8-2014

Comparative Proteomics Reveals Core vs. Unique Molecular Signatures for Dissimilatory Metal Reducing Bacteria Grown with Various Electron Acceptors

Xiaoxin Liu
University of Tennessee - Knoxville, xliu38@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Xiaoxin Liu entitled "Comparative Proteomics Reveals Core vs. Unique Molecular Signatures for Dissimilatory Metal Reducing Bacteria Grown with Various Electron Acceptors." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Robert L. Hettich, Major Professor

We have read this dissertation and recommend its acceptance:

Loren Hauer, Frank Loeffler, Kurt Lamour, Mircea Podar

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Comparative Proteomics Reveals Core vs. Unique Molecular Signatures for Dissimilatory Metal Reducing Bacteria Grown with Various Electron Acceptors

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Xiaoxin Liu
August 2014

DEDICATION

I dedicate this dissertation to my beloved parents, Yuqin Xiao and Yanqun Liu.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my mentor Dr. Robert Hettich for his support, guidance and dedication throughout my Ph.D. journey. I not only learned how to do science, but also received many "father's advice" from him. I am very grateful for all the encouragement he provided when I was at the lowest moment in my life. I remember the story he told me, rephrased: "I grew up in a farm, and my family grew corn. My father would use a cultivator to weed the corn. I sometimes helped him, but I could not drive the cultivator as straight as he did, and would bump the corn plants on the side. However, the bumped corn would correct itself and eventually became even stronger than the un-bumped ones. You know what, you are the bumped corn right now, and you will grow stronger!"

Secondly, I would like to thank my dissertation committee members: Dr. Frank Löffler, Dr. Loren Hauser, Dr. Mircea Podar, and Dr. Kurt Lamour for investing their time and effort to help me improve. Their comments and suggestions helped me build this dissertation. I am especially grateful for the opportunity to collaborate with Dr. Frank Löffler. It was a fantastic experience to work with his group. I admire his endless energy and brilliant ideas, and appreciate his scrutiny in editing manuscripts.

Thirdly, I would like to thank the Genome and Science Program at the University of Tennessee for education and assistance throughout my graduate life. I would also like to thank the staff in Organic Biochemical and Mass Spectrometry group in Oak Ridge National Laboratory for great working experiences.

Last but not least, I would like to thank my family and friends. I thank my parents who have always been there supporting me and loving me. I thank my fiancé, Jun Wang, for giving me unconditional support and assistance to this dissertation. Thank you my brother and sister, Dr. Changwen Chen and Connie Chen, who have been praying for me and helping me in my life.

ABSTRACT

Dissimilatory metal reducing bacteria (DMRB) are probably one of the most respiratory versatile microorganisms on earth. Their ability to use metals as terminal electron acceptor allows them to survive in severe environments (e.g. radionuclide contaminated soil). In addition to metals, many other organic and inorganic substrates can be utilized as electron acceptors for DMRB respiration, including fumarate, nitrate, oxygen, etc. Genome information for many DMRB species is available, which reveals large numbers of *c*-type cytochrome encoding genes present in their genomes. For example, the genomes of three DMRBs, *Anaeromyxobacter dehalogenans* strain 2CP-C, *Shewanella oneidensis* strain MR-1, and *Geobacter daltonii* strain FRC-32, contain 69, 40, and 72 putative *c*-type cytochrome genes, respectively. Although mutagenesis techniques have determined the respiratory roles of several *c*-type cytochromes, gene disruption for majorities of the putative *c*-type cytochromes does not generate visible phenotypical alterations, and is not able to functionally link them to specific respirational activities. Thus, comprehensive proteome characterization for DMRBs is needed to elucidate the molecular mechanisms underlying their respirational versatilities. In this dissertation, a mass spectrometry-based proteomics approach was used to interrogate the proteomes of *A. dehalogenans* strain 2CP-C, *S. oneidensis* strain MR-1, and *G. daltonii* strain FRC-32. The proteomic responses of DMRBs to a wide range of electron acceptors were tested in this dissertation, including soluble and insoluble ferric iron, manganese oxide, fumarate, nitrate, oxygen, and nitrous oxide. The in-depth proteomic characterizations comparatively revealed the *c*-type cytochrome profiles of DMRBs, providing evidence for the identities and expressions of putative *c*-type cytochromes, and established the linkage between specific electron acceptor and individual *c*-type cytochromes. The entire proteome complements of DMRBs were also characterized, generating metabolic maps reflecting pathway-level activities responding to various electron acceptors. The results identified the core proteome carrying out the essential cellular machineries for each tested DMRB, and demonstrated clearly elevated energy

metabolism for *A. dehalogenans* strain 2CP-C during respiration of metal electron acceptors. Comparative proteomics analysis between tested DMRB strains revealed the commonalities and differences of proteomic phenotypes displayed by different strains, and shed light into deeper understandings for DMRB metabolic activities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ATTACHMENT

Chapter 1

Introduction to MS-based Proteomics and Its Application to Environmental Microbiology

1.1 The "Omics" Tetralogy and Systems Biology

For centuries, research in biochemical, molecular, and cellular biology focused on the analysis of subsets of biomolecules and biochemical reactions in a given organism at a time, which achieved tremendous success in understanding cellular processes that reflect various aspects of living organisms. Life science research was revolutionized by the discovery of DNA (deoxyribonucleic acid) and its recognition as the carrier for genetic information [1-3]. Following the introduction of the double-helix structure of DNA, the genetic code embedded in the DNA sequence was deciphered, leading to the establishment of central dogma for molecular biology which describes a framework in which genetic information is transcribed from DNA to RNA (ribonucleic acid) and translated to proteins [4, 5]. Obtaining comprehensive genome sequences became desirable in that the whole genome sequence presents a blueprint of all potential functional activities of an organism. Ever since the completion of the first whole genome sequences of the bacterium *Haemophilus influenza* [6], the field of genomics has exploded, expanding our knowledge about life science in general, and has revolutionized how biological research is conducted. In particular, biological research embraced large-scale analyses through emerging 'transcriptomics', 'proteomics', and 'metabolomics' studies. While transcriptomics measures RNA expression in a given organism, proteomics characterizes the protein complement in a given biological system, and metabolomics captures the related

metabolites resulting from all cellular metabolic reactions. The burst of 'omics' studies compose a tetralogy which enabled the holistic view of biological systems such as a single living organism, a symbiotic system, or a microbial community. The system-level information offered by the "omics" studies was unachievable previously, giving rise to the discipline of "systems biology" (Figure 1.1).

Genomics is the very first chapter of the "omics" tetralogy, serving the basis for transcriptomics and proteomics. In the late 1970s, a powerful method (Sanger sequencing) for DNA sequencing was developed [7-9], providing a basis for enabling sequence-driven research. Continual technical refinement and improvement greatly reduced the cost and increased the speed and throughput for genome sequencing, allowing the whole genomes of more and more organisms being sequenced with high quality. The success of the first draft of human genome sequence [10, 11] is one of the greatest achievements in the 21st century, which unprecedentedly broadened the horizon toward the understanding of human being. Biomedical research tremendously benefited from this rich information-set, in that the genotyping of human health and diseases shed light into the underlying molecular mechanism for human disease (e.g. Huntington's disease is caused by genetic disorder [12]). Nowadays, complete genome sequence is widely available for numerous organisms ranging from *Prokaryote* to *Eukaryote*. Genomics not only provides insight into the genetic makeup for single organisms, but also reveals the genetic diversity, evolution, and metabolic potential for microbial communities (metagenomics) [13]. The success of direct extraction and analysis of the collective genomic information for microbial communities expanded the scope of microbiology by providing insight into uncultured microorganisms which could not be characterized

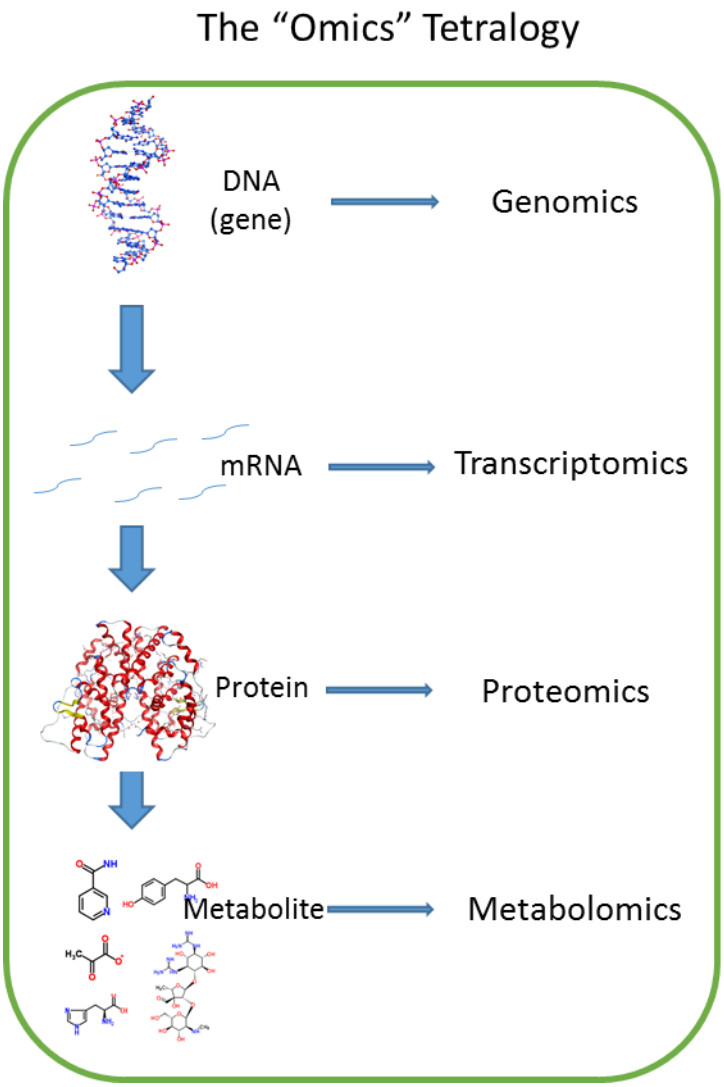


Figure 1.1. The “omics” tetralogy. The system-level information achieved by different levels of the “omics” sciences enable the comprehensive understanding of biological systems. Genomics describes the complete genetic makeup of a cell. Transcriptomics measures the complement of RNA transcripts expressed under certain conditions. Proteomics characterizes the entire repertoire of proteins present in the biological system. Metabolomics measures the suite of small molecules. Taken together, our understanding toward systems biology is enhanced through the “omics” tetralogy.

previously [14]. Metagenomics studies allow *in situ* characterization of the genomic elements in microbial consortia residing in diverse environmental settings such as ocean, soil, deep sea sediment, human gut, human oral cavity, etc. [15-19], and many of which cannot be cultured in the lab. Genome cataloguing describes all functional potentials of the biological system under study. Information contained in the genetic content must be utilized through the expression of genes, generating transcripts and proteins for functional activities at the cellular level. Thus, in the post-genomic era, great efforts are put into the study of all the transcripts (transcriptome) and proteins (proteome) present in a biological system.

Transcriptomics emerged as an endeavor to reveal the functional dynamics of the genome, since gene expression directly leads to RNA transcription. While the genome of an organism is always constant, gene activity is dynamic, and only subsets of the genome get transcribed for subsequent functional roles. Comprehensive interrogation of the identity and quantity for the complete set of transcripts in a cell directly reflects the activation of functional elements in the genome in a certain developmental state or physiological condition. Two different approaches are available for transcriptomics profiling, one is based on hybridization (e.g. microarray) and the other is based on sequencing techniques (e.g. RNA-seq). The hybridization-based approach is high-throughput and generally low-cost, is widely used for transcriptomics studies, and successfully profiled gene expression for various biological systems [20-22]. Development of next-generation sequencing made it feasible to achieve deep sequencing of the transcriptome with high resolution, providing the benefit of larger dynamic range, independence from existing genome information, and clear determination of transcription boundaries [23]. Transcriptomics provides direct measurement for gene activity,

but does not necessarily represent the functional aspect of gene expression, because the final product of gene expression is protein, not transcript. In addition, the constitutive expression and post-translational modifications of proteins, and proteins resulting from post-transcriptional modifications (e.g. alternative RNA splicing) could not be addressed by transcripts. In this regard, proteomics, generating the collective view of all the proteins, represents perhaps the most informative level of functional activities at the cellular level.

Proteomics developments started in late 20th century. In the early stage, the complete set of proteins in an organism is collected and separated by two-dimensional gel electrophoresis (2DE) [24, 25] which presents a proteome map containing spots of proteins with various charges and sizes. The identification of protein spots was resolved through the estimation of amino acid composition by matching the pI (isoelectric point) and molecular weight against the chemical properties of proteins in an existing database for a certain organism [26]. Quantification of protein spots could be achieved by various staining methods, as represented by staining intensities. When mass spectrometry (MS) was implemented for protein analysis, protein identification from gel spots became much more confident and reliable, owing to the determination of highly accurate protein mass by MS. The introduction of protein ionization methods, electrospray ionization (ESI) [27] and matrix assisted laser desorption/ionization (MALDI) [28], were the major driving forces for MS-based protein analysis. MS protein identification following 2DE has been a popular approach for proteomics studies, and was widely used for proteomic interrogation in many biological settings [29]. The utilization of liquid chromatography (LC) for protein and peptide separation outperforms 2DE in that separation time is no longer limited, the dynamic range of measurable proteins is tremendously

improved, and LC can directly connect in-line with MS [30]. The coupling of LC with MS (termed LC-MS) enabled proteomic characterization of complex samples (e.g. peptide mixture from tryptic digests of plant proteome [31]). Nowadays, LC-MS is the preferred method for MS-based proteomics analysis, providing an unparalleled platform for comprehensive characterization for protein sequence identification and quantification. With the continual improvement of MS instrumentation, proteome measurement can be achieved with higher speed, resolution, accuracy, and dynamic range. The successful application of LC-MS for proteomics analysis presents the protein-level functional view for metabolic activities in various biological systems, from single bacteria isolate, to higher order eukaryotes, and to complex microbial communities [32-37]. Greater details for MS-based proteomics will be discussed in section 1.2.

Metabolomics characterizes the complement of metabolites in a given biological system. Metabolites are small molecules being transformed during metabolism, which generally are the substrates, inhibitors, or products of biological catalytic reactions driven by proteins [38]. The information provided by metabolomics is the readout of biochemical activities at the molecular level, and offers a different perspective for the systematic functional overview for biological research. Characterization of metabolome, similarly to proteomics, heavily relies on mass spectrometry. Out of the four “omics” sciences described in this chapter, metabolomics is the one that does not directly tie to the genome information. Although still in its infancy, metabolomic studies have provided deep fundamental molecular insight for biological systems such as human, microbial communities, etc. [39, 40]. Together with proteomics, metabolomics presents informative details for the metabolic activities responsible for cellular functions.

The “omics” tetralogy provides different layers of information for systems biology: genomics for the total genetic potential, transcriptomics for gene activity elucidation, proteomics for functional representation, and metabolomics for direct signatures of metabolic reaction. Each field is under rapid development and enhancement, and is generating tremendous information assets for biological research. Although mechanics for the integration of the “omics” tetralogy remains elusive at present, the information output gathered from all aspects of “omics” research will propel our understanding about life science.

1.2 Proteomics - the Apogee of the “Omics” Science

As mentioned above, proteomics is best suited for the functional representation and physiological characterization of metabolic activities at the cellular level, as it directly measures proteins – the final gene functional product. Only through direct proteome characterization can we understand the identities, quantities, and modifications of gene functional carriers (proteins) accounting for the major metabolic units for cellular activities.

