MS/MS scans than the Orbitrap in full scans.
- Further, restricting the data set to only values between 10 and -10 ppm resulted in a further reduction in identified peptides and proteins.
- However, the rate of false positives also decreased with each reduction in peptides and proteins. The reduction of peptide and protein count was less than the reduction in false positives.
- Utilization of the high filter levels 1.8 (+1), 2.5 (+2) 3.5 (+3) in conjunction with data between <± 10 ppm from the LTQ Orbitrap reduced the false positive rate to below 0.1%.
- Furthermore, the low filter levels 1.5 (+1), 2.0 (+2) 3.0 (+3) in conjunction with LTQ Orbitrap data between <±10 ppm resulting in a false positive rate below 0.8%.

Figure 1

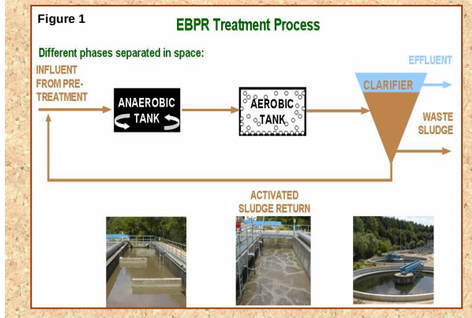


Figure 2

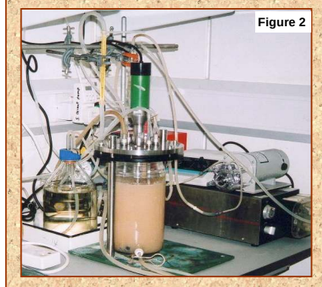


Figure 3

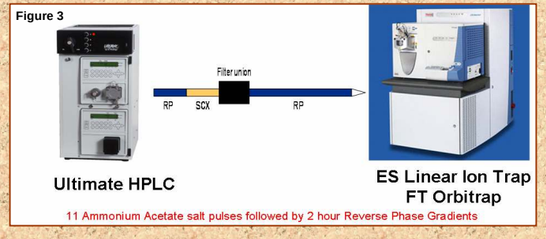


Table 1 — LTQ and Orbitrap EPBR Data

Instrument	Filter Level	Non-Redundant Protein Total Count	Non-Redundant Peptide Forward Count	False Positive (%)	Forward Count of <±10ppm Values (% of Total)	Forward Count of >±10ppm Values (% of Total)	Forward Count of Unresolved Values (% of Total)	False Positive (Values <±10ppm)	Average Forward PPM (Values <±10ppm)
LTQ	Low 1.5,2.0,3.0	2286	8960	17.8%					
LTQ	High 1.8,2.5,3.5	1170	5777	1.04%					
Orbitrap	Low 1.5,2.0,3.0	1884	7952	10.9%	76.1%	13.6%	10.3%	0.79%	4.2
Orbitrap	High 1.8,2.5,3.5	1112	5353	1.05%	84.7%	6.8%	8.4%	0.04%	4.4

Figure 4

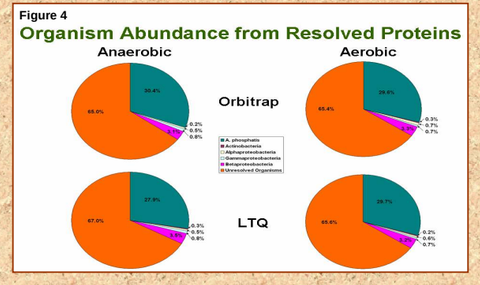


Figure 5

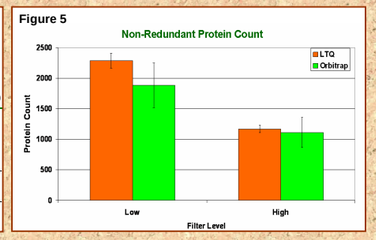


Figure 6

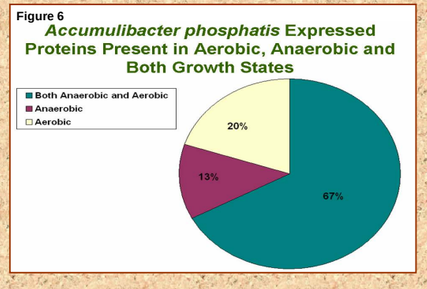


Figure 7

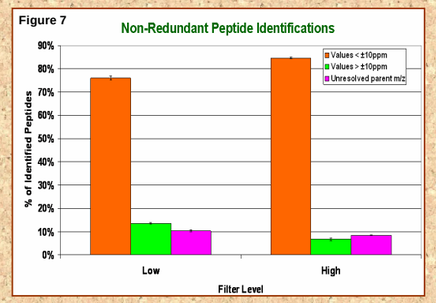


Figure 8

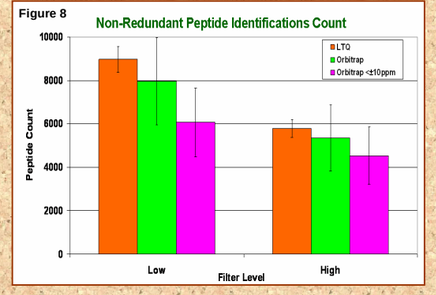


Figure 9

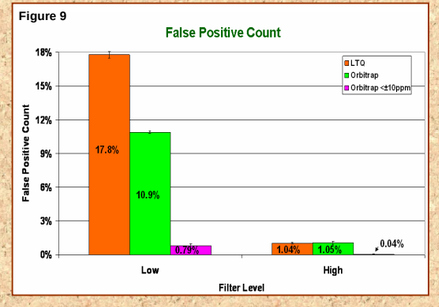


Figure 10

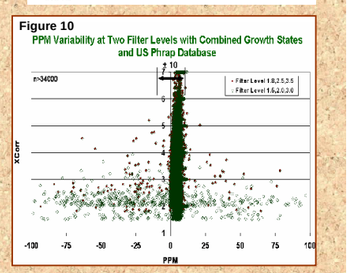
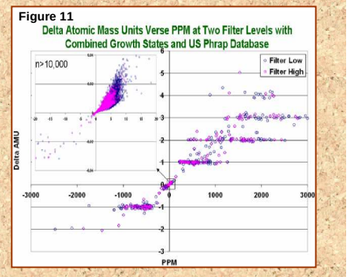


Figure 11



ACKNOWLEDGMENTS

- This research sponsored by U.S. Department of Energy under contract DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.
- We thank Dr. David Tabb and the Yates Proteomics Laboratory at Scripps Research Institute for DTASelect/Contrast software, the Institute for Systems Biology for proteome bioinformatics tools used in analysis of the MS data, and F. Larimer and M. Land of the ORNL Genome Analysis and System Modeling Group for computational resources for proteomic analysis.