Iowa Research Online
The University of Iowa's Institutional Repository

University of Iowa
Iowa Research Online

Theses and Dissertations

Fall 2009

Proteomic investigations of vinyl chloride-assimilating bacteria: from pure cultures to the environment

Adina Shiang Chuang
University of Iowa

Follow this and additional works at: https://ir.uiowa.edu/etd

Part of the Civil and Environmental Engineering Commons

Copyright 2009 Adina Shiang Chuang

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/347

Recommended Citation
Chuang, Adina Shiang. "Proteomic investigations of vinyl chloride-assimilating bacteria: from pure cultures to the environment." PhD (Doctor of Philosophy) thesis, University of Iowa, 2009.
https://doi.org/10.17077/etd.xkwd625l

Follow this and additional works at: https://ir.uiowa.edu/etd

Part of the Civil and Environmental Engineering Commons

المنارة للاستشارات

www.manaraa.com

PROTEOMIC INVESTIGATIONS OF

VINYL CHLORIDE-ASSIMILATING BACTERIA: FROM PURE CULTURES TO

THE ENVIRONMENT

by

Adina Shiang Chuang

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Civil and Environmental Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Assistant Professor Timothy E. Mattes

المنارة للاستشارات

www.manaraa.com

ABSTRACT

Vinyl chloride (VC) is a common groundwater pollutant and known human carcinogen that is commonly produced from the incomplete reductive dechlorination of tetrachloroethene and trichloroethene. A possible approach to clean-up VC-contaminated sites is with microorganisms that grow on VC and/or ethene as a carbon and energy source. However, little is known about the biochemical pathways involved in VC-assimilation within these strains and their distribution and activity *in situ* in the environment. This work uses mass-spectrometry-based proteomics to contribute to the understanding of these microbial communities in both pure cultures and in the environment.

The biochemical pathways of VC and ethene oxidation in *Nocardioides* sp. strain JS614 were studied using proteins identified with a peptide mass fingerprinting approach. New insights into a previously proposed pathway were made using mass spectrometry (MS)-based protein identifications, and potential protein biomarkers for the presence and activity of VC-assimilating bacteria in the environment were identified.

Techniques to extract and identify proteins from various environmental samples such as activated sludge, sediments, soils, and water samples were developed based on preliminary experiments with protein extraction from strain JS614. These techniques were extended to 14 ethene-enrichment cultures derived from VC-contaminated groundwater to detect the presence and diversity of VC- and/or ethene- assimilating bacteria with previously identified protein biomarkers.

VC-assimilating organisms can evolve under laboratory conditions from bacteria that grow on ethene but very little is known about the molecular changes involved. Tandem mass spectrometry and spectral counting were applied to three variants of ethene-grown *Mycobacterium* strain JS623 (a wild type and two VC-adapted variants) resulting in the identification of 174 differentially expressed proteins. The results of this

المنارة للاستشارات

www.manaraa.com

study suggest that VC-degradation in one VC-adapted variant (623-E) may be unique to the other studied variants.

Overall, this study demonstrates the use of proteomics to understand pollutant-degrading microbial communities. From a biochemical level, proteomics was applied as a tool to identify functional genes and their roles in metabolic pathways. At the ecological level, proteomic biomarkers were developed to understand the presence and activities of microbial communities in the environment.

Abstract Approved: ____________________________________
Thesis Supervisor

____
Title and Department

____
Date

المنارة للاستشارات
www.manaraa.com

PROTEOMIC INVESTIGATIONS OF

VINYL CHLORIDE-ASSIMILATING BACTERIA: FROM PURE CULTURES TO

THE ENVIRONMENT

by

Adina Shiang Chuang

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Civil and Environmental Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Assistant Professor Timothy E. Mattes

المنارة للاستشارات

www.manaraa.com

Copyright by

ADINA SHIANG CHUANG

2009

All Rights Reserved

المنارة للاستشارات

www.manaraa.com

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Adina Shiang Chuang

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Civil and Environmental Engineering at the December 2009
graduation.

Thesis Committee: ______________________________
Timothy Mattes, Thesis Supervisor

____
Subramanian Ramaswamy

____
Michelle Scherer

____
Jerald Schnoor

____
Lynn Teesch

المنارة للاستشارات
www.manaraa.com

To my family

المنارة للاستشارات

www.manaraa.com

ACKNOWLEDGMENTS

The past five years have been happily filled with the constant love and support of my best friend and now, husband, Mike, as I have completed this chapter of my life. He has been a calm and gentle listener to my frustrations and mistakes and has been the first to celebrate my achievements. He has patiently endured my absences for this work and always been waiting with a smile and two goofy dogs at his side when I get home. Coming home to him is my daily joy, and I can't wait to start our next adventure together.

My family makes me feel lucky every day, and I love them immensely. My parents were the first generation in our family to move to the United States from Singapore and Thailand and have given up their own personal successes and comforts to provide me with the opportunities of education and growing up in the United States. They have instilled a strong value for education within me and have always been supportive of my studies. My mother, Suree, and her husband, Don, and my sister, Desiree, have always made my trips home and our family vacations filled with love and laughter (and notably great food!)

Tim, who calls me a difficult student, has allowed me to find my own path in this research. He has patiently and enthusiastically supported me as I tried to get a grasp on everything involved in proteomics. His expertise on these bacteria and his willingness to teach me were invaluable to me. This work would have been impossible without his help. I have enjoyed him as a mentor and friend.

My committee, Jerry, Lynn, Michelle, and Rams, have challenged me to excel and provided with support and guidance which have made this work possible. I am thankful to them for sharing their different expertise for the improvement of these studies and graciously giving up their valuable time to evaluate this work.

المنارة للاستشارات
www.manaraa.com

Considerable analytical and technical support was provided by Yalan Li at the University of Iowa Proteomics Facility. I thank her for sharing her expertise and trusting me to not destroy any mass spectrometers.

I could not have found a better group to work with than the Mattes Mafia. Yang Oh, in addition to being an excellent researcher, is one of the funniest people I know and taught me the value of organization. Carmen has always been a friend I can turn to for good advice and a good laugh. She also helped me run 26 miles on a hot October day. Josh is the most curious person I know and as a result knows a lot about everything. Yi is an ideal coworker; she is always smiling and makes me feel lazy because she works so hard. Anne is a natural listener, and we've spent many hours in the lab talking about everything and anything. I am so fortunate to have had a chance to work with such bright individuals and to also have developed friendships with them (and their families) that I will treasure for life.

When my Master's advisor, Larry Nies, first described this program, he called it a "diamond in the rough." I am so glad he convinced me to come here as I cannot have imagined a better graduate experience. It is difficult to describe to others how uniquely awesome (!!!) this group really is. The students are some of the friendliest, most fun-loving, and kindest people I have ever met. Among these, Rob, Rachel, Cassie, Andres, Garrett, Drew, Carolyn, JV, Laura, and Chris have particularly filled this time with unforgettable memories and friendships. I am also indebted to the faculty (Craig, Gene, Jerry, Keri, Michelle, Rich, and Tim) and staff (Judy, Ginny, Angie, and Collin) who have created an environment which produces high quality research but treasures the energy found in happy, enriched, and well-taken care of students.

Finally, I thank the University of Iowa and the National Science Foundation who have funded my educational endeavors and provided me with unique opportunities to travel locally and worldwide, enriching all aspects of my life.

المنارة للاستشارات
www.manaraa.com

ABSTRACT

Vinyl chloride (VC) is a common groundwater pollutant and known human carcinogen that is commonly produced from the incomplete reductive dechlorination of tetrachloroethene and trichloroethene. A possible approach to clean-up VC-contaminated sites is with microorganisms that grow on VC and/or ethene as a carbon and energy source. However, little is known about the biochemical pathways involved in VC-assimilation within these strains and their distribution and activity *in situ* in the environment. This work uses mass-spectrometry-based proteomics to contribute to the understanding of these microbial communities in both pure cultures and in the environment.

The biochemical pathways of VC and ethene oxidation in *Nocardioides* sp. strain JS614 were studied using proteins identified with a peptide mass fingerprinting approach. New insights into a previously proposed pathway were made using mass spectrometry (MS)-based protein identifications, and potential protein biomarkers for the presence and activity of VC-assimilating bacteria in the environment were identified.

Techniques to extract and identify proteins from various environmental samples such as activated sludge, sediments, soils, and water samples were developed based on preliminary experiments with protein extraction from strain JS614. These techniques were extended to 14 ethene-enrichment cultures derived from VC-contaminated groundwater to detect the presence and diversity of VC- and/or ethene- assimilating bacteria with previously identified protein biomarkers.

VC-assimilating organisms can evolve under laboratory conditions from bacteria that grow on ethene but very little is known about the molecular changes involved. Tandem mass spectrometry and spectral counting were applied to three variants of ethene-grown *Mycobacterium* strain JS623 (a wild type and two VC-adapted variants) resulting in the identification of 174 differentially expressed proteins. The results of this

المنارة للاستشارات
www.manaraa.com

study suggest that VC-degradation in one VC-adapted variant (623-E) may be unique to the other studied variants.

Overall, this study demonstrates the use of proteomics to understand pollutant-degrading microbial communities. From a biochemical level, proteomics was applied as a tool to identify functional genes and their roles in metabolic pathways. At the ecological level, proteomic biomarkers were developed to understand the presence and activities of microbial communities in the environment.

www.manaraa.com

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES xi

CHAPTER I: INTRODUCTION........ 1
The Development of Proteomics 1
Mass Spectrometry 2
Mass Spectrometry Data Analysis........ 4
Metaproteomics 6
Objectives and Hypotheses........ 8
Thesis Overview 8

CHAPTER II: IDENTIFICATION OF POLYPEPTIDES EXPRESSED IN RESPONSE TO VINYL CHLORIDE, ETHENE, AND EPOXYETHANE IN *NOCARDIOIDES* SP. STRAIN JS614 USING PEPTIDE MASS FINGERPRINTING 10
Abstract........ 10
Introduction........ 10
Experimental Section........ 11
Results and Discussion 12
Conclusions........ 14

CHAPTER III: PROTEIN EXTRACTIONS FROM ENVIRONMENTAL SAMPLES 23
Abstract........ 23
Introduction........ 23
Experimental Section........ 26
Chemicals and Growth Conditions of Strain JS614 26
Environmental Samples........ 26
Extraction Buffers and Cell Lysis 27
Extraction of Proteins from Surface Waters and Sediments 28
Phenol and Nycodenz Extractions........ 29
Protein Extraction Evaluation........ 30
Trichloroacetic (TCA) Precipitation........ 30
Results and Discussion 30
Protein Extraction of Strain JS614 Cells 30
Protein Extraction from Activated Sludge 32
Protein Extraction from Soil........ 34
Protein Extraction from Surface Waters........ 37
Protein Extractions from Sediments........ 38
Protein Extraction from Groundwater Samples 40
Conclusions........ 40

CHAPTER IV: PROTEOMIC METHODS REVEAL POLYPEPTIDES ASSOCIATED WITH VINYL CHLORIDE- AND ETHENE-ASSIMILATING BACTERIA IN ENRICHED GROUNDWATER SAMPLES FROM A CONTAMINATED SITE 51
Abstract........ 51
Introduction........ 52

www.manaraa.com

Experimental Section 54
Environmental sample collection 54
Extraction of DNA and proteins from ethene enrichments 55
PCR, cloning, and DNA sequencing of EtnC and EtnE genes from ethene enrichments 56
Protein analysis using polyacrylamide gel electrophoresis, liquid chromatography, and mass spectrometry 57
Analysis of Mass Spectra and Identification of Proteins..........................59
Results...60
Detection of EtnC and EtnE in ethene enrichments 60
Evaluation of EtnC and EtnE as biomarkers of etheneotrophic VC-oxidizing bacteria 62
Detection of polypeptides associated with the aerobic VC and ethene biodegradation pathways in ethene enrichments 63
Discussion...64

CHAPTER V: PROTEOMIC INVESTIGATION OF *MYCOBACTERIUM* SP. STRAIN JS623 77
Abstract...77
Introduction...77
Vinyl Chloride-assimilating Bacteria 77
VC and Ethene Biodegradation Pathways...78
VC-adaptation in strain JS623 79
Mass Spectrometry 80
Experimental Section..81
Chemicals, Bacterial Strains, Media, and Culture Condition..................81
Protein Extraction 81
Mass Spectrometry Analysis 81
Results and Discussion 83
Conclusion 86

CHAPTER VI: ENGINEERING AND SCIENTIFIC SIGNIFICANCE.......................106

REFERENCES 111

APPENDIX A: CHAPTER I SUPPLEMENTARY INFORMATION 122

APPENDIX B. NYCODENZ EXTRACTION METHOD FOR SEDIMENT SAMPLES (PROVIDED BY LAURA BADTKE).....................................126

المنارة للاستشارات
www.manaraa.com

LIST OF TABLES

Table 2.1. JS614 proteins expressed in response to VC, ethene, and epoxyethane with statistical relevance ($P < 0.05$), identified by using MALDI-TOF MS PMF analysis......15

Table 2.2. Oligonucleotides used for RT-PCR......18

Table 2.3. Comparison of peptide masses resulting from theoretical trypsin digestion of closely related deduced protein sequences (Noc4814/Noc4841 and Noc4822/Noc4833) in *Nocardioides* sp. JS614[e]......19

Table 3.1. Protein concentrations measured from protein extracts of soil samples with JS614 cells in various buffers (n=3 for each buffer)......42

Table 3.2. Protein concentrations of extracts from half Sterivex filters through which a mixture of tap water and JS614 cells was filtered......43

Table 4.1 Analysis of ethene-enriched VC-contaminated groundwater samples (prepared on 10/03/2007)......68

Table 4.2 Analysis of proteins extracted from ethene-enriched VC-contaminated groundwater samples (prepared on 07/01/2008)......69

Table 4.3 EtnC peptides observed in VC-contaminated groundwater enrichment samples that were used to identify EtnC polypeptides homologous to those found in cultivated strains and/or environmental clones......71

Table 4.4. EtnE peptides observed in VC-contaminated groundwater enrichment samples that were used to identify EtnE polypeptides homologous to those expected from cultivated strains and/or environmental clones......73

Table 4.5 Proteins identified in enrichments that are associated with the VC/ethene biodegradation pathway in Nocardioides sp. strain JS614......75

Table 5.1 Total number of proteins, peptides, and spectra identified for each biological triplicate of JS623 W, E, and T sample......88

Table 5.2 Top three most abundant proteins identified for JS623 variants W, E, and T......89

Table 5.3 Proteins with statistically significant ($P<0.05$) differences in spectral count values between JS623 variants and most highly expressed in 623-E......90

Table 5.4 Proteins with statistically significant ($P<0.05$) differences in spectral count values between JS623 variants and most highly expressed in 623-T......96

Table 5.5 Proteins with statistically significant ($P<0.05$) differences in spectral count values between JS623 variants and most highly expressed in 623-W......97

Table 5.6 Proteins identified in JS623-E similar to proteins encoded by genes within the phenylacetic acid operon in *Pseudomonas* putida F1......98

المنارة للاستشارات
www.manaraa.com

Table A.1. Additional JS614 proteins expressed in response to VC, ethene, and epoxyethane that were identified with statistical relevance (P<0.05) using MALDI-TOF MS PMF analysis. 123

المنارة للاستشارات
www.manaraa.com

LIST OF FIGURES

Figure 2.1. Representative SDS-PAGE results with cell extracts from epoxyethane- (EPOX), vinyl chloride- (VC), ethene- (ETH), acetate- (ACE) grown JS614 cultures, and protein ladder (L)...20

Figure 2.2 A representative PMF of the probable AkMO alpha subunit (Noc4807) measured from a protein extract of a VC-grown JS614 culture 21

Figure 2.3. Diagram of plasmid-encoded region containing ethene/VC biodegradation genes in Nocardioides sp strain JS614...22

Figure 3.1. SDS-PAGE gel comparing freeze-thaw lysis (lanes 1-5, 20 µL extract/lane) to sonication (lanes 6-10, 2 µL extract/lane) in various protein extraction buffers. 44

Figure 3.2 1D SDS-PAGE gel comparing protein extractions from various lysis techniques of 10^8 VC-grown JS614 cells/ml:..45

Figure 3.3 SDS-PAGE of protein extracts (20 µL) from activated sludge using the phenol extraction method (lane 1), activated sludge using the freeze thaw method (lane 2), activated sludge using the sonication method (lane 3), activated sludge + JS614 cells using the sonication method (lane 4)......................46

Figure 3.4. SDS-PAGE of protein extracts from soil using the Nycodenz method followed by sonication (Lane 1) and sonication in Buffers A, B, C, and D, ultracentrifugation, and protein precipitation resuspended in Buffer E (lanes 2-5 contain sonication in Buffers A-D, respectively)..47

Figure 3.5. SDS-PAGE of ten protein extracts from Sand Lake using the UltraClean Soil DNA Isolation Kit...48

Figure 3.6. Results from XTandem! search of MS/MS spectra of peptides from protein from band in Lane 2 (Figure 3.4). 49

Figure 3.7. SDS-PAGE of protein extracts from groundwater samples obtained from VC-contaminated groundwater enrichment samples. 50

Figure 4.1 Workflow for protein identification in pure culture and enrichment samples...76

Figure 5.1. SDS-PAGE of cell extracts (30 µg protein/lane) prepared from ethene-grown JS623 wild-type (W), variant E, and variant T strains. 99

Figure 5.2 Venn diagrams showing unique protein identifications for JS623-W, -E, and –T with a minimum ProteinProbability of 95%, a minimum PeptideProbability of 95%, and at least 5 peptides identified. 100

Figure 5.3. Normalized spectral counts of proteins (identified with Uniprot ID) for 623-WT, 623-T, and 623-E...101

Figure 5.4 Quantitative scatter plot compares proteins shared between JS623 variants E and T. 104

www.manaraa.com

Figure 5.5 Quantitative scatter plot compares proteins shared between JS623 variants E and W. 105

Figure 6.1 Hypothetical pathway of VC- and ethene-assimilation in Nocardioides sp. JS614 110

المنارة للاستشارات

www.manaraa.com

CHAPTER I: INTRODUCTION

The Development of Proteomics

Proteomics is the systematic, large-scale study of proteins with the goal of providing details about the structure, function, and control of biological systems (54, 87). As proteins are involved in nearly all biological activities, its study is a valuable source of information. The properties of proteins (including their subcellular distribution, quantity, modification and interactions states, catalytic activity and structure) combined with their dynamic responses to internal and external perturbations are important for cellular characterization (46, 54, 56, 87, 88).

The growth of proteomics sprang from two parallel scientific movements. The first was the development of methods to study protein chemistry during the 1980s and early 1990s, particularly, the use of Edman degradation protein sequencing. Edman sequencing, though slow, provided a reliable and automated way to obtain protein sequence information from purified proteins and as a result linked the activity of the protein to its amino acid sequence and/or gene. The second movement towards the use of proteomics was the expansion and growth of DNA sequencing that was becoming available in the late 1990s. This growth resulted in increasing numbers of potential protein sequences becoming available within gene databases. As more and more gene sequences became identified, the need for *de novo* protein sequencing was replaced by the need for a method to rapidly identify sequence information from proteins and peptides and to correlate this information with available sequencing information. This led to the development of two main advancements central to proteomics, mass spectrometry (MS) and database search algorithms. With the development of these technologies, the critical gap between genetic potential and cellular function was filled (54, 74, 87).

المنارة للاستشارات
www.manaraa.com

Mass spectrometers have been used by analytical chemists since the early 1900s for analyzing small molecules. It wasn't until the late 1980s with the development of "soft ionization" techniques for peptides and proteins, mainly electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), that MS became regularly used for protein chemistry. By the mid-1990s, MS -based strategies had supplanted Edman degradation for amino acid sequencing of proteins (2, 3, 87, 98).

Mass Spectrometry

All mass spectrometers contain an ion source that ionizes analytes, a mass analyzer that measures the mass to charge ratio (m/z) of ions, and a detector that identifies the number of ions at each m/z value. As previously discussed, there are two methods commonly used to volatilize and ionize proteins or peptides, ESI and MALDI. After ionization, analytes pass through one or more mass analyzers, such as an ion trap, time-of-flight (TOF), quadrupole, and/or Fourier transform ion cyclotron (FT-MS) (3, 97). There are many types of mass spectrometers that support different analytical strategies, but all mass spectrometers identify proteins using the same working principle of measuring the mass of a protein or peptide with a high degree of accuracy. Even with high accuracy measurements, the mass of a protein alone is often not unique enough to identify a protein of interest. The solution to this problem is the use of an enzyme with known cleavage specificity (usually trypsin) which fragments a protein into its smaller peptides. The masses of these peptides can then be used to uniquely identify the protein (2, 3, 87, 98).

The idea of using its peptides to identify a protein is the principle behind the analytical technique known as peptide mass fingerprinting (PMF). PMF is commonly performed using MALDI-Time of Flight (ToF) MS. Digested peptides of a protein of interest are ionized and their masses measured, resulting in MS spectra (showing intensities of a range of m/z values) and a corresponding list of peptide fragment masses.

المنارة للاستشارات

www.manaraa.com

Using a database search algorithm (usually freely available or included with the purchase of a mass spectrometer), these masses are then compared to peptide masses from a database of known protein sequences which have been fragmented *in silico* by the same enzyme specificity as the protein of interest. The results are statistically analyzed to identify the best match for the targeted protein (52, 55, 72, 86, 87, 112).

PMF and MALDI-ToF MS are ideal for purified proteins or samples with limited complexity. However, for more complex mixtures or in-depth analysis of protein samples, tandem mass analysis should be used (69, 97). Tandem mass spectrometry (MS/MS) uses multiple steps of MS selection and analysis with some sort of ion fragmentation between steps. With MS/MS, not only is the precursor peptide mass measured but also individual peptide ions are separated, further fragmented into product ions, and these ions masses are also measured. The resulting MS/MS spectra gives information that is particularly useful for obtaining information about peptide sequence and/or structure. The ability of MS/MS spectra to represent the amino acid of a peptide and provide increased protein identification specificity is often why MS/MS is preferred over PMF. MS/MS, however, also requires significantly more sample preparation and analysis time (both sample and data analysis) compared to MS analysis (2, 3, 87, 98).

Using ESI with MS/MS is offers the advantage of being easily adaptable to liquid chromatography which allows for the separation of peptide mixtures prior to analysis by a mass spectrometer. ESI creates gaseous ionized molecules directly from a liquid solution. MALDI, on the other hand, uses energy transfer from a laser to “desorb” peptide ions from a solid matrix. The use of nano-liquid chromatography (nano-LC) and nanospray-ESI allows for LC-separated peptide mixtures to be sprayed directly into mass spectrometers at very low flow rates (nL/min) and with significantly increased sensitivity. Nanospray ESI-MS/MS connected to nano-LC has created opportunities for high sensitivity proteomic investigations of samples previously limited by sample availability and/or quantity (87, 98).

المنارة للاستشارات
www.manaraa.com

For the experiments described in this thesis, we used two specific kinds of mass spectrometers: a MALDI-ToF MS and a nanospray-ESI-linear ion trap quadruople (LTQ) MS/MS connected to an online nano-LC with an autosampler.

Mass Spectrometry Data Analysis

To identify proteins from MS or MS/MS spectra, database scoring algorithms must be used to compare experimentally observed peptide ion masses to theoretical peptide ion masses within a protein database. For PMF, the freely available program Mascot is often used to compare calculated peptide masses from select databases to experimental mass values (89). For MS/MS spectra, the database scoring algorithms are considerably more complex. Two of the most commonly used data analysis packages include SEQUEST and XTandem!.

SEQUEST is a licensed data analysis program that is often used for MS/MS spectra generated from Thermo Scientific mass spectrometers. Much like PMF data analysis programs, SEQUEST examines each generated MS/MS spectra individually and compares these to a theoretical list of peptide masses generated from a protein sequence database. It takes advantage of the fact that the peptide "parent mass" information is known from the MS/MS spectra and uses this information to select peptide sequences from the database whose masses are near the observed "parent mass." For each peptide in the protein database, SEQUEST predicts and compares the theoretical MS/MS spectra to the observed MS/MS spectra by how well the spectras "cross-correlate." Protein identifications from SEQUEST include statistically significant matches to a protein database based on cross correlation scores (comparing theoretical and observed MS/MS spectra) and delta correlation scores (comparing the top match to the next best match) (32).

XTandem! is a freely available MS/MS data analysis and scoring algorithm supported by the Global Proteome Machine Organization. XTandem!, like SEQUEST,

المنارة للاستشارات
www.manaraa.com

compares each individual MS/MS spectrum to all likely peptides in a database. However, unlike SEQUEST, XTandem! considers only the matching peaks in the MS/MS spectra (not the correlation of spectras) and assigns these matches with a "hyperscore". It then analyzes all hyperscores and assigns the statistical confidence that each match is not random (expectation value). Protein identifications from XTandem! include statistically significant matches to a protein database based on hyperscores (comparing theoretical and observed MS/MS spectra) and expectation values (comparing how unlikely a better match exist within the database) (22). Both XTandem! and SEQUEST have been used in numerous published proteomic studies and are well-accepted by the proteomics community. The choice of which data analysis program to use is usually based on the availability of the software and user preferences.

In this study, we also used the TransProteomic Pipeline (TPP), a freely available software package which includes PeptideProphet (63) and ProteinProphet (79). The PeptideProphet uses the scores calculated by other scoring algorithms, such as SEQUEST and XTandem!, and generates a single discriminant score for all MS/MS spectra. It then uses the distribution of these discriminant scores and Bayesian statistics to calculate the probability that a peptide sequence identification is correct. ProteinProphet uses the results from PeptideProphet to generate the probability that a protein identification based on peptide sequence matches is correct. TPP essentially converts scores like SEQUEST cross correlations scores and XTandem! expectation values to an easier-to-understand probability score. TPP is also integrated into the freely available Computational Proteomics Analysis System (CPAS) developed by LabKey Software Foundation (92). This program provides an excellent viewer to display, share, manipulate, and export MS/MS results. An alternative to the CPAS is Scaffold (Proteome Software, Portland, Oregon) which requires the purchase of a license. Scaffold is particularly useful for comparing protein identifications between multiple MS/MS analyses.

المنارة للاستشارات
www.manaraa.com

Metaproteomics

The advances of mass spectrometry, sequence databases, and database search tools created opportunities in the early 1990s to move beyond identifying purified proteins to identifying many of the proteins resolvable by electrophoresis gels. At the first 2-dimensional gel electrophoresis (2DE) meeting in 1994, the term “proteome” was defined as the protein complement to the genome (87). No one has yet successfully identified a full proteome for any species. This is unsurprising since the number of proteins within a proteome is formidable and much larger than the number of genes in a complementing genome; each gene can produce multiple proteins from post-translational modifications, sequence polymorphisms, or alternative splicing of transcripts (28).

The dynamic range of expressed protein does however offer advantages to studying microbial ecology in the environment. Unlike genomic studies, which capture only the metabolic potential of an environmental community, transcriptomic and proteomic approaches can reveal the functionality of specific microbial community members. The use of proteomics, unlike transcriptomics, is also not limited by the short half-life of RNA and bypasses the challenges related to low correlations between mRNA and protein expressions (44, 109). Proteomic analysis is an effective tool to obtain information about cellular responses to stressful environmental conditions and to the availability of different growth substrates. Cell maintenance and organization within different environmental communities can also be analyzed (74, 81, 109). Also, with the increasing availability of genomic sequences from bacteria within the environment (metagenomics), it is becoming easier to use proteomic techniques for the study of complex microbial communities in the environment (metaproteomics).

The use of metaproteomics for the functional analysis of environmental communities has been demonstrated by several published studies (6, 50, 62, 73, 82-84, 91, 93, 110). Within an acid drainage site, over 2,000 proteins in a microbial biofilm community were detected, identified, and analyzed using tandem mass spectrometry (91).

المنارة للاستشارات
www.manaraa.com

Differentially expressed proteins in microbial communities from activated sludge performing enhanced biological phosphorus removal (EBPR) and non-EBPR sludges were identified and compared using gel-based proteomics (110). The proteomic fingerprinting of extracellular proteins from dissolved organic matter revealed information about the distribution of bacteria in lake waters and forest soils (93). Metaproteomic studies have become critical complements to traditional genomic investigations of environmental microbial communities because of its potential to both identify novel functional genes and to relate diverse functions to specific microbial communities in complex environments (74, 81, 109).

A considerable challenge to metaproteomic investigations is the ability to extract proteins from complex environments, especially soils (6, 80, 109). Methods for metaproteomic analysis of environmental samples include three steps: extraction of microbial proteins from the environmental sample, resolution (or separation) of extracted proteins, and identification of proteins (usually with mass spectrometry) (74, 81). Protein solubilization and cell lysis are critical factors for effective proteomic analysis (80, 81, 99). The compatibility of downstream analysis should also be considered (69, 81); for example, if mass spectrometry is to be used, proteins should be resuspended in mass spectrometry-compatible buffers and should be at correct concentrations for effective identification. Also, because of the large number of proteins and the existence of these proteins in a variety of conformational states, the goals of analysis (i.e. targeted protein identification or all protein identification) should carefully be considered in developing protein extraction methodologies. In summary, method development for optimized protein extraction, separation, and identification is often a challenging aspect of proteomic studies (69, 81, 109).

www.manaraa.com

Objectives and Hypotheses

The main objective of this thesis was to investigate the use of proteomic techniques, mainly mass spectrometry, to identify proteins from environmentally-relevant bacteria in both pure cultures and environmental samples. Specifically, the ability to identify bacterial proteins involved in aerobic vinyl chloride degradation both in pure cultures and the environment was studied. This work explored four working hypotheses as follows:

1. The predicted vinyl chloride and ethene biodegradation pathways of *Nocardioides* sp. strain JS614 can be validated using mass spectrometry-based identification of proteins involved in its growth on vinyl chloride, ethene, and epoxyethane.
2. Protein extraction and identification is more effective from environmental samples containing increased numbers of bacterial cells and decreased concentration of humic acids.
3. Ethene- and/or vinyl-chloride assimilating bacteria in ethene-enriched groundwater samples can be detected using protein biomarkers.
4. Mass spectrometry-based protein identification and quantification can determine differences in vinyl-chloride adapation mechanisms between two different adapted vinyl-chloride assimilating variants of *Mycobacterium* sp. strain JS623.

Thesis Overview

This thesis is organized into four chapters (II-V) which correspond to the four hypotheses described above. Each chapter can stand alone and contains its own literature review, research objective, and discussion of results. An overview of the contents of each chapter is presented here.