The LC-MS-based proteomics is the current state-of-the-art technique for comprehensive proteome characterization, and can be divided into two different types: top-down and bottom-up (Figure 1.2). In the top-down approach, proteins are analyzed in the intact form, giving the opportunity to study protein structure (e.g. post-translational modification (PTM)), protein-protein interaction (e.g. polymerization), and protein diversity (e.g. splice variants). In a typical LC-MS-based top-down proteomics measurement, the mixture of whole proteins are separated based on hydrophobicity and/or charge in the mobile phase of LC. The

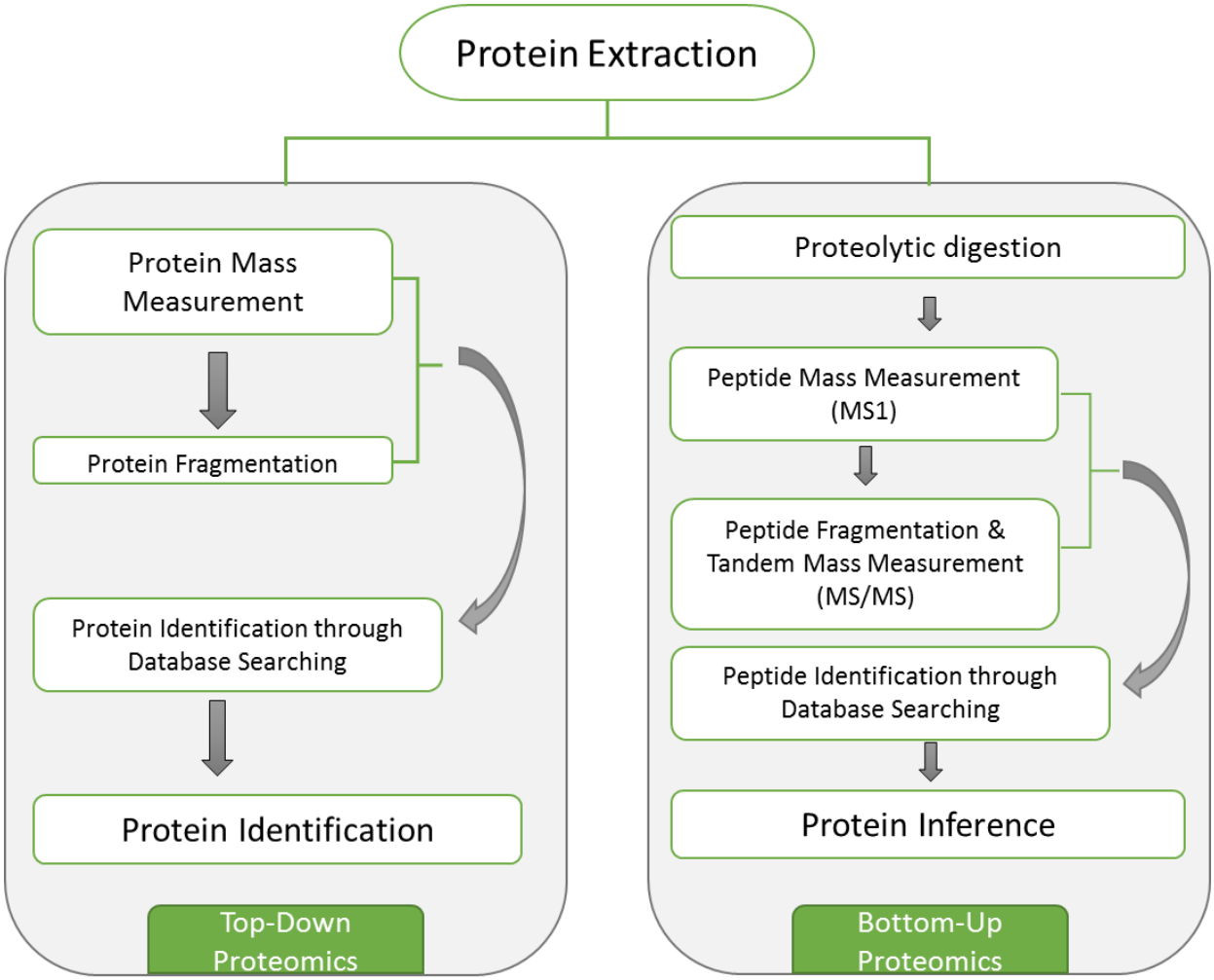


Figure 1.2. The top-down and bottom-up proteomics strategies. The top-down approach directly investigates intact proteins. The bottom-up approach measures digested proteins, and protein identification is achieved through identified peptide sequences.

precise m/z value (mass to charge ratio) of the intact proteins is measured by a MS instrument generally of high-resolution, such as Fourier Transform Iron Resonance (FTICR) or LTQ Orbitrap (Thermo Scientific, Inc.). The intact proteins could be fragmented inside the mass spectrometer, and the m/z values for the fragment ions could be measured, generating tandem mass (MS/MS) spectra. The exact mass of the intact protein, along with the fragmentation pattern presented in the MS/MS spectra, provides information for protein sequence identification. The accurate mass of the intact protein is also useful to distinguish very precise mass differences of protein variations (e.g. protein isoforms). Several limitations still exist for the application of top-down approach in large-scale deep proteomics interrogation. First of all, the mass range of proteins present in a biological system is large – the mass difference of the largest and the smallest protein in a proteome could be over several hundreds of kilo Daltons (kDa). Large protein molecules, such as those with a molecular mass greater than 50 kDa, propose difficulties for LC separation, ionization and MS detection [41]. Secondly, various protein modifications, such as PTMs and single nucleotide polymorphisms (SNPs), could induce changes to the molecular mass of intact proteins and challenge protein identification through top-down approach.

The bottom-up approach, also termed the “shotgun” approach, is more favorable for deep proteomic measurement at the present time. The concept of shotgun proteomics is very similar to that of shotgun genome sequencing, in which long DNA strands are divided into short segments, and the segment strands are sequenced individually and reassembled to present the complete DNA sequence. In shotgun proteomics, proteins are digested into peptides prior to separation and detection, and detected peptide sequences are used for protein identifications. The peptide mixture generated from proteome digestion greatly reduced the sample mass,

which is beneficial for ionization and subsequent MS analysis. For example, an average tryptic peptide contains 10 amino acids, which is about 1 kDa in molecular mass. The increased complexity of peptide mixture could be a nightmare for gel electrophoresis separation, but can be easily resolved through chromatographic separation. The m/z values for the separated and ionized peptides (parent/precursor mass) are measured by MS, and the precursor ions can be fragmented and analyzed to obtain MS/MS spectra. For peptide identification, a protein database derived from the corresponding genome database is exploited. Proteins in the database are *in silico* digested to generate predicted peptides which are subsequently fragmented *in silico* to obtain predicted fragmentation pattern. The experimental precursor mass could be used to search for predicted peptides, and the experimentally generated MS/MS spectra can be matched to predicted fragmentation patterns to determine peptide sequences. Once the peptide sequences are determined, they are traced back to the originating proteins. Many computational software algorithms have been developed to automate the peptide sequence determination and protein inference for bottom-up proteomics, such as SEQUEST, MyriMatch, DTASelect, IDPicker, etc. [42-45]. The bottom-up strategy is routinely performed for many large-scale proteomics studies, although it does contain a limitation incurred by its peptide-centric nature. For higher order eukaryotes (e.g. human, plant) whose genome contains many redundancies, sequence homologies among different proteins is very common. Proteolytic digestion of the homologous proteins could lead to redundant, homologous, or isobaric peptide species, making it difficult to identify and quantify the originating proteins. Despite this issue with the higher order eukaryotes, the bottom-up strategy is currently the

method of choice for large-scale proteomics studies and has achieved great success for proteomic characterizations for many biological systems [34, 36, 46-48].

In this dissertation, proteomic characterizations for microbial isolates were achieved using the bottom-up strategy, for which technical details will be provided in Chapter 2.

1.3 MS-based Proteomics Provides Insight into Environmental Microbiology

Microorganisms comprise the largest population on earth, accounting for 60% of the global biomass [49], and ubiquitously occupies every imaginable environment from soil, water, air to environments with extreme conditions (e.g. extreme pH, salinity, temperature, etc.) [50, 51], and to human bodies (e.g. microbial species present in human gut, oral cavity, skin surface, etc.) [52]. The activities of environmental microorganisms are very important for many ecological processes (e.g. the biogeochemical cycling of elements) as well as human health (e.g. pathogenic microbes cause disease conditions to human beings). Environmental microbiology, the study of environmental microorganisms, was pioneered by Beijerinck who developed the selective enrichment techniques which allowed cultivation of bacteria with specialized functions (e.g. nitrogen-fixation). The discovery of microbial biotransformation of nitrogen, carbon, sulfur, iron, and manganese by Beijerinck and Winogradsky initiated the field of environmental microbiology, and spurred research on environmental microorganisms [53, 54].

Nowadays, there are essentially two ways to study environmental microorganisms. One is to isolate microbial species of interest from the environment, and study various aspects of

the isolate under laboratory conditions to explore the physiological properties, metabolic activities, or the molecular mechanisms for specialized functions. In addition to studying single microbial species isolate, the laboratory-based strategy also includes the investigation of artificial mixtures of microbial isolates. In contrast, the other approach for environmental microbiological research is field-based, which investigates microorganisms (either isolates or communities) in their natural habitats, in an effort to unravel the microbial diversities in the ecological system, interactions and communications between microorganisms or with the environment, and the microbial functional contributions to the biosphere. For either strategy, laboratory-based or field-based, proteomics has played pivotal role in providing massive information to deepen our understandings about environmental microbiology.

The vast majority of environmental microbial proteomics studies are laboratory-based, focusing on model microorganisms with specialized properties of interest. Many microbial species are of interest due to their detoxification abilities, tolerance to extreme environmental conditions, respiratory versatilities, etc. Studies of their metabolic activities could lead to new biological discoveries (e.g. novel enzymes), or beneficial applications to the environment (e.g. bioremediation). Since next-generation sequencing made it quick and easy to achieve complete genome sequencing for numerous microbial isolate species, the available genome information made it straightforward to perform MS-based proteomics studies on microbial isolates. By investigating the repertoire of proteins expressed by microorganisms under monitored laboratory conditions, overall microbial physiological activities could be revealed, and could be exploited by comparative studies to uncover differential molecular processes responding to specific environmental perturbations. For example, the comparative proteomic analysis of

Shewanella oneidensis revealed proteins essential for the regulation of chromate transformation [55, 56]. Proteomics has been used to interrogate various microbial isolates and to characterize the molecular machineries for their unique traits, including the flexible metabolism of *Pseudomonas*, the methanogenesis property of *Methylococcus capsulatus*, the denitrification ability of *Paracoccus denitrificans*, etc [57-59]. The proteomic information achieved for numerous microbial isolates not only profiles the protein repertoire under laboratory conditions, but also provides detail for strategic design of subsequent biological and environmental studies and applications (e.g. gene knockout experiments, environmental manipulations) [49, 60].

Since microorganisms do not live in isolation, and their natural environmental conditions are almost impossible to emulate in laboratory, the study of microbial communities in their natural habitat became desirable. MS-based proteomic characterization for environmental microbial communities not only provides an *in situ* protein-level overview for the microbial community, but also allows investigation into the uncultivable microbial species which could not be achieved under laboratory conditions. Information about microbial community structure, diversity, functional activities can be achieved by metaproteomic characterization, and are of tremendous importance for the understanding of microbial interactions, adaptations, and ecological functions in the environment. The metaproteomic characterization of the acid mine drainage microbial biofilm community is a landmark for environmental community proteomics [46]. With more metagenomics studies providing genome information for environmental microbial communities, more and more metaproteomics studies have been conducted on microbial consortia in various environments

(e.g. water, soil) providing insight into many important research areas, such as wastewater treatment, bioremediation, global warming [61-63].

Despite the success applications of MS-based proteomics on environmental microbial communities, technical limitations impede the wide feasibility of field-based proteomics. Unlike microbial isolates whose genomes are widely available, metagenome information is limited and the quality may not be optimal, which directly affects the performance of metaproteomics identification. And strain-level differentiations are very difficult to achieve both at the metagenome and metaproteome level, due to the high complexity and unpredictability of environmental microbial communities. In addition, protein extraction from environmental samples is more difficult than from laboratory cultures because of interferences (e.g. humic acid in soil). In contrast, proteomic characterization for microbial isolates under laboratory conditions is a mature technique, and is routinely conducted in many research institutes. In this dissertation, the advantage of using MS-based proteomics on microbial isolates in laboratory conditions is exploited to provide insight into the metabolic activities of respirational versatile environmental microbial isolates.

1.4 Dissimilatory Metal Reducing Bacteria and Their Environmental Impact

Dissimilatory metal reducing bacteria (DMRB) are able to couple the dissimilatory reduction of metals with the oxidation of organic substrates or H_2, and the electron flow during this redox process directly relates to energy production [64]. The use of metal as terminal electron acceptor has been observed in microorganisms with diverse phylogeny, ranging from

Bacteria to Archaea domain. Most of the early identified DMRBs are within the delta subdivision of Proteobacteria such as *Geobacter*, *Desulfuromonas*, and *Pelobacter*, and later studies on microbial dissimilatory metal reduction extended DMRBs to the gamma and epsilon subdivisions of Proteobacteria (e.g. *Shewanella*, *Geospirillum*) as well as Acidobacteria (e.g. *Geothrix*) [65]. To date, more and more DMRBs are identified and their phylogeny spreads even broader throughout the Bacteria domain [66].

The diversity and ubiquity of DMRBs greatly impact the global biogeochemical distribution of metal elements. The valence transitions of metal elements can be abiotic or biotic, and the biotic enzymatic redox is considered the prevailing form of environmental metal transformation [67]. A large variety of metal elements could be respired by bacteria, including Fe(III), Mn(IV), U(VI), Se, Cr(VI) Hg(II), Tc(VII), V(V), Mo(VI), Cu(II), etc. [67], among which, dissimilatory reduction of Fe(III) is the most predominant, in that iron is the fourth most abundant element in earth crust [66]. The abundance of iron on early earth, as well as the prediction of ancestry iron-reducing microorganisms, suggests that microbial iron respiration may be one of the first forms of metabolic activities for life origin [68-71]. Most Fe(III) reducing bacteria are also able to reduce Mn(IV), and the environmental Mn(IV) reduction is readily accompanied with Fe(III) reduction [72, 73]. Thus, extensive research interests have been focused on the naturally occurring bacterial dissimilatory Fe(III) and Mn(IV) reduction and the environmental impacts [74]. Bacteria that utilize U(VI), Se(VI), Cr(VI), and Hg(II) as terminal electron accepter for respiration are also of great significance. Although not as widely distributed as iron, contamination with these elements proposes hazardous threats to environments due to their radiation or toxicity. And microbial reduction of these metal

elements could be exploited as a remediation strategy. For example, several *Geobacter* spp. are able to reduce U(VI) in contaminated soil and ground water, resulting in the immobilization of uranium which is amenable for subsequent removal and clean-up [75].

In addition to metal electron acceptors, many organic and inorganic compounds (e.g. fumarate, nitrate) could be utilized for respiration by DMRBs due to their respiratory versatilities. A range of electron donors could also be used as electron source for respiration, such as lactate, hydrogen, succinate, acetate, etc. The oxidation of organic matter coupled to dissimilatory reduction of metals is a very important mechanism for organic matter degradation in a variety of environments including aquatic sediments, submerged soils, and aquifers. The redox of various elements by DMRBs tremendously influences the geochemistry of metal and nutrient in the environment. As a result, water quality as well as soil composition could be interfered by microbial activities of DMRBs. In turn, plant growth and human food resource is also mediated by microorganisms. The ecological impacts of DMRBs are profound, in that their diverse respiratory activities significantly affect nutrient cycling, organic matter degradation, mineral dissolution, and weathering [66].

Many species within the *Geobacter* genus are identified as DMRBs such as *G. sulfurreducens* and *G. metallireducens*. *Geobacter* species are often used as model organisms for research on dissimilatory metal reduction because they are generally the most common and most abundant Fe(III) reducing bacteria in many soils and sediments where active iron reduction processes occur. In addition to Fe(III), the range of electron acceptors that could be respired by *Geobacter* species include Mn(IV), U(VI), Co(III), V(V), etc [76]. The electron donors

available for *Geobacter* respiration is also diverse, including acetate, hydrogen, humics, Fe(II), U(IV), aromatic compounds, and other fatty acids [76]. Due to the predominant Fe(III) reduction in natural occurrence, the physiological aspects about Fe(III) reduction by *Geobacter* spp. are the most studied. Fe(III) is highly insoluble at the circumneutral pH in most natural environments, and generally present in various oxide forms (e.g. goethite, hematite). In order to reduce environmental Fe(III) which is not readily accessible to the cell, *Geobacter* spp. must develop a special strategy to transport electrons to extracellular solid-phase Fe(III). Studies on the mechanism for extracellular electron transfer were mostly done for *G. sulfurreducens*. It is clear that *G. sulfurreducens*, and presumably other *Geobacter* species, produce electrically conductive pili structure, which is essential for Fe(III) reduction [77, 78]. The concept of metallic-like conductive pili responsible for long range electron transfer is still under investigation, but it has been proven that the pilin protein PilA is the essential structural protein required for extracellular Fe(III) reduction [79]. Reduction of Mn(IV) by *Geobacter* species is studied in less detail. But it is proposed that a generalized electron transfer mechanism, rather than specific and specialized energy conservation strategies, is developed by *Geobacter* spp. for the reduction of different metal ions [76].

Although much focus has been directed to *G. sulfurreducens* for the extracellular electron transfer mechanism of *Geobacter*, different species, and even different strains, have their unique physiological properties. For example, the *G. sulfurreducens* strain KN400 produces more current and reduces Fe(III) oxides at a faster rate than strain DL-1[80]. Thus, other species and strains of *Geobacter* need more extensive research to elucidate their individual functional roles in their natural environment. *G. daltonii* strain FRC-32 was isolated from the subsurface

sediments with mixed waste of heavy metal, radionuclides and hydrocarbon contamination. Its high abundance in the contaminated site, together with the ability to reduce metal electron acceptors such as U(VI) and Fe(III), attracted research interests to understand its contribution to the in-site metal reducing activities and contaminant remediation. In order to provide insight into the metabolic activities and molecular machineries of *G. daltonii* strain FRC-32, proteomic characterization of this strain is achieved in this dissertation and will be presented in Chapter 3 and 6.

The *Shewanella* genus is another group of intensively studied facultative bacteria for its versatile dissimilatory metal reducing capability. First isolated in 1931 from rotten butter, *Shewanella* were initially characterized as *Acromobacter*, but soon reclassified as *Pseudomonas* and then *Alteromonas* based on its phenotypic traits. The genus *Shewanella* was not established until 1985, when molecular taxonomical analysis based on 5S rRNA sequence were performed [81]. Our knowledge on *Shewanella* metal reducing capability started with the isolation of *Shewanella oneidensis* strain MR-1 from Lake Oneida, NY in 1988, where researchers found reduced form of manganese present in the aquatic environment in contrast to the commonly found oxidized form [82]. Lab cultural experiments in this study proved the direct link between manganese reduction and *S. oneidensis* anaerobic respiration. This finding triggered many following studies on the anaerobic respiration of *S. oneidensis* MR-1, especially on other possible external electron acceptors. Many other species of *Shewanella* were also isolated and identified from a variety of geographical locations, mostly aquatic habitats in marine and sedimentary environment [83]. In these studies *Shewanella* spp. demonstrated remarkable versatility on utilizing a broad range of electron acceptors including metals such as

Fe(III), U(VI), Cr(VI) as well as organic compound such as sulfite, thiosulfate, arsenate, fumarate etc. [83]. As a bacteria capable of growing with insoluble mineral oxide, *S. oneidensis* exhibits great potential for applications in bioremediation of heavy metal and radionuclide pollution sites as well as the development of microbial fuel cells. Studies with *S. oneidensis* strain MR-1 have revealed three possible routes that mediate the transfer of electrons to extracellular non-soluble mineral oxide (electron acceptors). Outer-membrane cytochromes MtrC and OmcA were initially attributed with the capability of electron export via direct contact with extracellular solid surfaces [84]. *S.oneidensis* strain MR-1 also secretes soluble riboflavins to serve as "electron shuttles" to enable electron exchange to external surfaces without direct interaction [85]. Besides, *S. oneidensis* MR-1 is also capable of producing electronically conductive nanowires to connect cells as well to biofilm attached surfaces. The electron conductivity, however is lost upon the mutating of out-membrane cytochrome MtrC and OmcA genes [86]. Overall these findings suggested that *Shewanella* species possess sophisticated metabolic systems to harness diverse sources of external electron acceptors and hence empower the organism to survive in different harsh environments.

Anaeromyxobacter dehalogenans strain 2CP-C is recent addition to the DMRB family. Isolated from the Cameroon rain forest soil [87], *A.dehalogenans* strain 2CP-C has shown to be able to use a wide range of respiratory substrates including Fe(III), Mn(IV), fumarate, nitrate, oxygen, U(VI), ortho-substituted chlorophenols, etc. [87-89]. *Anaeromyxobacter* populations are also major participants in the U(VI), Fe(III), and chlorophenol reduction processes carried out in contaminated sites [90-92]. The chlororespiring activity of strain 2CP-C can be paralleled with constitutive Fe(III) reduction [87], and U(VI) reduction can be affected by the respiration of

nitrate, nitrite, and Fe(III) [88], suggesting the presence of competitive interactions between different terminal electron-accepting processes in the environment. Thus, *A.dehalogenans* strain 2CP-C is an excellent model organism for research on the interference of competing respiratory substrates. In addition, the potential utilization and manipulation of *A.dehalogenans* strain 2CP-C for in situ bioremediation of radionuclides and chlorophenols provides a promising strategy for soil and water quality improvement. The studies on *A.dehalogenans* strain 2CP-C mostly focused on its phylogeny and physiology, and there is a knowledge gap for molecular mechanisms contributing to its respiratory versatility. Further molecular-level exploration is needed to reveal whether *A.dehalogenans* strain 2CP-C has a unique electron transfer system, or a similar strategy is developed as with *Geobacter* (pili strategy) or *Shewanella* (cytochrome strategy). This dissertation provides insight into the proteomic compositions of *A.dehalogenans* strain 2CP-C in Chapter 3 and 4.

1.5 *c*-Type Cytochromes Plays Important Role in Respiratory Versatilities of DMRBs

One common characteristic of most DMRBs is the large number of *c*-type cytochrome encoding genes present in their genomes. For example, in *G. sulfurreducens* genome, 111 genes potentially encode putative *c*-type cytochrome proteins [93], which is the highest number reported by far. Although detailed mechanisms for metal reduction is under investigation, the wealth of *c*-type cytochromes in DMRB is thought to be involved in the reduction of electron acceptors.

c-Type cytochromes are proteins characterized by the covalent attachment of *c*-type heme on the conserved heme-binding motif CXXCH in the protein sequence. Heme is a very common biological cofactor, typically existing as the prosthetic group for a protein. The heme molecule is essentially protoporphyrin IX with iron; and the two vinyl groups of the *c*-type heme molecule are linked to the two cysteine residues in the CXXCH motif via thioether bonds (Figure 1.3) [94]. In prokaryotes, many *c*-type cytochromes contain multiple *c*-type hemes in a single protein, each attaching to a distinct CXXCH motif. The *c*-type heme is thought to contribute to the stabilization of the protein as well as constructing "hardwires" for electron flow between molecules [95-98].

c-Type cytochromes play very important roles in various bacterial respiration pathways [98]. They participate in the electron transfer chain, which typically generates electron flow and the proton gradient across the cell membrane, which can be used for ATP synthesis, providing energy for various cell activities. Energy generation through metal respiration is a key feature for DMRBs as mentioned in the previous chapter. The functional roles of DMRB *c*-type cytochromes during electron transfer and terminal metal reduction for energy conservation is the question that many studies are trying to answer.

The functional roles of many *c*-type cytochromes in *Shewanella* species have been extensively studied, and the molecular basis involving multiple *c*-type cytochromes has been proposed for the extracellular electron transfer process [99]. CymA (a teraheme *c*-type cytochrome) oxidizes menaquinol in the inner membrane and transfers electrons to the periplasm [100-102]. MtrA and MtrB form a porin-cytochrome complex on the outer

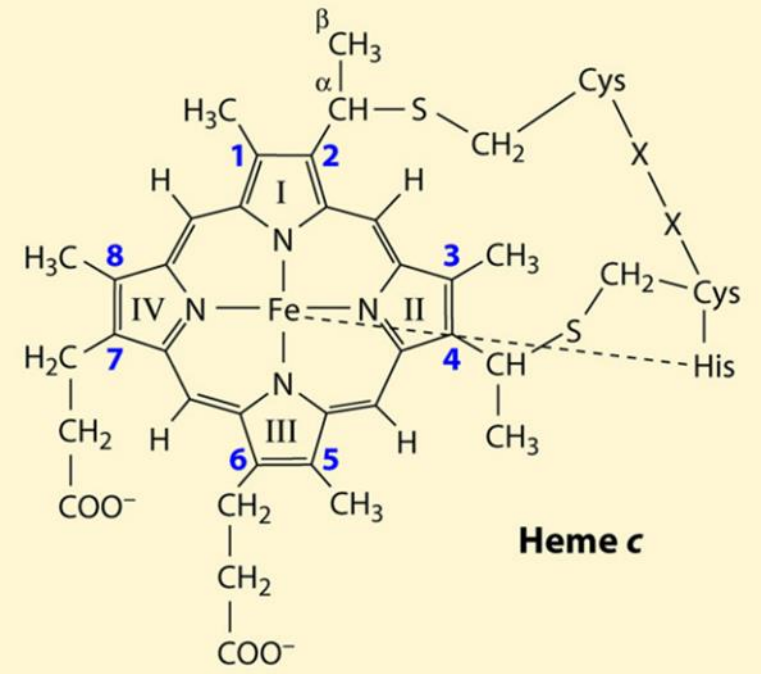


Figure 1.3. The molecular structure of c-type heme covalently binding to the conserved CXXCH motif on a polypeptide chain [94].

membrane that passes electrons from the inside of the outer membrane to the outside [102]. The solid extracellular electron acceptor (e.g. Fe(III) oxide) is then reduced, most likely by receiving electrons from extracellular multi-heme cytochromes MtrC, MtrF, and OmcA [102]. As mentioned in the previous section (1.4), *S. oneidensis* strain MR-1 could produce conductive pili structured "nanowire" possibly relating to metal reduction. And the cytochrome MrtC and OmcA is essential for the conductivity of the "nanowire" [103].

As for *Geobacter* species, the *c*-type cytochrome OmcC was shown to play important role for Fe(III) reduction[104]. Three outer membrane *c*-type cytochromes OmcS, OmcT, and OmcE are important elements for metal oxide reduction but not essential for the respiration of chelated metal species [105]. And the cytochrome OmpJ are required for the reduction of Fe(III) and Mn(IV) [106]. Unlike *S. oneidensis* strain MR-1, studies on the involvement of *c*-type cytochromes to the conductive pili structure of *G. sulfurreducens* showed that electron transfer alongside pili is not mediated by *c*-type cytochromes [77, 78]. Instead of aiding the pili conductivity, cytochrome OmcS is proposed to facilitate the electron transfer from pili to Fe(III) oxide [107]. Multiheme *c*-type cytochromes of *Geobacter* species are hypothesized to function as capacitors which are able to temporarily store electrons when extracellular electron acceptors are not readily available [108, 109].

The specific electron transfer processes for DMRBs are still elusive, but taken together all the studies on various *c*-type cytochromes in different DMRB organisms, it is obvious that the high abundance of diverse *c*-type cytochromes present in all DMRBs is reflecting the highly branched respiratory systems. In this dissertation, the three DMRBs under investigation (*A.*

dehalogenans strain 2CP-C, *G. daltonii* strain FRC-32, and *S. oneidensis* strain MR-1) all have high *c*-type cytochrome gene content in their genome based on genome sequence analysis. And majority of the predicted *c*-type cytochromes contain very high numbers of heme-binding motifs per protein [110, 111]. However, many of the putative *c*-type cytochromes, especially for *A. dehalogenans* strain 2CP-C and *G. daltonii* strain FRC-32, are not validated for their expression, interaction, and specific function. In order to gain insight into the utilization of *c*-type cytochromes during respiration, the comprehensive proteomic profiling of their *c*-type cytochrome repertoires is conducted and presented in Chapter 3.

1.6 MS-based Proteomics Offers Insight into DMRB Studies Unachievable by Other Techniques

The studies about DMRBs mainly have been focused on the ecological and physiological properties, such as their population diversities and abundances in field samples, efficiency in metal reduction, current production, utilization of different electron acceptors and donors, etc. The environmental relevance of metal reduction processes by DMRBs is profound, and a significant number of bacterial species have been recovered and enriched in laboratories from various important environments (e.g. *G. sulfurreducens* and *S. oneidensis* as mentioned earlier). The isolation of DMRBs has led to deeper understanding of bacterial utilization and competition for different electron donors and acceptors, and revealed metabolic processes of individual species contributing to the environmental ecology.

While our knowledge about the physiological and ecological aspects of DMRBs is advanced, more effort needs to be directed toward the elucidation of precise mechanisms involved. With the advent of high throughput next-generation sequencing leading to the bloom of microbial whole genome sequencing, the complete genetic makeup for many isolated DMRBs are easily accessible. Functional predictions for DMRB genes were made through comparative genome analysis, and revealed many unprecedented potential metabolic features. For example, the genome of *G. sulfurreducens* and *A. dehalogenans* suggested motility function and aerobic metabolism, both of which were subsequently validated experimentally [93, 111, 112]. Among all the identified DMRBs, the complete genomes for *S. oneidensis* and *G. sulfurreducens* were the earliest available [93, 113]. Thus, the understanding about the molecular basis for respiratory activities of *S. oneidensis* and *G. sulfurreducens* are the most extensive due to well-constructed genetic systems [114, 115]. With the available genetic systems, progress has been made in identifying specific genes that are essential for selected biological processes using gene manipulation strategies, and many genes responsible for specific cellular functions were characterized. For example, some *c*-type cytochromes are essential for metal reduction, as mentioned in the previous section. However, mutant strain generation is technically difficult and time consuming. In addition, disruption of a particular gene often does not provide discernable phenotype, as exemplified by the mutagenesis studies on *S. oneidensis* and *G. sulfurreducens* [116, 117]. The metal respiration ability of *S. oneidensis* and *G. sulfurreducens* is only hampered and not eliminated by the disruption of a few *c*-type cytochrome genes [116, 117]. More comprehensive interrogations to the DMRB cellular activities and the detailed mechanisms are needed.

In the post-genomics era, transcriptomics and proteomics enable the functional overview of gene expression. Several transcriptomics and proteomics studies have been done on *S. oneidensis* and *G. sulfurreducens* [118]. The microarray analysis on *S. oneidensis* mRNAs identified 121 genes showing differential expression changes growing with different electron acceptors [119]. The gene *fur* of *S. oneidensis*, a homolog to the iron uptake regulator in *E. coli*, was identified to be an important element in the regulation of energy metabolism and iron assimilation through transcriptomics and gel-based proteomics analysis [120, 121]. As discussed in section 1.1 and 1.2, transcriptomics is one step away from the final gene product – proteins, and comprehensive proteomics interrogation could provide the direct systematic functional overview for cellular activities. The unfractionated MS-based proteomic studies on *S. oneidensis* explored the proteome expression, and provided experimental evidence for the expression of a significant proportion of the uncharacterized hypothetical proteins [122-124]. Proteomics characterization also provided insight into the global cellular activities of *S. oneidensis* and *G. sulfurreducens* toward different growth environments, stress conditions, etc. [125-128]. In all, characterization of the molecular metabolic activities provided by proteomics analysis could not be achieved through any other techniques. While *S. oneidensis* and *G. sulfurreducens* have been extensively explored, many other DMRBs with significant environmental influences should also be interrogated. The MS-based proteomics analysis in this dissertation provided the metabolic overview for *A. dehalogenans* strain 2CP-C, *G. daltonii* strain FRC-32, and *S. oneidensis* strain MR-1, and comparative analysis revealed differential proteome expression profiles as well as differential *c*-type cytochrome utilizations.