Chapter II presents a proteomic investigation of a pure culture by comparing the enzymes expressed by the vinyl-chloride degrader, *Nocardioides* sp. strain JS614, under different growth conditions in order to elucidate enzymes involved in its vinyl chloride and ethene biodegradation pathway. A key result of this chapter is the identification of potential protein biomarkers for vinyl-chloride degradation in strain JS614. This chapter is published in *Applied Environmental Microbiology* (17).

المنارة للاستشارات
www.manaraa.com

Chapter III examines different methods to extract and identify proteins from strain JS614 and from various environmental samples including surface water, groundwater, soil, sediments, and activated sludge. Protein extraction methods were evaluated based on their ability to extract proteins that could eventually be identified by MS/MS. The results of this work helped to direct proteomic investigations of environmental samples towards ethene-enriched vinyl chloride-contaminated groundwater samples. Portions of this chapter detailing the optimization of protein extraction from strain JS614 will be submitted for publication in *Journal of Microbiological Methods.*

Chapter IV describes optimized methods to extract and identify select protein biomarkers (using MS/MS) for VC-degradation from enriched cultures of vinyl chloride-contaminated groundwater samples. This chapter builds on the results of previous Chapters II and III. VC-degrading biomarkers identified in Chapter II were targeted in our proteomic investigations of enriched groundwater samples, and the methods developed in Chapter III were adapted and extended for protein identification by MS/MS for enriched groundwater samples. This chapter will be submitted for publication in *Journal of Microbiological Methods* (September, 2009).

Chapter V explores the use of MS/MS to compare the proteomes of three different strains *Mycobacterium* sp. JS623, a wild-type strain and two strains adapted for growth on VC. Semi-quantitative measurements of protein expression based on spectral counting were performed. Based on this information, protein expression differences between the wild type and two VC-adapted strains were identified, and VC-adaptation mechanisms for the two mutant strains were proposed.

المنارة للاستشارات
www.manaraa.com

CHAPTER II: IDENTIFICATION OF POLYPEPTIDES EXPRESSED IN RESPONSE TO VINYL CHLORIDE, ETHENE, AND EPOXYETHANE IN *NOCARDIOIDES* SP. STRAIN JS614 USING PEPTIDE MASS FINGERPRINTING[1]

Abstract

Enzymes expressed in response to vinyl chloride, ethene, and epoxyethane by *Nocardioides* sp. strain JS614 were identified using a peptide mass fingerprinting (PMF) approach. PMF provided insight concerning vinyl chloride biodegradation in strain JS614 and extends the use of MALDI-TOF MS as a tool to enhance characterization of biodegradation pathways.

Introduction

Vinyl chloride (VC), a known human carcinogen (14) and groundwater contaminant (101), is often generated in groundwater by the incomplete reduction of chlorinated solvents. Diverse bacterial genera, including *Mycobacterium* (19, 47), *Nocardioides* (19, 77), *Ochrobactrum* (23), *Pseudomonas* (23, 106), and *Ralstonia* (31) use both VC and ethene as carbon and energy sources. Several strains appear to use the same enzymes to metabolize both VC and ethene (21, 47, 77). Alkene monooxygenase (AkMO) oxidizes VC to chlorooxirane (47, 106) and ethene to epoxyethane (19, 26, 27, 77). Epoxyalkane:coenzyme M transferase (EaCoMT) participates in further metabolism of both epoxyethane (21, 24, 77) and chlorooxirane (21). An unknown number of enzymatic steps catalyze the conversion of these epoxides to acetyl-CoA (26). Elucidating the remaining enzymes and intermediates of aerobic VC and ethene biodegradation will facilitate development of molecular tools for detecting and

---

[1] Published in Chuang, A. S., and T. E. Mattes. 2007. Identification of polypeptides expressed in response to vinyl chloride, ethene, and epoxyethane in *Nocardioides* sp. strain JS614 by using peptide mass fingerprinting. Appl. Environ. Microbiol. 73:4368-72.

المنارة للاستشارات
www.manaraa.com

differentiating VC- and ethene-assimilating bacteria in the environment. The completion of the *Nocardioides* sp. strain JS614 genome sequence (http://genome.ornl.gov/microbial/noca/) provides opportunities to use new approaches to identify enzymes involved in VC and ethene biodegradation. In this study, proteomic techniques were used to rapidly and accurately identify enzymes expressed in response to VC, ethene, and epoxyethane in strain JS614.

Experimental Section

Chemicals, media, growth conditions, and protein extraction methods are described elsewhere (77). SDS-PAGE analysis was performed by the method of Laemmli (65) with extracts (10-50 μg protein) from VC-, ethene-, epoxyethane-, and acetate-grown cells. Polyacrylamide gels were stained with Bio-Safe Coomassie Blue (Bio-Rad Laboratories, Inc.). Visual inspection (Figure 2.1) revealed several polypeptides expressed in response to ethene, epoxyethane, and VC that were not expressed in response to acetate, suggesting they were directly involved in VC, ethene, and epoxyethane metabolism. Polypeptide bands from all lanes in each numbered section of the gel were excised and digested with bovine trypsin (Promega Corp.). The resulting monoisotopic peptide fragment masses were analyzed with a Bruker BiflexIII Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight (MALDI-ToF) mass spectrometer in positive-ion/reflector mode using an a-cyano-4-hydroxycinnamic acid matrix (58, 86). Peptide mass fingerprints (PMFs) of digested polypeptide bands (Figure 2.2) were compared to the PMF of a control gel fragment that had no contact with cell extracts, and matching masses within a 0.3 Da mass tolerance were excluded from analysis. The remaining peptide masses were compared to all bacteria in the NCBInr database using the following parameters in Mascot (89): zero missed cleavages, carbamidomethyl (fixed modification), methionine oxidation (variable modification), and a 0.3 Da mass tolerance. Results were evaluated using the Mowse probability based scoring algorithm (86).

المنارة للاستشارات
www.manaraa.com

Results and Discussion

PMF analysis revealed 7 polypeptides expressed in response to VC, ethene, and epoxyethane with statistically significant Mascot MOWSE scores ($P < 0.05$) from at least two independent PMF analyses of a single growth condition (Table 2.1). None of these polypeptides were observed in extracts from acetate-grown JS614 cultures (data not shown). However, our PMF data do not rule out the possibility that proteins reported in Table 1 are present in acetate-grown cells at very low abundance. Eleven additional polypeptides expressed in response to VC, ethene, and/or epoxyethane were identified ($P < 0.05$) in either just one PMF analysis or only once under a particular growth condition (Appendix A, Table A.1).

The polypeptides reported in Table 2.1 mapped to a plasmid-encoded region of the JS614 genome sequence (Figure 2.3) that contained known VC/ethene biodegradation genes (77) suggesting that they are likely to participate in VC/ethene metabolism. The identification of AkMO subunit (EtnC), EaCoMT (EtnE), putative CoM biosynthesis protein ComA, and a probable CoA transferase in cell extracts from VC- and ethene-grown cells is consistent with previous reports (77) confirming the robustness of this proteomics approach. Five proteins not previously reported to be involved in VC/ethene metabolism were also identified. Reverse-transcription (RT)-PCR assays were conducted to provide independent confirmation of proteomics results. Total RNA was extracted and purified, and RT-PCR was performed as described in (76, 77) using 0.1 ng RNA template. JS614 genomic DNA, extracted as described previously (19), was used as the positive RT-PCR control. Sterile water was the negative RT-PCR control, the 16S rRNA gene was the positive RNA control, and the RpoB gene was the positive mRNA control. Using primers described in Table 2, RT-PCR products of the expected size were observed with RNA extracted from ethene-, epoxyethane-, and VC- grown cells but not acetate-grown cells (data not shown). RT-PCR products were observed in all RNA extracts

www.manaraa.com

when 16S rDNA and rpoB primers were used, and products were not observed in control PCRs with RNA template and no reverse transcriptase (data not shown).

Our JS614 genome-assisted proteomics approach has led us to reformulate hypotheses concerning the VC/ethene biodegradation pathway in strain JS614. Previously, we proposed that a putative reductase (encoded by Noc4815, Acc. No. ZP_00655559) participated in regeneration of free CoM (77). In this study, we identified a protein (encoded by Noc4827, Acc. No. ZP_00655570) expressed in response to ethene, VC, and epoxyethane with 53% amino acid identity (E=2e-156) to the 2-oxopropyl-CoM reductase/carboxylase from the propene-oxidizing Xanthobacter sp. Py2 (Acc No. Q56839) (4) (Table 2.1, Fig. 2.1). By analogy to the strain Py2 propene biodegradation pathway, the Noc4827 product could participate in carboxylating a VC/ethene pathway metabolic intermediate with concomitant regeneration of free CoM. Based on BLAST homologies, we propose that the Noc4815 product is a CoA reductase rather than a previously predicted CoM reductase. We also previously hypothesized that a putative short-chain dehydrogenase (encoded by Noc4814, Acc. No. ZP_00655558) carried out a two-step dehydrogenation of 2-hydroxyethyl-CoM to 2-carboxymethyl-CoM (77). Here, we provide circumstantial evidence (Fig. 2.1, Table 2.1) that a putative dual function alcohol/acetaldehyde dehydrogenase encoded by Noc4822 (Acc. No. ZP_00655566) and/or Noc4833 (Acc. No. ZP_00655575) participates in the JS614 VC/ethene pathway. Detailed biochemical studies are required to determine the role of these proteins in VC/ethene metabolism.

Previously, we reported that JS614 harbors 2 alleles of the EaCoMT gene etnE . Bioinformatic analysis of the JS614 genome sequence has revealed a cluster of 14 additional genes that appear to have been recently duplicated, including Noc4841 (Acc. No. ZP_00655584, 90% aa identity to Noc4814) and Noc4833 (86% aa identity to Noc4822) (Fig. 2.3). Theoretical trypsin protein digests obtained from ExPASy PeptideMass (43, 108) showed several peptide masses unique to each protein (data

www.manaraa.com

available online, http://aem.asm.org/cgi/content/full/73/13/4368/DC1). Significant MOWSE scores were obtained when these unique masses were input into MASCOT (data not shown), indicating that PMF can distinguish between these highly similar proteins.

Conclusions

Although MALDI-ToF MS is increasingly being used in microbiology, there are limited publications demonstrating its application in biodegradation research (45, 53). Here, we used MALDI-ToF MS to identify 11 proteins expressed in response to VC and ethene by strain JS614. Unique peptide masses were used to distinguish between mixtures of highly similar expressed proteins. To our knowledge, this is the first extensive application of PMF in conjunction with a completed genome sequence to identify enzymes translated in response to a xenobiotic pollutant. The ability to monitor and distinguish microbial involvement in pollutant degradation is a significant obstacle for bioremediation. An understanding of aerobic VC and ethene biodegradation pathways and the ability to detect the involved expressed proteins are important initial steps towards monitoring VC and ethene biodegradation in the environment.

www.manaraa.com

Table 2.1. JS614 proteins expressed in response to VC, ethene, and epoxyethane with statistical relevance ($P < 0.05$), identified by using MALDI-TOF MS PMF analysis.

Excised gel section[a]	Growth Substrate[b]	Significant MASCOT hits[c]	Predicted Mass (kDa)[d]	Observed Mass (kDa)[e]	MOWSE Score[f]	E Value[g]	% coverage[h]	BBM[i]	MM/ TPMS[j]	Gene No.[k]
1	epoxyethane	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	60.47	74	142	1.6E-08	37%	61	14/49	4827
1	VC	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	60.47	79	101	2.5E-04	28%	58	12/55	4827
1	ethene	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	60.47	73	141	7.5E-07	30%	60	14/41	4827
2	epoxyethane	Probable alkene monooxygenase alpha subunit	57.78	62	178	1.5E-06	44%	59	19/70	4807
2	VC	Probable alkene monooxygenase alpha subunit	57.78	65	142	1.5E-08	36%	75	19/66	4807
2	ethene	Probable alkene monooxygenase alpha subunit	57.78	65	126	9.0E-03	33%	56	13/56	4807
3	epoxyethane	alcohol/acetaldehyde dehydrogenase	51.78	54	138	9.5E-08	46%	60	15/72	4833
3	VC	alcohol/acetaldehyde dehydrogenase	51.78	55	119	3.3E-06	41%	61	14/77	4833

المنارة للاستشارات
www.manaraa.com

Table 2.1 (continued)

3	ethene	alcohol/acetaldehyde dehydrogenase	51.78	53	115	1.2E-04	41%	62	14/68	4833
3	epoxyethane	bifunctional protein: alcohol/acetaldehyde dehydrogenase	50.02	54	149	3.0E-06	42%	59	14/52	4822
3	VC	bifunctional protein: alcohol/acetaldehyde dehydrogenase	50.02	55	129	5.7E-03	40%	59	14/66	4822
3	ethene	bifunctional protein: alcohol/acetaldehyde dehydrogenase	50.02	53	112	6.5E-04	35%	64	14/61	4822
4	epoxyethane	epoxyalkane: coenzyme M transferase	40.68	45	166	1.9E-08	50%	60	13/40	4810
4	VC	epoxyalkane: coenzyme M transferase	40.68	46	163	2.4E-08	49%	60	13/46	4810
4	ethene	epoxyalkane: coenzyme M transferase	40.68	45	150	2.1E-04	45%	62	13/47	4810
5	epoxyethane	Short-chain dehydrogenase/reductase SDR	25.85	31	140	3.4E-07	51%	63	11/46	4814
5	VC	Short-chain dehydrogenase/reductase SDR	25.85	33	158	1.5E-09	59%	64	12/40	4814
5	ethene	Short-chain dehydrogenase/reductase SDR	25.85	31	145	2.0E-06	53%	61	10/35	4814
5	epoxyethane	Short-chain dehydrogenase/reductase SDR	26.20	31	149	1.2E-07	59%	61	12/45	4841

المنارة للاستشارات
www.manaraa.com

Table 2.1 (continued)

5	VC	Short-chain dehydrogenase/reductase SDR	26.20	33	170	1.2E-09	61%	64	12/40	4841
5	ethene	Short-chain dehydrogenase/reductase SDR	26.20	31	163	1.5E-09	60%	63	12/38	4841

Notes:

a Gel sections were excised as shown in Fig. 2.1.

b Growth conditions and methods are described in the text.

c Most significant Mascot hits to current gene annotations in the NCBInr database. FAD, flavin adenine dinucleotide.

d Theoretical mass reported by Mascot.

e Estimated from results of SDS-PAGE (Fig. 2.1).

f Mowse score reported by Mascot.

g E value reported by Mascot.

h Percent coverage of amino acids to matched peptides reported by Mascot.

i Mowse score of the next best Mascot match (NBM) not associated with JS614.

j MM, mass matches; TPMS, total number of peptide masses searched.

k Gene number in finished JS614 genome sequence corresponding to Mascot hit.

l All proteins shown were found in at least two independent analyses, and the results reported are averaged values.

المنارة للاستشارات
www.manaraa.com

Table 2.2. Oligonucleotides used for RT-PCR

Oligonucleotides (5'-3')	Sequence (5' to 3')	Gene No.a	Expected Product Size (bp)	Reference
CoM-F3	GCTCTCAAGATGTGCTTCTGCCAACCA	Noc4810	831	(77)
CoM-R3	CGGTGCGTCCGACCTCGTAGTTCAG	Noc4810	831	(77)
etnC_F	CTTGAAACCGTCCACGAGAAGAG	Noc4807	1481	(77)
etnC_R	AGCGGGTCCTTGATCTCGTACTT	Noc4807	1481	(77)
scdehd_F	GCCATAGGCGTACTTGACCTG	Noc4814	560	present study
scdehd_R	ATGGGGTACTTCACGAGTGTCT	Noc4814	560	present study
carb_F	GGAGTACGGCTCCTTCTACCA	Noc4827	832	present study
carb_R	GTTGAGGAACAACTCATTCATCTC	Noc4827	832	present study
dehy_F	ATCACGTCAATACTGGGGATGT	Noc4833	703	present study
dehy_R	AGGACAAGGTCGTCAAGATCAA	Noc4833	703	present study
27f	AGAGTTTGATC(C/A)TGGCTCAG	n/a	1482	(66)
1492r	TACGG(C/T)TACCTTGTTACGACTT	n/a	1482	(66)
rpoB_F	GCTTCGGGTTGAAGTAGTAGTTGT	Noc0711	802	(76)

المنارة للاستشارات
www.manaraa.com

Table 2.3. Comparison of peptide masses resulting from theoretical trypsin digestion of closely related deduced protein sequences (Noc4814/Noc4841 and Noc4822/Noc4833) in *Nocardioides* sp. JS614[e].

Gene number	Total No. of Peptide Masses[a]	No. of Unique Peptide Masses[b]	Shared Peptide Masses[c]	% Coverage (a.a.)[d]
Noc4814	21	9	12	46%
Noc4841	23	11		
Noc4822	39	21	18	15%
Noc4833	39	21		

Notes:

a Total peptide masses generated from an in silico trypsin digestion.

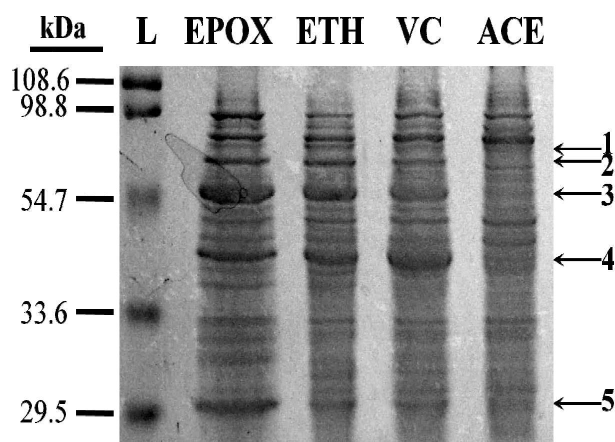
b Number of unique peptide masses resulting from an in silico trypsin digestion.

c Peptide masses shared between theoretical digests of either Noc4814 and Noc4841 or Noc4822 and Noc4833.

d Percent coverage of either protein by shared peptide masses. The analysis did not include peptide masses that resulted from variable modifications (e.g., methionine oxidation). aa, amino acids.

e See material available online, http://aem.asm.org/cgi/content/full/73/13/4368/DC1, for more details.

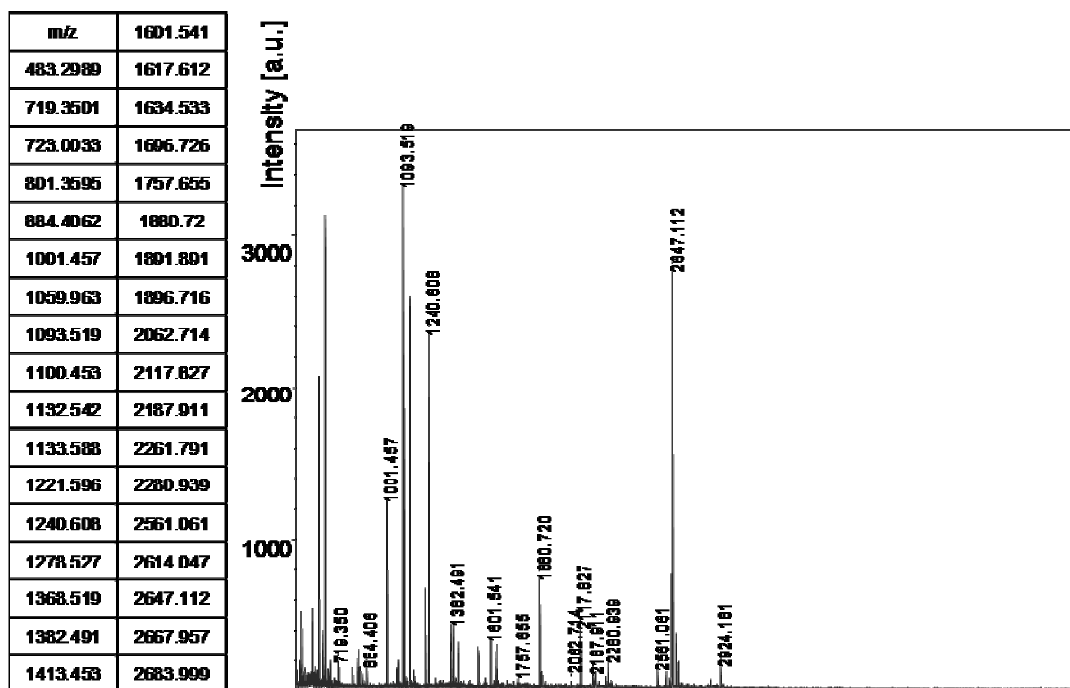
المنارة للاستشارات
www.manaraa.com



Note: Numbers and arrows indicate sections of the gel that were excised for MALDI-TOF MS analysis and correspond to results presented in Table 2.1.

Figure 2.1. Representative SDS-PAGE results with cell extracts from epoxyethane- (EPOX), vinyl chloride- (VC), ethene- (ETH), acetate- (ACE) grown JS614 cultures, and protein ladder (L).

www.manaraa.com



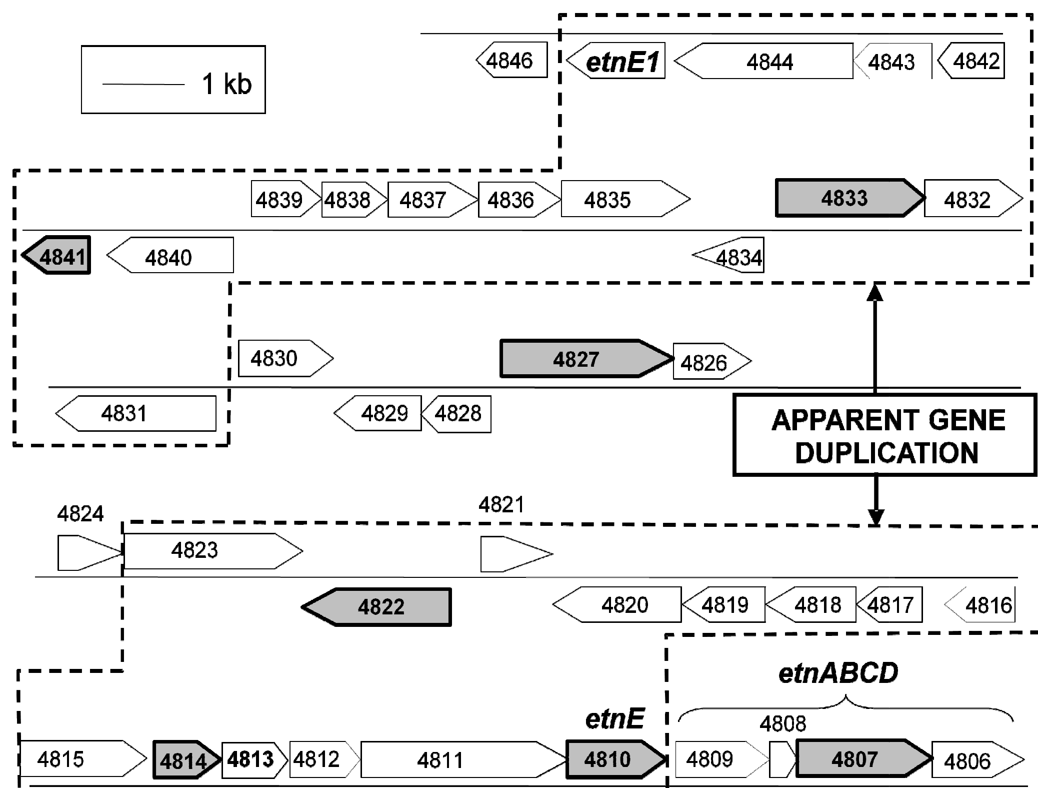
m/z	1601.541
483.2989	1617.612
719.3501	1634.533
723.0033	1696.726
801.3595	1757.655
884.4062	1880.72
1001.457	1891.891
1059.963	1896.716
1093.519	2062.714
1100.453	2117.827
1132.542	2187.911
1133.588	2261.791
1221.596	2280.939
1240.608	2561.061
1278.577	2614.047
1368.519	2647.112
1382.491	2667.957
1413.453	2683.999

Note: The mass spectrum and a complete list of peptide masses that make the up the PMF are shown, and "a.u." refers to atomic unit.

Figure 2.2 A representative PMF of the probable AkMO alpha subunit (Noc4807) measured from a protein extract of a VC-grown JS614 culture

المنارة للاستشارات

www.manaraa.com



Notes: Genes expressed and translated into protein in response to epoxyethane, VC, or ethene as identified by MALDI-TOF MS and PMF analysis are shaded gray. Genes 4831-4844 and the second EaCoMT allele etnE1 appears to be the result of a duplication of genes 4810-4823 evidenced by high amino acid similarities and the conservation of gene orientation. For example, genes 4814 and 4822 share 90 and 86% amino acid similarity with genes 4841 and 4833, respectively.

Figure 2.3. Diagram of plasmid-encoded region containing ethene/VC biodegradation genes in Nocardioides sp strain JS614.

المنارة للاستشارات
www.manaraa.com

CHAPTER III: PROTEIN EXTRACTIONS FROM ENVIRONMENTAL SAMPLES

Abstract

Effective protein extraction is critical for the successful identification of proteins using mass spectrometry. For environmental samples, protein extraction can be particularly challenging because the type of environmental matrix and targeted proteins for extraction can affect protein extraction efficiency. In this study, the ability to extract proteins from a variety of environmental samples was examined. The separation of bacterial cells from environmental matrices prior to cell lysis was found to greatly improve protein extraction efficiency. In soils and sediments, chemical density gradients combined with centrifugation could be used for separating bacterial cells, but great care had to be taken to minimize the co-extraction of interfering compounds such as humic acids which affects downstream analyses such as MS. Separating bacterial cells from water samples, on the other hand, was much easier as it often only required filtration or centrifugation, and the presence of humic acids was minimal. In general, protein extraction methods will still require specific optimization for different environmental samples. Several options for protein extraction in soils, sediments, surface waters, groundwaters, and activated sludge are presented in this study and can be applied to future proteomic investigations of different environmental samples.

Introduction

Successful identification of proteins from the environment relies on the effective extraction of targeted proteins from environmental samples. A number of published extraction techniques have been applied to environmental samples with varying amounts of success.

Two of the first methods introduced for the extraction of proteins from environmental samples were the boiling and freeze thawing of samples in various buffers

www.manaraa.com

as described by Ogunseitan (83). These methods use extreme temperatures to degrade cell walls resulting in cell lysis and the subsequent release of proteins. Ogunseitan (83) reported that the boiling method was preferred for the extraction of proteins from wastewater (10-30 µg protein/mL wastewater) whereas the freeze thaw method was more effective for soils and sediments (20-50 µg protein/g soil). He also noted that soil and sediment protein extractions included the coextraction of interfering compounds (i.e., humic substances) which would interfere with protein quantification and protein analysis. Focusing on less complex samples for his later experiments, Ogunseitan extracted proteins from wastewater sludge samples (~1.17 mg protein/mL wastewater) and freshwater samples (~0.07 mg protein/mL freshwater) using a combination of centrifugation and sonication techniques (82). His methods were extended by Singleton et al. (84) to identify proteins (mercuric reductases) in unenriched and enriched freshwater samples. In 2003, these methods were again modified and used to extract proteins (up to 35 µg protein/g) from cadmium-contaminated soils (96). A review and evaluation of several methods of protein extraction from bacterial cells and freshwater samples was performed (73) where 79-105 µg protein/mL was reported from freshwater samples.

When coextraction of interfering compounds occurs, it is often necessary to separate targeted proteins from the environmental matrix (most often soils or sediments). Nycodenz density gradient extractions demonstrated the separation of bacterial cells from soil samples for DNA extractions(7, 8). This method was later applied by (74) to extract proteins from freshwater bacterial communities (79-105 µg protein/mL). An alternative method using phenol rather than Nycodenz lysed cells in 0.1 M NaOH and subsequently separated proteins from the soil matrix using a phenol extraction (6). Both of these indirect protein extraction methods enable the separation of humic organic matter from the protein extraction, thus minimizing humic acid-associated interferences in downstream protein quantification and analysis.

المنارة للاستشارات
www.manaraa.com

Depending on the proteins targeted by the extraction, lysis conditions and mediums vary widely. In general, cell lysis is usually performed by both chemical and physical means, such as the use of highly denaturing chemicals (i.e., urea or guanidine) combined with beading beating, sonication, or freeze-thawing to disrupt bacterial cell walls. Additionally, extraction buffers for cell lysis are optimized to stabilize protein activity and maximize protein solubility by adjusting pH, buffer capacity, and the presence of detergents, reducing agents, chelators, and proteolytic inhibitors.

The objective of this chapter was to optimize methods to extract proteins from VC-degrading bacteria in environmental samples. To accomplish this, we applied select protein extraction methods reported in literature to pure cultures of VC-degrading strain JS614. Strain JS614 was chosen as a model organism for proteomic investigations because of i) its relevance to environmental biotechnology applications, ii) its genome has been completely sequenced and annotated (Genback Accession CP000508-CP000509), and iii) previous proteomic investigations using peptide mass fingerprinting and MALDI-ToF mass spectrometry (MS) have successfully revealed several key enzymes involved in the JS614 VC biodegradation pathway (17). Pure culture studies were used to select the best extraction methods to extend to environmental samples (with and without the addition of JS614). As these methods were applied to environmental samples, our goal was to evaluate the possibility of identifying protein biomarkers using tandem mass spectrometry. As instrument time on the tandem mass spectrometer is limited and expensive (at the time of this study, $50/hour x ~2 hours/sample x ~10 samples/protein extraction = $1000/protein extraction), these preliminary studies directed us towards the most promising future proteomic experiments for VC-degrading protein biomarkers in environmental samples.

www.manaraa.com

Experimental Section

Chemicals and Growth Conditions of Strain JS614

Chemicals, media, and growth conditions are described in (77). Strain JS614 cells (ATCC BAA-499) were grown in a minimal salts medium (MSM) with ethene (0.2-0.3 mM) or VC (1.3-2.1 mM) in the laboratory. Cells were harvested by centrifugation (7500xg, 5 min., 4°C) during the mid-exponential growth phase (OD_{600}=0.25-0.35) for protein extraction. All reagants were of high purity and were purchased from Fisher Scientific unless otherwise indicated.

Environmental Samples

Environmental samples consisted of activated sludge, soil, surface water, sediment, and groundwater samples. Where noted, cells of strain JS614 were added to environmental samples for evaluating the efficiency of protein extraction.