1.7 Overview of This Dissertation

This dissertation is focused on the development and application of a high-performance MS-based proteomics technique to enhance the understanding of environmental microbiology. The particular focus of the research presented in this dissertation is to comprehensively characterize the proteomes of three different species of DMRBs in an effort to reveal the molecular mechanisms underlying the cellular metabolic activities in relation to their respiratory versatilities.

Chapter 2 will provide the in-depth proteomics experimental design for achieving comprehensive DMRB proteome characterizations, including details related to cell culture, proteome sample preparation methods, mass spectrometry instrumentation and experimental setup, protein identification strategy, and data normalization. Chapter 3 will demonstrate deep protein expression profiling for the putative *c*-type cytochromes in *A. dehalogenans* strain 2CP-C, *G. daltonii* strain FRC-32, and *S. oneidensis* strain MR-1, providing the accurate details about the utilization of different sets of *c*-type cytochromes in response to different electron acceptors. The systematic proteome characterizations for *A. dehalogenans* strain 2CP-C will be presented in Chapter 4, which compares the regulation of different metabolic pathways toward the respiration of different electron acceptors. Chapter 5 focuses on the comprehensive proteomics characterization for *G. daltonii* strain FRC-32 and *S. oneidensis* strain MR-1 under different electron-accepting conditions. And the proteome expression and represented metabolic pathways of *G. daltonii* strain FRC-32 and *S. oneidensis* strain MR-1 is compared and contrasted to the proteome detection of *A. dehalogenans* strain 2CP-C. Chapter 6 will conclude

the scientific achievements throughout this dissertation and the contributions made to the environmental microbiology field. The current status and on-going development of environmental microbiology through MS-based proteomics will also be discussed in Chapter 6.

Chapter 2

Experimental Design, Instrumentation and Data Analysis

2.1 Experimental Overview

The shotgun proteomics approach was employed for studies presented throughout this dissertation (Figure 2.1). Since the focus of this dissertation was on the proteomic response of DMRBs to the utilization of different electron acceptors, the overall workflow of the shotgun proteomics approach started with sample collection which obtained cells grown under various electron-accepting conditions. The next step was to effectively extract the proteome from the collected cell pellets, which is crucial to present the comprehensive repertoire of proteins for the following mass spectrometry measurement. In order to achieve efficient cell lysis and whole proteome preservation and extraction, unfractionated sample preparation methods employing detergent or chaotropic agent were chosen. Next, extracted proteins were enzymatically digested into peptides using trypsin, and the tryptic peptide mixtures were separated using multidimensional high performance liquid chromatography (HPLC), which was directly interfaced with a mass spectrometer. The separated peptides were brought into gas-phase through ionization by electrospray, and the mass to charge values (m/z) and relative intensities of the peptide ions were measured in the mass analyzer and detector, generating full spectra (MS1). Peptide ions representing the most abundant peaks in the full spectra were selected for fragmentation by collision induced dissociation (CID), and the fragmented ions where measured for their m/z values and relative intensities, generating the tandem mass spectrum (MS/MS). Peptide sequences were determined using the spectrum matching algorithm

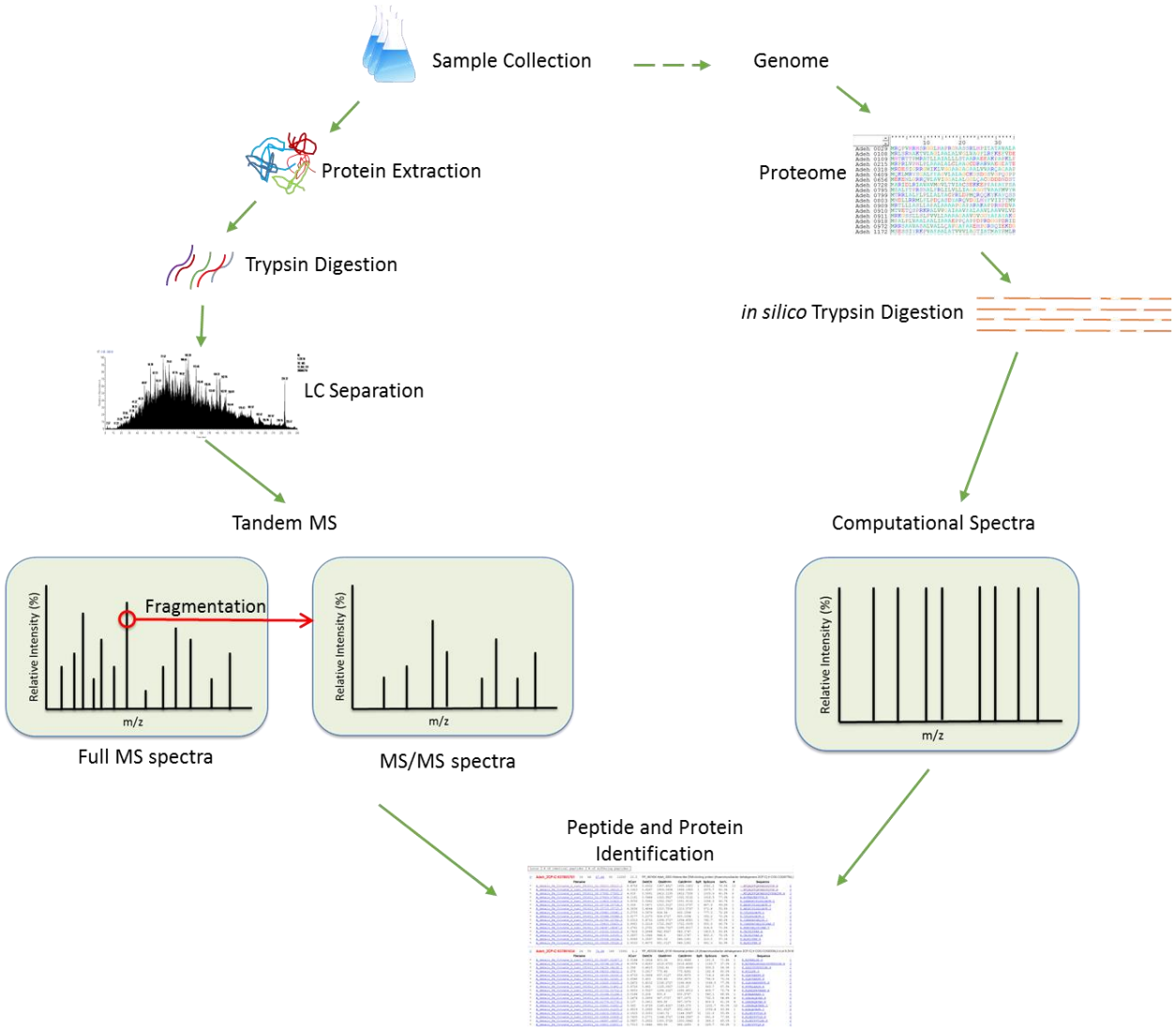


Figure 2.1. Shotgun proteomics workflow.

SEQUEST, which utilized predicted peptide MS/MS spectra to infer the measured peptide sequences. Protein identification was achieved by matching the identified peptide sequences to the corresponding protein sequences, and spectral counts were used as the measure for protein abundances. Using the shotgun proteomics approach outlined above, the proteomes of tested DMRBs were characterized and quantified. The overall experimental design and specific techniques used for each experimental step were chosen with careful considerations. Technical details for each step are provided in the following section.

2.2 Sample Collection

Three dissimilatory metal reducing bacterial strains were tested throughout this dissertation, including *Anaeromyxobacter dehalogenans* strain 2CP-C (ATCC BAA-259), *Geobacter daltonii* strain FRC-32 (DSM 22248), and *Shewanella oneidensis* strain MR-1 (ATCC BAA-1096). Cell culture growth and collection was conducted by our collaborator in Dr. Frank Löffler's lab in The University of Tennessee, Knoxville. For shotgun LC-MS/MS experiments on microbial isolates, typically >1mg cell pellets contain sufficient protein amounts for accurate detection. To minimize biological variations, the tested strains were grown in triplicates, which also provided the convenience to perform statistical analysis after experimental measurements. Generally, technical replications are achieved by repetitive MS measurements for the same sample, which counteracts the technical variations introduced by separate experiments. Due to the large sample set tested in this dissertation, technical replications were performed on several samples and demonstrated high qualities of replication.

Each bacterial strain was cultured independently in biological triplicates in 160 ml glass serum bottles containing 100 ml reduced bicarbonate-buffered mineral salts medium with a N_2/CO_2 headspace (80:20, vol/vol) as described previously [129, 130]. The mineral salts in the synthetic medium are as following (per liter): NaCl, 1g; $MgCl_2 \cdot 6H_2O$, 0.5g; KH_2PO_4, 0.2g, NH_4Cl; 0.3g, KCl, 0.3g; $CaCl_2 \cdot 2H_2O$, 0.015g; resazurin, 1mg; trace element solution A, 1ml; trace element solution B, 1 ml; $Na_2S \cdot 9H_2O$, 0.048g; L-cysteine, 0.035 g; $NaHCO_3$, 2.52 g; vitamin solution [131], 10 ml; pyruvate, 20 mM; and 3 Cl-4-HBA,1mM. Trace element solution A contained the following (per liter): HCl (25% [wt/wt] solution), 10 ml; $FeCl_2 \cdot 4H_2O$, 1.5 g; $CoCl_2 \cdot 6H_2O$, 0.19 g; $MnCl_2 \cdot 4H_2O$, 0.1 g; $ZnCl_2$, 70 mg; H_3BO_3, 6 mg; $Na_2MoO_4 \cdot 2H_2O$, 36 mg; $NiCl_2 \cdot 6H_2O$, 24 mg; and $CuCl_2 \cdot 2H_2O$, 2 mg. Trace element solution B contained (per liter): Na_2SeO_3, 6mg; $Na_2WO_4 \cdot 2H_2O$, 8 mg; and NaOH, 0.5 g. The serum bottles were closed with butyl rubber stoppers (Geo-Microbiol Technologies, Inc., Ochelata, OK) and crimped with aluminum seals (Wheaton, Millville, NJ). Different growth conditions for each bacterial strain were generated by adding different electron donors and acceptors into the growth medium, details of which are provided below.

A total of eight different growth conditions with various electron acceptors were tried for each bacterial strain, designated as “FeOOH”, “Fe citrate”, “MnO_2”, “fumarate”, “nitrate”, “tryptic soy broth”, “oxygen”, and “N_2O” (Figure 2.2). Cell growth of strain 2CP-C was achieved with all eight culture conditions, whereas seven conditions were tested for strain MR-1, and strain FRC-32 could only grow under four tested conditions (Figure 2.2). Acetate (5 mM) or lactate (5 mM) was provided as electron donors for all conditions except “tryptic soy broth”. For the “FeOOH”, “Fe citrate”, “MnO_2”, “fumarate”, and “nitrate” conditions, the following

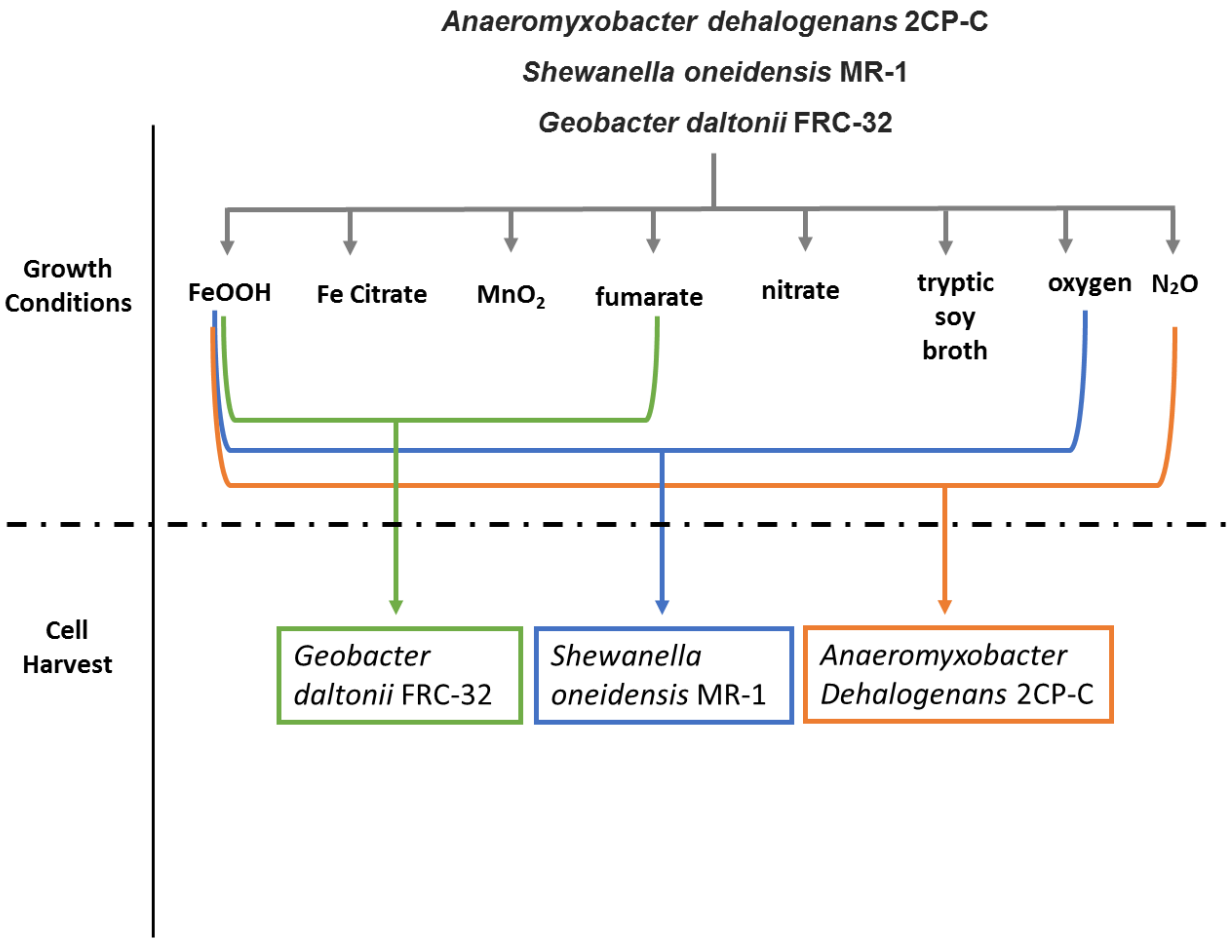


Figure 2.2. Cell growth conditions tested for *A. dehalogenan* 2CP-C, *S. oneideisns* MR-1, and *G. daltonii* FRC-32.

electron acceptors were added, respectively: ferric oxyhydroxide (20 mM, nominal), ferric citrate (50 mM), manganese oxide (20 mM, nominal), fumarate (10 mM), and nitrate (2 mM). For the "tryptic soy broth" condition, 1/10 diluted tryptic soy broth medium was used in place of the mineral salt medium, providing undefined electron donor(s), and fumarate (10 mM) was added as the electron acceptor. When oxygen was used as electron acceptor for the "oxygen" condition, the medium volume was reduced to 50 ml, and 3% (vol/vol) oxygen was added in the headspace as electron donor, by injecting the appropriate volume of filter-sterilized air into the serum bottles. As shown in Figure 2.1, the "N_2O" growth condition was achieved only for strain 2CP-C. In the "N_2O" condition, strain 2CP-C was grown in the synthetic medium with Na_2S omitted, acetate (5mM) provided as electron donor, and 2ml of 99% N_2O gas (Aldrich) provided as electron acceptor by sterile filtering the gas into the serum bottle. Consecutive spikes with N_2O gas were carried out a total of 10 times. The consumption of N_2O was determined by measuring the headspace of the cultures with a 3000A MicroGC gas analyzer (Agilent Technologies; Column: Plot Q; Carrier gas: He). Briefly, a 1 ml headspace sample was taken under sterile and anoxic condition and manually injected into the 3000A MicroGC.

In each culture, cells were grown at 30°C, shaken at 90 rpm until mid to late exponential phase, as determined by OD (optical density) and qPCR (for fumarate growth) metrics, and the entire volume was harvested by centrifugation (4,700 x g for 20 min at 4°C). Cell pellets were instantly frozen and stored at -80°C.

2.3 Cell Lysis and Proteome Extraction

Once bacterial cells are collected, the next step is to lyse the cells and extract all the proteins to prepare proteome samples for mass spectrometry interrogation. Two types of proteome preparation methods were used throughout this dissertation: 1) SDS-TCA method, and 2) Guanidine-HCl method (Figure 2.3). Both methods are effective in cell lysis and total proteome extraction.

2.3.1 The SDS-TCA Method

The SDS-TCA method was originally developed for improved proteome extraction from complex environmental soil microbes [132]. The use of the strong detergent sodium dodecyl sulfate (SDS), together with heat, was proven to be efficient in cell lysis and provides effective recovery of highly hydrophobic proteins such as proteins located in the cell membrane [32, 133, 134]. Since a major focus of DMRB research is on the *c*-type cytochrome proteins, most of which are predicted to be membrane-associated or periplasmic, the SDS-TCA method was chosen to provide effective proteome extraction including the *c*-type cytochrome proteins.

For cells grown with Fe citrate, MnO_2, fumarate, nitrate, tryptic soy broth, oxygen, and N_2O, the SDS-TCA method was applied (Figure 2.3). The cell pellet collected was re-suspended in 1.5 ml cell lysis buffer (5% (w/v) SDS, 50 mM Tris-HCl, pH 8.5; 0.15 M NaCl; 0.1 mM ethylenediaminetetraacetic acid (EDTA); 1 mM $MgCl_2$; 50 mM dithiothretiol (DTT)), and subjected to a boiling water bath for 10 min to achieve cellular lysis. Cell lysate was collected after centrifugation (21,000 x g for 15 min) and chilled 100% trichloroacetic acid (TCA) was added to a final concentration of 25% (v/v) and samples were incubated overnight at -10°C.

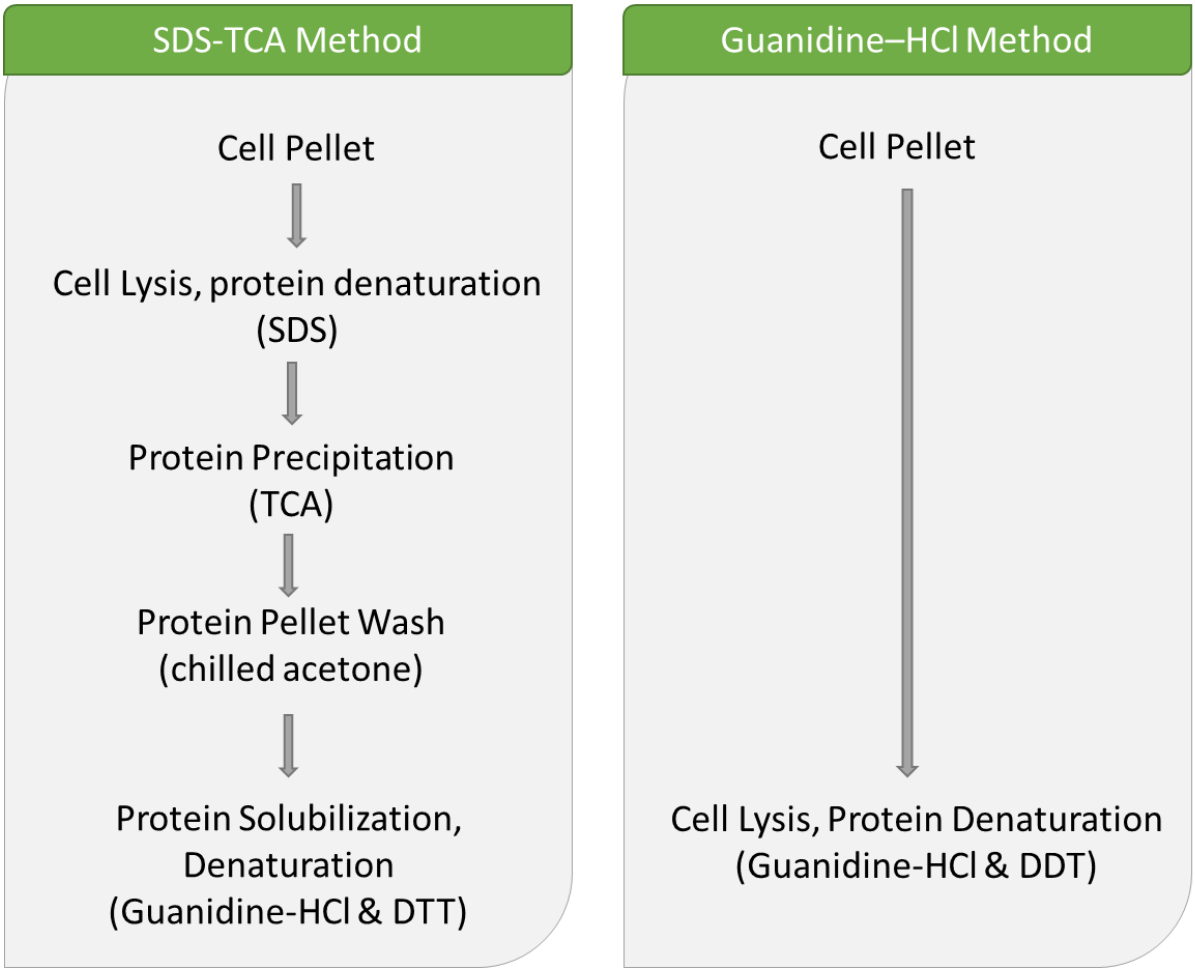


Figure 2.3. Cell lysis and proteome extraction methods. The SDS-TCA method (left) uses detergent SDS for cell lysis and protein denaturation, which provides effective extraction for hydrophobic proteins. The Guanidine-HCl method (right) directly use guanidine-HCl solution for cell lysis and protein denaturation, and contains less preparation steps which reduce the chance of sample loss.

Proteins were precipitated by centrifugation at 21,000 x g for 20 min, and the supernatant was discarded. Cell pellet were washed thrice with chilled acetone and centrifuged at 21,000 x g for 10 min after each acetone wash. Protein pellets obtained, were air-dried and solubilized in guanidine/DTT buffer (6 M guanidine HCl, 10 mM DTT in Tris $CaCl_2$ buffer (50 mM Tris, 10 mM $CaCl_2$, pH 7.6)) and incubated at 40°C for 4 hrs prior to proteolytic digestion.

Attempt to use the SDS-TCA method for FeOOH grown cultures were made, but did not yield effective proteome extraction. We suspect that after cell lysis and protein denaturation by SDS and heat, proteins adsorbed onto the FeOOH solid particles and did not present in the cell lysate collected through centrifugation. Thus, the Guanidine-HCl method described below was applied on cells grown with FeOOH.

2.3.2 The Guanidine-HCl Method

The Guanidine-HCl method has been the traditional approach for microbial isolate proteome sample preparation, which has achieved many successes for the comprehensive microbial proteome characterizations [35, 57, 135]. As mentioned above, in this dissertation, the Guanidine-HCl was used for the proteome extraction of FeOOH grown cells.

In this approach, cell lysis and protein denaturation and reduction was achieved simultaneously by re-suspending collected cell pellets in 2 ml guanidine/DTT buffer (6 M guanidine HCl, 10 mM DTT in Tris $CaCl_2$ buffer (50 mM Tris, 10 mM $CaCl_2$, pH 7.6)) with vigorous vortex. Cell lysate was incubated at 40°C for 4 hrs prior to direct proteolytic digestion.

Comparing to the SDS-TCA method, the Guanidine-HCl method contains fewer steps, and does not involve centrifugation process, which greatly reduces the chance of protein loss during sample preparation steps. For the FeOOH grown cultures, effective proteome recovery was achieved though the Guanidine-HCl as shown by the results in Chapter 3, 4, and 5.

2.4 Protein Digestion

Following protein extraction using either the SDS-TCA or Guanidine-HCl methods (described above), the protein solution was diluted 6-fold using Tris $CaCl_2$ buffer, pH 7.6 in order to reduce the guanidine concentration, and proteolytic digestion was performed using sequencing-grade trypsin (Promega, Madison, WI), which enzymatically cleaves at the C-terminal side of lysine and arginine residues unless followed by a proline residue. Generally the spacing of lysine and arginine in the majority of proteins could generate tryptic peptides with amenable length for MS analysis. Tryptic peptides generally carry a positive charge on both the N- and C-terminals, could be separated by liquid chromatography, and are able to get ionized efficiently by electrospray. Prior to trypsin addition, the total concentration of proteins was quantified using the bichinchoninic acid (BCA) assay [136]. And optimal amount of trypsin was used according to the obtained protein amount (trypsin:protein = 1:100 (w/w)). To ensure effective tryptic digestion, half of the optimal amount of trypsin was added with overnight incubation at 37°C, followed by a second addition of the remaining amount of trypsin and incubation at 37°C for 4 hrs.

After proteins are enzymatically digested into peptides, the peptide solution was desalted using C-18 solid-phase extraction (SepPak, Waters, Milford, MA), and was concentrated and solvent exchanged into acidified water (0.1% formic acid in HPLC grade water; v/v) by vacuum centrifugation using a Savant SpeedVac instrument (ThermoFisher Scientific, Waltham, MA). All peptide samples were filtered by Durapore PVDF filters, 0.45 µm (Millipore) and stored at -80 °C before analysis by mass spectrometry.

2.5 Liquid Chromatography

After trypsin digestion, the sample complexity is significantly increased, in that each extracted proteins are digested into multiple peptides. To resolve the highly complex peptide mixture, high-performance liquid chromatography (HPLC) was employed to separate numerous peptide species before interrogation by mass spectrometer. The separation strategy is critical for good quality mass spectrometry proteomics characterization. For complex mixture of cell proteome, multidimensional HPLC that employs orthogonal separations are needed to provide strong resolving power [137, 138]. In order to improve peptide separation power, the multidimensional protein identification technology (MudPIT) was applied to all proteome samples in this dissertation [139].

Two-dimensional liquid chromatography is utilized in the MudPIT method, through which, peptides are separated by charge and hydrophobicity. Using this approach, strong cation-exchange (SCX) resins and reverse-phase resins are integrated in a microcapillary chromatographic column. While peptides are separated by charge during SCX chromatography,

reserve-phase offers peptides separation by hydrophobicity. As illustrated in Figure 2.4, a “back column” is built by packing a fused silica microcapillary column (150 µm inner-diameter and 360 µm outer-diameter; Polymicro Technologies; Phonex, AZ) with ~5 cm SCX resins (Luna 5µm particle size; 100 Å pore size; Phenomenex, Torrance, CA). Peptide samples were loaded onto the “back column” using a pressure pump. After pressure-loading, peptides were bond to the SCX resins through ionic interaction, and the “back column” was subject to an offline-wash by connecting to a quaternary HPLC pump (Ultimate 3000 HPLC, Dionex, Sunnyvale, CA) which flow through the “back column” with a 15 min gradient from low to high organic (100% solvent A to 50% solvent B) to remove any lingering salt and SDS. Solvent A contained 95% H_2O, 5% acetonitrile, and 0.1% formic acid (vol/vol). Solvent B was highly organic, which contained 30% H_2O, 70% acetonitrile, 0.1% formic acid (vol/vol). Then the “back column” was equilibrated using a 5 min gradient from 50% solvent B to 100% solvent A, followed by another 5 min flow of 100% solvent A.

After the offline wash and equilibration, the peptides-loaded “back column” was interfaced to a “front column” via a PEEK union and a 0.5 µm inline filter (Upchurch Scientific, Oak Harbor, WA). The “front column” contained a pointy tip serving as the nanospray emitter (15±1 µm tip diameter, 100 ID, New Objective, Woburn, MA), and was packed with 12-15 cm C-18 reverse-phase resins (Aqua 5 µm particle size; 125 Å pore size; Phenomenex, Torrance, CA) using a pressure pump (Figure 2.4).

In this dissertation, each peptide sample was subject for an eleven-step chromatographic separation strategy, with each step lasting for around two hours. During each

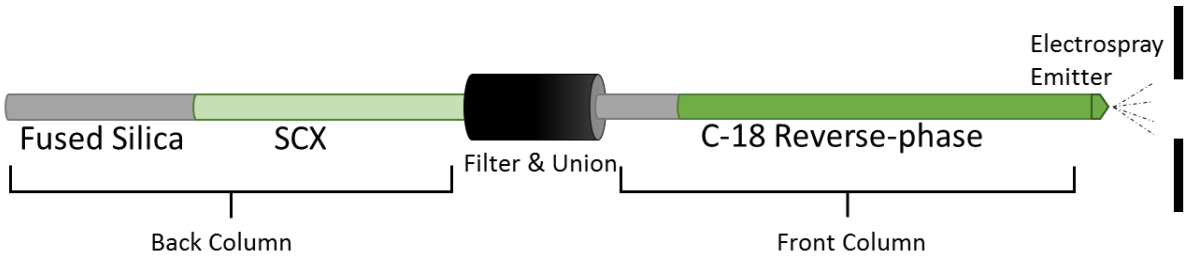


Figure 2.4. Schematic diagram of a MudPIT experiment setup. A back column is constructed by packing a fused silica microcapillary tube with strong cation exchange (SCX) material. The front column contains a pointy tip which functions as the electrospray emitter. The C-18 reserve-phase material is packed into the front column. The front and back columns are connected by filter and union.

of the first ten steps, the ionic interaction between the SCX resins and the binding peptides was challenged by a short (typically 5 min) salt pulse (solvent C: 500 mM ammonium acetate in solvent A), causing a proportion of the bonded peptides eluting from the SCX section of the "back column" into the C-18 resins in the "front column". A long gradient (roughly 100 min) from 100% solvent A to 50% solvent B was applied after the salt disruption to further separate peptides in the reserve-phase by hydrophobicity. The salt pulse concentration was increased through each incremental step, resulting in peptides with different charge states separated during each step. From step one to ten the following concentration of solvent C was applied respectively: 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%. In the eleventh step, no salt pulse was introduced, and the gradient started from 100% solvent A to 100% solvent B. The HPLC separation was directly coupled online with the following electrospray and mass spectrometry analysis. As peptides eluting off the "front column" over the course of the chromatographic separation, the separated peptides were subsequently ionized and introduced into the mass spectrometer.

2.6 Ionization

Ionization techniques made it feasible to bring large, non-volatile biological molecules into gas-phase for mass spectrometer measurement. Two types of "soft ionization" methods were most used for proteomics analysis: matrix-assisted laser desorption ionization (MALDI) [140] and electrospray ionization (ESI) [27]. For MALDI ionization, the analyte (peptides or proteins) are fixed in a substrate or "matrix", and a laser beam is used to trigger desorption of

the analyte leading to ionization. And the ions generated through MALDI are generally singly charged. In contrast, ESI is performed by applying a strong electric field to an aqueous solution flowing through a capillary needle, and the proteins or peptides in the liquid are directly ionized through protonation. Thus, liquid chromatography separation could be readily interfaced with ESI, and the real-time separation, ionization and MS measurement could be integrated for peptides solution analysis. The ions generated by ESI are generally multiply charged, allowing the *m*/*z* (mass to charge ratio) of large molecules to be accommodated by the mass range (the highest and lowest *m*/*z* that could be measured) of the mass analyzer. Thus, ESI was chosen for peptide ionization in this dissertation.