Activated sludge samples were obtained from the University of Iowa Wastewater Treatment Plant (Iowa City, IA) and used immediately. Top soil was sampled from Orion, Illinois in August, 2008. All soil samples were passed through a #20 sieve and stored at 4°C. Surface water protein extractions from Sand Lake in Marshall County, Iowa, were obtained from Joshua Livermore. Livermore pumped surface water samples (100-500 mL) through Sterivex-GP 0.22 µm filters (Millipore Corp). Sediment proteins were obtained from Laura Badtke. Sediment samples were from Indiana Harbor, Indiana, and were sampled on August, 2006, as described by (75).

VC-contaminated groundwater enrichment cultures (100 ml) were obtained from Bioremediation Consulting, Inc. (Watertown, MA). VC-contaminated groundwater, collected from several monitoring wells at a site in MA on 07/01/2008, was used to prepare ethene-fed enrichment cultures. Previous work has shown that ethene-oxidizing bacteria capable of VC biodegradation are present within the VC plume at this site (39). Enrichment cultures were prepared in sterile 160 mL serum bottles and contained 40 mL

المنارة للاستشارات
www.manaraa.com

of groundwater and 60 ml of a mineral salts media (containing 40 ppm N, 70 ppm P, 166 ppm K, 27 ppm S, 20 ppm Mg, 5 ppm Ca, 0.06 ppm Fe, 218 ppm PO_4, 81 ppm SO_4, and 177 ppm NO_3, and a trace metals mixture at pH 6.9). Ethene was added to an initial concentration of 2.845 mM. After 63-64 days of incubation, the ethene concentration (determined by EPA method 5021A). Visual inspection of turbidity in the bottles was also conducted. Based on these analyses, the enrichments were declared either positive or negative for the presence of ethene-oxidizing bacteria (called "etheneotrophs"). Following the completion of these analyses, the enrichments were shipped to our laboratory for proteomic analysis where bacterial cells and solids were harvested by adding Tween80 (0.01%) to each sample and centrifuging (10,000xg, 5 min, 4°C). The remaining pellet was washed once with 50 mM Tris/10 mM $CaCl_2$ [pH 7.6], centrifuged (10,000xg, 5 min, 4°C), resuspended in 600 μL 50 mM Tris/10 mM $CaCl_2$, pH 7.6, transferred to a microfuge tube, and stored at -80°C.

Extraction Buffers and Cell Lysis

Several cell lysis buffers were evaluated including: *Buffer A* - 9 M Urea, 70 mM CHAPS, 100 mM DTT, 0.5% Triton X100, from (73); *Buffer B* - 20 mM Tris-HCl, 1 mM DTT, protease inhibitor cocktail P2714 (Sigma Aldrich, St. Louis, MO) from (82, 84) *Buffer C* - 50 mM Tris HCl, 10% sucrose, 2 mM DTT, 4 mM EDTA, 0.1% Brij 58, pH 7.59, protease inhibitor cocktail (83, 96); *Buffer D* - 1.5 volumes (w/v) of cold 0.3 M K_2HPO_4, plus 0.5 volumes of 0.3 M EDTA, pH 8.0 (80); *Buffer E* – 20 mM KH_2PO_4, pH 7.0.

Using the above described buffers, cell lysis was performed using one or a combination of the following lysis techniques: (1) homogenization of cells with 1 mL of 0.1 mm zirconium beads using the Mini Bead Beater 8 (BioSpec Products) for 3 bead beating cycles (1 minute followed by cooling on ice for 1 minute), (2) sonication of cells placed in an ethanol ice bath with 10 second pulses for 3 minutes at the 45% amplitude

المنارة للاستشارات
www.manaraa.com

setting using a 3-mm microtip on a Sonic Dismembrator 550 (Fisher Scientific, Tustin, CA), (3) freeze thawing of cells through 4 cycles of freezing in liquid nitrogen followed by thawing at 30°C, (4) disruption of cells in 3 cycles through a cold (~4°C) French press mini-cell (124,000 kPa) (Thermo Scientific, Waltham, MA), or (5) chemical lysis of pelletted sample using 1 mL of 6 M guanidine and 10 mM DTT in 50 mM Tris/10 mM $CaCl_2$ [pH 7.6], 60°C. To avoid contamination betweens samples, the French Press cylinder and Sonic Dismembrator tip were rinsed thoroughly with ethanol, Windex (SC Johnson), and deionized water before reuse. Following all lysis, the soluble fraction of protein extracts was separated by centrifugation (21,000xg, 5 min, 4°C) unless otherwise specified.

Extraction of Proteins from Surface Waters and Sediments

Protein extracts from surface water samples were prepared by Joshua Livermore. Half of each Sterivex filter containing filtrate from surface waters was used for DNA extraction with the UltraClean Soil DNA Isolation Kit (MoBio Laboratories, Inc.) using a bead beating method for cell lysis (1 minute followed by cooling on ice for 1 minute, repeated twice). Protein removal was performed as a protein precipitation step per the manufacturer's instructions with Solution S2 provided within the UltraClean Soil DNA Isolation Kit (based on MSDS data, a proprietary acetate-based salt). For DNA extraction, the protein pellet is usually discarded; however, Livermore supplied these protein pellets for proteomic investigation. The protein pellets were stored at -20°C.

Protein extracts from sediment samples were prepared by Laura Badtke. Badtke separated cells from sediments using a modified Nycodenz method (8) with two different lysis buffers, the NaCl buffer described in (7) and the sodium hexametaphosphate buffer described in (8). Detailed methods are provided in Appendix B. After separation of cells from sediments, proteins were extracted through precipitation using the UltraClean

المنارة للاستشارات
www.manaraa.com

Soil DNA Isolation Kit, the same method as used for the surface water samples from Joshua Livermore. Sediment protein pellets were stored at -20°C.

Phenol and Nycodenz Extractions

For select samples where coextraction of interfering compounds was problematic, bacterial proteins were separated from the environmental matrix using one of the methods described in detail by (6) or (7). Briefly, the “phenol extraction (6)” was used to treat 5 g of soil with 10 mL of 0.1 M NaOH for 30 minutes followed by centrifugation (16,000xg, 10 min, 20°C). A portion of the supernatant (6 mL) was then mixed with 16 mL liquid phenol (10 g phenol and 1 mL water) and 10 mL deionized water and shaken for 1 hour at 20°C and then centrifuged (14,000xg, 10 min, 20°C) to separate the upper aqueous and lower phenol layers. The phenol phase (15 mL) was collected and washed with deionized water, and proteins precipitated with a five-fold volume of 0.1 M ammonium acetate overnight.

For the “Nycodenz extraction (7)”, Nycodenz (9 mL) was added to 25 mL of a slurry prepared from the homogenization of a soil sample (40 g) and disruption buffer (140 mL, 0.2 M NaCl, 50 mM Tris-HCl [pH 8.0]). Nycodenz was carefully inserted below the homogenate with a syringe and needle, and the sample was then centrifuged (10,000xg, 20 min, 4°C) in a Beckman SW28 swing-out rotor. After centrifugation, the sample contained from bottom to top: soil precipitate, an aqueous layer, a faint white band containing bacterial cells, and a nycodenz layer. Depending on the number of bacterial cells within the sample, the resolution of the white layer varied. The volume of sample containing the white band was removed carefully with a pipette, centrifuged (10,000xg, 4°C, 6 min), washed with phosphate buffered saline, centrifuged again, and resuspended in the appropriate buffer for protein extraction by cell lysis methods as described above.

المنارة للاستشارات
www.manaraa.com

Protein Extraction Evaluation

Protein extraction was evaluated by the measuring the amount of protein extracted by the Bradford Assay (11) (Coomassie Plus - Better Bradford Assay, Pierce Protein Research Products, Rockford, IL) and/or visualization on SDS-PAGE. For SDS-PAGE, protein extracts (maximum volume of 22 µL) were denatured (95°C, 5 minutes) in Laemmli Sample Buffer per manufacturer's instructions (Biorad Laboratories, Hercules, CA) and separated by mass and visualized by SDS-PAGE using the method of Laemmli (65). Polypeptides were loaded onto polyacrylamide gels (Criterion Tris-HCl Gel, 12.5%, Biorad Laboratories, Hercules, CA), and electrophoresis was allowed to proceed to completion (200 V, 55 min).

Trichloroacetic (TCA) Precipitation

In some cases, proteins were precipitated out of solution using TCA. To do so, ice-cold TCA (20% v/v) was added to samples and allowed to precipitate for 30 minutes (or overnight) at 4°C. After precipitation, protein was centrifuged (10,000xg, 5 min., 4°C). Pellet was washed two times with acetone. Finally, acetone was removed by vacuum centrifugation.

Results and Discussion

Protein Extraction of Strain JS614 Cells

Development of a suitable cell lysis method involved resuspending ~10^{8} VC- or ethene-grown JS614 cells into various buffers and lysing them via mechanical (french press, sonication, bead beating, or freeze-thawing) or chemical (9 M urea or 6 M guanidine) means to investigate the efficiency of protein release while also minimizing contaminants or chemicals that would interfere with downstream mass spectrometry analyses.

المنارة للاستشارات
www.manaraa.com

In general, the method used for cell lysis had a much greater effect on protein extraction efficiency than the choice of lysis buffer. Among the extraction buffers used in this study, there was not a preferred buffer based on the amount of protein extracted as observed qualitatively on an SDS-PAGE gel (Figure 3.1). Different buffers and chemical lysis agents (e.g., *Buffer A* (contains 9 M urea), *Buffer D* (contains 0.3 M EDTA), and 6 M guanidine) required that protein extractions be diluted, precipitated, and/or interfering solvents removed from prior to quantification and mass spectrometry analysis. Attempts to precipitate proteins using either trichloroacetic acid (TCA) or acetone resulted in insoluble pellets and/or significant protein loss. We also considered that the practical costs of making large volumes of buffers, especially *Buffer A*, did not provide significant increases in protein extraction. Therefore we used 20 mM K_2HPO_4 [pH 7.0] for extraction of protein from JS614 cells. The 50 mM Tris/10 mM $CaCl_2$ [pH 7.6] buffer was equally as effective for extracting proteins from JS614 and has been used previously for protein extraction from environmental samples with subsequent mass spectrometry analysis (91). The Tris-$CaCl_2$ buffer was used for select environmental samples as noted.

Among the cell lysis techniques, the most effective for ~10^8 JS614 cells/ml was the French Press method which resulted in the most abundant and consistent protein extractions (2.7 ± 0.4 mg protein/mL, n=3 extractions). In this study, bead beating methods were limited to the maximum sample volume that could be homogenized by the described bead beater (~ 2 mL) and measured protein concentrations were more variable than other methods (2.4 ± 1.1 mg protein/mL, n=3 extractions). Using guanidine for chemical lysis of JS614 cells (0.9 ± 0.2 mg protein/mL, n=3 extractions) or combining guanidine lysis with sonication (1.3 ± 0.7 mg protein/mL, n=3 extractions) was not as effective as sonication alone (1.9 ± 0.3 mg protein/mL) and required significant dilution prior to downstream applications. The freeze thaw lysis method was the least effective among the lysis methods studied (0.15 ± 0.04 mg protein/mL). Protein extractions from

المنارة للاستشارات
www.manaraa.com

pure JS614 cultures were further visualized with SDS-PAGE (Figure 3.2) and these results supported measurements of protein concentration. Extracts resulting from French press, sonication, and bead beating required dilution (at least ten fold) for comparison to freeze thaw and guanidine extracts. Darker bands (more protein) on SDS-PAGE were observed for protein extracts lysed by sonication, French press, and bead beating methods, while the lightest bands (less protein) were observed for freeze thawing and guanidine protein extracts.

From these results, we selected French Press and sonication as our preferred lysis methods with either 20 mM K_2HPO_4 [pH 7.0] or 50 mM Tris/10 mM $CaCl_2$ [pH 7.6] buffers for proteins from visible cell pellets (such as those from centrifuged pure cultures or water samples). If particulates are present within a sample, the French Press method cannot be used without damaging the French Press cylinder, and lysis through sonication is recommended.

Protein Extraction from Activated Sludge

Activated sludge samples were obtained from the local wastewater treatment plant. A 200 mL sample was centrifuged (7000xg, 10 min, 4°C), and the pellet was treated with 10 mL of 0.1 M NaOH and phenol extraction as described in (6). Unexpectedly, no protein pellet was ever visible after protein precipitation steps during the extraction process, but the extraction was carried through completion. For comparison, additional samples of activated sludge (200 mL each) were pelletted, and protein extracted with sonication and freeze thaw lysis methods in 1 mL of *Buffer E.* JS614 cells (10^7 cells, 50 mL) were also added to an activated sludge sample (200 mL), pelletted, and protein extracted with sonication. Protein extractions were visualized on a SDS-PAGE gel (Figure 3.3). No bands were observed in protein extractions using the phenol extraction method which is unsurprising since we could not visually identify any protein precipitate from the extraction (lane 1, Figure 3.3). It is possible that the 0.1 M

المنارة للاستشارات
www.manaraa.com

NaOH buffer was not effective for lysing the cells in our samples. Proteins extracted from freeze-thawed activated sludge samples resulted in a “smearing” effect within the lane (Figure 3.3, lane 2). Sonicated protein extracts from activated sludge without and with the addition of JS614 cells (Figure 3.3, lanes 3 and 4, respectively) resulted in a less intense “smearing” effect on SDS-PAGE and the observation of faint protein bands where JS614 cells were added (Figure 3.3, lane 4). Considering the addition of JS614 cells, we did not observe as significant protein extraction in our samples as expected (compared to protein extracts from pure cultures of JS614 (lane 5, Figure 3.3)). In future experiments, we recommend increasing the amount of JS614 cells added to samples. Further studies evaluating the phenol extraction method should also include a mechanical lysis step such as sonication rather than relying on chemical lysis alone.

Sections within the 45-80 kDa region of select lanes on the gel (as indicated on Figure 3.3) were excised for select samples and processed for protein identification with tandem mass spectrometry and the protein identification program SEQUEST (32). Searching the translated EMBL (trEMBL) database for bacteria, no conclusive protein identifications could be made. Protein identifications required at least 2 peptide identifications per protein with significant scores (XCorr values of at least 1.8 (+1), 2.5 (+2), 3.5 (+3)). For freeze thaw extracted proteins, single peptide matches to proteins from *Prochlorococcus* marinus, *Francisella* philomiragia, and *Nitratiruptor* sp. strain SB155-2 were found using rigorous scoring filters (z=2, XCorr>2.5). However, these matches were not considered significant since only one peptide was identified per protein. Strain JS614 proteins were identified in bands A (15 proteins), B (8 proteins), and C (1 protein) as shown in Figure 3.3. Comparatively, if the same proteins had been excised from pure cultures of JS614, we are usually able to identify between 20-70 proteins with similar methods.

Based on these results, we concluded that more protein is necessary to reevaluate extraction methods and to recover conclusive protein identifications from activated

www.manaraa.com

sludge samples. Furthermore, the protein database used to identify protein matches requires consideration before selection. For proteomic investigations, it is important to use the most practical database available for protein identifications. For our activated sludge protein identifications, we used the trEMBL database which contains over 4 million sequences (compared to ~2,000 protein sequences for JS614). The scoring algorithm used for protein identifications analyzes the best protein match in comparison to all other possible proteins in the database (32). Thus, a larger protein database increases the search time and also increases the number of sequences that could be false matches to the spectra in question. For future experiments, we concluded that, if possible, a targeted proteomic investigation would be more effective than a discovery approach like that used for activated sludge samples.

Protein Extraction from Soil

Preliminary efforts to extract proteins directly from soil resulted in brown substances within extracts which interfered with both the Bradford assay and SDS-PAGE analysis (particularly when using *Buffers A* and *D*). To remove these interfering compounds, an ultracentrifugation step (80,000xg, 1 hour, 4°C) and protein precipitation step (TCA precipitation) was amended to all protein extraction methods for soil samples. We also found that the addition of strain JS614 cells (10^{8}–10^{9} cells) to soil samples ensured that sufficient protein was present within samples to evaluate protein extraction methods.

Strain JS614 cells were added to soil samples collected from Orion, IL. Protein extraction was performed with sonication in various buffers. Previously, we had found that with JS614 cells, the choice of lysis method was more important than buffer choice and that sonication should provide effective lysis for samples with particulates. In the soil matrix, microbial communities might interact differently with soil particles than in pure cultures (i.e. cells adhere to particulates); therefore, we tested a range of buffers for

www.manaraa.com

protein extraction effectiveness. Each buffer (2 mL, *Buffers A-E*) was combined with 2 g of soil and 10^9 cells of JS614 and sonicated. Samples without the addition of JS614 cells in *Buffer E* were also sonicated. Following protein extraction, ultracentrifugation and protein precipitation was performed.

After sonication and ultracentrifugation, for *Buffers A* and *D*, we observed a brown pellet and brown-yellow supernatant, and for *Buffers B*, *C,* and *E*, we observed a brown pellet and clear supernatant. It is assumed that brown pellets contained the soil matrix and that proteins were contained in supernatants. After overnight protein precipitation of supernatants, a brown, very loose pellet (oily in viscosity) was observed in *Buffer A* samples, nothing in *Buffer B* and *Buffer C*, and white creamy pellets in *Buffers D* and *E*. Protein pellets were resuspended in *Buffer E*. In general, we found it difficult to resuspend protein precipitates, and it is likely that protein loss resulted from incomplete resuspension of pellets. We had experienced similar problems when resuspending proteins precipitated from pure strain JS614 cells. Protein concentrations were estimated with the Bradford Assay (Table 3.1). Protein concentrations varied widely within triplicate samples for proteins extracted in *Buffers A-D*. These inconsistencies were most likely related to interferences of the brown precipitate (*Buffer A*) and/or protein losses during the resuspension of precipitated proteins. Proteins concentrations from proteins extracted in *Buffer E* were the most consistent supporting the use of *Buffer E* for future extractions. SDS-PAGE analysis was performed to visualize protein extractions, and streaking was observed for extracts in *Buffers A – D* (lane 2-5, Figure 3.4). A dark protein band was observed in extracts in *Buffers C* and *D* (lanes 4 and 5, Figure 3.4). The combination of inconsistent protein concentration measurements and streaks during SDS-PAGE indicate that interfering substances, most likely from the incomplete removal of humic organic matter, were present in our protein extractions. Even with the addition of JS614 cells, we did not observe many proteins on SDS-PAGE.

المنارة للاستشارات
www.manaraa.com

We also tested protein extractions from similar samples with bead beating lysis rather than sonication (1 g soil, 1 mL buffer). Results were similar to sonicated extracts in that protein concentration measurements varied widely for identically treated samples (data not shown). However, for bead beaten extracts, we did not observe any indication of protein extraction on SDS-PAGE (i.e. no bands). These results validated that sonication extraction methods are preferred over bead beating methods for soil samples and samples with high particulate content.

Protein extractions from soil using these methods were relatively unsuccessful because of the presence of interfering compounds. We determined that methods to separate bacterial cells and/or proteins from the soil matrix would be necessary for the identification of proteins from these samples.

Using the Nycodenz extraction method, we used 40 g of soil from Orion, Illinois and added 10^9 JS614 cells and disruption buffer. After mixing, Nycodenz was added to the slurry and centrifuged. A white band of bacterial cells could be observed between the aqueous and nycodenz layer. This layer was removed, pelletted, and washed as described in (7). The pellet was resuspended in 500 µL of *Buffer E* and sonicated as described above. The resulting protein extract was visualized on SDS-PAGE (Lane 1, Figure 3.4). As shown, clear protein bands can be observed using the Nycodenz method. The protein band pattern is very similar to that of strain JS614, and it is likely that these proteins are from the addition of JS614 cells and not from the soil sample. Further experiments would be required to identify the observed protein bands.

We also attempted the phenol extraction with 5 grams of soil and the addition of 10^9 JS614 cells. Similarly to our activated sludge samples, we were unable to obtain any protein precipitates using this method and determined that the Nycodenz extraction method to separate cells was more promising for our experiments.

We were able to extract proteins from a soil sample with the fresh addition of a significant number of bacterial cells using the Nycodenz extraction method. However, in

المنارة للاستشارات
www.manaraa.com

soil from the environment, it is unlikely that samples would contain such highly enriched populations of a single species of bacteria. Nonetheless, this experiment shows that the Nycodenz method could be used to separate some cells from the soil matrix. Further study of this method from environmental samples would be necessary to evaluate the ability of protein identifications from soil.

Protein Extraction from Surface Waters

Protein extracts from ten surface water samples were given to us from Joshua Livermore. Protein extracts were in the form of a protein precipitate pellet which we resuspended in *Buffer E* and analyzed with SDS-PAGE. Three bands were observed (Figure 3.5, Lanes 2, 5, and 6). Two of these bands (Figure 3.5, Lanes 2 and 6) were excised and analyzed for protein identification with tandem mass spectrometry and the program XTandem! (22). Using all available bacterial sequences on the XTandem! database (October 5, 2008), we found significant protein matches to the alkanal monooxygenase alpha chain (11 peptides, 48% amino acid coverage) and alkanal monooxygenase beta chain (10 peptides, 52% amino acid coverage) from *Vibrio* harveyi (ATCC BAA-1116) from the protein band excised from lane 2 (Figure 3.5 and 3.6). These protein identifications were the first significant protein identifications from an environmental sample in our study. We attributed our success with these samples to either an effective extraction method and/or the ability to concentrate bacterial cells on a filter.

We tested the UltraClean Soil DNA Isolation Kit for its extraction efficiency with pure cultures of JS614. We found that the protein precipitation step (using the S2 buffer in the kit) was rather inefficient compared to TCA precipitation; there was much less precipitate formed using the kit. Further studies with filtered surface water samples should be conducted using other protein extraction and/or precipitation methods.

المنارة للاستشارات
www.manaraa.com

As a preliminary study of using filters for protein extraction, we added 10^9 JS614 cells to 1 L of tap water, stirred for 15 minutes at room temperature, and then passed the mixture through Sterivex filters (500 mL/filter). Half of each filter was then sonicated in 2 mL of buffer (*Buffers A-E*) and centrifuged (10,000xg, 30 min, 4^oC) The supernatant was removed. As we had nearly 2 mL of supernatant, we concentrated the protein extract using Amicon Ultra-4 10kDa Centrifugal Filter Units (Millipore Corp.). The concentrated filtrate contained protein extracts (>10 kDa). Protein concentrations were measured with the Bradford Assay (Table 3.2). Controls of tap water without JS614 were also analyzed, and no protein was measured in these samples. Interestingly, protein extracts in *Buffer E* were the least effective and consistent while protein extracts in *Buffers A* and *B* were the most effective and consistent. We visually observed break down of the filter material in these buffers during the extraction. It is likely that the components of these buffers help to break down the filters and thus release the attached cells more effectively than *Buffer E*. If further proteomic investigations into filtered water samples were used, more research into the effects of lysis buffer should be conducted. These preliminary studies demonstrate that filtered water samples would be good candidates for targeted proteomic investigations.

Protein Extractions from Sediments

Protein extracts were prepared from sediment samples from Indiana Harbor by Laura Badtke. She was interested in protein identification within these samples using tandem mass spectrometry but was limited by time and funding for her project. We decided to run her proteins on SDS-PAGE gel for subsequent analysis by tandem mass spectrometry; however, instead of running protein extracts for a full hour on the gel, proteins were run only briefly (5-10 minutes) until proteins were “in” the gel. In this case, the primary purpose of SDS-PAGE was to clean the sample prior to tandem mass spectrometry rather than to provide separation of proteins. The advantage of this

www.manaraa.com

experiment is that only one band needed to be excised and analyzed for each protein extract rather than multiple bands within a region of interest.

In total, 4 protein extracts from sediments were provided and analyzed with "brief" SDS-PAGE. Each sample was excised as a single band and analyzed by tandem mass spectrometry and the protein identification software SEQUEST (32) and the TransProteomic Pipeline (TPP) (63) (Seattle Proteome Center, Seattle, WA). The TPP program converts SEQUEST scores into a score that is based on the probability that the match is correct, making it easier to communicate the significance of data to scientists outside the field. The trEMBL bacteria database was used for protein identifications. Significant protein identifications required the TPP probability to be greater than 0.90 and at least two peptides identified for each protein. Among the sediment samples, only one sample contained significant protein identifications. Eight proteins were identified, a 60 kDa chaperonin from *Sorangium* cellulosum, a putative branched-chain amino acid transport system substrate-binding protein from *Polaramonas* sp. JS666, three translation elongation factor Tu proteins, and two RNA polymerases.

These results suggest that protein identifications from sediment samples can be made using Nycodenz separation as described in (8) and buffer described in (7). No protein identifications were made from extracts using the other tested buffer, sodium hexametaphosphate, and Badtke commented that the cell layer interface between Nycodenz and aqueous layers was easier to identify using the Nycodenz buffer rather than the sodium hexametaphosphate buffer. Given our initial success, more research is warranted to study the ability to extract proteins from sediment samples using these methods. In this experiment, we mostly identified proteins that are likely to be highly abundant in any sample (i.e. elongation factors and RNA polymerases). For future studies, assuming catabolic proteins are targeted, protein separation and concentration are recommended with "complete" SDS-PAGE and Amicon Ultra-4 10kDa Centrifugal Filter Units (previously used for surface water samples).

المنارة للاستشارات
www.manaraa.com

Protein Extraction from Groundwater Samples

Cell pellets prepared from groundwater samples from Bioremediation Consulting were sonicated or French pressed for protein extraction, depending on the amount of particulates visually observed in the sample. If particulates were present, proteins were extracted with sonication; in no particulates were present, proteins were extracted with the French press. Cell pellets were resuspended in 600 µL 50 mM Tris/10 mM CaCl2 [pH 7.6], and proteins extracted. Protein extracts (20 µL each) were analyzed with SDS-PAGE (Figure 3.7). Protein bands were observed for the three groundwater enrichment samples shown indicating that enough protein is likely to be present for successful protein identification with tandem mass spectrometry.

An additional advantage to these samples is that because they are enrichment samples for VC- and/or ethene-degrading bacteria, we can use a targeted proteomics approach for VC-degrading protein biomarkers. Protein identifications can be made using a much smaller database compared to the previously used trEMBL bacteria database. The presence of "enriched" VC-degraders and a smaller database increases the probability of obtaining significant protein identifications. The initial success with protein extraction in these groundwater samples combined with the opportunity to practically use a targeted proteomics approach led us to select these samples and others from this site for further tandem mass spectrometry analysis (Chapter IV).

Conclusions

Protein extraction from several environmental samples was demonstrated. In the most successful cases, bacterial cells were separated from the environmental matrix prior to protein extraction. In soil and sediment samples, the Nycodenz density gradient was used to remove bacterial cells. In water samples, filtration or centrifugation was used to separate bacterial cells. We found with soils in particular, interfering substances (most likely organic matter or humic acids) made the evaluation and concentration of protein

المنارة للاستشارات
www.manaraa.com

extraction difficult. We had success identifying proteins from water samples using our extraction methods and tandem mass spectrometry. In general, we found that sonication can effectively lyse most cells, and SDS-PAGE is a versatile tool that provides both sample clean-up and protein separation. Although SDS-PAGE is limited by the amount of volume that can be loaded into gel wells, protein concentration steps such as protein precipitation or filter centrifugation can be used to improve results.

Based on these preliminary results, protein extraction and subsequent identification is possible for environmental samples but will require optimization for the specific samples being studied. In general, targeted proteomic investigations are much simpler than discovery proteomic investigations because of the simplified protein analysis of a smaller database. It is also extremely helpful to use environmental samples where some information about microbial contents is known. Environmental samples that are enriched with targeted proteins are ideal for proteomic investigations because firstly, one may use a smaller targeted protein database for protein identifications, and secondly, more protein is available for extraction from such samples.

المنارة للاستشارات

www.manaraa.com

Table 3.1. Protein concentrations measured from protein extracts of soil samples with JS614 cells in various buffers (n=3 for each buffer).

	Protein Concentration (μg/μL)		
Soil with JS614 cells in Buffer A	382.8	±	99.3 (n=3)
Soil with JS614 cells in Buffer B	70.5	±	29.7 (n=3)
Soil with JS614 cells in Buffer C	66.4	±	24.6 (n=3)
Soil with JS614 cells in Buffer D	240.1	±	497.2 (n=3)
Soil with JS614 cells in Buffer E	61.5	±	2.55 (n=3)
Soil without JS614 cells in Buffer E	84.0	±	8.6 (n=3)

Note: Lysis was performed with sonication.

المنارة للاستشارات

www.manaraa.com

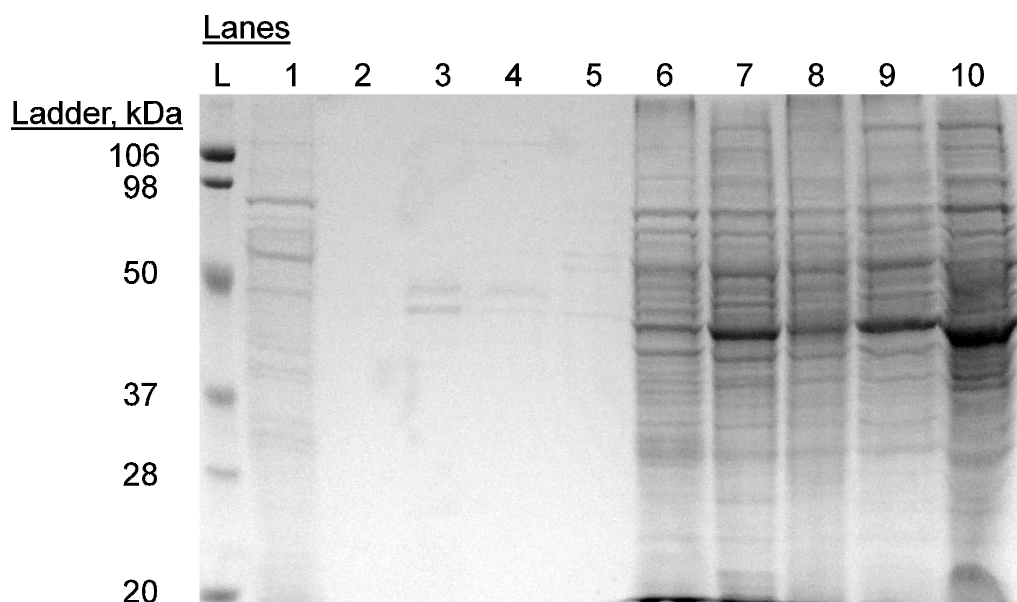
Table 3.2. Protein concentrations of extracts from half Sterivex filters through which a mixture of tap water and JS614 cells was filtered.

	Protein Concentration (μg/μL)		
Water and JS614 cells in Buffer A	298.2	±	39.3 (n=3)
Water and JS614 cells in Buffer B	246.5	±	26.8 (n=3)
Water and JS614 cells in Buffer C	268.6	±	69.1 (n=3)
Water and JS614 cells in Buffer D	102.2	±	94.6 (n=3)
Water and JS614 cells in Buffer E	119.8	±	118.2 (n=3)

Note: Extracts were concentrated in centrifugal filter units.

المنارة للاستشارات

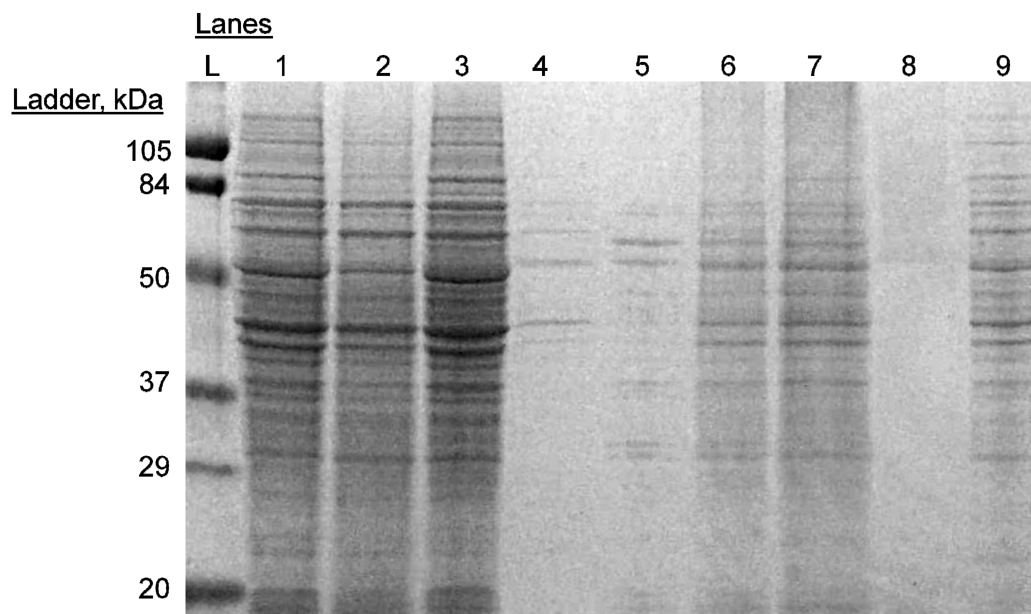
www.manaraa.com



Notes: Protein was extracted from 10^8 cells of VC-grown JS614 resuspended in 1 mL of buffer A, B, C, D, or E as described here. Buffer A - 9 M Urea, 70 mM CHAPS, 100 mM DTT, 0.5% Triton X100 (lanes 1 and 6); Buffer B - 20 mM Tris-HCl, 1 mM DTT, protease inhibitor cocktail P2714 (lanes 2 and 7); Buffer C - 50 mM Tris HCl, 10% sucrose, 2 mM DTT, 4 mM EDTA, 0.1% Brij 58, pH 7.59, protease inhibitor cocktail (lanes 3 and 8); Buffer D - 1.5 volumes (w/v) of cold 0.3 M K2HPO4, plus 0.5 volumes of 0.3 M EDTA, pH 8.0 (lanes 4 and 9); Buffer E – 20 mM KH2PO4, pH 7.0 (lanes 5 and 10).

Figure 3.1. SDS-PAGE gel comparing freeze-thaw lysis (lanes 1-5, 20 µL extract/lane) to sonication (lanes 6-10, 2 µL extract/lane) in various protein extraction buffers.

المنارة للاستشارات
www.manaraa.com

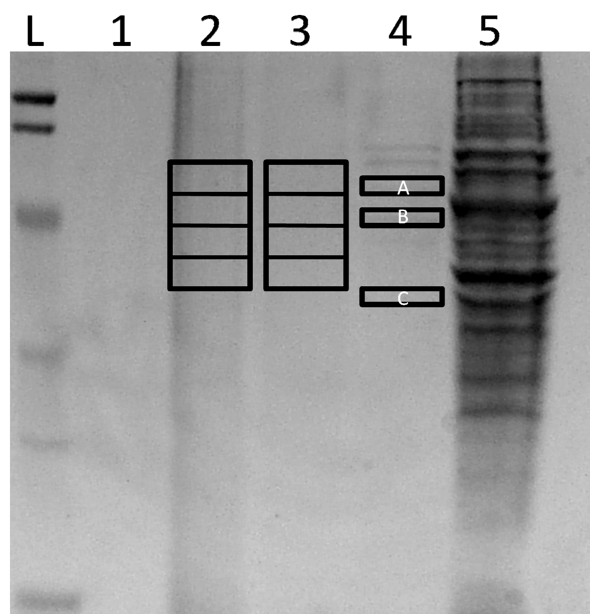


Notes: French press (lane 1, 2 µL extract, 5.7 µg/µL), sonication (lane 2, 2 µL extract, 4.4 µg/µL), bead beating (lane 3, 2 µL extract, 7.5 µg/µL), french press (lane 4, 0.2 µL extract, 0.6 µg/µL), sonication (lane 5, 0.2 µL extract, 0.4 µg/µL), bead beating (lane 6, 0.2 µL extract, 0.8 µg/µL), freeze thawing (lane 7, 20 µL extract, 0.8 µg/µL), guanidine (lane 8, 20 µL extract, 0.02 µg/µL), guanidine and sonication (lane 9, 20 µL extract, 0.2 µg/µL). Note that guanidine extracts required a six-fold dilution prior to SDS-PAGE; concentrations shown represent concentration of protein loaded for SDS-PAGE.)

Figure 3.2 1D SDS-PAGE gel comparing protein extractions from various lysis techniques of 10^8 VC-grown JS614 cells/ml:

المنارة للاستشارات

www.manaraa.com



Notes: Protein extract from JS614 cells (10 µL) using the French Press method (lane 5) are also shown for comparison. Boxed region indicates where bands were cut for tandem mass spectrometry analysis. Strain JS614 proteins were identified for proteins extracted from bands shown as A, B, and C.

Figure 3.3 SDS-PAGE of protein extracts (20 µL) from activated sludge using the phenol extraction method (lane 1), activated sludge using the freeze thaw method (lane 2), activated sludge using the sonication method (lane 3), activated sludge + JS614 cells using the sonication method (lane 4).

www.manaraa.com

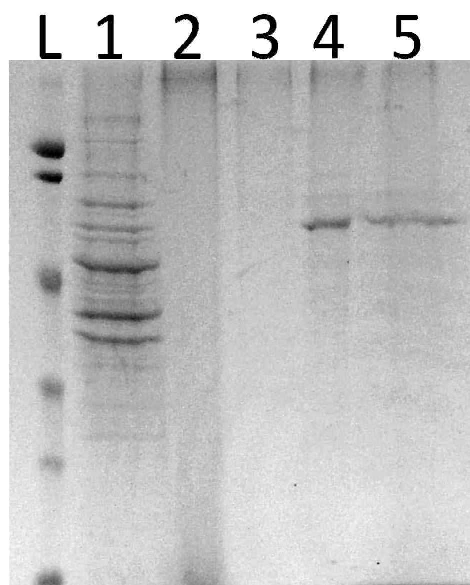
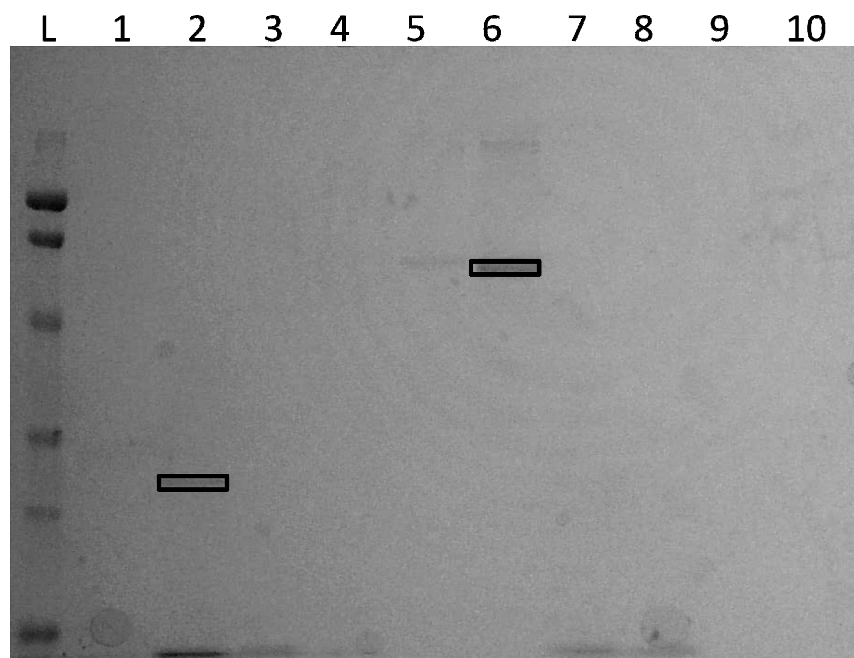


Figure 3.4. SDS-PAGE of protein extracts from soil using the Nycodenz method followed by sonication (Lane 1) and sonication in Buffers A, B, C, and D, ultracentrifugation, and protein precipitation resuspended in Buffer E (lanes 2-5 contain sonication in Buffers A-D, respectively).

المنارة للاستشارات
www.manaraa.com



Note: Each protein extract is from half a filter which surface waters were passed through. Indicated bands were excised for protein identification with tandem mass spectrometry.

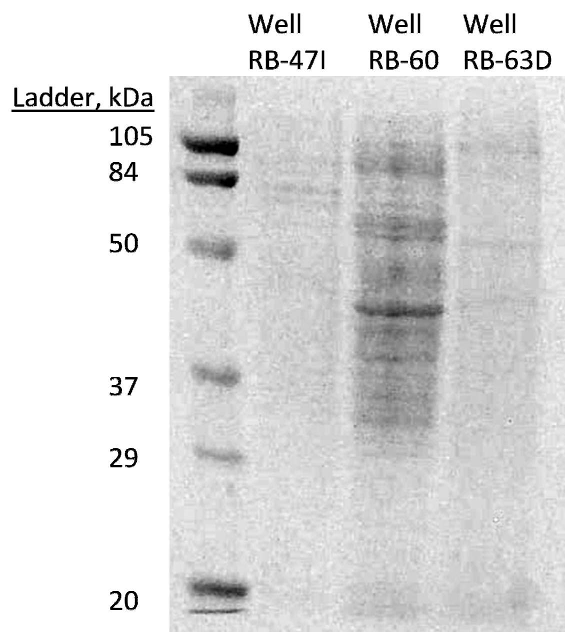
Figure 3.5. SDS-PAGE of ten protein extracts from Sand Lake using the UltraClean Soil DNA Isolation Kit.

المنارة للاستشارات

www.manaraa.com

Figure 3.6. Results from XTandem! search of MS/MS spectra of peptides from protein from band in Lane 2 (Figure 3.4).

المنارة للاستشارات
www.manaraa.com



Note: Well RB-47I contains 1.7 µg, Well-RB60 contains 11.8 µg, and Well RB-63D contains 1.2 µg protein.

Figure 3.7. SDS-PAGE of protein extracts from groundwater samples obtained from VC-contaminated groundwater enrichment samples.

www.manaraa.com

CHAPTER IV: PROTEOMIC METHODS REVEAL POLYPEPTIDES ASSOCIATED WITH VINYL CHLORIDE- AND ETHENE-ASSIMILATING BACTERIA IN ENRICHED GROUNDWATER SAMPLES FROM A CONTAMINATED SITE[2]

Abstract

Contamination of groundwater with vinyl chloride (VC), a known human carcinogen, is a common environmental problem at plastics manufacturing, dry cleaning, and military sites. One approach to the clean-up of VC-contaminated sites is via aerobic biodegradation with microorganisms that grow on VC and/or ethene as a carbon and energy source (called VC- and ethene-assimilating bacteria). Application of proteomic techniques to samples from polluted environments (e.g. VC-contaminated sites) can reveal the metabolic functionality of specific microbial community members and therefore holds the potential to benefit the practice of bioremediation. Mass spectrometry-based proteomic methods for identifying protein biomarkers of aerobic, VC- and ethene-assimilating bacteria (alkene monooxygenase alpha subunit (EtnC) and epoxyalkane:CoM transferase (EtnE)) were developed and applied to ethene enrichment cultures derived from VC-contaminated groundwater. EtnC and EtnE peptides were identified in 9 of the 14 enrichments analyzed by these proteomic methods. BLAST analysis revealed that 4 EtnC and 4 EtnE peptides were unique to deduced EtnC and EtnE sequences found in 3 different cultivated strains. In addition, 6 *etnE* sequences retrieved from enrichment samples RB-63I and RB-73 matched with expressed EtnE polypeptides observed in enrichment samples RB-58I, RB-60, and RB-73. These results suggest that

[2] Planned for submission for publication in *Journal of Microbiological Methods* in September, 2009 with coauthors including Yang Oh Jin, Laura Schmidt, Yalan Li, Samuel Fogel, and Timothy Mattes.

www.manaraa.com

there is a diverse population of VC- and/or ethene-assimilating bacteria present at this particular VC-contaminated site.

Introduction

Prior to the genomic era, understanding microbial community structure and function relied on studies of microbial isolates and PCR-based analysis of rRNA and/or functional genes from environmental samples. Relatively recent advances in DNA sequencing technology have facilitated a proliferation of microbial genome and metagenomic sequences. Application of microbial genomic and metagenomic techniques has led to novel insights about the structure and function of microbial communities. The ever increasing availability of microbial genome and metagenomic sequences in concert with advances in microarray and mass spectrometry technologies has initiated the "post-genomic" era, whereby transcriptomic and proteomic methodologies are now being applied to the study of microbial communities. Postgenomic methods have the potential to reveal metabolic functionality, rather than metabolic potential, of specific microbial community members. Proteins are particularly well-suited as bioindicators of microbial activity in the environment because they are directly involved in enzyme-catalyzed reactions, cell maintenance, and cellular response to environmental perturbations. Proteomic approaches are also not limited by the short half-life of RNA and bypass the challenges related to low correlations between mRNA and protein expressions that are typically associated with transcriptomics (44, 109).

Several studies have used proteomics approaches to characterize microbial communities in different environments (reviewed in (74, 109)). These environments include a natural microbial biofilm (91), water (6, 62, 73, 82-84), soil (6, 83, 93, 96), and activated sludge (50, 110). The ability to identify proteins in environmental samples has created unique opportunities to study microbial communities involved in the bioremediation of contaminated environmental sites. However, application of proteomics

المنارة للاستشارات
www.manaraa.com

to bioremediation problems is still a burgeoning research area. Here, we examine microbial communities in ethene-enriched groundwater samples from a vinyl chloride (VC)-contaminated site.

VC, a known human carcinogen (13) and common groundwater contaminant (101), is frequently generated in groundwater by incomplete reductive dechlorination of the widely used chlorinated solvents tetrachloroethene (PCE) and trichloroethene (TCE), also common groundwater contaminants (101). At some sites, VC produced under anaerobic conditions will migrate into aerobic groundwater along with dissolved ethene and methane, also formed under anaerobic conditions. However, reports of rapid VC degradation in aerobic groundwater are common (25, 29, 67). At least three groups of microorganisms could participate in the observed aerobic attenuation of VC. These include the methanotrophs (38, 105) and the "etheneotrophs" (also called ethene-assimilating bacteria) (40), both of which cometabolize VC during aerobic growth on their primary substrates (methane and ethene, respectively). A specialized group of ethene-assimilating *Mycobacterium, Nocardioides, Pseudomonas, Ralstonia* and *Ochrobactrum* strains can carry out growth-coupled VC oxidation (also called VC-assimilating bacteria). VC-assimilating bacteria appear to be widespread in the environment and have been isolated from soil (19, 30), groundwater (19, 23), sediment (47), and sewage sludge samples (19, 47, 106).

Both VC- and ethene-assimilating bacteria employ the enzymes alkene monooxygenase (AkMO; EtnABCD) and epoxyalkane:coenzyme M transferase (EaCoMT; EtnE) during the initial steps of the aerobic microbial VC and ethene biodegradation pathways (20, 21, 24, 47, 60, 77). AkMO converts ethene to epoxyethane and VC to chlorooxirane. These epoxide intermediates are subsequently metabolized by EaCoMT (21, 24, 77).

The objective of this study was to explore the application of proteomics techniques for detecting protein biomarkers (e.g. EtnC and EtnE) indicative of the

المنارة للاستشارات
www.manaraa.com

presence and functionality of aerobic VC- and ethene-assimilating bacteria in the environment. We developed proteomic methods that facilitated identification of EtnC, EtnE and other polypeptides from VC-contaminated groundwater that was subjected to microbial enrichment with ethene as a carbon and energy source for approximately 60 days. Bioinformatic analysis of EtnC and EtnE sequences revealed a diversity of functional etheneotrophic bacteria in these enrichments.

Experimental Section

Environmental sample collection

VC-contaminated groundwater, collected from several monitoring wells at a site in MA on 10/03/2007 and 07/01/2008, was used to prepare ethene-fed enrichment cultures. Previous work has shown that ethenotrophic bacteria capable of VC biodegradation are present within the VC plume at this site (39). Enrichment cultures were prepared in sterile 160 mL serum bottles and contained 40 mL of groundwater and 60 ml of a mineral salts media (containing 40 ppm N, 70 ppm P, 166 ppm K, 27 ppm S, 20 ppm Mg, 5 ppm Ca, 0.06 ppm Fe, 218 ppm PO_4, 81 ppm SO_4, and 177 ppm NO_3, and 1 ml of a trace metals mixture at pH 6.9). The trace metals mixture consisted of $ZnSO_4$*7 H_2O (7 mg/ml), $MnCl_2$*4 H_2O (2 mg/ml), H_3BO_3 (1.5 mg/ml), $CoCl_2$*6 H_2O (5 mg/ml), $CuCl_2$*2 H_2O (1 mg/ml), $NiCl_2$*6 H_2O (1 mg/ml) and Na_2MoO_4*$2H_2O$ (3 mg/ml). Ethene was added to an initial concentration of 2.85 mM. After 63-64 days of incubation, headspace samples (100 ul) were analyzed by gas chromatography in accordance with EPA method 5021A. Both flame ionization detection (for ethene measurements) and thermal conductivity detection (for % oxygen and % CO_2 measurements) were used. Indirect evidence of microbial growth in the enrichments was determined by visual inspection of turbidity. Based on these analyses, the enrichments were declared either positive or negative for the presence of etheneotrophs. Following

www.manaraa.com

the completion of these analyses, the enrichments were shipped to our laboratory for molecular biology and proteomic analyses.

Extraction of DNA and proteins from ethene enrichments

Solids were harvested from the enrichments by adding Tween80 (0.01% final concentration) to each 100 mL sample, and then centrifuging (10,000xg, 5 min, $4^{o}C$). For DNA extraction, pellets from 10/03/2007 enrichments were resuspended in 600 uL STE buffer and subjected to a beadbeating and phenol/chloroform/isoamyl alcohol extraction procedure carried out essentially as described previously (61). For protein extraction, the pellet was washed once with 50 mM Tris/10 mM $CaCl_2$ [pH 7.6], centrifuged (10,000xg, 5 min, $4^{o}C$), resuspended in 600 uL 50 mM Tris/10 mM $CaCl_2$ [pH 7.6], transferred to a microfuge tube, and stored at $-80^{o}C$ until further treatment, as described below. Proteins were extracted from pellets of 07/01/2008 enrichments that were resuspended in 600 µL 50 mM Tris/10 mM $CaCl_2$ [pH 7.6] and lysed using one of the following methods: (1) sonication of cells placed in an ethanol ice bath with 10 second pulses for 3 minutes at the 45% amplitude setting using a 3-mm microtip on a Sonic Dismembrator 550 (Fisher Scientific, Tustin, CA), (2) disruption of cells in 3 cycles through a cold (~$4^{o}C$) French press mini-cell (124,000 kPa) (Thermo Scientific, Waltham, MA), or (3) chemical lysis of pelletted sample using 1 mL of 6 M guanidine and 10 mM DTT in 50 mM Tris/10 mM $CaCl_2$ [pH 7.6], $60^{o}C$. Protein extraction methods were chosen based on the visible presence of particulates in resuspended pellets. If particulates were visible, proteins were extracted via sonication; if no particulates were visible, proteins were extracted via either French press or guanidine lysis. Based on these criteria, protein in the majority of the samples (n=10) was extracted by sonication. Protein in the remaining 4 samples was extracted by either French press or guanidine lysis (Table 4.1, Fig. 4.1).

المنارة للاستشارات
www.manaraa.com

To avoid contamination betweens samples, the French Press and Sonic Dismembrator were rinsed thoroughly between extractions with ethanol, Windex (SC Johnson, Racine, WI), and deionized water. For guanidine-treated samples, a 6-fold dilution with 50 mM Tris/10 mM $CaCl_2$, pH 7.6, was performed prior to SDS-PAGE or trypsin digestion. Following lysis, the soluble fraction of protein extracts was separated by centrifugation at 21,000xg for 5 minutes at 4°C. Extracted proteins were quantified using the Coomassie Plus – The Better Bradford Assay (Pierce Protein Research Products, Rockford, IL) against BSA standards resuspended in the appropriate buffer.

PCR, cloning, and DNA sequencing of EtnC and EtnE
genes from ethene enrichments

EtnE genes were amplified with CoM-F1L and CoM-R2E primers (expected size 891 bp) (20) with the thermocycling parameters described in (21). EtnC genes were amplified with NVC105 and NVC106 primers (expected size 360 bp) using the thermocycling parameters described in (18). PCR products of the appropriate size were purified with the Qiagen PCR purification kit, cloned into the pDRIVE vector (Qiagen) and transformed into competent *E. coli* (New England BioLabs). Transformants were grown on LB+kanamycin (50 ng/µl) with 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside (X-gal) (40 ng/µl), and Isopropyl-β-D-Thiogalactopyranoside (IPTG) (100 mM) at 37 °C overnight. Single white colonies were transferred into LB+Kanamycin (50 ng/µl) broth, and incubated at 37 °C overnight with shaking at 250 rpm. Recombinant vectors were purified with the QIAprep Spin Miniprep Kit (Qiagen) and sequenced with the M13 forward primer at the University of Iowa DNA facility. Retrieved sequences were analyzed by BLAST (5), compared with the PHYLIP DNA distance algorithm (35), and deposited into Genbank under accession nos. GQ847806-GQ847821.

www.manaraa.com

Protein analysis using polyacrylamide gel electrophoresis,
liquid chromatography, and mass spectrometry

The majority of protein extracts were denatured, separated by mass, and visualized by SDS-PAGE using the method of Laemmli (65). Extracts were denatured in Laemmli Sample Buffer (Biorad Laboratories, Hercules, CA) for 5 minutes at 95°C per manufacturer's instructions. Polypeptides were loaded onto polyacrylamide gels (Criterion Tris-HCl Gel, 12.5%, Biorad Laboratories, Hercules, CA), and electrophoresis was allowed to proceed to completion (200 V, 55 min) where clear bands of polypeptide separation were observed. We excised 4 2x10x1 mm^3 gel slices near the 58 kDa region and 1 4x10x1 mm^3 gel slice where we expect to observe the EtnC polypeptide and EtnE polypeptide, respectively, (17) and subjected these slices to in-gel tryptic digestion. Tools used for gel excisions (i.e. scalpels and tweezers) were decontaminated with ethanol, Windex, and deionized water between samples. For each gel analyzed, a blank section of gel containing no protein was analyzed as a control, and spectra found in this control were excluded from our protein identification analyses.

After excision, gel slices were washed with 50% 100 mM NH_4HCO_3/50% acetonitrile (ACN) and then with 50% 10 mM NH_4HCO_3/50% ACN to remove Coomassie stain. Gel pieces were dried by centrifugal evaporation and treated with 20 mM DTT (1 hour, 56°C) followed by treatment with a 55 mM iodoacetamide solution (1 hour, 25°C, dark conditions) to prevent formation of disulfide bonds. Gel slices were then washed with two cycles of first 25 mM NH_4HCO_3 followed by 50% 25 mM NH_4HCO_3 /50% ACN and then dried to completeness in a vacuum centrifuge (Savant SpeedVac Concentrator SPD111V-115, Thermo Scientific). Sequencing-grade trypsin (1:50 w/w) (Promega Corp., Madison, WI) in 25 mM NH_4HCO_3 was added to samples for overnight digestion at 37°C. The following day, the supernatant (containing peptides) was removed for MS analysis. To improve peptide extraction efficiency, digested gel

المنارة للاستشارات
www.manaraa.com

slices were sonicated, washed with 50% H_2O /50% ACN, and all supernatants combined. All samples were concentrated by vacuum centrifugation to approximately 5 µL.

As an alternative to SDS-PAGE separation, two protein extracts were digested with trypsin in solution, and the resulting peptides separated with strong cation exchange (SCX) chromatography. The soluble protein fraction was digested with trypsin overnight (1:50 w/w) at 37°C, treated with 20 mM DTT for 1 hour at 37°C, desalted and exchanged into 0.1% formic acid in HPLC-grade water using SepPak Lites C_{18} (Waters Corp., Milford, MA), and concentrated by centrifugal evaporation to approximately 500 uL. Peptides were filtered through an Ultrafree-MC centrifugal filter unit (Millipore Corp., Billerica, MA), separated using SCX on UltraMicrospin or Microspin columns (Nest Group, Inc., Southboro, MA), and eluted with a step gradient of ACN/ammonium formate buffers (*Buffer 1*: 25% ACN/75% 10 mM ammonium formate, pH 4.0; *Buffer B*: 25% ACN/75% 200 mM ammonium formate). The elution steps were 0%, 10%, 20%, 30%, 40%, 50%, 60%, 100% Buffer B in Buffer A. Peptide fractions (50-100 uL) were desalted with PepClean C-18 Spin Columns (Thermo Scientific) and concentrated by centrifugal evaporation to approximately 5-10 uL for mass spectrometry analysis.

All resulting peptides were then subjected to liquid chromatography coupled to electrospray-tandem mass spectrometry (LC-ES-MS/MS). Peptides were loaded onto a New Objective ProteoPepII column (5 µm, RP-C18 resin, New Objective, Woburn, MA) using an Eksigent nanoLC autosampler connected to a Thermo Finnigan LTQ XL mass spectrometer. All samples were analyzed using a 75-minute reverse phase gradient from 95% H_2O/5% ACN 0.1% formic acid (FA) to 30% H_2O/70% ACN/0.1% FA. The LTQ mass spectrometer was operated as follows: nanospray voltage (2.4 kV), heated capillary temperature (200°C), full scan m/z range (400-2000). Data dependent MS/MS spectra were collected with the following parameters: 8 MS/MS spectra for the most intense ions from the full scan (minimum signal required=3000.0, isolation width=3.0) with 35% collision energy for collision-induced dissociation. Dynamic exclusion of the same

المنارة للاستشارات
www.manaraa.com

abundant peptides was enabled with a repeat count of 1 and an exclusion duration of 1 minute or 3 minutes. LC columns were washed several times with the above described gradient or replaced between sample sets to avoid contamination between runs. MS/MS spectra were visually inspected between samples to further minimize the possibility of crossover contamination.

Analysis of Mass Spectra and Identification of Proteins

A protein sequence database containing all currently known complete and partial AkMO subunit sequences, EaCoMT sequences, and environmental sequences encoding partial EtnC and EtnE genes from the present study was constructed and named the "biomarker database." The "biomarker database" contains the following: complete EtnC and EtnE sequences from VC-assimilating *Nocardioides* sp. JS614, *Mycobacterium* sp. strains JS60 and JS623 (Genbank Accession No. FJ602754) (19, 60); partial EtnE sequences from VC-assimilating *Mycobacterium* sp. strains JS61, JS616, JS617, JS619 (19), ethene-assimilating *Mycobacterium* sp. strains JS622, JS624, and JS625 (20), VC-assimilating *Pseudomonas* sp. strain AJ, and *Ochrobactrum* sp. strain TD (24); complete EaCoMT (XecA) sequences from propene-assimilating *Gordonia rupertinctus* strain B-276 (Genbank Accession No. AAL28081) and *Xanthobacter* strain Py2 (103); partial EtnC sequences from ethene-assimilating *Mycobacterium* sp. strains NBB1, NBB2, NBB3, and NBB4 (18); 70 soluble di-iron monooxygenase alpha subunit sequences retrieved from environmental samples in a previous study (18); 7 sequences from cloned *etnC* PCR products; 8 sequences from cloned *etnE* PCR products. Additionally, a protein sequence database containing the 4,905 entries of all predicted proteins encoded in the VC-assimilating strain JS614 genome (Genbank Accession CP000508-CP000509), the only complete genome of a VC-assimilating bacterium, was used in our bioinformatic analyses.

Protein identifications were made by comparing MS/MS spectra to theoretical

www.manaraa.com

spectra generated from either the JS614 genome or the biomarker database. For SEQUEST searches (32), MS/MS spectra Thermo RAW files were searched using Thermo Bioworks v3.1.1 (Thermo Scientific) using the following parameters: enzyme: trypsin, up to 4 missed cleavages allowed, parent mass tolerance 3.0 AMU, and fragment mass tolerance 0.5 AMU. For gel-treated samples, peptide masses in SEQUEST searches included a fixed modification of carbamidomethylation. SEQUEST *.dta and *.out files were further analyzed by PeptideProphet (63) and ProteinProphet (79) included in the Trans-Proteomic Pipeline (TPP) software package v4.1.1 (Seattle Proteome Center, Seattle, WA) and results filtered and viewed in Labkey Server Computational Proteomics Analysis System (CPAS) (92). Protein identifications were considered significant if all of the following criteria were met: (1) at least two peptides with PeptideProbability scores equal to or greater than 0.90 were identified; (2) ProteinProphet probability was equal to or greater than 0.90; (3) SEQUEST XCorr values of 2 (+1), 2 (+2), 2 (+3); and (4) SEQUEST delta Cn values equal or greater to 0.08. For SDS-PAGE samples, protein identifications from single excised gel slices were scored independently in TPP, and for SCX-separated samples, peptide results for all SCX fractions within a sample were combined before scoring in TPP.