During ESI, the flow rate of sample solution is kept low (generally below 10 µl/min). Currently, the most common ESI is the nano-electrospray (nano-ESI) which has a typical flow rate of 20-50 nl/min, and generates droplets less than 200 nm in diameter [141]. The advantages of nano-ESI, compared to conventional ESI, include less sample consumption, longer stable signal for accurate measurement, better ion desolvation, and higher tolerance to high aqueous solvents and salt contamination [142-144]. In this dissertation, nano-ESI is employed for all MS experiments.

During the nano-ESI process, the sample solution slowly flowing through the capillary tube is sprayed out from the needle emitter under the strong electric field. And the following process through which peptides are transferred from the liquid phase to the gas phase is considered to have three steps: 1) generation of charged droplets from the high-voltage electrospray emitter; 2) evaporation of solvent in the charged droplet and disintegration of

large droplets into smaller droplets; 3) gas-phase ion forming from the small droplets (Figure 2.5). As the liquid eluting out from the tip of the emitter, a cone-shaped spray (termed: Taylor cone) is formed due to surface tension and electric force [145]. Subsequently, the Taylor cone is dispersed into a plume of charged droplets under the influence of the strong electric field. The charged droplets could each contain tens of thousands of charged ions with like charges distributing on the surface of the droplet [146]. While surface tension tries to maintain the spherical shape of the droplets, the like charges on the droplet surface cause Coulomb force of repulsion which tries to break down the spherical shape [146-149]. As charged droplets move in the electric field toward the heated capillary, the solvent starts to evaporate from heating, and the size of the droplets shrink. Eventually, the surface tension of the shrinking droplet could no longer combat the Coulomb force of repulsion, reaching the "Rayleigh limit" [150], and leading to the "Coulomb fission" where large droplets break down into smaller charged droplets. The solvent evaporation and Coulomb fission repeatedly occur, which generate more and more small offspring droplets with reduced size. In the end, ions are brought into gas-phase from the droplet. The exact mechanism for gas-phase ion formation is under debate. Two prevailing hypothesis are the "charge residue model" and the "ion evaporation model". The charge residue model proses that the solvent evaporation and Coulomb fission repeatedly occur until only one ion is present in the droplet, and the ion is released by droplet desolvation [151-153]. On the contrary, the ion evaporation model suggests that direct ion emission occur under the electric field when the droplet diameter is below 10 nm [154, 155].

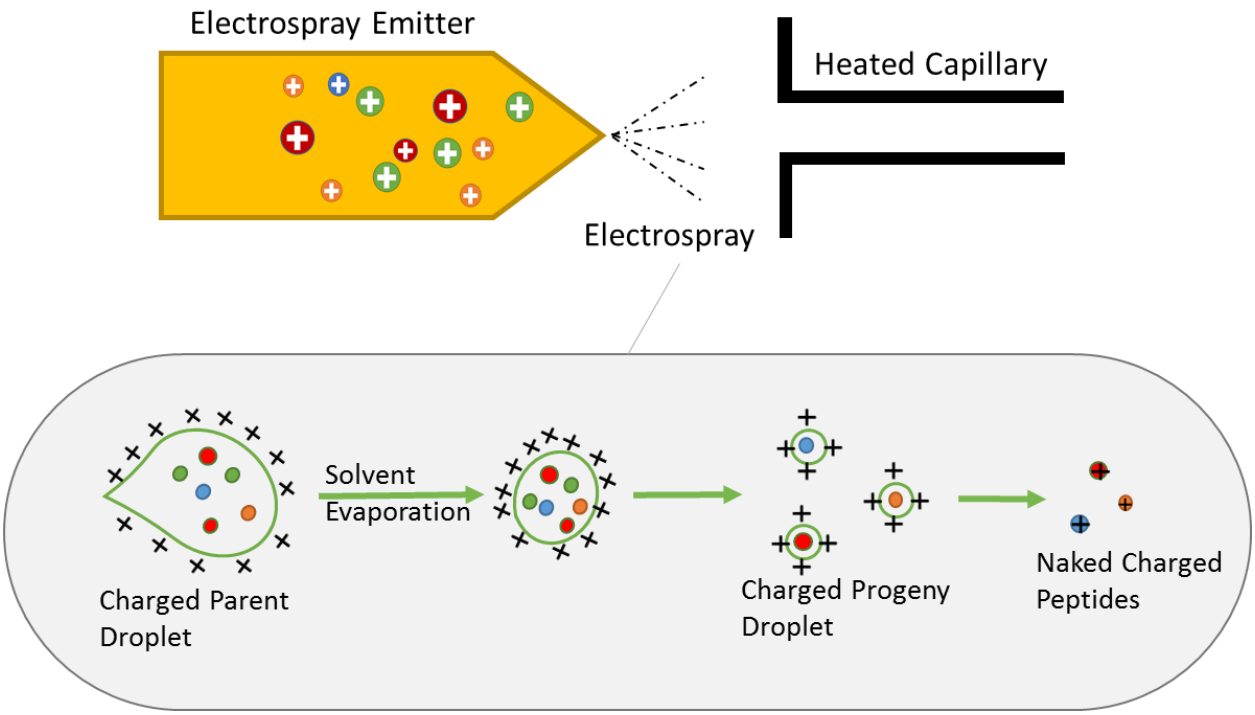


Figure 2.5. Illustration of the electrospray ionization process.

2.7 Mass Analyzer and Detector

There are three major components in a mass spectrometer instrument: the ion source (discussed above in 2.6), mass analyzer, and detector. After ions are brought into the gas-phase in the ion source, the mass analyzer resolves the *m*/*z* (mass to charge ratio) values of the ions, and detector measures the signal intensity. There are various types of mass analyzer and detector developed, and each has its unique figures of merit: 1) mass accuracy: the *m*/*z* measurement error divided by the expected *m*/*z* value, usually measured in parts-per-million (ppm); 2) resolving power: the ability to distinguish two peaks of close *m*/*z* values, typically measured in FWHM (full width of the peak at half its maximum height); 3) dynamic range: the signal ratio of the most abundant and least abundant detections; 4) mass range: the range between the highest and lowest *m*/*z* values that could be measured; 5) sensitivity: the ability to detect the increase/decrease of signal intensity from the increase/decrease of the analyte concentration; 6) speed: reflected by the number of collected spectra in a given time frame. The analytical figures of merit are very important features for mass spectrometer instruments, and should be carefully considered when choosing the appropriate instrument for desired experiments. For example, comprehensive characterization of proteomes requires instrument with high speed and sensitivity, such as that of an ion trap mass spectrometer. In contrast, high resolution is more important for intact protein measurement, which requires protein isotopes resolution achieved by the chosen instrument (e.g. FR-ICR).

The mass spectrometer instrument used in this dissertation is a two-dimensional linear ion trap quadrupole, LTQ-XL (Thermo Scientific, Inc.) [156]. The linear ion trap consists an array

of four hyperbolic metal rods. Two types of RF (radio frequency) voltages are applied on the metal rods: DC (direct current) and AC (alternating current). By adjusting the DC and AC voltages, ion movements could be manipulated in an axial fashion in the center space between the rods. The linear ion trap could selectively isolate ions with desired *m*/*z*, and can eject ions to the detectors placed on the sides of the ion trap. The ion trapping volume, high sensitivity, and fast speed of LTQ-XL made it widely used as the optimal platform for the proteomics characterization for microbial isolates.

Another type of mass analyzer is the orbitrap [157], which could be combined with linear ion trap for improved mass accuracy and resolution. In contrast to the utilization of RF voltages for ion confinement, the orbitrap uses a central spindle-shaped electrode and an outer barrel-shaped electrode to trap ions in between. The trapped ions orbit the central electrode, and the *m*/*z* values are measured through the oscillation frequencies [158]. The LTQ-Orbitrap-XL (Thermo Scientific, Inc.) is a hybrid mass spectrometer which contains a LTQ mass analyzer and an orbitrap analyzer (Figure 2.6). The analytical figures of merit for LTQ-XL and LTQ-Orbitrap-XL are presented in Table 2.1. When performing tandem MS measurement (introduced below in 2.7), the full scan is typically generated through the orbitrap which resolves highly complex peptide ion mixtures, and the tandem scan for peptide fragments with lower complexity is generally performed in the linear ion trap with high scan speed.

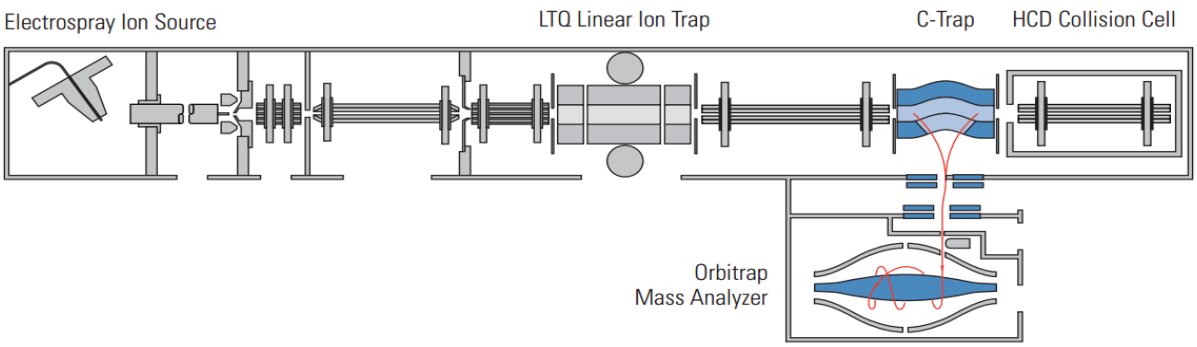


Figure 2.6. Schematic of the hybrid mass spectrometer LTQ-Orbitrap-XL. Two mass analyzers, the LTQ linear ion trap and the orbitrap, are combined for advanced resolution, accuracy, sensitivity and scan speed. Image obtained from www.thermoscientific.com.

Table 2.1. Analytical figures of merit comparison for LTQ-XL and LTX-Orbitrap-XL.

	Mass Accuracy	Resolution (FWHM)	Mass Range (m/z)	MS/MS Sensitivity
LTQ-XL	0.1 Da	1000-2000	15-200; 50-2000; 200-4000	femtomole
LTQ-Orbitrap-XL	<1 ppm	7500-over 100,000 at m/z 400	50-2000; 200-4000	attomole - femtomole

2.8 Tandem Mass Spectrometry

In this dissertation, the tandem mass spectrometry (MS/MS) technique is used for proteomics identification. When the peptides are separated and ionized, their *m*/*z* values (mass to charge ratio) and intensities could be measured by the mass spectrometer, generating the full scan mass spectrum (MS1 spectrum). The measured peptide ions could be selected for gas-phase fragmentation inside the mass spectrometer, and *m*/*z* values and intensities of the resulting fragments could be measured, providing information for the peptide sequence. The ions selected for fragmentation are called the parent ions or precursor ions, and the spectra corresponding to the fragmented ions are the tandem spectra (called MS/MS or MS2 spectra).

Many fragmentation strategies are available for peptide analysis, such as collision induced dissociation (CID) [159], electron capture dissociation (ECD) [160], electron transfer dissociation (ETD) [161], heated capillary dissociation (HCD) [162], multi-stage activation (MSA) [163], pulsed-Q dissociation (PQD) [164], etc. For large scale peptide sequencing, the CID approach is most commonly used. During CID, the selected precursor ion is accelerated and collided with neutral inert gas molecules (e.g. helium, nitrogen, or argon), and the kinetic energy gained is redistributed as internal vibrational energy within the molecule causing its dissociation [165]. The fragmentation primarily occurs along the peptide backbone, and three types of the bonds are generally cleaved: the NH-CH, CH-CO, and CO-NH bonds. The charge of the parent ion could be retained on either the C-terminus or N-terminus. Depending on the cleavage position and charge retention, fragmented ions are classified into a-, b-, c-, x-, y-, z-

types (Figure 2.7). The polypeptide chain fragmented by CID could break at any of the amide peptide bonds, and generally b- and y- ions are the most observed fragment ion types [166].

Since the full scan measures thousands of ions with different *m*/*z* values at different intensities, it is time-consuming to fragment all the ions present in MS1. Thus, only the most abundant peaks in the MS1 are generally selected for fragmentation. To avoid repeatedly sampling the same abundant ions with high intensities in consecutive MS1 scans, the dynamic exclusion is applied, which prohibits the fragmentation of a measured parent ion within a time period. For proteomics measurement of complex protein or peptide mixtures, the dynamic exclusion allows the high abundance and low abundance analytes to be measured.

2.9 Informatics

Once the experimental MS spectra are collected, peptide sequence identification could be achieved by searching a proteome database using computational algorithms. Several search algorithms have been developed, such as SEQEUST [42], Myrimatch [44], DBDigger [167], etc. In this dissertation, all protein identifications were done by the spectrum-matching approach (Figure 2.8) achieved by database searching using the SEQUEST algorithm.

During the SEQUEST search, all the collected MS/MS spectra are processed to remove the low abundant signals and noise. Only peaks with the top two hundred relative abundances are retained, in order to ensure overall high quality of the spectra for better search accuracy. On the other hand, SEQUEST performs *in silico* trypsin digestion to all the proteins contained in

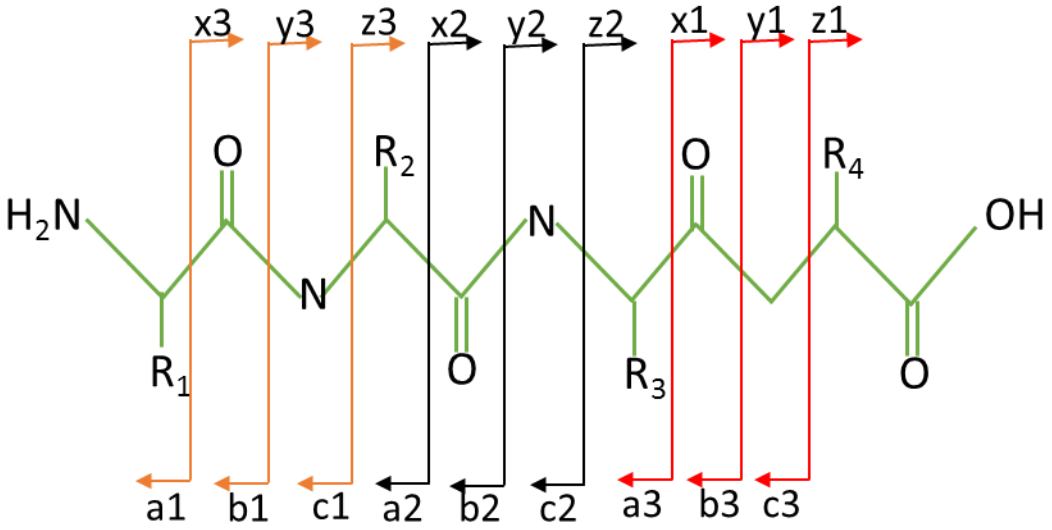


Figure 2.7. Schematic of the peptide ion fragmentation. Different types of fragment ions are generated depending on the charge retention. When charge is retained on the N-terminus, a-, b-, and c- types of ions are generated. And when charge is retained on the C-terminus, x-, y-, and z- type ions are generated. The fragment ions are named numerically from N-terminus to C-terminus for a-, b-, and c-type ions, and numerically reversed for x-, y-, and z- type ions.