Results

Detection of EtnC and EtnE in ethene enrichments

PCR analysis of DNA extracted from 9 ethene-enriched VC-contaminated groundwater samples (prepared on 10/03/2007) was conducted to determine if *etnC* and *etnE* were present. Of these enrichments, 7 tested positive for etheneotrophs using the culture based technique described above (Table 4.1). A band of the expected size of the *etnE* PCR product (891bp) was detected in 8 of the samples, while a band of the expected size of the *etnC* PCR product (360 bp) was only detected in 5 samples.

المنارة للاستشارات
www.manaraa.com

We hypothesized that if the microbes harboring these genes were functional in these enrichments, then the expressed polypeptides, EtnC and EtnE, will be detectable using proteomic techniques. To test this hypothesis, proteins were extracted from a total of 14 ethene-enriched VC-contaminated groundwater samples (prepared on 07/01/2008) and subsequently analyzed by a proteomics workflow developed in our laboratory (Fig. 1). Of these enrichments, 12 tested positive for the presence of etheneotrophs using the culture-based technique described above (Table 4.2). Initial mass spectrometry analysis of protein extracts suggested that EtnC and EtnE were present in several of the enrichment samples (data not shown).

To confirm the identity and extend available *etnC* and *etnE* sequence information for proteomic investigations, PCR bands of the expected size were cloned and sequenced from selected 10/03/2007 enrichment samples. This resulted in 7 partial *etnC* and 5 partial *etnE* sequences from the Well RB-73 enrichment sample and 3 partial *etnE* sequences from the Well RB-63I enrichment sample. Analysis of the 7 partial (357 bp) *etnC* sequences from sample RB-73 with a PHYLIP DNA distance similarity table indicated that they are 98.3-99.7% identical to each other, and 91.4-92.0% identical to the top BLAST hit (soluble di-iron monooxygenase alpha subunit from uncultured clone L10 derived from an ethene enrichment of freshwater sediment; Genbank Acc. No. DQ264663). The sequences were also 91-92.3% identical to the *etnC* in ethene-assimilating *Mycobacterium flavescens* NBB1 (Genbank Acc. No. DQ264721) (18). Analysis of the 5 partial (865 bp) *etnE* sequences from sample RB-73 indicated that they are 99.0-99.8% identical to each other, and 94% identical to the partial *etnE* from VC-assimilating *Mycobacterium* strain JS61. Analysis of the 3 partial (873 bp) *etnE* sequences from sample RB-63I indicated that they are 99.8-99.9% identical to each other and 97.2% identical to *etnE* (873 bp) from VC-assimilating *Nocardioides* sp. strain JS614. These sequences were included in the "biomarker database" that was used to analyze peptide MS/MS spectra from protein extracts described above.

المنارة للاستشارات
www.manaraa.com

MS/MS spectra were searched (with SEQUEST) against the "biomarker database" for AkMO and EaCoMT protein sequences. This resulted in the identification of EtnC and EtnE polypeptides in 3 and 9 enrichment samples, respectively, that were homologous to deduced EtnC and EtnE sequences from 12 different cultivated strains of VC-, ethene-, and propene-assimilating bacteria (Tables 4.3 and 4.4). In addition, EtnE peptides identified in samples RB-58I, RB-60, and RB-73 (from 07/01/2008) matched with the inferred products of the partial EtnE gene sequences we retrieved from the 10/03/2007 RB-63I and RB-73 enrichments (Table 4.4).

Evaluation of EtnC and EtnE as biomarkers of etheneotrophic VC-oxidizing bacteria

We evaluated the uniqueness of the EtnC and EtnE peptides observed in the enrichment samples to determine the effectiveness of our proteomics approach for biomarker discovery and to provide insight into the diversity of the expressed polypeptides in the enrichments. Peptide uniqueness was evaluated by subjecting the EtnC and EtnE peptides we identified to a BLASTp search of the NCBI non-redundant protein sequence database. In the Well RB-46 enrichment sample, two EtnC peptides currently unique to the EtnC from strain JS614 (NANLAEPR and VYGALDSNVR) were identified (Table 4.3). Additionally, two EtnC peptides currently unique to the EtnC from strain JS60 (YDWGFDYARPDP and LEYEVKDPLK) were identified in the RB-60 and RB-73 enrichment samples (Table 4.3). The remaining EtnC peptides observed in the RB-60 and RB-73 samples were shared by two or more EtnC sequences (and not unique to the EtnC from a particular strain) present in the biomarker database (Table 4.3).

EtnE peptides currently unique to VC-assimilating strain JS614 (DFVAGIIDVK, AVLEFVPAER, and WYDGSGFATFPK) and propene-assimilating strain B-276 (LRALADGAAIVRKELAAK) were identified in groundwater ethene enrichments from

المنارة للاستشارات
www.manaraa.com

Wells RB-58I, RB-63D, RB-63I, and RB-64I (Table 4.4). Multiple EtnE peptides currently unique to strain JS614 were identified in the samples from Wells RB-63I and RB-64I suggesting that the JS614 EtnE protein or a protein with a similar sequence is present in the vicinity of these wells. In the enrichment sample from Well RB-73, we identified the peptide, LHALADGAAIVR, which is currently an expected peptide common to the EtnE sequences found in VC-assimilating strains JS60 and JS623. However, the peptide VLDTNITALNYEVGR was also observed in the RB-73 enrichment sample. A trypsin digest of the deduced EtnE gene product found in strains JS623 and JS60 would not yield this peptide. Instead, it is an expected tryptic peptide of the deduced EtnE gene product found in strains JS622 (an etheneotroph) and JS21 (a VC-assimilator). The relative uniqueness of these two peptides suggests that at least two distinct ethene-assimilating strains were present and active in the RB-73 ethene enrichment. Overall, peptide identifications from proteins extracted from enrichment samples indicate that multiple ethene-assimilating (and possibly VC-assimilating) strains were present and active in these enrichments.

Detection of polypeptides associated with the aerobic VC and ethene biodegradation pathways in ethene enrichments

The *Nocardioides* sp. strain JS614 genome, currently the only complete genome of a VC-assimilating microorganism, contains protein sequences of all VC and ethene biodegradation pathway enzymes employed by JS614. We hypothesized that by subjecting MS/MS spectra retrieved from ethene enrichments to a SEQUEST search against the JS614 genome we would find evidence of a functional VC/ethene biodegradation pathway in the enrichments. Polypeptides in 7 of the enrichments were matched to the products of the following JS614 genes: Noca_4807 (*etnC*), Noca_4810 (*etnE*), Noca_4812, Noca_4813, Noca_4814, Noca_4815, Noca_4820, Noca_4822, Noca_4827, Noca _4833, and Noca_4841 (Table 4.5). In JS614, these genes reside in a

www.manaraa.com

large cluster that also includes genes encoding EaCoMT (*etnE*) and AkMO (*etnABCD*). In addition to *etnC* and *etnE*, Noca_4814, Noca_4822, Noca_4827, Noca_4833, and Noca_4841were found to be expressed in response to VC, ethene, and epoxyethane in strain JS614 (17) and therefore further indicate the presence of a functional VC and/or ethene biodegradation pathway in the enrichment samples.

Discussion

Detecting the presence and functionality of aerobic, VC- and ethene-assimilating microorganisms in environmental samples with molecular biology and proteomic tools could be valuable for monitoring aerobic VC bioremediation in the field, but such tools remain under development. For example, in one study, *etnC* was undetectable in un-enriched (and unpolluted) environmental samples but became detectable in samples that were enriched with ethene (18), suggesting that *etnC* could be a useful biomarker of VC-oxidizing bacteria at contaminated sites. In another study, EtnE genes were amplified directly from streambed sediments and VC-fed sediment microcosms from a PCE-contaminated site undergoing reductive dechlorination to VC and ethene (1). These results suggest that the presence of *etnE* at a VC-contaminated site could also indicate the presence of VC-oxidizing bacteria. Here, PCR analysis of ethene enrichments for the presence of both *etnC* and *etnE* suggested that, with the exception of the RB-64I sample, the presence of the EtnE gene correlated well with the culture-based determination of etheneotroph presence or absence (Table 4.1). On the other hand, the presence of the EtnC gene was difficult to determine conclusively due to the presence of several bands in the PCRs that were not of the expected size (data not shown).

In this study, we have demonstrated that detecting biomarkers of VC-assimilating and ethenotrophic bacteria in environmental samples with proteomic methods is feasible. PCR amplification and sequencing of EtnC and EtnE genes from select enrichment samples further validate the proteomic methods developed here. It should be noted that

المنارة للاستشارات
www.manaraa.com

these proteomic methods were applied to ethene-enriched groundwater samples from a VC-contaminated site in an effort to minimize the potential effects of low protein abundance in un-enriched environmental samples. These methods will provide a foundation for future proteomic studies in un-enriched environmental samples (e.g. groundwater). In these future experiments, protein extraction from environmental samples should be optimized for the identification of catabolic protein biomarkers. Because VC/ethene-assimilating cell numbers will likely be much lower in un-enriched samples than in enriched samples, methods to concentrate microbial cells and/or proteins (e.g. by filtration) might be necessary.

Cell lysis and protein extraction methods (i.e. guanidine, sonication, and French press lysis) were successfully applied to the ethene-enriched groundwater samples and resulted in protein identifications of the targeted EtnC and EtnE biomarkers and other proteins that appear to be involved the strain JS614 VC biodegradation pathway. However, because the initial amount of protein available for extraction among the samples was unknown, it is not possible to directly compare the results of protein extraction techniques for enrichment samples.

Interestingly, 4 EtnC and 4 EtnE peptides were found to be unique to inferred *etnC* and *etnE* gene products sequenced from 3 different cultivated strains. Therefore, the proteomics methods developed and applied in this study has revealed that at least 3 distinct etheneotrophic strains are active in these enrichments. This significant finding further suggests that there is a diversity collection of active etheneotrophs present at this particular VC-contaminated site. However, it is important to note that that the EtnC and EtnE polypeptides we have identified are not likely to be identical to those expressed in cultivated strains. Because the amount of proteomic data required to identify a protein only covers a portion of the polypeptide (18-47% amino acid coverage for the EtnC and EtnE polypeptides identified in this study) the amino acid sequence variability in the unsampled regions of the polypeptide remains unknown. It should also be noted that we

www.manaraa.com

did not attempt to distinguish between VC- and ethene-assimilating bacteria in this study. This is mainly because ethene-assimilating bacteria can evolve into VC-assimilating bacteria in both pure cultures fed VC and in VC enrichments (61, 107). Even though genetic changes in EtnE have been observed in response to this adaptation (60), additional research is required to determine if there are protein biomarkers that distinguish between these two closely related bacterial groups.

In this study, a combination of two protein identification algorithms, SEQUEST (32) and TPP containing PeptideProphet (63) and ProteinProphet (79), were used to interpret tandem mass spectrometry data. For a valid protein identification, we required that peptides meet specific XCorr and deltaCn filter cutoffs in SEQUEST as well as specific peptide and protein probabilities in TPP. Additionally, we required that two peptides from the same protein meet both SEQUEST and TPP filtering cutoffs. If filtering parameters were less rigorous, more peptide identifications (and more protein identifications) would be valid within our samples. Also, more protein identification of sequences encoded by etnC and etnE clones which were too short to satisfy all currently applied constraints would be valid. Given the small size of the biomarker database, we chose to use a conservative threshold for protein identifications and supported biomarker protein identifications with indirect evidence of microbial growth within enrichments samples and PCR studies determining the presence of biomarker genes. In order to identify the maximum number of proteins within environmental samples, future work should be directed toward identifying the optimal threshold for protein identifications in environmental samples using the biomarker database.

A limitation of our targeted protein biomarker approach is the extent to which sequencing information is available in public databases for catabolic genes of interest, in this case, *etnC* and *etnE*. Protein identifications in ethene enrichment cultures were made using available protein sequences in the UniProt databases from cultivated VC-, ethene-, and propene-assimilating strains. In addition, by sequencing of PCR products amplified

www.manaraa.com

from enrichment samples we were able to add to the EtnC and EtnE sequence database. The data currently available remains useful for indicating microbial functionality within a sample, revealing the diversity of expressed polypeptide biomarkers, and inferring the phylogeny of the microorganisms that are participating in biodegradation. Continued sequencing of functional genes and/or metagenomic libraries from VC-contaminated environments will provide additional database support for making effective protein identification from environmental samples.

المنارة للاستشارات

www.manaraa.com

Table 4.1 Analysis of ethene-enriched VC-contaminated groundwater samples (prepared on 10/03/2007).

Sample	EtnC gene	EtnE gene	Etheneotrophs
RB-46	+	+	+
RB-47D	+	+	+
RB-58D	-	+	+
RB-58I	-	-	-
RB-63D	-	+	+
RB-63I	+	+	+
RB-64I	-	+	-
RB-73	+	+	+
RV-EE	+	+	+

Note: Presence “+” or absence “-“ of EtnC and EtnE genes was determined by PCR while the presence “+” or absence “-“ of ethenotrophs was determined from the culture-based technique described in section 2.1.

المنارة للاستشارات
www.manaraa.com

Table 4.2 Analysis of proteins extracted from ethene-enriched VC-contaminated groundwater samples (prepared on 07/01/2008).

Sample	Lysis method	Protein separation method	Total Protein Loaded (µg)	Total JS614 Protein IDs	EtnC Biomarker	EtnE Biomarker	Ethene-otrophs[b]
I-A	GU	SCX	8.6	0	-	-	+
RB-46	GU	SCX	29.5	17[a]	+	+	+
RB-47D	SN	Gel	0.5	0	-	-	-
RB-47I	SN	Gel	1.7	11[a]	-	+	+
RB-52I	SN	Gel	1.0	1	-	+	+
RB-58D	SN	Gel	1.3	0	-	-	+
RB-58I	SN	Gel	1.5	3	-	+	+
RB-60	SN	Gel	11.8	12[a]	+	+	+
RB-61	SN	Gel	0.9	0	-	-	-
RB-63D	FP	Gel	1.2	0	-	+	+
RB-63I	SN	Gel	2.0	5	-	+	+
RB-64I	SN	Gel	2.4	1	-	+	+
RB-73	FP	Gel	5.5	0	+	+	+
RV-EE	SN	Gel	1.3	0	-	-	+

المنارة للاستشارات
www.manaraa.com

Notes:

Proteins were extracted using French press (FP), sonication (SN), or guanidine (GU) lysis methods.

When gel separations were used, proteins were excised from regions of the gel where, based on a previous study (17) the EtnE (~40 kDa) and the EtnC (~57 kDa) were observed to migrate.

Unique protein and peptide identifications from LC-ES-MS/MS analysis of excised SDS-PAGE slices (gel) or from LC-ES-MS/MS of proteins extracted using SCX spin columns (SCX) using the "biomarker database"are shown.

Protein IDs were filtered with PeptideProphet probabilities of at least 0.90; ProteinProphet probabilities of at least 0.90; XCorr values of at least 2.0 (+1), 2.0 (+2), 2.0 (+3); deltaCn values of at least 0.08; and at least two unique peptide identifications.

[a]Results combined from duplicate analysis on LC-ES-MS/MS. Duplicates were performed because initial run on LC-ES-MS/MS was not optimized (i.e. HPLC leaks, ESI spray problems, etc.).

[b]Presence "+" or absence "-" of ethenotrophs was determined from the culture-based technique described in text.

المنارة للاستشارات
www.manaraa.com

Table 4.3 EtnC peptides observed in VC-contaminated groundwater enrichment samples that were used to identify EtnC polypeptides homologous to those found in cultivated strains and/or environmental clones.

Sample	Source Strain	Peptides
RB-46	JS614	***NANLAEPR***
		VYGALDSNVR
RB-60	JS623	GYAAMESEK
		LEVVHEK
		TLIGQPHLNAER
		VYGGLDSNVR
RB-60	JS60	TLIGQPHLNAER
		VYGGLDSNVR
		YDWGFDYARPDPK
RB-60	NBB1	FDGMDLADVILALGYVR
		LSPFGLKPPAR
RB-73	JS623	GYAAMESEK
		HTQIEVALR
		LSPFGLKPPAR
		MWTIDDIR
		QGYAGQMLDEVR
		TLIGQPHLNAER
		VYGGLDSNVR
RB-73	JS60	HTQIEVALR
		LEYEVKDPLK
		LSPFGLKPPAR

المنارة للاستشارات
www.manaraa.com

Table 4.3 (continued)

		MWTIDDIR
		QGYAGQMLDEVR
		TLIGQPHLNAER
		VYGGLDSNVR
		YDWGFDYARPDPK
RB-73	JS614	MWTIDDIR
		QGYAGQMLDEVR

Notes:

EtnC matches were filtered according to the following criteria: PeptideProphet probability of at least 0.90; ProteinProphet probability of at least 0.90; XCorr value of at least 2.0 (+1), 2.0 (+2), 2.0 (+3); delta Cn of at least 0.08; and at least two unique peptide identifications.

Peptides in bold italics are unique to the deduced EtnC sequence from the indicated source organism.

www.manaraa.com

Table 4.4. EtnE peptides observed in VC-contaminated groundwater enrichment samples that were used to identify EtnE polypeptides homologous to those expected from cultivated strains and/or environmental clones.

Sample	Source Strain or Clone Sequence	Peptides
RB-46	JS623, JS614, JS60	ALADGAAIVR GEFVYDAISR
RB-47I	JS614	MISAGKLRALADGAAIVR STVTETADEVADR
RB-52I	B-276, JS614	LGLSTDCGLINLPR STVTETADEVADR
RB-58I	JS614	AFNEDFK ***AVLEFVPAER*** DFVAGIIDVK ELGMAIAK GEFVYDAISR LGLSTDCGLINLPR STVTETADEVADR
RB-58I	RB-63I-E03	AFNEDFK DFVAGIIDVK ELGMAIAK TGPDDLKPLVDNNWEK
RB-58I	B-276	AFNEDFK ALADGAAIVR LGLSTDCGLINLPR STVTETADEVADR
RB-58I	AJ	ELGMAIAK GEFVYDAISR
RB-60	RB-73-E08, RB-73-E09, RB-73-E10, JS622, JS61	GEFVYDAISR VLDTNITALNYEVGR
RB-63D	B-276	***LRALADGAAIVRKELAAK*** STVTETADEVADR
RB-63D	TD, JS60, JS621	GEFVYDAISR KATNKPVK
RB-63I	B-276	AFNEDFK ALADGAAIVR
RB-63I	JS614	AFNEDFK ALADGAAIVR ***AVLEFVPAER*** ***DFVAGIIDV*** LGLSTDCGLINLPR STVTETADEVADR ***WYDGSGFATFPK***

المنارة للاستشارات
www.manaraa.com

Table 4.4 (continued)

RB-64I	JS614	ALADGAAIVR ***AVLEFVPAER*** ***DFVAGIIDV*** GEFVYDAISR STVTETADEVADR
RB-73	JS623	AAAIFPK GEFVYDAISR LHALADGAAIVR
RB-73	RB-73-E08, RB-73-E09, JS622, JS61	AAAIFPK GEFVYDAISR VLDTNITALNYEVGR
RB-73	RB-73-E10, TD, AJ, JS625, JS621, JS619	AAAIFPK GEFVYDAISR
RB-73	RB-73-E05, RB-73-E06	AAAIFPK VLDTNITALNYEVGR
RB-73	JS60	GEFVYDAISR LHALADGAAIVR

Notes:

EtnE matches were filtered according to the following criteria: PeptideProphet probability of at least 0.90; ProteinProphet probability of at least 0.90; XCorr value of at least 2.0 (+1), 2.0 (+2), 2.0 (+3); delta Cn of at least 0.08; and at least two unique peptide identifications.

Peptides in bold italics are unique to the deduced EtnE sequence from the indicated source organism.

المنارة للاستشارات
www.manaraa.com

Table 4.5 Proteins identified in enrichments that are associated with the VC/ethene biodegradation pathway in Nocardioides sp. strain JS614.

Sample	Uniprot Accession Number	Protein Function Annotation	JS614 Gene Number
RB-46, RB-60	Q5U9K9	Probable alkene monooxygenase alpha subunit, EtnC	Noca_4807
RB-46, RB-52I, RB-58I, RB-63I, RB-64I	Q5U9J8	Epoxyalkane: coenzyme M transferase, EtnE	Noca_4810
RB-46	A1SC62, A1SC91	Coenzyme A transferase	Noca_4813
RB-46	Q5U9K1	Putative short-chain dehydrogenase	Noca_4814
RB-60, RB-63I	A1SC71	Succinate-semialdehyde dehydrogenase (NAD(P)(+))	Noca_4822
RB-46	A1SC76	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Noca_4827
RB-46, RB-60	A1SC82	Succinate-semialdehyde dehydrogenase (NAD(P)(+))	Noca_4833

Notes:

Spectra were searched against the proteins encoded by the JS614 genome.

A successful protein identification required a PeptideProphet probability of 090; ProteinProphet probability of at least 0.90; XCorr values of at least 2.0 (+1), 2.0 (+2), 2.0 (+3); dCn values of at least 0.08; and at least two unique peptides.

Protein IDs are UniProt accession numbers (www.uniprot.org).

Gene numbers were generated during machine annotation of the JS614 genome sequence (genome.ornl.gov/microbial/noca).

المنارة للاستشارات

www.manaraa.com

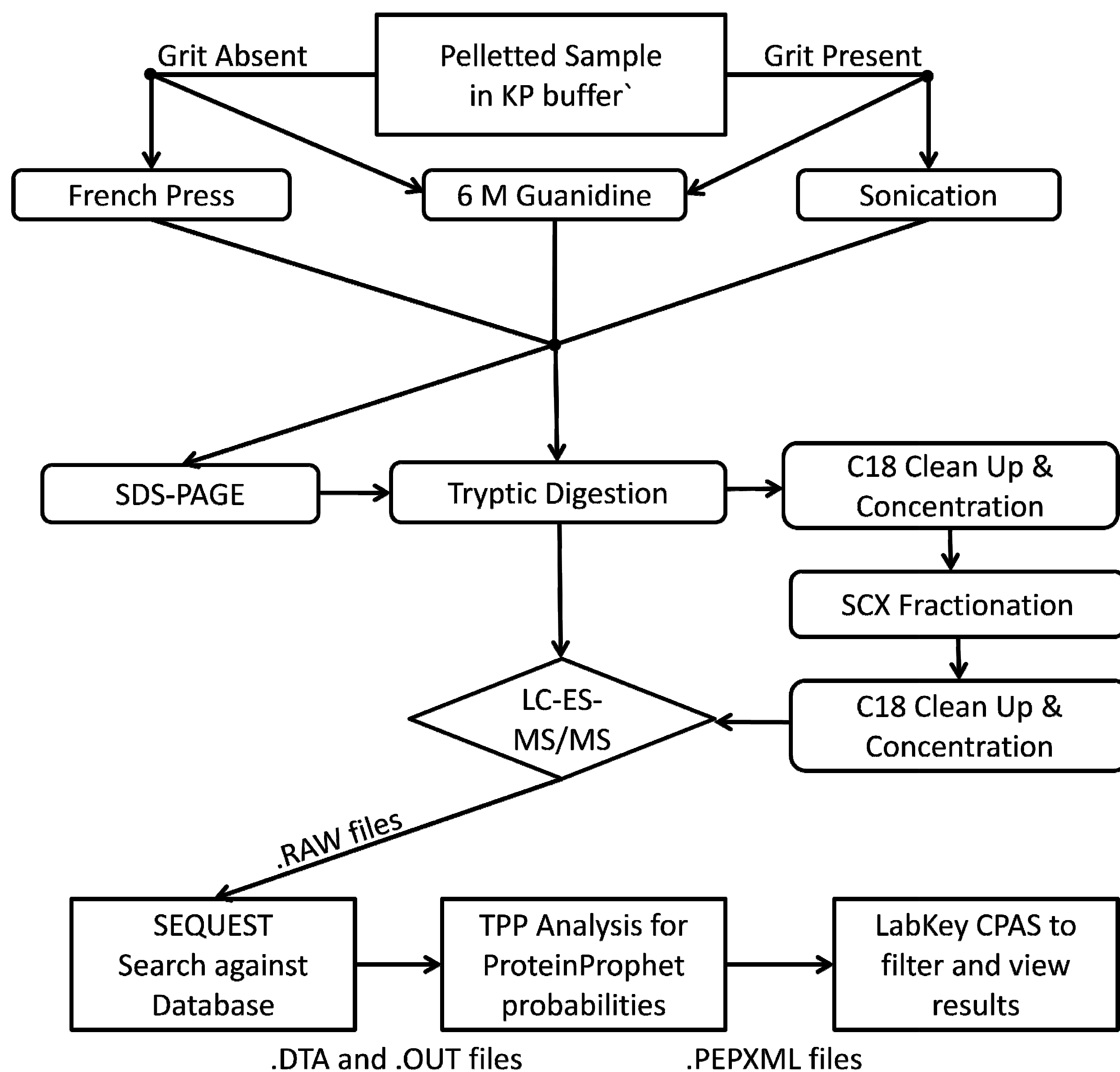


Figure 4.1 Workflow for protein identification in pure culture and enrichment samples.

المنارة للاستشارات
www.manaraa.com

CHAPTER V: PROTEOMIC INVESTIGATION OF *MYCOBACTERIUM* SP. STRAIN JS623

Abstract

Currently, very little is known about bacterial vinyl chloride (VC)- and ethene-assimilation pathways. Tandem mass spectrometry and spectral counting were applied to the study of three variants of *Mycobacterium* strain JS623, one wild type and two adapted for growth on vinyl chloride (VC). Overall, 174 total proteins for all three JS623 variants were identified. Among these, 85 proteins were found to be differentially expressed ($p<0.05$) within one strain. The methods introduced in this study rapidly and efficiently compared protein expressions between three bacterial variants and identified several interesting proteins which can be studied in the future for their involvement in unique adaptation to growth on VC.

Introduction

Vinyl Chloride-assimilating Bacteria

Vinyl chloride (VC) is a common groundwater pollutant (101) and known human carcinogen (64) that is commonly produced from the incomplete reductive dechlorination of tetrachloroethene and trichloroethene, chlorinated solvents often used in plastics and dry cleaning solvent manufacturing (101). The treatment of VC-contaminated sites by bacteria that can biodegrade VC has been demonstrated to be a practical and potentially cost-effective alternative to traditional "pump and treat" site cleanup options (12, 67, 71). Several such bacteria have been isolated from chlorinated-ethene-contaminated sites by enrichment on VC as a carbon and energy source. These so-called VC-assimilators include aerobic bacterial strains of *Mycobacterium* (19, 47, 48), *Nocardioides* (19), *Pseudomonas* (23, 106, 107), *Ochrobactrum (23)*, and *Ralstonia* (31, 95). Recent research suggests that VC-assimilating bacteria, all of which can also assimilate ethene, are closely related to ethene-assimilators that only cometabolize VC (i.e. "etheneotrophs") (19-21, 77) and can even evolve spontaneously from the

المنارة للاستشارات
www.manaraa.com

latter (61, 107). Much of this research has focused on providing a better understanding of the biodegradation pathways involved in VC- and ethene-assimilating bacteria and the relationships between VC-assimilating microbial communities with the goal of improving their effectiveness in bioremediation applications. Specifically, this information could lead to the development of specialized, improved strains of bacteria (i.e. extended substrate range or increased specific activity) that would be better suited for clean-up of contaminated sites (10, 41, 57, 78, 90, 102, 104).

VC and Ethene Biodegradation Pathways

Studies of VC-assimilating *Mycobacterium* strain JS60 (21) and *Nocardioides* sp. JS614 (77) indicate that there are two large gene clusters involved in the metabolism of VC and ethene, one encoding for enzymes that metabolize alkenes and epoxides and the other that appears to encode biosynthesis of coenzyme M (CoM). Within the first cluster are genes encoding an alkene monooxygenase (AkMO, EtnABCD) and an epoxyalkane:coenzyme M transferase (EaCoMT, EtnE). AkMO is a soluble, di-iron-containing monooxygenase that initiates the metabolism of ethene or VC and yields epoxyethane or chlorooxirane (19, 47). These highly reactive, toxic and mutagenic epoxide products (51) are subsequently metabolized by EtnE, a zinc-containing transferase (20, 21, 24, 77). AkMO and EaCoMT share homology to enzymes involved in propene assimilation. AkMO is most closely related to propene monooxygenase from *Gordonia* rubriperinctus B-276 (Amo/PmoABCD, 60% amino acid identity to alpha subunit of *Mycobacterium* strain JS623), while EtnE is related to the EaCoMTs found in propene-assimilating *Xanthobacter* Py2 and *Gordonia* rubriperinctus B-276 (45% and 72% amino acid identity with strain JS623, respectively) (60). In addition to sharing homology to propene-assimilating strains, ethene-assimilating strains are genetically similar to each other. For example, the EtnE of *Pseudomonas* putida AJ shares a 99% amino acid identity with VC-assimilating *Mycobacterium* rhodesiae strain JS60 (60).

www.manaraa.com

Beyond these enzymatic reactions and intermediates, little is known about the VC- or ethene-assimilation pathways. Hydroxyethyl-CoM (21) and acetyl-CoA (26) are the only known intermediates in the ethene pathway. Clues to downstream reactions in the VC- and ethene-assimilation pathways can be found in existing physiological and biochemical data (21, 24, 77), proteomics and RT-PCR studies (17, 21, 77), and by analogy to the propene catabolic genes and propene assimilation pathway in *Xanthobacter* Py2 (Genbank Acc. CP000781-CP000782) (34). Collectively, this information suggests that the ethene pathway intermediate hydroxyethyl-CoM is converted into 2-ketoethyl-CoM via a short chain dehydrogenase/reductase family alcohol dehydrogenase. In contrast, the proposed VC pathway intermediate 2-chloro-2-hydroxyethyl-CoM could spontaneously decompose into 2-ketoethyl-CoM. A bifunctional reductase/carboxylase could subsequently attack 2-ketoethyl-CoM to form malonate semialdehyde and reduced CoM. The pathway would be completed with the conversion of a malonate semialdehyde to malonate by a dehydrogenase and the transformation of malonate into acetyl-CoA via a CoA synthetase/CoA transferase cycle (49).

VC-adaptation in strain JS623

Recent experiments have demonstrated the adaptation of pure cultures of aerobic ethene-assimilating bacteria to growth on VC (61, 107) and present a unique opportunity to study VC- and ethene-assimilating pathways. A study of the ethene-assimilating wild-type *Mycobacterium* strain JS623 (JS623-W) and two mutant VC-assimilating JS623 strains (ethene-grown variant E (JS623-E) and trypticase soy agar glucose-grown variant T (JS623-T)) have shown that both genetic changes and physiological responses (i.e. enzyme expression changes) to varying growth substrates are associated with VC adaptation (60). For example, both JS623-E and JS623-T displayed significantly higher EaCoMT activities than JS623-W. However, analysis of EtnE expression levels in ethene-grown strain JS623 variants (JS623-W, JS623-E, and JS623-T) by comparing EtnE band intensity in SDS-PAGE could not explain the observed increase of EaCoMT activity in VC-adapted strains. Interestingly, these same SDS-PAGE protein profiles

www.manaraa.com

also revealed that protein expression in JS623-E is markedly different than JS623-W and JS623-T (Figure 5.1). MALDI-TOF MS analysis of several highly expressed bands in the JS623-E lane suggested that a putative methylmalonate semialdehyde dehydrogenase was up-regulated in JS623-E. This observation is consistent with the proposed downstream VC and ethene pathways described above.

Mass Spectrometry

Technical advances in high performance liquid chromatography (LC) systems connected to tandem mass spectrometers with high resolution and fast scanning rates have made it possible to identify hundreds to thousands of proteins in sample (2, 3, 28, 68, 69, 87, 97). Furthermore, recent studies have shown that peptide ion counts across mass spectral analysis of control samples is highly reproducible (9, 15, 16) and that there is a relationship between protein abundance and mass spectral sampling statistics such as *sequence coverage* (the total coverage of a protein sequence by the peptides identified) (37), *peptide count* (the total number of peptides identified from a protein) (42), and *spectral count* (the total number of MS/MS spectra taken on peptides for a protein) (70, 85) in LC-tandem mass spectrometry (MS/MS) proteomics. The advantages of using sampling statistics to estimate protein abundance, also called label-free comparative quantification, is that it does not require additional experimental procedures and can be performed across many samples simultaneously (70, 111, 113). Spectral counting, even more so than protein sequence coverage or peptide counts, has been shown to be highly reproducible and sensitive to protein abundance change (70, 113). The principle behind spectral counting is that the more abundant the peptide, the more likely it will be selected for MS/MS analysis. Thus, the number of spectra observed for peptides of a given protein correlates to the abundance of that protein within the sample. Spectral counts of proteins identified within one sample are normalized (by the total spectral count across all proteins) and statistical tests are then used to verify the significance of the hypothesized protein abundance change betweens samples (70, 113). Spectral counting has been shown to distinguish protein abundance changes of over 2-fold

المنارة للاستشارات
www.manaraa.com

between samples (70, 85) but may overestimate protein ratios when spectral counts are low (0-4 counts) (70, 113).

Experimental Section

Chemicals, Bacterial Strains, Media, and Culture Condition

Vinyl chloride (VC) (99.5%) was from Fluka and ethene (99%) was from Airgas. All other chemicals were either reagent grade or molecular biology grade.

Three strain JS623 variant cultures (Wild Type (JS623-W), Variant E (JS623-E), and Variant T (JS623-T)) were grown in a minimal salts medium (MSM) with ethene (0.25 mM) as previously described (60, 61). JS623-W cultures cannot use VC as a sole carbon and/or energy source (20, 61). VC-adapted strain JS623 variants were grown either on ethene (JS623-E) or one-tenth-strength trypticase soy agar + 1% glucose (JS623-T) prior to their adaptation to VC (61). Cells were harvested by centrifugation (7500xg, 5 min., 4°C) during the mid-exponential growth phase (OD_{600}=0.25-0.35) for protein extraction.

Protein Extraction

Cell extracts of JS623 variants (biological triplicates) were prepared as described previously (76, 77) with the following modifications. Harvested cells were washed once with dipotassium phosphate (KP) buffer (20 mM), suspended in 3 ml MGD buffer [50 mM MOPS, 10% glycerol, 1 mM DTT (pH 7.2)] at an OD_{600}~50, and lysed by a mini French pressure cell (124,000 kPa, three cycles) at 4 °C. Extracted proteins were quantified using the Coomassie Plus – The Better Bradford Assay (Pierce Protein Research Products, Rockford, IL) against BSA standards.

Mass Spectrometry Analysis

Protein extracts (100 µg) were adjusted to a pH of 8.0 with 1 M Tris, pH 8.0. Extracts were subsequently treated with 10 mM dithiotreitol (DTT) at 65°C for 10 minutes and 10 mM iodoacetamide (IAA) solution at 30°C for 30 minutes. Protein was then digested by sequencing

المنارة للاستشارات
www.manaraa.com

grade trypsin (Promega V5111) at a substrate:trypsin ratio of 50:1 at 37°C overnight. Digested peptides were desalted and exchanged into 0.1% formic acid in HPLC-grade water using SepPak Lites C_{18} (Waters Corp., Milford, MA), and dried to completion with centrifugal evaporation.

Digested peptides were sent for mass spectrometry analysis to the University of California-Davis Proteomics Core Lab. The Paradigm MG4 HPLC System (Michrom Bio Resources, Auburn, CA) coupled with a Thermo Finnigan LTQ ion trap mass spectrometer (Thermo, San Jose, CA) was used for separation and analysis. Twenty micrograms of each digested sample was loaded onto a trap column (Zorbax 300SB-C18, 5 µm, 0.3 mm × 5 mm: Agilent, Santa Clara, CA) and desalted. Peptides were then eluted from the trap, separated with a reverse-phase Michrom Magic C18AQ (200 um x 150 mm) capillary column at a flow rate of 2 µL per min. Peptides were eluted using a 240 min gradient of 2% B to 35% B over 160 minutes, 35% B to 80% B for 63 minutes, held at 80% B for 2 minute, 80% B to 5% B in 1 minute, and re-equilibrated for 14 minutes at 5%B (A = 0.1% Formic Acid, B= 100% Acetonitrile) and directly sprayed into the mass spectrometer. The mass spectrometer was operated with a spray voltage 1.8kV, a heated capillary temperature of 180°C, and a full scan range of 350-1400 m/z. Data dependent MS/MS spectra were collected with the following parameters: 10 MS/MS spectra for the most intense ions from the full scan (minimum signal required = 500.0, isolation width = 2.0) with 35% collision energy for collision-induced dissociation. Dynamic exclusion of the same abundant peptides was enabled with a repeat count of 2 and an exclusion duration of 1 minute.

All MS/MS samples were analyzed using X! Tandem (36) (www.thegpm.org; version TORNADO (2008.02.01.2)) . X! Tandem was configured, assuming the digestion enzyme trypsin, to search an abridged database containing 138,163 protein sequence entries. The database includes all predicted gene products from the VC-assimilating *Nocardioides* sp. strain JS614 genome (Genbank Acc. CP000508-CP000509). Additionally, the database was also populated with expected gene products resulting from keyword searches of *Mycobacterium*,

المنارة للاستشارات
www.manaraa.com

Pseudomonas putida, *Xanthobacter* Py2, *Ochrobactrum*, and *Rhodococcus* rhodochrous within the UniProt Consortium Protein Knowledgebase (www.uniprot.org, 4/19/2009). X! Tandem was searched with a fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 1.8 Da. Iodoacetamide derivative of cysteine was specified in X! Tandem as a fixed modification. Variable modifications specified were: Deamidation of asparagine, oxidation of methionine, sulphone of methionine, tryptophan oxidation to formylkynurenin of tryptophan and acetylation of the N-terminus.

Scaffold (version Scaffold_2_04_00, Proteome Software Inc., Portland, OR) (94) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (63). Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 5 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (79). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Semi-quantitative spectrum counts (70) calculated using the Scaffold software package (94) were average normalized between samples. ANOVA ($p<0.05$) statistical analysis was applied on a protein by protein basis to isolate significantly changing protein expressions.

Results and Discussion

Protein expression profiles of JS623-E, JS623-T, and JS623-W previously compared by SDS-PAGE were validated with MS/MS-based spectral counting techniques estimating protein expression in these variants. The same amount of digested proteins from ethene-grown JS623-E, JS623-T, and JS623-W were analyzed, and proteins identified with MS/MS. Subsequently, statistical analyses of spectral counts of identified proteins from triplicate studies of each variant were used to determine protein expression differences. Overall, we identified (to the best available homologous protein) between 46 to 173 proteins per sample in extracts from strain

المنارة للاستشارات
www.manaraa.com

JS623 variants (Table 5.1). The most abundant proteins identified in JS623-E, JS623-T, and JS623-W were most similar to expected gene products in other *Mycobacterium* and *Pseudomonas* strains (Table 5.2). In total, using MS/MS and our abridged database, we were able to compare 174 protein identifications (each protein had a minimum ProteinProbability of 95%, a minimum PeptideProbability of 95%, and at least 5 peptides identified).

Protein identifications for JS623-W, -T, and –E were compared and are presented in a Venn diagram to show unique and shared proteins among the variants (Figure 5.2). Note that if the constraints for protein identification (i.e. ProteinProbability, PeptideProbability, and number of peptides) were changed, the number of protein identifications for each strain would also change. Interestingly, we identified 96 proteins that were apparently unique to JS623-E whereas none were uniquely identified to JS623-T and JS623-W. Normalized spectral counts of proteins identified in more than one strain were compared (Figure 5.3). Quantitative scatter plots were also used to compare protein expressions between different strains. These graphs display each protein as a point on a two dimensional scatter plot where the horizontal axis depicts the normalized spectral count for a protein in one variant and the vertical axis is the normalized spectral count in another variant. Proteins with similar abundances in the two compared variants are represented as data points near the 45-degree line, and data points outside the dashed lines represent proteins with spectral counts that are at least two standard deviations away from the average spectral count of the compared strain. Quantitative scatter plots comparing proteins expressions between JS623-E and -T (Figure 5.4) and JS623-E and -W (Figure 5.5) reveal several proteins that were identified by MS/MS which are significantly ($p<0.05$) different in JS623-E compared to JS623-W and JS623-T. Overall, a total of 85 differentially expressed proteins ($p<0.05$) between at least two variants (E, T, and W) were noted. We found 72, 6, and 7 proteins were most highly expressed in 623-E, 623-T, and 623-W, respectively (Tables 5.3, 5.4, and 5.5). The putative methylmalonate semialdehyde dehydrogenase previously identified by MALDI-TOF MS analysis (60) was found in our MS/MS analysis to be most abundant in 623-E (Table 5.2) but failed ANOVA analysis of significant expression differences between 623-W and

المنارة للاستشارات
www.manaraa.com

JS623-E (p=0.065). Interestingly, many more proteins were identified as differentially expressed in 623-E than 623-W and 623-T suggesting that the VC-degradation in 623-E may be unique among the variants studied.

Among the proteins identified in variant 623-E, we identified several proteins homologous to proteins encoded by genes located on a large phenylacetic acid (Paa) operon in *Pseudomonas* putida F1 (Genbank Acc. No. CP000712). Proteins encoded by ten of the fifteen genes of this operon appeared to be upregulated in JS623-E (Table 5.6). It is unclear from our data whether these proteins are present at undetectable amounts in the other JS623 variants or unique to 623-E. Additionally, without a completed JS623 genome, it is impossible to rule out undetected contamination of 623-E strains by other bacterial strains. Regardless, the large number of Paa degradation proteins found in our 623-E strains is interesting and deserves further investigation. A study of PaaI involvement in aerobic phenylacetate degradation suggests that it may release reduced coenzyme A (CoA) by acting as a thioesterase (100). This result suggests that this operon may be involved in CoA recycling in strain JS623.

Our results suggest that there are pathway differences between strain JS623 and strain JS614 validating previous results suggesting that strain JS614 is unique among VC-assimilators. Not only is strain JS614 the only *Nocardioides* documented to degrade VC, but it has also been characterized by its relatively high VC yield coefficient, high VC utilization rate, and unique VC starvation sensitivity (19, 76). Unlike strain JS623, the completed genome of strain JS614 contains no Paa genes. Furthermore, we found only two proteins homologous to strain JS614 in our study of three JS623 variants, a succinate semi-aldehyde dehydrogenase (Noca_4822) and short chain dehydrogenase (Noca_4814). The low number of shared homologous proteins combined with the lack of carboxylase-like proteins, like those involved in propene-oxidation in strain Py2 (33, 34) and ethene-oxidation in strain JS614 (17), suggest that strains JS614 and JS623 use different VC-assimilation pathways. It should be noted that a completed JS623 genome is necessary to further validate these observations. As JS614 is the only VC-

المنارة للاستشارات
www.manaraa.com

assimilating bacterium with a completed genome sequence, this result highlights the value of obtaining other VC-assimilating genomes.

The expression of EtnE in JS623-E, -T, and –W was estimated using MS/MS-based spectral counting and compared to previous SDS-PAGE results (60) (Figure 5.1). Band intensities from SDS-PAGE experiments estimated with 623-W as a reference found that EtnE expression in 623-E and 623-T were 0.6 ± 0.1 (n=3) and 1.6 ± 0.7 (n=3), respectively (59). MS/MS spectral counting estimates the expression of EtnE based on the number of spectra identified for peptides of EtnE. With the analyses of triplicate samples of JS623-E, -T, and –W, we estimated with 623-W as a reference that EtnE expression in JS623-T (2.7 ± 0.5) was significantly higher than that observed in JS623-E (0.5 ± 0.4). These results are consistent with SDS-PAGE analysis of EtnE band intensities that follow the pattern of JS623-E < JS623-W < JS623-T (Figure 5.1) (59).

Previously, it has been reported that EaCoMT activities in VC-adapted variants are higher than in JS623-W, with JS623-T and JS623-E having ten-fold and two-fold increases in activity, respectively (60). Previous SDS-PAGE results (59, 60) and the presented MS/MS-spectral counting analysis of protein extracts of JS623-E indicate that the higher EaCoMT activity observed in this strain is not due to the overexpression of the EtnE polypeptide. Instead, as previously suggested, increased EaCoMT activity in JS623-E may be the result of downstream reactions influencing the rate of CoM recycling or missense mutations in the EtnE gene (60). In JS623-T, previously observed increased EaCoMT activity (60) can, however, be attributed to the presence of more EtnE polypeptide (in addition to previously observed higher CoM levels (60)).

Conclusion

The results of this study preliminarily identified many proteins involved in ethene-growth of all three strain JS623 variants and effectively utilized mass spectrometry-based proteomic techniques to not only validate previously observed enzyme expression levels but also quickly compare arrays of protein expression in three bacterial variants. Previously, only six proteins

المنارة للاستشارات
www.manaraa.com

from VC-grown JS623 cultures have been identified using MALDI-ToF MS (59). The methods presented here demonstrate significantly more (>100) protein identifications in only one MS/MS analysis. Furthermore, as we are only identifying between 0.9 to 9.7% of the available spectra using our abridged database, there are potentially many more possible protein identifications if a complete JS623 genome were available. Even with limited database support, we were still able to identify many proteins expressed from VC-adapted JS623-E that are significantly different than those of JS623-T and JS623-W. We propose that these proteins expressed in JS623-E, which are different than those identified in JS623-T, participate in assimilation of VC and ethene pathway metabolites into central metabolism. The results presented in this study thus rapidly provide a list of potential enzymes for further study that may be involved in unique adaptations to growth on VC that were not previously identified by studying isolated variants.

It should be noted, however, that this proteomic approach to studying strain JS623 provide insights into but do not give conclusive information about the pathway. The lack of a completed JS623 genome sequence limits protein identifications to the best available homologous protein in our database and makes it challenging to use proteomic evidence to ensure that contamination of strains is not occurring. However, despite these weaknesses, the methods described have rapidly given direction to future research that may be performed on the biological mechanisms involved in the mechanisms involved with VC-adaptation by highlighting differences between the JS623 variant E strain and variant T and wild-type strains.

www.manaraa.com

Table 5.1 Total number of proteins, peptides, and spectra identified for each biological triplicate of JS623 W, E, and T sample.

Category	Replicate	Protein IDs	Peptide IDs	Spectra	% Spectra Identified
Wild Type	W1	63	1362	71682	1.90%
Wild Type	W2	56	501	55760	0.90%
Wild Type	W3	54	645	15115	4.27%
Variant T	T1	77	1202	69802	1.72%
Variant T	T2	63	1239	72000	1.72%
Variant T	T3	46	620	14768	4.20%
Variant E	E1	173	2727	76232	3.58%
Variant E	E2	170	2616	69621	3.76%
Variant E	E3	121	1453	15042	9.66%

المنارة للاستشارات

www.manaraa.com

Table 5.2 Top three most abundant proteins identified for JS623 variants W, E, and T.

Strain	Protein	UniProt Acc. No.	Source Organism	Normalized Spectral Counts
Wild Type	Putative alkene monooxygenase alpha subunit	B9VTP1_9MYCO	*Mycobacterium* strain JS623	94 ±41
Wild Type	Putative alkene monooxygenase beta subunit	B9VTP1_9MYCO	*Mycobacterium* strain JS623	93 ±40
Wild Type	60 kDa chaperonin 1	CH601_MYCSJ	*Mycobacterium* strain JLS	81 ±37
Variant T	60 kDa chaperonin 1	CH601_MYCSJ	*Mycobacterium* strain JLS	102 ±16
Variant T	Putative alkene monooxygenase alpha subunit	B9VTP1_9MYCO	*Mycobacterium* strain JS623	105 ±23
Variant T	Epoxyalkane coenzyme M transferase	B9VTP6_9MYCO	*Mycobacterium* strain JS623	145 ±25
Variant E	Methylmalonate-semialdehyde dehydrogenase	A5VY45_PSEP1	*Pseudomonas* putida strain KT2440	157 ±125
Variant E	Ring-opening enzyme	Q88HT3_PSEPK	*Pseudomonas* putida strain KT2440	104 ±35
Variant E	Beta-ketoadiply CoA thiolase	Q88HS3_PSEPK	*Pseudomonas* putida strain KT2440	72 ±2

www.manaraa.com

Table 5.3 Proteins with statistically significant ($P<0.05$) differences in spectral count values between JS623 variants and most highly expressed in 623-E.

Protein ID	Source Organism	Acc. No.	Normalized Spectral Counts								
			Variant E			Variant T			Wild Type		
Ring-opening enzyme	*Pseudomonas* putida (strain KT2440)	Q88HT3_PSEPK	104.0	±	34.7	2.3	±	4.0	0.0	±	0.0
Beta-ketoadipyl CoA thiolase PhaD	*Pseudomonas* putida (strain KT2440)	Q88HS3_PSEPK	71.7	±	1.5	0.0	±	0.0	0.0	±	0.0
Elongation factor Tu-B	*Pseudomonas* putida (strain KT2440)	EFTU2_PSEPK	44.7	±	9.5	1.0	±	1.7	0.0	±	0.0
60 kDa chaperonin	*Pseudomonas* putida (strain F1 / ATCC 700007)	CH60_PSEP1	39.0	±	16.5	1.0	±	1.0	1.0	±	1.0
Phenylacetate-CoA oxygenase, PaaI subunit	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W3A8_PSEP1	33.0	±	17.7	0.3	±	0.6	0.0	±	0.0
Translation elongation factor 2 (EF-2/EF-G)	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5VXP4_PSEP1	32.0	±	4.6	0.0	±	0.0	0.0	±	0.0
DNA-directed RNA polymerase subunit beta	*Pseudomonas* putida (strain F1 / ATCC 700007)	RPOB_PSEP1	24.7	±	8.6	0.0	±	0.0	0.0	±	0.0
Chaperone protein dnaK	*Pseudomonas* putida (strain KT2440)	DNAK_PSEPK	24.3	±	17.0	0.0	±	0.0	0.0	±	0.0
Aconitase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W5Z7_PSEP1	23.0	±	1.0	0.0	±	0.0	0.0	±	0.0
OmpF family protein	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W6L5_PSEP1	23.0	±	6.6	0.0	±	0.0	0.0	±	0.0
Succinyl-CoA ligase [ADP-forming] subunit beta	*Pseudomonas* putida (strain F1 / ATCC 700007)	SUCC_PSEP1	22.3	±	13.6	0.0	±	0.0	0.0	±	0.0

المنارة للاستشارات
www.manaraa.com

Table 5.3 (continued)

3-hydroxyacyl-CoA dehydrogenase PaaC	*Pseudomonas* putida (strain KT2440)	Q88HS1_PSEPK	22.0	±	7.0	0.7	±	1.2	0.0	±	0.0
Phosphoenolpyruvate synthase	*Pseudomonas* putida (strain W619)	B1J594_PSEPW	19.7	±	13.4	0.0	±	0.0	0.0	±	0.0
Glyceraldehyde-3-phosphate dehydrogenase, type I	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W6F7_PSEP1	19.3	±	7.5	0.0	±	0.0	0.0	±	0.0
DNA-directed RNA polymerase subunit beta'	*Pseudomonas* putida (strain F1 / ATCC 700007)	RPOC_PSEP1	18.7	±	7.2	0.0	±	0.0	0.0	±	0.0
Phenylacetate degradation probable enoyl-CoA hydratase paaB	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W3A1_PSEP1	18.3	±	4.6	0.0	±	0.0	0.0	±	0.0
Aldehyde dehydrogenase family protein	*Pseudomonas* putida (strain KT2440)	Q88QE9_PSEPK	18.0	±	3.6	0.0	±	0.0	0.0	±	0.0
Phenylacetate-CoA oxygenase/reductase, PaaK subunit	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W3B0_PSEP1	17.7	±	3.5	0.0	±	0.0	0.0	±	0.0
Extracellular ligand-binding receptor	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W9Q5_PSEP1	16.7	±	3.8	0.3	±	0.6	0.0	±	0.0
Putative uncharacterized protein	*Pseudomonas* putida (strain KT2440)	Q88PT2_PSEPK	16.3	±	9.8	0.0	±	0.0	0.0	±	0.0
Branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein	*Pseudomonas* putida (strain KT2440)	Q88NR4_PSEPK	15.3	±	5.1	0.0	±	0.0	0.0	±	0.0
Polyribonucleotide nucleotidyltransferase	*Pseudomonas* putida (strain F1 / ATCC 700007)	PNP_PSEP1	15.0	±	4.6	0.0	±	0.0	0.0	±	0.0
Putative uncharacterized protein	*Pseudomonas* putida (strain KT2440)	Q88PQ1_PSEPK	14.7	±	2.1	0.0	±	0.0	0.0	±	0.0
Phenylacetic acid degradation protein PaaD	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W3A3_PSEP1	14.3	±	3.1	0.0	±	0.0	0.0	±	0.0

المنارة للاستشارات
www.manaraa.com

Table 5.3 (continued)

Isocitrate lyase	*Pseudomonas* putida (strain W619)	B1JBC5_PSEPW	14.3	±	9.9	0.0	±	0.0	0.0	±	0.0
Phenylacetate-CoA ligase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W3A5_PSEP1	13.7	±	3.2	0.0	±	0.0	0.0	±	0.0
Succinyl-CoA ligase [ADP-forming] subunit alpha	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W115_PSEP1	13.7	±	6.4	0.0	±	0.0	0.0	±	0.0
Putative uncharacterized protein	*Pseudomonas* putida (strain KT2440)	Q88LY2_PSEPK	13.3	±	4.7	0.0	±	0.0	0.0	±	0.0
Electron transfer flavoprotein, alpha/beta-subunit-like protein	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W0Z7_PSEP1	12.3	±	3.8	0.0	±	0.0	0.0	±	0.0
Electron transfer flavoprotein, alpha subunit-like protein	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W0Z8_PSEP1	12.0	±	2.0	0.7	±	1.2	0.0	±	0.0
Elongation factor Ts	*Pseudomonas* putida (strain F1 / ATCC 700007)	EFTS_PSEP1	10.7	±	1.2	0.0	±	0.0	0.0	±	0.0
Redoxin domain protein	*Pseudomonas* putida (strain GB-1)	B0KPM9_PSEPG	10.7	±	3.1	0.0	±	0.0	0.0	±	0.0
Aspartate semialdehyde dehydrogenase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W6Y4_PSEP1	10.7	±	6.0	0.0	±	0.0	0.0	±	0.0
Ring-oxidation complex protein 1	*Pseudomonas* putida (strain KT2440)	Q88HS5_PSEPK	10.3	±	4.0	0.0	±	0.0	0.0	±	0.0
OmpA/MotB domain protein	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W884_PSEP1	10.3	±	5.8	0.0	±	0.0	0.0	±	0.0
Enoyl-CoA hydratase/isomerase	*Pseudomonas* putida (strain KT2440)	Q88HR9_PSEPK	9.0	±	0.0	0.0	±	0.0	0.0	±	0.0
Peptidoglycan-associated lipoprotein	Pseudomonas putida (strain GB-1)	B0KTI6_PSEPG	9.0	±	1.0	0.0	±	0.0	0.0	±	0.0

المنارة للاستشارات
www.manaraa.com

Table 5.3 (continued)

UDP-glucose pyrophosphorylase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W1T2_PSEP1	8.7	±	1.2	0.7	±	1.2	0.0	±	0.0
Thioredoxin	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5WAT2_PSEP1	8.7	±	2.9	0.0	±	0.0	0.0	±	0.0
Cold-shock domain family protein	*Pseudomonas* putida (strain KT2440)	Q88P68_PSEPK	8.3	±	3.2	0.0	±	0.0	0.0	±	0.0
Dihydrolipoyl dehydrogenase	*Pseudomonas* putida (strain KT2440)	Q88FB1_PSEPK	8.3	±	3.2	0.0	±	0.0	0.0	±	0.0
Aspartyl-tRNA synthetase	*Pseudomonas* putida (strain F1 / ATCC 700007)	SYD_PSEP1	8.0	±	5.2	0.0	±	0.0	0.0	±	0.0
DNA-directed RNA polymerase, alpha subunit	*Pseudomonas* putida (strain GB-1)	B0KK91_PSEPG	7.7	±	2.5	0.3	±	0.6	0.0	±	0.0
D-lysine aminotransferase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W2G1_PSEP1	7.3	±	1.2	0.0	±	0.0	0.0	±	0.0
Ketol-acid reductoisomerase	*Pseudomonas* putida (strain F1 / ATCC 700007)	ILVC_PSEP1	7.3	±	1.5	0.0	±	0.0	0.0	±	0.0
Serine hydroxymethyltransferase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5VYA6_PSEP1	7.3	±	2.5	0.0	±	0.0	0.0	±	0.0
Branched chain amino acid aminotransferase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W2P3_PSEP1	7.3	±	3.1	0.0	±	0.0	0.0	±	0.0
Extracellular solute-binding protein, family 3	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5VX63_PSEP1	7.3	±	4.9	0.0	±	0.0	0.0	±	0.0
D-3-phosphoglycerate dehydrogenase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5WAM0_PSEP1	7.0	±	2.0	0.0	±	0.0	0.0	±	0.0
Cysteine synthase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W018_PSEP1	7.0	±	1.7	0.0	±	0.0	0.0	±	0.0
PpiC-type peptidyl-prolyl cis-trans isomerase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W3R9_PSEP1	7.0	±	1.7	0.0	±	0.0	0.0	±	0.0

المنارة للاستشارات
www.manaraa.com

Table 5.3 (continued)

Transketolase	*Pseudomonas* putida (strain KT2440)	A5W9Z8_PSEP1	7.0	±	2.0	0.0	±	0.0	0.0	±	0.0
Pyruvate kinase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W8M5_PSEP1	7.0	±	2.6	0.0	±	0.0	0.0	±	0.0
Phosphoenolpyruvate carboxykinase [ATP]	*Pseudomonas* putida (strain W619)	PCKA_PSEPW	7.0	±	4.4	0.0	±	0.0	0.0	±	0.0
Peptidyl-prolyl cis-trans isomerase	Pseudomonas putida (strain GB-1)	B0KUU2_PSEPG	6.7	±	0.6	0.0	±	0.0	0.0	±	0.0
Extracellular solute-binding protein, family 1	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5VXK6_PSEP1	6.7	±	0.6	0.0	±	0.0	0.0	±	0.0
NusA antitermination factor	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W988_PSEP1	6.7	±	1.2	0.0	±	0.0	0.0	±	0.0
Glycyl-tRNA synthetase beta subunit	*Pseudomonas* putida (strain KT2440)	SYGB_PSEPK	6.7	±	1.5	0.0	±	0.0	0.0	±	0.0
Extracellular solute-binding protein, family 3	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5VX06_PSEP1	6.3	±	3.5	0.0	±	0.0	0.0	±	0.0
Phosphoglycerate kinase	*Pseudomonas* putida (strain KT2440)	PGK_PSEPK	6.3	±	3.5	0.0	±	0.0	0.0	±	0.0
Citrate synthase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W106_PSEP1	6.0	±	1.0	0.0	±	0.0	0.0	±	0.0
Phosphoribosylformylglycinamidine cyclo-ligase	*Pseudomonas* putida (strain F1 / ATCC 700007)	PUR5_PSEP1	6.0	±	1.7	0.0	±	0.0	0.0	±	0.0
Succinate dehydrogenase subunit A	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W109_PSEP1	6.0	±	2.6	0.3	±	0.6	0.0	±	0.0
Sulfite reductase (Ferredoxin)	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W5P0_PSEP1	5.7	±	1.2	0.0	±	0.0	0.0	±	0.0
Fructose-bisphosphate aldolase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W9Z3_PSEP1	5.3	±	1.5	0.0	±	0.0	0.0	±	0.0

المنارة للاستشارات
www.manaraa.com

Table 5.3 (continued)

Glutamate dehydrogenase (NAD)	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W6M4_PSEP1	5.3 ± 6.1	0.0 ± 0.0	0.0 ± 0.0
Malate synthase G	*Pseudomonas* putida (strain KT2440)	MASZ_PSEPK	5.0 ± 0.0	0.3 ± 0.6	0.0 ± 0.0
Transaldolase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W6D8_PSEP1	5.0 ± 2.6	0.0 ± 0.0	0.0 ± 0.0
ABC transporter, ATP-binding protein	*Pseudomonas* putida (strain KT2440)	Q88Q24_PSEPK	4.7 ± 1.5	0.0 ± 0.0	0.0 ± 0.0
Thioredoxin reductase	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5VYL2_PSEP1	4.7 ± 1.5	0.0 ± 0.0	0.0 ± 0.0
Argininosuccinate synthase	*Pseudomonas* putida (strain F1 / ATCC 700007)	ASSY_PSEP1	4.3 ± 1.5	0.0 ± 0.0	0.0 ± 0.0
RNAse E	*Pseudomonas* putida (strain F1 / ATCC 700007)	A5W723_PSEP1	3.0 ± 1.7	0.0 ± 0.0	0.0 ± 0.0

المنارة للاستشارات
www.manaraa.com

Table 5.4 Proteins with statistically significant ($P<0.05$) differences in spectral count values between JS623 variants and most highly expressed in 623-T.