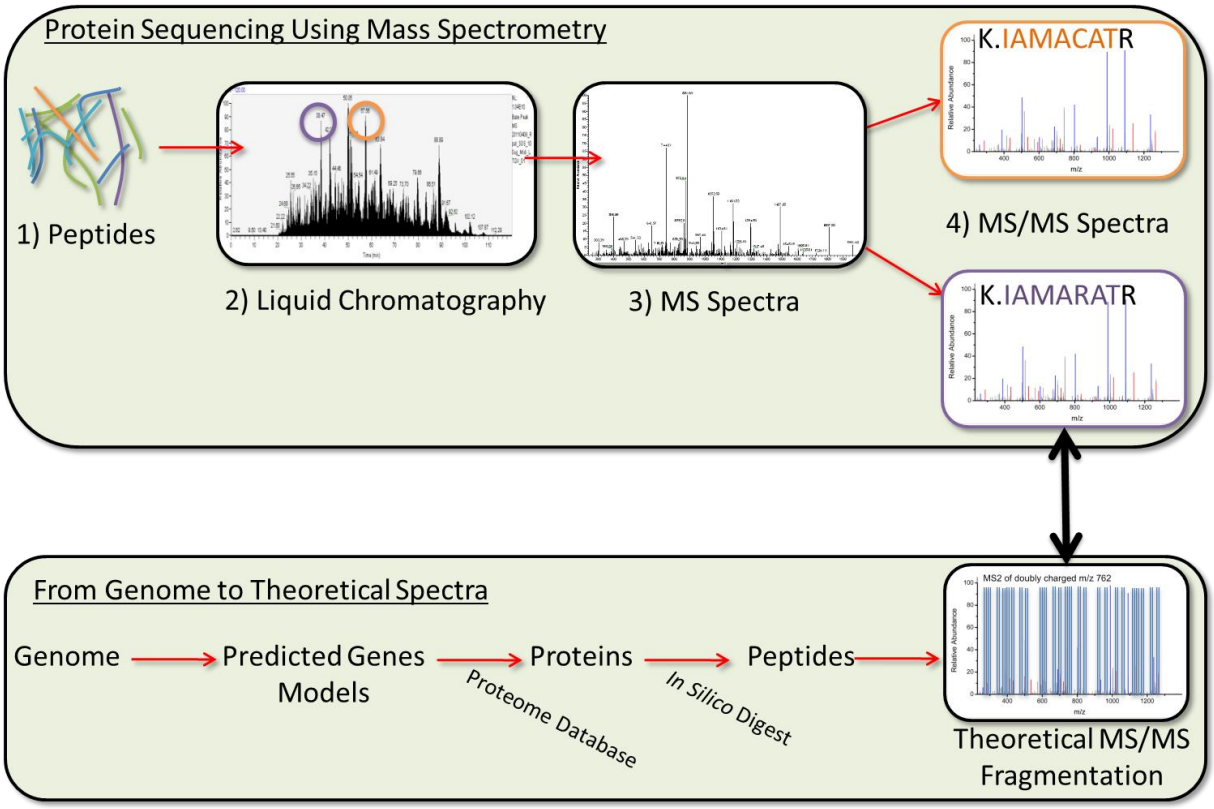


Figure 2.8. Peptide identification through spectrum matching. In a shotgun proteomics analysis, protein mixtures are enzymatically digested into peptides. Then the generated small peptides are sent to liquid chromatography (LC) for separation. Peptides are separated by hydrophobicity and electric charge. Separated peptides are measured by tandem mass spectrometry (MS/MS) generating spectrum providing information of mass to charge ratio and abundance about the peptides. Genome information is utilized to predict all the possible proteins that could be measured. And computationally digested peptides and their theoretical fragmentation spectra are generated. Peptide sequence identification is achieved by matching the experimental results to theoretical spectra.

the predicted proteome database, and computationally generates theoretical fragmentation patterns for predicted tryptic peptides (theoretical MS/MS spectra). Predicted peptides having a molecular mass approximate to the measured parent ion (within a user defined mass window) are selected as candidates for the following spectrum matching. The theoretical MS/MS spectra of the candidate peptides are serving as barcodes to match to the processed experimental MS/MS spectrum corresponding to the parent ion. The quality of the peptide spectra matching (PSM) is determined by a correlation score, XCorr. The XCorr is weighed by the total intensity of matched ions, the continuity of the fragment sequences, and the number of measured ions relative to the predicted number of fragments. Additionally, another score, DeltaCN, is calculated as the difference between the XCorr scores for the best and the second best matched spectrum. The DeltaCN and XCorr scores together indicate the confidence of the peptide spectrum matching.

After the peptide sequences are identified by SEQUEST, the DTASelect software [43] was used for protein inference. DTASelect evaluates all the PSMs scores achieved through SEQUEST, filters the scores with user defined criteria, and match all the filtered PSMs to the corresponding proteins. In this dissertation, standard criteria were applied which required the XCorr scores to be above 1.8, 2.5, and 3.5 for +1, +2, and +3 charged peptide ions, respectively [168]. And the DeltCN filter was set at 0.08. In addition, for each identified protein, at least two peptide sequences much be identified. The DeltCN output lists all identified proteins and their corresponding measured peptides, together with the scoring matrices (Figure 2.9).

Locus Key:

Validation Status	Locus	Sequence Count	Spectrum Count	Sequence Coverage	Length	MolWt	pI	Descriptive Name

Similarity Key:

Locus	# of identical peptides	# of differing peptides

U **Adeh_2CP-C:637865910** 20 399 95.1% 142 14763 6.8 YP_467553 Adeh_4353 protein of unknown function DUF583 [Anaeromyxobacter dehalogenans 2CP-C] # COG:COG1664(M) # pI:6.30 MW:14763

	Filename	XCorr	DeltCN	ObsM+H+	CalcM+H+	SpR	SpScore	Ion%	#	Sequence	
*	A dehalo Fe Citrate run1 030711 04.07026.07026.2	5.7779	0.5147	2083.3728	2084.3372	1	1876.7	78.9%	3	K.REELTSIPAASGDLNALLGR.G	2
*	A dehalo Fe Citrate run1 030711 03.07305.07305.2	5.0031	0.5572	1928.1327	1928.1497	1	739.7	72.2%	4	R.EELTSIPAASGDLNALLGR.G	2
*	A dehalo Fe Citrate run1 030711 04.08472.08472.2	3.4338	0.5896	1660.1327	1657.8206	1	1250.1	75.0%	30	R.GSEFEGKLTFEGTVR.I	2
*	A dehalo Fe Citrate run1 030711 05.00225.00225.2	5.9583	0.6521	2233.5728	2234.5547	1	2219.7	66.7%	43	R.IDGKFTGTIVTNDVLVVGEGAK.V	2
*	A dehalo Fe Citrate run1 030711 04.08481.08481.3	5.6742	0.5467	2234.1555	2234.5547	1	2412.2	42.9%	156	R.IDGKFTGTIVTNDVLVVGEGAK.V	3
*	A dehalo Fe Citrate run1 030711 03.08631.08631.2	4.6746	0.5588	1822.2327	1821.0807	1	1947.6	73.5%	97	K.FTGTIVTNDVLVVGEGAK.V	2
*	A dehalo Fe Citrate run1 030711 09.00308.00308.2	2.4991	0.1594	2172.3328	2169.4607	1	381.5	42.5%	1	K.VSAEITCGTVIVHGEINGNVR.A	2
*	A dehalo Fe Citrate run1 030711 12.03522.03522.1	2.4817	0.4155	1115.47	1116.2651	1	315.9	66.7%	2	K.NAVELHHPAK.M	1
*	A dehalo Fe Citrate run1 030711 12.03478.03478.2	3.2996	0.2789	1115.9927	1116.2651	1	812.4	88.9%	3	K.NAVELHHPAK.M	2
*	A dehalo Fe Citrate run1 030711 10.04806.04806.2	4.2156	0.5393	1576.1327	1575.8817	1	1387.2	76.9%	7	K.MRGNIEAPSLMVEK.G	2
*	A dehalo Fe Citrate run1 030711 04.04995.04995.1	2.8445	0.5051	1288.5	1288.5017	1	517.7	59.1%	5	R.GNIEAPSLMVEK.G	1
*	A dehalo Fe Citrate run1 030711 04.04972.04972.2	3.5492	0.4149	1289.2327	1288.5017	1	762.9	72.7%	5	R.GNIEAPSLMVEK.G	2
*	A dehalo Fe Citrate run1 030711 05.06569.06569.2	4.8109	0.6441	2234.9126	2234.5725	1	2090.0	65.0%	3	R.GNIEAPSLMVEKGVIFEGQSK.M	2
*	A dehalo Fe Citrate run1 030711 03.03507.03507.2	2.6936	0.288	966.0927	965.0941	1	816.7	81.2%	1	K.GVIFEGQSK.M	2
*	A dehalo Fe Citrate run1 030711 04.03526.03526.1	2.1727	0.4357	966.49	965.0941	1	424.2	62.5%	6	K.GVIFEGQSK.M	1
*	A dehalo Fe Citrate run1 030711 04.02195.02195.1	3.081	0.3778	1106.4	1107.2671	1	516.6	72.2%	6	K.MEALDKNASK.A	1
*	A dehalo Fe Citrate run1 030711 04.02102.02102.2	3.0469	0.3693	1109.2126	1107.2671	1	1346.8	88.9%	7	K.MEALDKNASK.A	2
*	A dehalo Fe Citrate run1 030711 10.04632.04632.3	4.9058	0.4705	2319.4453	2319.7246	1	2863.7	52.2%	8	K.MEALDKNASKAAPAPAPAVAAVKP.-	3
*	A dehalo Fe Citrate run1 030711 11.04023.04023.2	3.5605	0.4858	1231.0927	1231.4807	1	865.6	76.9%	7	K.AAPAPAPAVAAVKP.-	2
*	A dehalo Fe Citrate run1 030711 12.04246.04246.1	2.2037	0.0918	1232.72	1231.4807	1	749.5	57.7%	5	K.AAPAPAPAVAAVKP.-	1

U **Adeh_2CP-C:637863933** 89 866 85.9% 548 58035 5.4 YP_465601 Adeh_2394 chaperonin GroEL [Anaeromyxobacter dehalogenans 2CP-C] # COG:COG0459(O) # pI:5.30 MW:58035

	Filename	XCorr	DeltCN	ObsM+H+	CalcM+H+	SpR	SpScore	Ion%	#	Sequence	
*	A dehalo Fe Citrate run1 030711 12.04039.04039.2	2.5548	0.374	1068.4927	1069.2083	3	274.4	81.2%	1	K.EIAFHQPAR.E	2
3	A dehalo Fe Citrate run1 030711 06.02602.02602.1	1.817	0.0949	601.39	601.7239	1	176.0	87.5%	1	R.EAILR.G	111111
*	A dehalo Fe Citrate run1 030711 05.07863.07863.1	3.1501	0.4543	1553.51	1554.8271	1	649.2	53.3%	2	R.GVQTLAEAVAVTLGPK.G	1
*	A dehalo Fe Citrate run1 030711 07.07549.07549.3	5.1319	0.4971	1554.5955	1554.8271	1	2206.4	56.7%	7	R.GVQTLAEAVAVTLGPK.G	3
*	A dehalo Fe Citrate run1 030711 05.07850.07850.2	4.4702	0.3532	1557.4528	1554.8271	1	1878.7	73.3%	19	R.GVQTLAEAVAVTLGPK.G	2
*	A dehalo Fe Citrate run1 030711 12.05584.05584.2	4.4162	0.431	1768.3927	1768.0665	1	1807.3	64.7%	29	R.GVQTLAEAVAVTLGPKGR.N	2
*	A dehalo Fe Citrate run1 030711 12.05554.05554.3	3.6603	0.4783	1768.9154	1768.0665	1	1132.8	51.5%	1	R.GVQTLAEAVAVTLGPKGR.N	3
	A dehalo Fe Citrate run1 030711 12.03428.03428.1	1.9942	0.2726	914.49	915.0806	24	80.9	50.0%	4	K.GRNVVIEK.S	11
	A dehalo Fe Citrate run1 030711 09.02637.02637.2	2.4964	0.1827	916.9927	915.0806	1	756.6	92.9%	1	K.GRNVVIEK.S	22
	A dehalo Fe Citrate run1 030711 06.02366.02366.1	2.1938	0.2354	702.44	701.8412	1	313.8	80.0%	12	R.NVVIEK.S	11
*	A dehalo Fe Citrate run1 030711 07.04296.04296.2	4.2562	0.0802	1620.4327	1620.8864	1	1883.6	75.0%	1	R.NVVIEKSYGAPTITK.D	2
*	A dehalo Fe Citrate run1 030711 05.03269.03269.1	1.916	0.2442	937.35	938.0685	1	419.6	68.8%	2	K.SYGAPTITK.D	1
*	A dehalo Fe Citrate run1 030711 06.04105.04105.2	4.1603	0.1246	1608.3127	1608.8322	1	2084.8	76.7%	7	K.SYGAPTITKDGVTVAK.E	2
*	A dehalo Fe Citrate run1 030711 05.06501.06501.3	4.945	0.6225	2437.4353	2437.7498	1	901.9	42.0%	3	K.SYGAPTITKDGVTVAKEIELESK.F	3

Figure 2.9. An example of the DTASelect output. DTASelect output file lists all the detected proteins (highlighted in red) and their corresponding peptides (highlighted in blue). All scoring matrices for peptide identification and protein inference are presented.

2.10 Label-free Quantification and Data Normalization

For label-free mass spectrometry analysis, protein abundances are measured by spectral count [169, 170], which is the number of times all the corresponding peptides are detected during a MS experiment. The more abundant proteins and peptides have higher chance to be detected in mass spectrometry measurement, thus spectral count is a semi-quantitative indicator for the relative abundances of proteins and peptides in the sample. Proteins with longer length potentially generate more peptides after proteolysis, which could lead to increased spectral counts. In order to consider the spectral count bias toward longer proteins, the normalized spectra abundance factor (NSAF) [171] is generally used for data normalization. With NSAF normalization, the measured spectral count (SpC) is divided by the length of the protein (L), and the normalized spectral count is calculated by dividing the SpC/L of an identified protein by the sum of SpC/L for all identified proteins in a MS experiment. The NSAF normalization also corrects the experimental variability to a certain degree, in that the total normalized spectral counts is constant for all experiments. Thus, normalized spectral counts of proteins identifications from different MS measurements (e.g. cell growth under different conditions, measurements done in different dates) could be compared.

Chapter 3

Mass Spectrometry-based Proteomic Profiling Reveals Core and Electron Acceptor Specific *c*-Type Cytochromes in Dissimilatory Metal Reducing Bacteria

A version of this chapter is submitted for publication as the following journal article:

> Xiaoxin Liu, Silke Nissen, Karuna Chourey, Frank Löffler, Robert Hettich. "Core and Electron Acceptor-specific *c*-Type Cytochromes in Metabolically Versatile Metal-Reducing Bacteria". Submitting to *The ISME Journal* 2014.

Xiaoxin Liu contributed to the proteome sample preparation, mass spectrometry experiments performing, biological data analysis and visualizations, and manuscript writing.

3.1 Introduction

Shewanella spp., *Geobacter* spp. and *Anaeromyxobacter dehalogenans*, are renowned for having well established respiratory chains that allow them to function under many different environmental conditions by utilizing a variety of electron acceptors and donors to sustain energy production. This is exemplified in studies of the microbial diversity in the uranium contaminated aquifer at the Integrated Field Research Challenge site at Oak Ridge, TN, which revealed that *Anaeromyxobacter sp.*, *Shewanella* spp. and *Geobacter* spp. are key players for the *in-situ* bioremediation of uranium [172]. In addition to soluble electron acceptors like U(VI), these microbes are also able to reduce solid metals, such as iron oxide and manganese oxide[105, 173, 174]. In cases where the electron acceptor is insoluble, or soluble but not able to permeate into the cell (e.g. Fe citrate), the bacteria needs to have an extended electron

transfer chain to transmit electrons from the inner membrane through the periplasm to the outer membrane or even to the outside of the cell [175].