Protein ID	**Source Organism**		**Normalized Spectral Counts** Variant E	Variant T	Wild Type
Epoxyalkane coenzyme M transferase mutant 2	*Mycobacterium* sp. JS623	B9VTP6_9MYCO	28.0 ± 22.9	144.7 ± 25.4	53.0 ± 22.9
60 kDa chaperonin 1	*Mycobacterium* gilvum (strain PYR-GCK)	CH601_MYCGI	41.3 ± 30.7	103.7 ± 21.1	84.7 ± 19.4
60 kDa chaperonin 2	*Mycobacterium* smegmatis (strain ATCC 700084 / mc(2)155)	CH602_MYCS2	7.3 ± 5.5	22.0 ± 2.6	18.3 ± 0.6
Succinyl-CoA ligase [ADP-forming] subunit beta	*Mycobacterium* gilvum (strain PYR-GCK)	SUCC_MYCGI	5.3 ± 4.7	15.7 ± 4.2	9.3 ± 2.5
Isocitrate lyase	*Mycobacterium* avium (strain 104)	A0QLM2_MYCA1	1.7 ± 1.5	12.7 ± 2.9	5.3 ± 5.5
Adenosylhomocysteinase	*Mycobacterium* sp. (strain JLS)	SAHH_MYCSJ	0.3 ± 0.6	7.7 ± 2.3	6.0 ± 3.6

المنارة للاستشارات
www.manaraa.com

Table 5.5 Proteins with statistically significant ($P<0.05$) differences in spectral count values between JS623 variants and most highly expressed in 623-W.

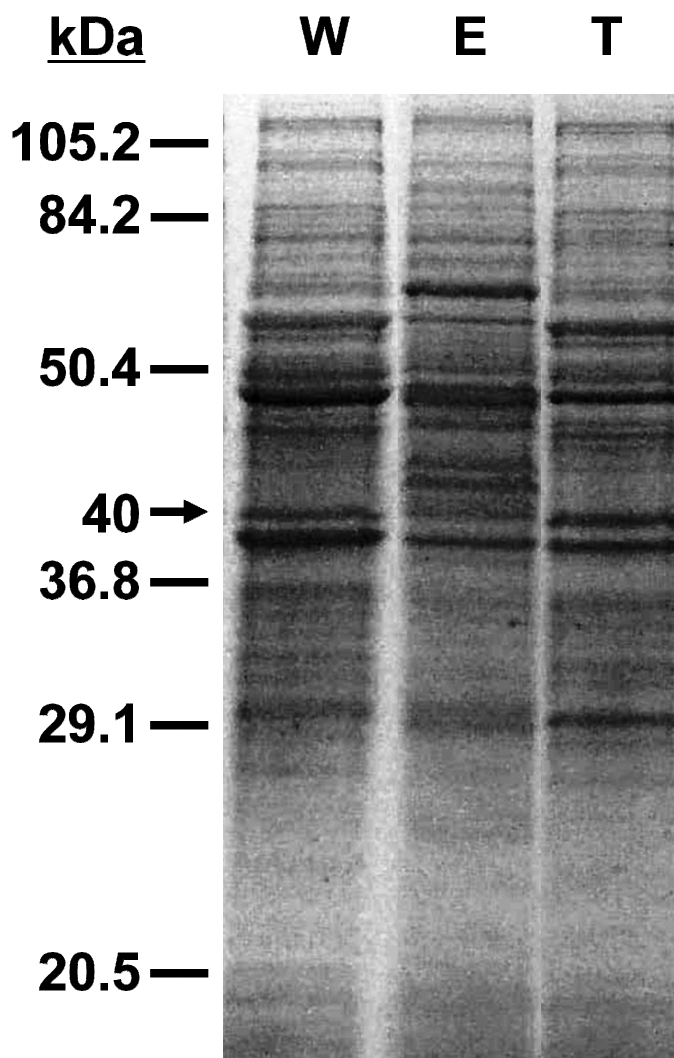
Protein ID	**Source Organism**		**Normalized Spectral Counts**		
			Variant E	**Variant T**	**Wild Type**
Succinate-semialdehyde dehydrogenase (NAD(P)(+))	*Nocardioides* sp. (strain BAA-499 / JS614)	A1SC71_NOCSJ	5.3 ± 2.9	15.7 ± 9.1	26.0 ± 8.9
Short-chain dehydrogenase/reductase SDR	*Mycobacterium* sp. (strain KMS)	A1U911_MYCSK	3.7 ± 3.2	7.3 ± 1.5	10.7 ± 2.5
Catalase-peroxidase	*Mycobacterium* sp. (strain JLS)	KATG_MYCSJ	1.7 ± 1.5	6.3 ± 2.1	9.3 ± 1.5
Transcription termination factor Rho	*Mycobacterium* gilvum (strain PYR-GCK)	A4T8L9_MYCGI	2.7 ± 2.5	7.3 ± 2.1	9.3 ± 1.2
DivIVA protein	*Mycobacterium* smegmatis (strain ATCC 700084 / mc(2)155)	A0R006_MYCS2	1.7 ± 1.5	6.3 ± 1.2	8.3 ± 3.8
MihF	*Mycobacterium* vanbaalenii (strain DSM 7251 / PYR-1)	A1T8H3_MYCVP	1.7 ± 1.5	3.0 ± 2.6	7.3 ± 2.5
Glyceraldehyde-3-phosphate dehydrogenase, type I	*Mycobacterium* smegmatis (strain ATCC 700084 / mc(2)155)	A0QWW2_MYCS2	0.7 ± 0.6	4.0 ± 1.0	6.0 ± 2.6

المنارة للاستشارات
www.manaraa.com

Table 5.6 Proteins identified in JS623-E similar to proteins encoded by genes within the phenylacetic acid operon in *Pseudomonas* putida F1.

Gene (F1 number)	**Possible function/annotation**	**JS623-E**
paaY	phenylacetic acid degradation protein	
paaA	enoyl-CoA hydratase-isomerase	
paaB	enoyl-CoA hydratase-isomerase	**X**
paaC	3-hydroxyacyl-CoA dehydrogenase	**X**
paaD	phenylacetic acid degradation protein	**X**
paaE	beta-ketoadipyl CoA thiolase	**X**
paaF	phenylacetate-CoA ligase	**X**
paaG	phenylacetate-CoA oxygenase	**X**
paaH	phenylacetate-CoA oxygenase	**X**
paaI	phenylacetate-CoA oxygenase	**X**
paaJ	phenylacetate-CoA oxygenase	
paaK	phenylacetate-CoA oxygenase/reductase	**X**
paaL	acetate permease	
paaM	outer membrane porin	
paaN	bifunctional aldehyde dehydrogenase/enoyl-CoA hydratase (also called ring opening enzyme)	**X**

المنارة للاستشارات
www.manaraa.com



Note: The EtnE polypeptide has previously identified from the 40 kDa band in a VC-grown JS623 culture (modified from (59)).

Figure 5.1. SDS-PAGE of cell extracts (30 µg protein/lane) prepared from ethene-grown JS623 wild-type (W), variant E, and variant T strains.

www.manaraa.com

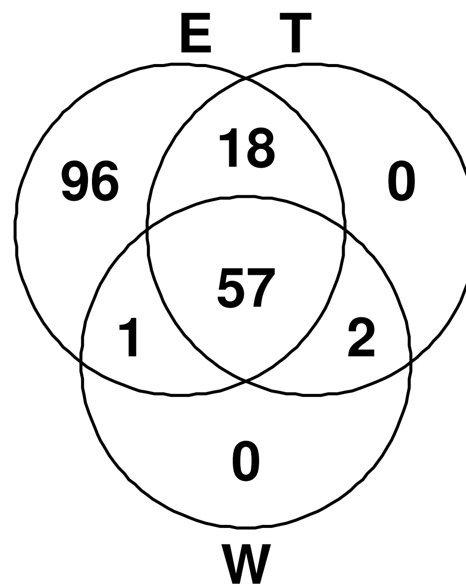


Figure 5.2 Venn diagrams showing unique protein identifications for JS623-W, -E, and –T with a minimum ProteinProbability of 95%, a minimum PeptideProbability of 95%, and at least 5 peptides identified.

المنارة للاستشارات

www.manaraa.com

Figure 5.3. Normalized spectral counts of proteins (identified with Uniprot ID) for 623-WT, 623-T, and 623-E.

المنارة للاستشارات

www.manaraa.com

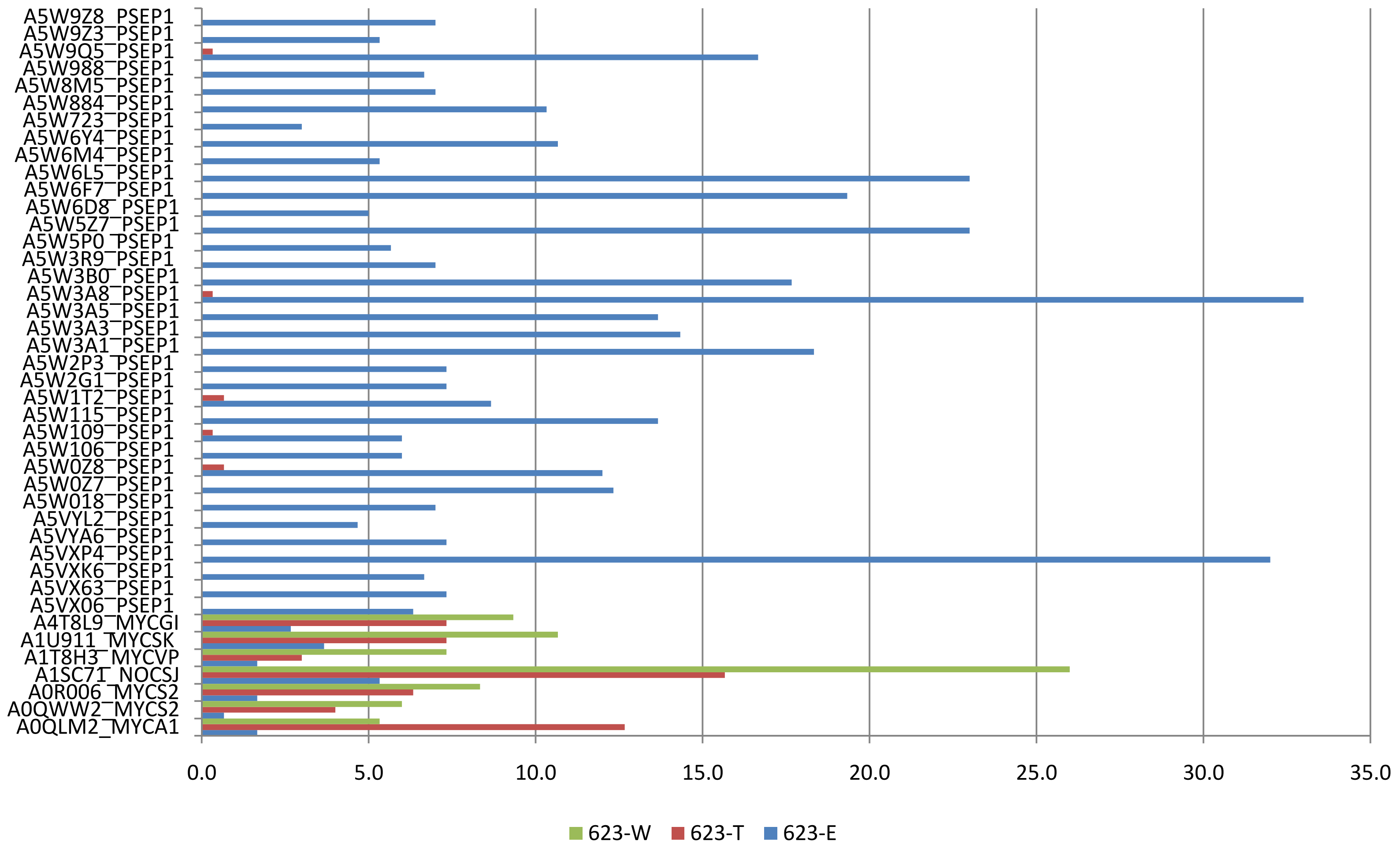


Figure 5.3 (continued)

المنارة للاستشارات
www.manaraa.com

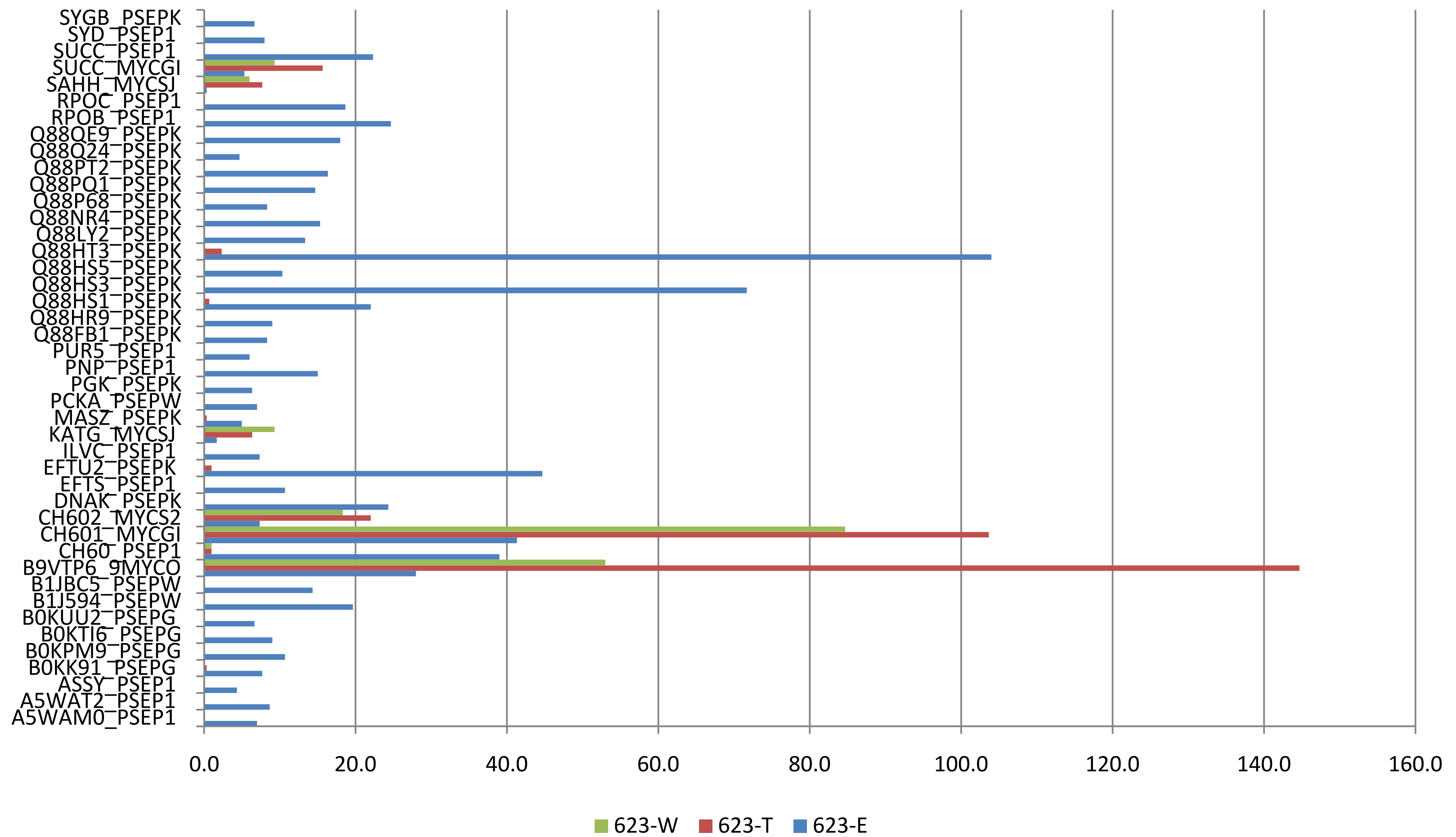
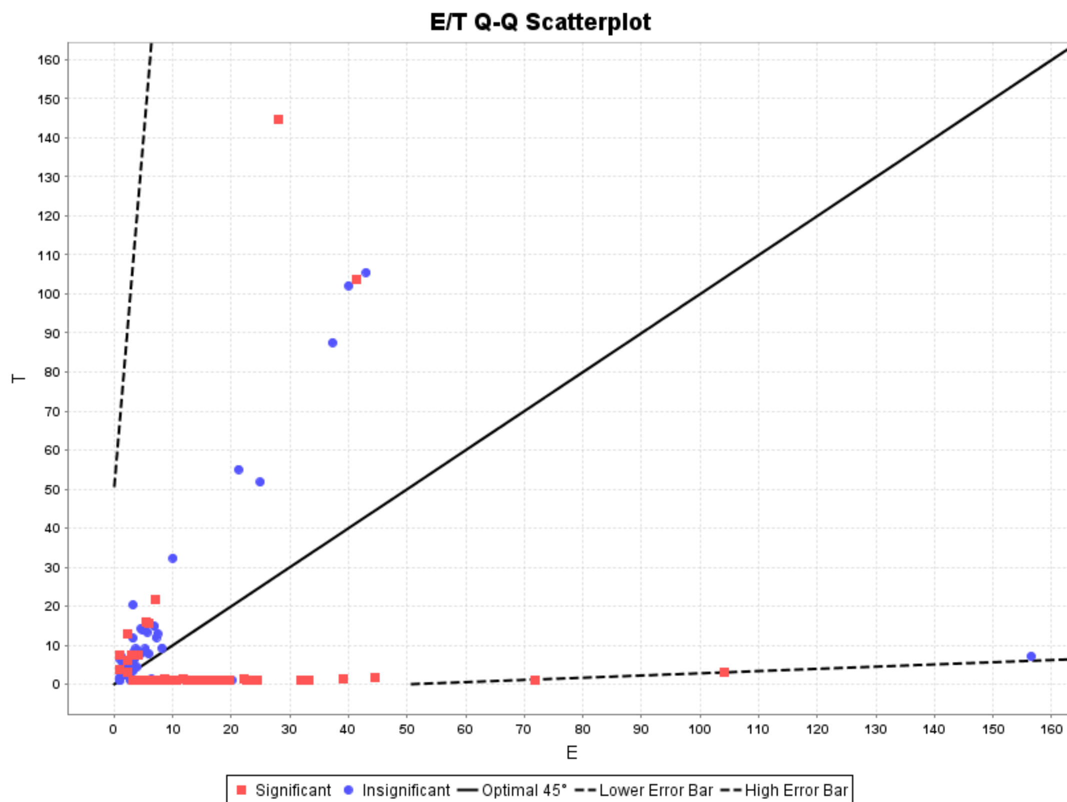


Figure 5.3 (continued)

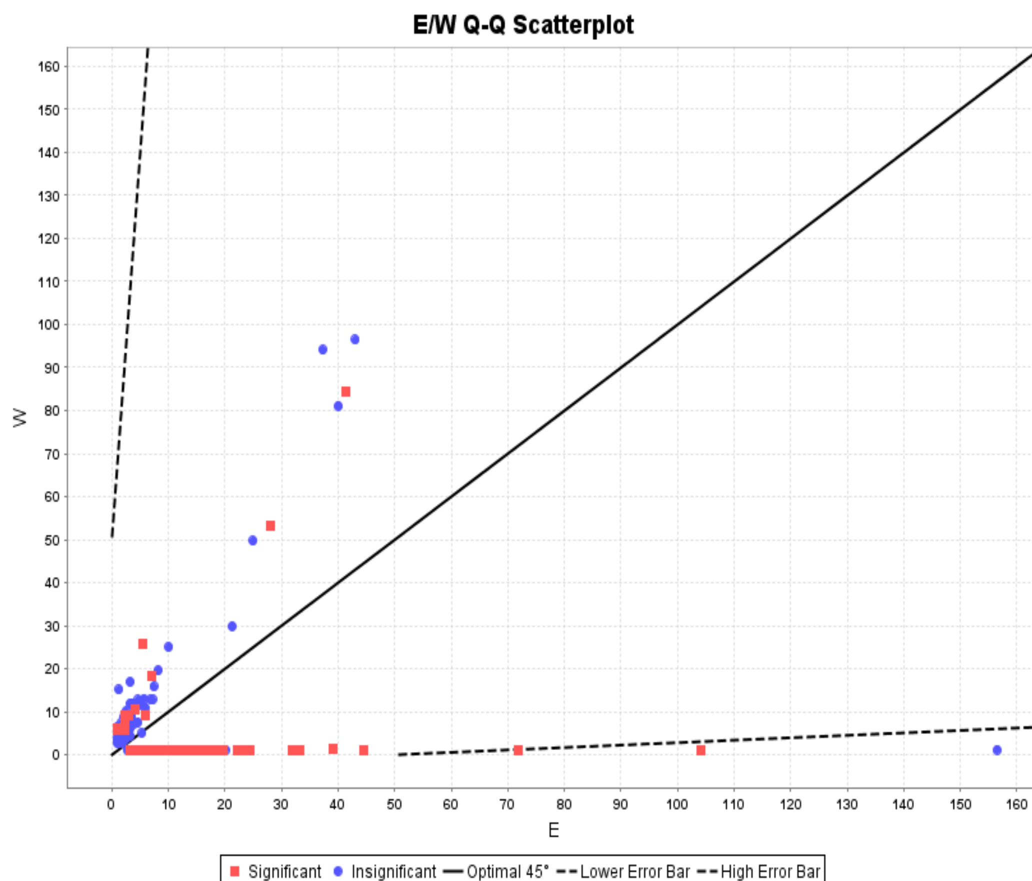
المنارة للاستشارات
www.manaraa.com



Notes: Graph axes represent normalized spectral counts for each variant. The optimal 45° line represents proteins with identical spectral counts in variants E and T. Proteins with similar abundances are shown near this line. Dotted lines represent spectral counts with more than two standard deviations difference. Red and blue points represent proteins that have significant ($p<0.05$) and insignificant differences ($p\geq0.95$), respectively, between JS623 variants E, T, and W.

Figure 5.4 Quantitative scatter plot compares proteins shared between JS623 variants E and T.

www.manaraa.com



Notes: Graph axes represent normalized spectral counts for each variant. The optimal 45° line represents proteins with identical spectral counts in variants E and T. Proteins with similar abundances are shown near this line. Dotted lines represent spectral counts with more than two standard deviations difference. Proteins plotted outside these lines are differentially expressed. Red and blue points represent proteins that have significant and insignificant differences, respectively, between JS623 variants E, T, and W.

Figure 5.5 Quantitative scatter plot compares proteins shared between JS623 variants E and W.

المنارة للاستشارات

www.manaraa.com

CHAPTER VI: ENGINEERING AND SCIENTIFIC SIGNIFICANCE

This work contributes to the understanding of microbial communities in both pure cultures and in the environment through the study of their proteins. The use of proteomics in this study has helped to provide insights into microbial VC-assimilation, specifically, the biochemical pathways of vinyl chloride (VC) and ethene oxidation (Chapter II); the diversity and ecology of protein biomarkers for ethene- and/or VC-assimilation in ethene groundwater enrichments (Chapter III and IV); and mechanisms involved in adaptation to growth on VC (Chapter V).

The functional profiling of the proteins of the VC-assimilating strain, *Nocardioides* sp. strain JS614, was used to revisit previous studies of the VC and ethene oxidation pathway. Several polypeptides involved in VC- and ethene-assimilation were rapidly identified using peptide mass fingerprinting and matrix-assisted laser desorption ionization time-of-flight (MALDI-ToF) mass spectrometry (MS). Among these, EtnC and EtnE were found to be highly expressed and consistently identified with MS, suggesting that they would be ideal candidates as biomarkers of VC-assimilation in the environment. In addition to identifying genes of functional importance, this study helped to provide amendments to the proposed metabolic processes involved in VC- and ethene oxidation (Figure 6.1), such as a previously unexpected downstream carboxylase involvement, which can be investigated in the future.

As a result of the success of MALDI-ToF MS to identify proteins in pure cultures of strain JS614, the study of its proteome was extended to liquid chromatography (LC) separation and electrospray ionization tandem mass spectrometry (ESI-MS/MS). Using LC-ESI-MS/MS, hundreds more proteins were identified, and similarly to MALDI-ToF MS, VC-assimilating biomarkers of interest, EtnC and EtnE, were consistently identified.

The advantage of protein functional indicators, such as EtnC and EtnE, is that they can directly monitor microbial ethene- and/or VC-assimilation in the environment

المنارة للاستشارات
www.manaraa.com

by revealing physiological responses to changes in the environment with explicit functional data. In this study, the ability to extract such proteins from a variety of environmental samples was examined. The separation of bacterial cells from environmental matrices prior to cell lysis was found to greatly improve protein extraction efficiency. In soils and sediments, chemical density gradients combined with centrifugation could be used for separating bacterial cells, but great care had to be taken to minimize the co-extraction of interfering compounds such as humic acids which affects downstream analyses such as MS. Separating bacterial cells from water samples, on the other hand, was much easier as it often only required filtration or centrifugation, and the presence of humic acids was minimal. In general, protein extraction methods will still require specific optimization for different environmental samples. Several options for protein extraction in soils, sediments, surface waters, groundwaters, and activated sludge are presented in this study and can be applied to future proteomic investigations of different environmental samples.

For this study, we specifically targeted polypeptides involved in VC-assimilation in the environment. Ethene-enriched groundwater samples from a VC-contaminated site were ideally suited for investigating the ability to detect protein biomarkers for VC-assimilating bacteria. The biomarkers EtnC and EtnE were identified in several groundwater enrichment samples and suggested the presence of a variety of ethene- and /or VC-assimilating bacteria. The methods described in this study can be further extended to proteins extracted *in situ* to characterize the dynamics and sustainability of microbial ethene- and/or VC-assimilation within the environment. The next logical step of this research would be to identify EtnC and EtnE directly from the groundwater samples at the VC-contaminated site. As the results of this study suggested that the presence of several strains of ethene- and/or VC-assimilating can be detected using proteomic techniques, future work could include using semi-quantitative MS/MS

www.manaraa.com

techniques, such as spectral counting (as described in Chapter V), to establish the relative contribution of diversity to VC-assimilation in contaminated samples.

The availability of MS/MS allows for the in-depth and routine characterization of protein mixtures as demonstrated by the study of the proteome strain JS614. Using LC-ESI-MS/MS, hundreds of proteins in three variants of *Mycobacterium* strain JS623 were also identified. By comparing three different JS623 strains, one wild type and two adapted for growth on VC, genes of functional importance to VC adaptation were identified. The functional profiling of these three strains revealed insights that were not apparent by examining the proteome of only one strain. Not only were proteins identified but their expression levels in the different VC-adapted strains were compared using spectral counting methods. Two markedly different protein profiles between the two mutant VC-adapted strains were observed and suggested that two different adaptation mechanisms were involved. Using the list of identified proteins and their expression levels, hypotheses about the biochemical pathways involved in VC-adaptation were made. Based on these proposed pathways, future investigations into the reactions involved in VC-adaption can be prioritized.

This study demonstrates the use of proteomics to understand pollutant-degrading microbial communities. From a biochemical level, proteomics can be used as a tool to identify elusive functional genes and their roles in metabolic pathways. At the ecological level, proteomic biomarkers can be developed to understand the presence and activities of microbial communities in the environment. With appropriate protein biomarkers, proteomics could be used to investigate the effects of biodiversity on ecosystem functioning as a whole and to track genes involved in other complex metabolic pathways (i.e. carbon and/or nitrogen cycle). It is a valuable tool to improve the understanding of the activities of catabolic enzymes involved in environmental bioremediation and to provide insight into the ecotoxicological effects of chemical pollutants and their cycling within the environment.

المنارة للاستشارات
www.manaraa.com

Ultimately, proteomics combined with gene sequence data, provides answers about the molecular basis of bacterial cells (or any cells) at any point in time or in any environment. This information can be used to direct and accelerate research efforts of molecular phenomena through determining the existence of proteins in complex mixtures. The identification of important proteins can be extended to further protein-focused studies including protein structure (protein purification), enzyme-substrate reactions (kinetic assays), enzyme activity (MS with standards, Western blotting), and protein-protein interactions (immunoaffinity chromatography).

The future of proteomics lies in the ability to reproducibly resolve all the peptides (proteins) within a mixture and to accurately identify or sequence these proteins. Advances in the sensitivity and versalitily of high-throughput MS/MS and data interpretation software continually increase the number of proteins within a sample which can be identified. Future improvements to proteomics will grow from first, an increase in the numbers of sequenced genomes that can be used to identify proteins, and second, the ability to automatically acquire and analyze an increasing number of spectra of peptides resolved from complex mixtures (similar to the large-scale sequencing of genomes). The potential of proteomics is thus directly fueled by the continual growth of the genomics field, and the application of proteomics to rapidly and reliability answer questions about the expression and activity of proteins under any condition ensure its position as an emergent and powerful field.

المنارة للاستشارات
www.manaraa.com

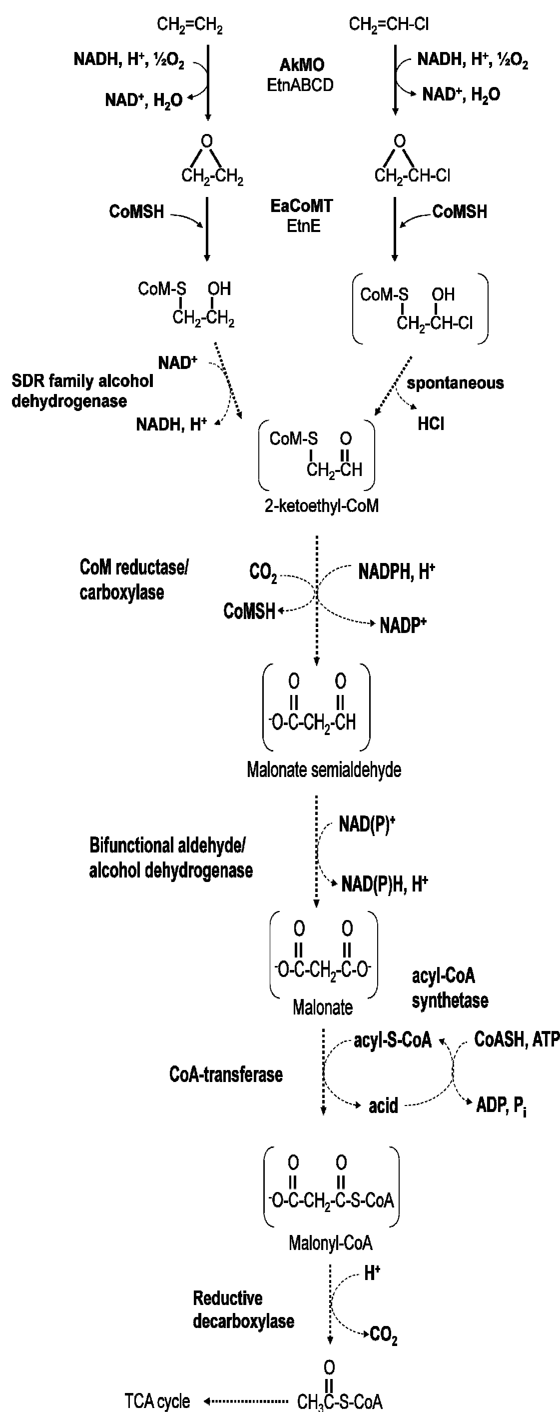


Figure 6.1 Hypothetical pathway of VC- and ethene-assimilation in Nocardioides sp. JS614

المنارة للاستشارات
www.manaraa.com

REFERENCES

1. **Abe, Y., R. Aravena, J. Zopfi, B. Parker, and D. Hunkeler.** 2009. Evaluating the fate of chlorinated ethenes in streambed sediments by combining stable isotope, geochemical and microbial methods. J. Contam. Hydrol. **107:**10-21.

2. **Aebersold, R., and D. R. Goodlett.** 2001. Mass spectrometry in proteomics. Chem. Rev. **101:**269-95.

3. **Aebersold, R., and M. Mann.** 2003. Mass spectrometry-based proteomics. Nature **422:**198.

4. **Allen, J. R., and S. A. Ensign.** 1997. Characterization of three protein components required for functional reconstitution of the epoxide carboxylase multienzyme complex from *Xanthobacter* strain Py2. J. Bacteriol. **179:**3110-5.

5. **Altschul, S., W. Gish, W. Miller, E. Myers, and D. Lipman.** 1990. Basic Local Alignment Search Tool. J. Mol. Biol. **215:**403-410.

6. **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:**224-34.

7. **Berry, A. E., C. Chiocchini, T. Selby, M. Sosio, and E. M. Wellington.** 2003. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiol. Lett. **223:**15-20.

8. **Bertrand, H., F. Poly, V. T. Van, N. Lombard, R. Nalin, T. M. Vogel, and P. Simonet.** 2005. High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. J. Microbiol. Methods **62:**1-11.

9. **Bondarenko, P. V., D. Chelius, and T. A. Shaler.** 2002. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. Anal. Chem. **74:**4741-9.

10. **Bosma, T., J. Damborsky, G. Stucki, and D. B. Janssen.** 2002. Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. Appl. Environ. Microbiol. **68:**3582-3587.

المنارة للاستشارات
www.manaraa.com

11. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248-54.

12. **Bradley, P. M., F. H. Chapelle, and J. T. Wilson.** 1998. Field and laboratory evidence for intrinsic biodegradation of vinyl chloride contamination in a Fe(III)-reducing aquifer. J. Contam. Hydrol. **31:**111-127.

13. **Bucher, J. R., G. Cooper, J. K. Haseman, C. W. Jameson, M. Longnecker, F. Kamel, R. Maronpot, H. B. Matthews, R. Melnick, R. Newbold, R. W. Tennant, C. Thompson, and M. Waalkes.** 2001. Ninth report on carcinogens. U.S. Dept. of Health and Human Services, National Toxicology Program. [http://ehis.niehs.nih.gov/roc/ninth/known/vinylchloride.pdf].

14. **Bucher, J. R., G. Cooper, J. K. Haseman, C. W. Jameson, M. Longnecker, F. Kamel, R. Maronpot, H. B. Matthews, R. Melnick, R. Newbold, R. W. Tennant, C. Thompson, and M. Waalkes.** 2005. Report on Carcinogens, Eleventh Edition. U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program.

15. **Chelius, D., and P. V. Bondarenko.** 2002. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J. Proteome Res. **1:**317-23.

16. **Chelius, D., T. Zhang, G. Wang, and R. F. Shen.** 2003. Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry. Anal. Chem. **75:**6658-65.

17. **Chuang, A. S., and T. E. Mattes.** 2007. Identification of polypeptides expressed in response to vinyl chloride, ethene, and epoxyethane in *Nocardioides* sp. strain JS614 by using peptide mass fingerprinting. Appl. Environ. Microbiol. **73:**4368-72.

18. **Coleman, N. V., N. B. Bui, and A. J. Holmes.** 2006. Soluble di-iron monooxygenase gene diversity in soils, sediments and ethene enrichments. Environ. Microbiol. **8:**1228-1239.

19. **Coleman, N. V., T. E. Mattes, J. M. Gossett, and J. C. Spain.** 2002. Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. Appl. Environ. Microbiol. **68:**6162-6171.

20. **Coleman, N. V., and J. C. Spain.** 2003. Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading *Mycobacterium* strains. Appl. Environ. Microbiol. **69:**6041-6046.

المنارة للاستشارات
www.manaraa.com

21. **Coleman, N. V., and J. C. Spain.** 2003. Epoxyalkane:Coenzyme M transferase in the ethene and vinyl chloride biodegradation pathways of *Mycobacterium* strain JS60. J. Bacteriol. **185:**5536-5545.

22. **Craig, R., and R. C. Beavis.** 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics **20:**1466-1467.

23. **Danko, A. S., M. Luo, C. E. Bagwell, R. L. Brigmon, and D. L. Freedman.** 2004. Involvement of linear plasmids in aerobic biodegradation of vinyl chloride. Appl. Environ. Microbiol. **70:**6092-6097.

24. **Danko, A. S., C. A. Saski, J. P. Tomkins, and D. L. Freedman.** 2006. Involvement of coenzyme M during aerobic biodegradation of vinyl chloride and ethene by *Pseudomonas putida* strain AJ and *Ochrobactrum* sp. strain TD. Appl. Environ. Microbiol. **72:**3756-3758.

25. **Davis, J. W., and C. L. Carpenter.** 1990. Aerobic degradation of vinyl chloride in groundwater samples. Appl. Environ. Microbiol. **56:**3878-3880.

26. **de Bont, J. A. M., and W. Harder.** 1978. Metabolism of ethylene by *Mycobacterium* E 20. FEMS Microbiol. Lett. **3:**89-93.

27. **de Bont, J. A. M., M. M. Atwood, S. B. Primrose, and W. Harder.** 1979. Epoxidation of short-chain alkenes in *Mycobacterium* E20: the involvement of a specific monooxygenase. FEMS Microbiol. Lett. **6:**183-188.

28. **Domon, B., and R. Aebersold.** 2006. Challenges and opportunities in proteomics data analysis. Mol. Cell. Proteomics **5:**1921-6.

29. **Edwards, E. A., and E. E. Cox.** 1997. Field and laboratory studies of sequential anaerobic-aerobic chlorinated solvent biodegradation, p. 261-265. *In* B. C. Alleman, and Leeman, A. (ed.), In Situ and On-Site Bioremediation, vol. 3. Battelle Press, Columbus, OH.

30. **Elango, V., A. Liggenstoffer, and B. Fathepure.** 2006. Biodegradation of vinyl chloride and *cis*-dichloroethene by a *Ralstonia* sp. strain TRW-1. Appl Microbiol Biotechnol **72:**1270.

31. **Elango, V. K., A. S. Liggenstoffer, and B. Z. Fathepure.** 2006. Biodegradation of vinyl chloride and *cis*-dichloroethene by a *Ralstonia* sp. strain TRW-1. Appl. Microbiol. Biotechnol. **72:**1270-5.

المنارة للاستشارات
www.manaraa.com

32. **Eng, J. K., A. L. McCormack, and J. R. Yates.** 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. **5:**976-989.

33. **Ensign, S. A.** 2001. Microbial metabolism of aliphatic alkenes. Biochemistry **40:**5845-53.

34. **Ensign, S. A., and J. R. Allen.** 2003. Aliphatic Epoxide Carboxylation. Annu. Rev. Biochem. **8:**8.

35. **Felsenstein, J.** 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.

36. **Fenyo, D., and R. C. Beavis.** 2003. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal. Chem. **75:**768-74.

37. **Florens, L., M. P. Washburn, J. D. Raine, R. M. Anthony, M. Grainger, J. D. Haynes, J. K. Moch, N. Muster, J. B. Sacci, D. L. Tabb, A. A. Witney, D. Wolters, Y. Wu, M. J. Gardner, A. A. Holder, R. E. Sinden, J. R. Yates, and D. J. Carucci.** 2002. A proteomic view of the *Plasmodium* falciparum life cycle. Nature **419:**520-6.

38. **Fogel, M. M., A. R. Taddeo, and S. Fogel.** 1986. Biodegradation of chlorinated ethenes by a methane-utilizing mixed culture. Appl. Environ. Microbiol. **51:**720-724.

39. **Fogel, S., J. F. Begley, and C. R. LeBlanc.** 2005. Presented at the 8th International Symposium on In Situ and On-site Bioremediation, Baltimore, MD.

40. **Freedman, D. L., and S. D. Herz.** 1996. Use of ethylene and ethane as primary substrates for aerobic cometabolism of vinyl chloride. Water Env. Res. **68:**320-328.

41. **Furukawa, K., H. Suenaga, and M. Goto.** 2004. Biphenyl dioxygenases: Functional versatilities and directed evolution. J. Bacteriol. **186:**5189-5196.

42. **Gao, J., G. J. Opiteck, M. S. Friedrichs, A. R. Dongre, and S. A. Hefta.** 2003. Changes in the protein expression of yeast as a function of carbon source. J. Proteome Res. **2:**643-9.

المنارة للاستشارات
www.manaraa.com

43. **Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch.** 2005. Protein Identification and Analysis Tools on the ExPASy Server. *In* J. M. Walker (ed.), The Proteomics Protocols Handbook.

44. **Gygi, S. P., Y. Rochon, B. R. Franza, and R. Aebersold.** 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. **19:**1720-30.

45. **Halden, R. U., D. R. Colquhoun, and E. S. Wisniewski.** 2005. Identification and phenotypic characterization of *Sphingomonas wittichii* strain RW1 by peptide mass fingerprinting using matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. **71:**2442-51.

46. **Harrigan, G. G., and R. Goodacre.** 2003. Metabolic profiling : its role in biomarker discovery and gene function analysis. Kluwer Academic, Boston, Mass.

47. **Hartmans, S., and J. A. M. de Bont.** 1992. Aerobic vinyl chloride metabolism in *Mycobacterium aurum* L1. Appl. Environ. Microbiol. **58:**1220-1226.

48. **Hartmans, S., J. A. M. de Bont, J. Tramper, and K. C. A. M. Luyben.** 1985. Bacterial degradation of vinyl chloride. Biotechnol. Lett **7:**383-388.

49. **Hayaishi, O.** 1955. Enzymatic decarboxylation of malonic acid. J Biol Chem **215:**125-36.

50. **Henriques, I. D., D. S. Aga, P. Mendes, S. K. O'Connor, and N. G. Love.** 2007. Metabolic footprinting: a new approach to identify physiological changes in complex microbial communities upon exposure to toxic chemicals. Environ. Sci. Technol. **41:**3945-51.

51. **Henschler, D.** 1994. Toxicity of Chlorinated Organic Compounds: Effects of the Introduction of Chlorine in Organic Molecules. Angew. Chem. Int. Ed. Engl. **33:**1920-1935.

52. **Henzel, W., T. Billeci, J. Stults, S. Wong, C. Grimley, and C. Watanabe.** 1993. Identifying Proteins from Two-Dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Databases. Proc. Natl. Acad. Sci. USA **90:**5011-5015.

53. **Hickey, W. J., and G. Sabat.** 2001. Integration of matrix-assisted laser desorption ionization-time of flight mass spectrometry and molecular cloning for the identification and functional characterization of mobile ortho-halobenzoate oxygenase genes in *Pseudomonas aeruginosa* strain JB2. Appl. Environ. Microbiol. **67:**5648-55.

المنارة للاستشارات
www.manaraa.com

54. **James, P.** 1997. Protein identification in the post-genome era: the rapid rise of proteomics. Q. Rev. Biophys. **30:**279-331.

55. **James, P., M. Quadroni, E. Carafoli, and G. Gonnet.** 1993. Protein identification by mass profile fingerprinting. Biochem. Biophys. Res. Commun. **195:**58-64.

56. **Janson, J.-C., and L. Rydén.** 1989. Protein purification : principles, high resolution methods, and applications, New York, N.Y.

57. **Janssen, D. B., I. J. T. Dinkla, G. J. Poelarends, and P. Terpstra.** 2005. Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. Environ. Microbiol. **7:**1868-1882.

58. **Jensen, O. N., A. V. Podtelejnikov, and M. Mann.** 1997. Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. Anal. Chem. **69:**4741-50.

59. **Jin, Y. O.** 2009. Evolutionary Mechanisms of Adaptation to Vinyl Chloride by Ethene-Assimilating *Mycobacterium* strains. Ph.D. Dissertation.

60. **Jin, Y. O., A. Chuang, S. Cheung, N. Coleman, and T. Mattes.** Submitted. Missense mutations in epoxyalkane coenzyme M transferase are associated with adaptation to vinyl chloride in *Mycobacterium* sp. strain JS623.

61. **Jin, Y. O., and T. E. Mattes.** 2008. Adaptation of aerobic, ethene-assimilating Mycobacterium strains to vinyl chloride as a growth substrate. Environ. Sci. Technol. **42:**4784-9.

62. **Kan, J., T. E. Hanson, J. M. Ginter, K. Wang, and F. Chen.** 2005. Metaproteomic analysis of Chesapeake Bay microbial communities. Saline Syst. **1:**7.

63. **Keller, A., A. I. Nesvizhskii, E. Kolker, and R. Aebersold.** 2002. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. Anal. Chem. **74:**5383-5392.

64. **Kielhorn, J., C. Melber, U. Wahnschaffe, A. Aitio, and I. Mangelsdorf.** 2000. Vinyl chloride: still a cause for concern. Environ. Health Perspect. **108:**579-88.

65. **Laemmli, U. K.** 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature **227:**680-685.

المنارة للاستشارات
www.manaraa.com

66. **Lane, D. J.** 1991. 16S/23S rRNA sequencing, p. 177-203. *In* E. Stackbrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley and Sons, Chichester, U.K.

67. **Lee, M. D., J. M. Odom, and R. J. J. Buchanan.** 1998. New perspectives on microbial dehalogenation of chlorinated solvents: insights from the field. Annu. Rev. Microbiol. **52:**423-452.

68. **Lin, D., D. L. Tabb, and J. R. Yates, 3rd.** 2003. Large-scale protein identification using mass spectrometry. Biochim. Biophys. Acta. **1646:**1-10.

69. **Link, A., P. Andrews, and J. LaBaer.** 2008. Proteomics: A Cold Spring Harbor Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

70. **Liu, H., R. G. Sadygov, and J. R. Yates, 3rd.** 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. **76:**4193-201.

71. **Lorah, M. M., and M. A. Voytek.** 2004. Degradation of 1,1,2,2-tetrachloroethane and accumulation of vinyl chloride in wetland sediment microcosms and in situ porewater: biogeochemical controls and associations with microbial communities. J. Contam. Hydrol. **70:**117.

72. **Mann, M., P. Højrup, and P. Roepstorff.** 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol. Mass Spectrom. **22:**338-345.

73. **Maron, P. A., C. Mougel, S. Siblot, H. Abbas, P. Lemanceau, and L. Ranjard.** 2007. Protein extraction and fingerprinting optimization of bacterial communities in natural environment. Microb. Ecol. **53:**426-34.

74. **Maron, P. A., L. Ranjard, C. Mougel, and P. Lemanceau.** 2007. Metaproteomics: a new approach for studying functional microbial ecology. Microb. Ecol. **53:**486-93.

75. **Martinez, A., K. Norstrom, K. Wang, and K. C. Hornbuckle.** 2009. Polychlorinated biphenyls in the surficial sediment of Indiana Harbor and Ship Canal, Lake Michigan. Environ Int.

76. **Mattes, T. E., N. V. Coleman, A. S. Chuang, A. J. Rogers, J. C. Spain, and J. M. Gossett.** 2007. Mechanism controlling the extended lag period associated with vinyl chloride starvation in *Nocardioides* sp. strain JS614. Arch. Microbiol. **187:**217-26.

المنارة للاستشارات
www.manaraa.com

77. **Mattes, T. E., N. V. Coleman, J. M. Gossett, and J. C. Spain.** 2005. Physiological and molecular genetic analyses of vinyl chloride and ethene biodegradation in *Nocardioides sp.* JS614. Arch. Microbiol. **183:**95-106.

78. **McGuinness, M., C. Ivory, N. Gilmartin, and D. N. Dowling.** 2006. Investigation of substrate specificity of wildtype and mutant *BphK*^{LB400} (a glutathione S-transferase) from *Burkholderia* LB400. Int. Biodeterior. Biodegrad. **58:**203-208.

79. **Nesvizhskii, A. I., A. Keller, E. Kolker, and R. Aebersold.** 2003. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. Anal. Chem. **75:**4646-4658.

80. **Ogunseitan, O.** 2006. Soil Proteomics: Extraction and Analysis of Proteins from Soil, Soil Biology, vol. 8.

81. **Ogunseitan, O., J. LeBlanc, and E. Dalmasso.** 2001. Microbial Community Proteomics, p. 125-140, Environmental Molecular Microbiology: Protocols and Applications. Horizontal Scientific Press, Wymondham, UK.

82. **Ogunseitan, O. A.** 1997. Direct extraction of catalytic proteins from natural microbial communities. J. Microbiol. Methods **28:**55.

83. **Ogunseitan, O. A.** 1993. Direct extraction of proteins from environmental samples. J. Microbiol. Methods **17:**273.

84. **Ogunseitan, O. A.** 1998. Protein method for investigating mercuric reductase gene expression in aquatic environments. Appl. Environ. Microbiol. **64:**695-702.

85. **Old, W. M., K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevinsky, K. A. Resing, and N. G. Ahn.** 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol. Cell Proteomics **4:**1487-502.

86. **Pappin, D. J., P. Hojrup, and A. J. Bleasby.** 1993. Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. **3:**327-32.

87. **Patterson, S. D., and R. H. Aebersold.** 2003. Proteomics: the first decade and beyond. Nat. Genet. **33 Suppl:**311-23.

88. **Peng, J., and S. P. Gygi.** 2001. Proteomics: the move to mixtures. J. Mass Spectrom. **36:**1083-91.

المنارة للاستشارات
www.manaraa.com

89. **Perkins, D. N., D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell.** 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20:**3551-3567.

90. **Pries, F., A. J. van den Wijngaard, R. Bos, M. Pentenga, and D. B. Janssen.** 1994. The role of spontaneous cap domain mutations in haloalkane dehalogenase specificity and evolution. J. Biol. Chem. **269:**17490-4.

91. **Ram, R. J., N. C. Verberkmoes, M. P. Thelen, G. W. Tyson, B. J. Baker, R. C. Blake, 2nd, M. Shah, R. L. Hettich, and J. F. Banfield.** 2005. Community proteomics of a natural microbial biofilm. Science **308:**1915-20.

92. **Rauch, A., M. Bellew, J. Eng, M. Fitzgibbon, T. Holzman, P. Hussey, M. Igra, B. MacLean, C. W. Lin, A. Detter, R. Fang, V. Faca, P. Gafken, H. Zhang, J. Whiteaker, D. States, S. Hanash, A. Paulovich, and M. W. McIntosh.** 2006. Computational Proteomics Analysis System (CPAS): An Extensible, Open-Source Analytic System for Evaluating and Publishing Proteomic Data and High Throughput Biological Experiments J. Proteome Res. **5:**1024-1024.

93. **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:**335-43.

94. **Searle, B. C., M. Turner, and A. I. Nesvizhskii.** 2008. Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. J. Proteome. Res. **7:**245-53.

95. **Singh, H., F. E. Loffler, and B. Z. Fathepure.** 2004. Aerobic biodegradation of vinyl chloride by a highly enriched mixed culture. Biodegradation **15:**197-204.

96. **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.

97. **Siuzdak, G.** 2003. The Expanding Role of Mass Spectrometry in Biotechnology. MCC Press, San Diego.

98. **Siuzdak, G.** 1996. Mass Spectrometry for Biotechnology. Academic Press, San Diego.

99. **Solaiman, Z., M. A. Kashem, and I. Matsumoto.** 2007. Environmental Proteomics: Extraction and Identification of Protein in Soil, Soil Biology, vol. 11.

المنارة للاستشارات

www.manaraa.com

100. **Song, F., Z. Zhuang, L. Finci, D. Dunaway-Mariano, R. Kniewel, J. A. Buglino, V. Solorzano, J. Wu, and C. D. Lima.** 2006. Structure, Function, and Mechanism of the Phenylacetate Pathway Hot Dog-fold Thioesterase PaaI. J. Biol. Chem. **281:**11028-11038.

101. **Squillace, P. J., M. J. Moran, W. W. Lapham, C. V. Price, R. M. Clawges, and J. S. Zogorski.** 1999. Volatile organic compounds in untreated ambient groundwater of the United States, 1985-1995. Environ. Sci. Technol. **33:**4176-4187.

102. **Suenaga, H., M. Mitsuoka, Y. Ura, T. Watanabe, and K. Furukawa.** 2001. Directed evolution of biphenyl dioxygenase: Emergence of enhanced degradation capacity for benzene, toluene, and alkylbenzenes. J. Bacteriol. **183:**5441-5444.

103. **Swaving, J., C. A. G. M. Weijers, A. J. J. van Ooyen, and J. A. M. de Bont.** 1995. Complementation of *Xanthobacter* Py2 mutants defective in epoxyalkane degradation, and expression and nucleotide sequence of the complementing DNA fragment. Microbiology **141:**477-484.

104. **van der Meer, J. R.** 1997. Evolution of novel metabolic pathways for the degradation of chloroaromatic compounds. Antonie Van Leeuwenhoek **71:**159-178.

105. **van Hylckama Vlieg, J. E. T., W. de Koning, and D. B. Janssen.** 1996. Transformation kinetics of chlorinated ethenes by *Methylosinus trichosporium* OB3b and detection of unstable epoxides by on-line gas chromatography. Appl. Environ. Microbiol. **62:**3304-312.

106. **Verce, M. F., R. L. Ulrich, and D. L. Freedman.** 2000. Characterization of an isolate that uses vinyl chloride as a growth substrate under aerobic conditions. Appl. Environ. Microbiol. **66:**3535-3542.

107. **Verce, M. F., R. L. Ulrich, and D. L. Freedman.** 2001. Transition from cometabolic to growth-linked biodegradation of vinyl chloride by a *Pseudomonas* sp. isolated on ethene. Environ. Sci. Technol. **35:**4242-4251.

108. **Wilkins, M. R., I. Lindskog, E. Gasteiger, A. Bairoch, J. C. Sanchez, D. F. Hochstrasser, and R. D. Appel.** 1997. Detailed peptide characterization using PEPTIDEMASS--a World-Wide-Web-accessible tool. Electrophoresis **18:**403-8.

109. **Wilmes, P., and P. L. Bond.** 2006. Metaproteomics: studying functional gene expression in microbial ecosystems. Trends Microbiol. **14:**92.

المنارة للاستشارات
www.manaraa.com

110. **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:**217-26.

111. **Wong, J. W., M. J. Sullivan, and G. Cagney.** 2008. Computational methods for the comparative quantification of proteins in label-free LCn-MS experiments. Brief Bioinform. **9:**156-65.

112. **Yates, J. R., 3rd, S. Speicher, P. R. Griffin, and T. Hunkapiller.** 1993. Peptide mass maps: a highly informative approach to protein identification. Anal. Biochem. **214:**397-408.

113. **Zhang, B., N. C. VerBerkmoes, M. A. Langston, E. Uberbacher, R. L. Hettich, and N. F. Samatova.** 2006. Detecting differential and correlated protein expression in label-free shotgun proteomics. J. Proteome. Res. **5:**2909-2918.

المنارة للاستشارات
www.manaraa.com

APPENDIX A: CHAPTER I SUPPLEMENTARY INFORMATION

المنارة للاستشارات
www.manaraa.com

Table A.1. Additional JS614 proteins expressed in response to VC, ethene, and epoxyethane that were identified with statistical relevance ($P<0.05$) using MALDI-TOF MS PMF analysis.

www.manaraa.com

Growth Substrate[a]	Significant MASCOT matches[b]	Predicted Mass (kDa)[c]	Observed Mass (kDa)[d]	MOWSE Score[e]	E Value[f]	% coverage[g]	NBM[h]	MM/ TPMS[i]	Gene No[j]
Chromosomally-encoded proteins									
ethene	electron transfer flavoprotein, alpha subunit	32.4	34	94	9.2E-04	54%	72	9/89	1816
epoxyethane	hypothetical protein Noca_3467	23.7	31	98	3.6E-04	46%	-	8/42	3467
epoxyethane*	hypothetical protein Noca_3572	25.5	31	95	1.1E-02	50%	67	9/52	3572
ethene	glutamine synthetase, type I	53.4	74	85	1.4E-02	28%	66	10/65	2295
Plasmid-encoded proteins									
vinyl chloride	probable alkene monooxygenase beta subunit	39.3	34	133	2.4E-07	42%	58	10/32	4809
ethene	probable alkene monooxygenase beta subunit	39.3	36	136	5.8E-08	44%	-	18/126	4809
ethene*	glutaconate CoA-transferase	30.3	31	130	1.5E-05	70%	65	13/95	4812
ethene	coenzyme A transferase	33.2	32	116	5.8E-06	54%	66	15/110	4842
ethene	glutaconate CoA-transferase	31.5	31	128	3.6E-07	73%	-	15/117	4843
ethene	Phosphosulfolactate synthase	28.0	34	81	1.7E-02	44%	-	8/65	4839
vinyl chloride	(2R)-phospho-3-sulfolactate synthase, ComA	28.1	34	84	8.8E-03	40%	69	7/41	4816
ethene	(2R)-phospho-3-sulfolactate synthase, ComA	28.1	34	95	6.6E-04	49%	-	9/65	4816
acetate	(2R)-phospho-3-sulfolactate synthase, ComA	28.1	34	86	6.2E-03	40%	-	7/39	4816
epoxyethane	luciferase family protein	32.6	44	118	3.6E-06	59%	-	12/70	4832
vinyl chloride	luciferase family protein	32.6	44	88	3.5E-03	49%	68	10/89	4832
ethene*	luciferase family protein	32.6	37	111	3.2E-05	59%	72	11/83	4832

المنارة للاستشارات
www.manaraa.com

Table A.1 (continued)

Notes:

With the exception of coenzyme A (CoA), all other proteins observed in acetate lanes are not shown. Previous RT-PCR results (18) support the presence of Coenzyme A (ComA) in acetate-, ethene-, epoxyethane-, and VC-grown cells.

[a]Growth substrate using methods described in (77)

[b]Most significant Mascot hits to current gene annotations in the NCBInr database

[c] Theoretical mass reported by Mascot

[d] Observed mass estimated by SDS-PAGE

[e] Mowse score reported by Mascot

[f] E-value reported by Mascot

[g] Amino acid percent coverage of matched peptides reported by Mascot

[h] Next Best Mascot Match (NBM) result not associated with JS614

[i] Number of peptide mass matches (MM) / Total number of peptide masses searched (TPMS)

[j] Gene number in finished JS614 genome sequence corresponding to MASCOT hit. (*observed in two independent analyses and reported as averaged values)

المنارة للاستشارات
www.manaraa.com

APPENDIX B. NYCODENZ EXTRACTION METHOD FOR SEDIMENT SAMPLES (PROVIDED BY LAURA BADTKE)

المنارة للاستشارات
www.manaraa.com

Add 100 g of sediment to 150 ml Nycodenz buffer (.2M NaCl, 50mM Tris-HCl pH 8.0) or .2% sodium hexametaphosphate in 250 ml Warring blender for 1 minute on low speed. Divide solution into five centrifuge tubes and centrifuge in swing-out rotor at 700x g (2000 rpm) for 15 min, 10° C. Supernatant is still dirty with soil particulate with many of the samples, but more clear with clay-like samples. Supernatant was filtered into three clean centrifuge tubes through two 4-ply gauze sheets lining a funnel and then centrifuged at 7500x g (6920 rpm) for 20 min, 10° C. The clear supernatant was poured off and the pellet often still looked quite dirty. The three pellets were each resuspended in about 8 ml of sterile .8% sodium chloride solution. The pellets resuspended easily with vortexing. The three resuspended pellets were carefully poured into one clean centrifuge tube containing 10 ml of Nycodenz solution. When adding the sediment samples the interface with the Nycodenz solution would become very blurred and mix together (with both buffers), but with the fresh soil sample the interface between the two remained defined. The solution was centrifuged at 1400x g (9650 rmp) for 40 mins, 10° C in the swing-out rotor. After centrifugation, soil particles were on the bottom of the tube beneath the Nycodenz. The cell layer on top of the Nycodenz was very difficult to distinguish and almost clear with the sodium hexametaphosphate buffer, but the interface was usually a bit dirty and easy to distinguish with the Nycodenz buffer. With clay-like soils the cells at the interface were whitish and very easy to distinguish. About 5 ml of interface liquid (cell layer) was pipetted from right above the Nycodenz layer and placed in a new centrifuge tube. 15 ml of sterile water was added and cells were centrifuged at 7500x g (6920 rpm) for 20 min, 10° C. Pellet was usually whitish and small. Supernatant was removed and cells resuspended in another 15 ml of sterile water with vortexing. Cell solution centrifuged at 7500x g (6920 rpm) for 20 min, 10° C. Supernatant removed and cells resuspended in 900 µl KP buffer.

All 900 µl of resuspended cells were added directly to the MoBio kit. The kit procedure for increased yields was followed exactly until the Solution 2 was added. The

المنارة للاستشارات

www.manaraa.com

protein was then left to precipitate at least 30 minutes, but usually overnight. The protein was then pelleted as directed in the MoBio kit and resuspended in 20 µl of KP buffer.

المنارة للاستشارات
www.manaraa.com