DNA sequencing has revealed that the genomes of these genera harbor many *c*-type cytochrome genes. *c*-Type cytochromes are electron transfer proteins and are characterized by the presence of *c*-type hemes, which are covalently attached in specific binding motifs. The most common *c*-type heme-binding motifs are CXXCH, CXXXCH, and CXXXXCH [176]. *Shewanella oneidensis* strain MR-1 has a genome encoding for a total of 42 *c*-type cytochromes [110], and *Geobacter sulfurreducens* has a repertoire of over one hundred *c*-type cytochromes [93]. The structures and functions of selected *c*-type cytochromes of *S. oneidensis* strain MR-1 have been characterized by a variety of studies. For example, CymA transfers electrons from the inner membrane of *S. oneidensis* to the periplasm [100, 101, 177]. MtrA and MtrB form a porin-cytochrome complex on the outer membrane that passes electrons from the inner face of the outer membrane to the outside [102]. The solid extracellular electron acceptor is reduced, possibly by receiving electrons from extracellular multi-heme cytochromes MtrC, MtrF, and OmcA [102]. While there is limited information on *c*-type cytochromes in *Geobacter daltonii* strain FRC-32, another member of the *Geobacter* genus, *Geobacter sulfurreducens* strain PCA, has been extensively studied. There are 111 *c*-type cytochromes encoded by *G. sulfurreducens*, one of which contains 27 heme-binding motifs [93]. A current model of extracellular metal ion reduction by *G. sulfurreducens* highlights the type IV pili structure, which is electrically conductive [107]. It is believed that the pili receive electrons from the OmcE and OmcS complex and transport electrons directly to the electron acceptor [99]; however, a detailed characterization of this process has not been conducted. Although there are large numbers of

putative *c*-type cytochrome genes present in the genomes of *S. oneidensis* and *G. sulfurreducens*, only a few have been validated and functionally characterized (as mentioned above). There are 68 *c*-type cytochrome encoding genes in lesser characterized *Anaeromyxobacter dehalogenans* strain 2CP-C, one of which contains as many as 40 heme motifs [111]. The expression and function of these *c*-type cytochromes of *A. dehalogenans* are virtually unknown. This suggests a critical need to better characterize factors influencing expression of the *c*-type cytochromes of *A. dehalogenans* and then compare and contrast them with the *c*-type cytochromes of *S. oneidensis* and *G. daltonii*.

To date, most studies on *c*-type cytochromes have focused on the analysis of the function and structures of a few specific types. The specificity of a *c*-type cytochrome for a certain electron acceptor is usually determined by mutant studies/gene-knockout experiments in which the knock-out mutants are examined for a discernible phenotype [117, 178]. However, eliminating a particular *c*-type cytochrome often does not provide a phenotype marker, which may indicate a large functional overlap between different *c*-type cytochromes in which the role of a specifically knocked-out cytochrome may be substituted by another one, resulting in no appearance of a phenotype marker. Thus, it is essential to study the complete repertoire of *c*-type cytochromes in order to obtain a more accurate and comprehensive overview. A few large-scale studies inspected *c*-type cytochrome gene expression at the transcriptome level for *Shewanella* and *Geobacter* [179-181]. These studies discovered that a few *c*-type cytochrome gene transcripts were elevated under metal-reducing conditions. However, *c*-type cytochrome proteins, not genes or transcripts, are the final functional units which directly constitute the electron transfer chain and carry out bacterial respirational ability. It is imperative to study the

cytochrome proteins to characterize their expression and functional differences. Mass spectrometry is currently the method of choice to characterize protein expression in a high throughput and accurate manner. For example, a number of studies illustrated how mass spectrometry can be employed to characterize proteome expression profiles of a range of organisms, such as *Pseudomonas, Ignicoccus*, *Populus* [36, 37, 182]. Several proteomic studies have been conducted on *S. oneidensis* MR-1, but all of which only tested a few growth conditions and did not focus on the *c*-type cytochrome proteins [123, 126, 183].

In this study, we used high performance liquid chromatography coupled with tandem mass spectrometry to characterize the *c*-type cytochrome identities, abundance changes, and differential expression profiles of *Anaeromyxobacter dehalogenans* strain 2CP-C, *Shewanella oneidensis* strain MR-1, and *Geobacter daltonii* strain FRC-32 grown with a wide variety of different electron acceptors. In particular, this study is designed to be a global *c*-type cytochrome characterization for the aforementioned three strains with a focus on *A. dehalogenans* strain 2CP-C. The results reveal that the microbes adjust their *c*-type cytochrome expressions in response to the available electron acceptor. A number of "core" *c*-type cytochromes were identified that are expressed under all tested growth conditions, regardless of the electron acceptor provided and are likely essential for general 'housekeeping' activities. Additionally, "unique" *c*-type cytochromes were only detected in response to certain electron acceptors. These cytochromes may be used as specific biomarkers, which are required for the utilization of specific electron acceptors.

3.2 Material and Methods

Bacterial Strains and Culture Conditions. In this chapter, *Anaeromyxobacter dehalogenans* strain 2CP-C, *Geobacter daltonii* strain FRC-32, and *Shewanella oneidensis* strain MR-1 were grown in the following conditions as described in Chapter 2: FeOOH, Fe citrate, fumarate, MnO_2, nitrate, tryptic soy broth, oxygen. Please refer to Chapter 2 for detailed methods.

Cell Lysis and Protein Extraction. The SDS-TCA method and Guanidine-HCl method was used for cell lysis and protein extraction according to the growth conditions. Details about the procedures are provided in Chapter 2.

Peptide Analysis by LC-MS/MS. Peptides were analyzed by two-dimensional liquid chromatography (Ultimate HPLC System, LC Packings, a division of Dionex, San Francisco, CA) coupled online with a linear ion trap mass spectrometer (LTQ XL, ThermoFisher Scientific, San Jose, CA). A total of 53 runs were collected, which provided biological triplicates for all population. Peptides samples were loaded onto 5 cm Strong Cation Exchange (SCX) resins (Phenomenex, Torrance, CA) packed column, followed by offline wash to remove any lingering salt and contaminant and then connected online to a front column (PicoTip emitter, New Objective, Woburn, MA) packed with ~12 cm Reverse-Phase (RP) resins (Phenomenex, Torrance, CA). A 12-step 24-h Multidimensional Protein Identification Technology (MudPIT) was used to analyze peptides as previously described [139]. Briefly, a total of 11 incremental salt pulses were applied to separate peptides by charge. At each step, a gradient of solvent A (95% H_2O, 5% acetonitrile, 0.1% formic acid (v/v)) to solvent B (30% H_2O, 70% acetonitrile, 0.1% formic acid (v/v)) further separates peptides based on their hydrophobicity. LTQ XL was operated in a data

dependent mode and one full MS scan was followed by five MS/MS scans. The m/z isolation width was set to 3 m/z. Dynamic exclusion repeat was set at 1 with a duration time of 60 sec.

Database Searching and Data Normalization. Experimental MS/MS spectra collected for *A. dehalogenans* strain 2CP-C, *S. oneidensis* strain MR-1, and *G. daltonii* strain FRC-32 were searched against FASTA protein databases specific to the respective bacteria. The FASTA protein database for each strain consists of corresponding predicted proteome database obtained from Joint Genome Institute (JGI, http://genome.jgi.doe.gov/, downloaded in July, 2011) and common contaminants such as trypsin, keratin, etc. A decoy database consisting of reversed sequences from corresponding proteome database was also appended to the FASTA protein database in order to calculate False Discovery Rate (FDR) at the protein level. The SEQUEST search algorithm was used for peptide identification. Identified peptides were filtered and assembled into proteins by DTASelect [43]. The DeltCN filter was set to 0.08. The cross-correlation score filtering process started with conservative criteria as following: XCorr: +1 = 1.8, +2 = 2.5, +3 = 3.5. The XCorr criteria were slightly altered to adjust FDR to ~1% for each sample. At least two peptides were required per protein identification, and at least one peptide had to be unique to that protein. Raw spectral counts were normalized by protein lengths and the total spectral counts using Normalized Spectral Abundance Factor (NSAF) as previously described [184]. Normalized spectral counts are then uniformly multiplied by a factor of 100,000 for better readability and comparability to raw spectral counts. Adjusted normalized spectral count is denoted as nSpC throughout this paper. The nSpCs for biological triplicates were averaged for each growth condition.

***c*-Type Cytochrome Prediction.** Putative *c*-type cytochromes of *A. dehalogenans* strain 2CP-C, *S. oneidensis* strain MR-1 and *G. daltonii* strain FRC-32 were predicted by searching their respective protein databases followed by BLASTp against NCBI database. This process was automated using a customized perl script. Two criteria were used by the script to predict *c*-type cytochromes: 1) the protein sequence must contain at least one of the following common *c*-type cytochrome motifs: CXXCH, CXXXCH and CXXXXCH; 2) BLASTp result of this protein must contain at least one annotated *c*-type cytochrome.

Cellular Localization Prediction. Cellular localizations of the putative *c*-type cytochromes were predicted based on their protein sequences using Psortb.3.0 [185]

3.3 Results

3.3.1 *c*-Type Cytochrome Identification

A total of 69, 40, and 72 putative *c*-type cytochromes were predicted for *A. dehalogenans* strain 2CP-C, *S. oneidensis* strain MR-1, and *G. daltonii* strain FRC-32, respectively (Table 3.1). With results pooled across all growth conditions, the LC-MS/MS approach detected 53, 25, and 38 *c*-type cytochromes for strain 2CP-C, strain MR-1, and strain FRC-32, respectively (Table 3.1). Strain 2CP-C and strain MR-1 grown with nitrate as electron acceptor yielded the highest number of identified *c*-type cytochromes (37 and 20, respectively), while most *c*-type cytochromes (35) for strain FRC-32 were detected with fumarate provided as electron acceptor (Table 3.1). Cells grown with FeOOH yielded the least number of measureable *c*-type

Table 3.1. *c*-Type cytochrome identification for *A. dehalogenans* 2CP-C, *S. oneidensis* MR-1 and *G. daltonii* FRC-32 grown under various redox conditions.

	***A. dehalogenans* 2CP-C**	***S. oneidensis* MR-1**	***G. daltonii* FRC-32**
Number of predicted *c*-type cytochromes	69	40	72
Number of total detected *c*-type cytochromes	53	25	38
Growth conditions	**Number of detected *c*-type cytochromes under different conditions**		
FeOOH	11	8	8
Fe citrate	27	19	10
MnO_2	25	20	21
Nitrate	37	20	N/A
Tryptic Soy Broth	29	16	N/A
Fumarate	33	18	35
Oxygen	30	19	N/A

N/A: *Geobacter* growth could not be achieved under these conditions.

cytochromes for all three bacteria (Table 3.1), which may due to the slow growth resulting in less variety of proteins expressed by the microbes under FeOOH amendment conditions.

A hierarchical clustering of the *c*-type cytochrome abundance patterns in strain 2CP-C cultures in response to different growth conditions revealed a remarkable level of expression differentiation (Figure 3.1). The *c*-type cytochrome abundance patterns of the three metal electron acceptor conditions (FeOOH, MnO_2, Fe citrate; the three left-most columns) clustered together, while oxic growth (far right column) was the most distant from all anoxic conditions (Figure 3.1). Heat maps were also generated for *c*-type cytochromes in strain MR-1 and FRC-32 (Figure 3.2 and 3.3), and also revealed fined-tuned expression patterns that were somewhat distinct from strain 2CP-C. Although only four conditions were tested for strain FRC-32, the three conditions with metal electron acceptors clustered together, showing distinct *c*-type cytochrome expression patterns to growth with fumarate (Figure 3.3). For strain MR-1, the cytochrome expression patterns under metal electron acceptor conditions did not show similar trends, and were not clustered together as with strain 2CP-C and FRC-32 (Figure 3.2). All hierarchical clustering approaches reveal that the expression profiles of *c*-type cytochromes can be grouped into three categories: “core” (present in all conditions), “unique” (represented in only one condition), and non-detected.

3.3.2 “Core” *c*-Type Cytochromes

Proteomic measurements revealed several *c*-type cytochromes that were constitutively expressed under all growth conditions, suggesting that these constitute the “core” cytochromes that are required under all growth conditions. Six “core” cytochromes were detected in strain

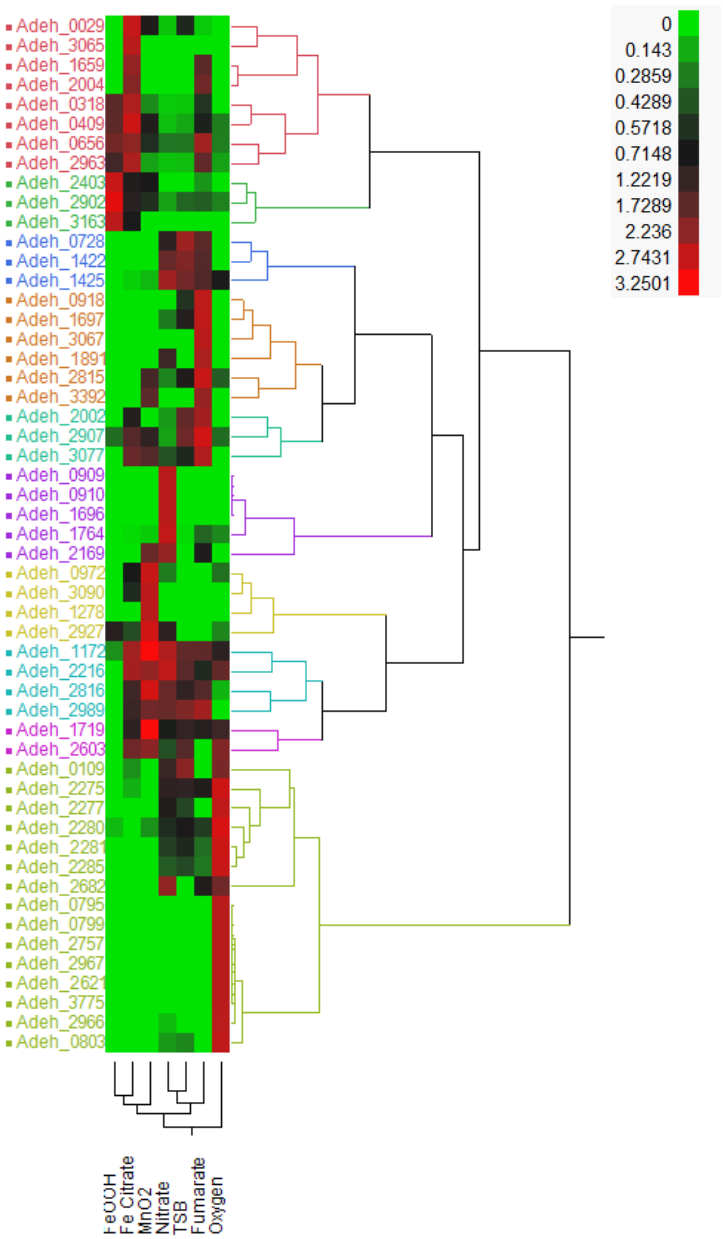


Figure 3.1. Hierarchical clustering of detected c-type cytochromes in *A. dehalogenans* 2CP-C cultures grown under various electron-accepting conditions. Hierarchical clustering and the heatmap of protein abundances were generated using JMP genomics 6.0 (SAS, Cary, NC). Hierarchical clustering was conducted using Ward algorithm. Each row was scaled so that the variance of each row is 1. The green-black-red scale on the top-right indicates scaled nSpCs (normalized spectral counts) for all *c*-type cytochromes.

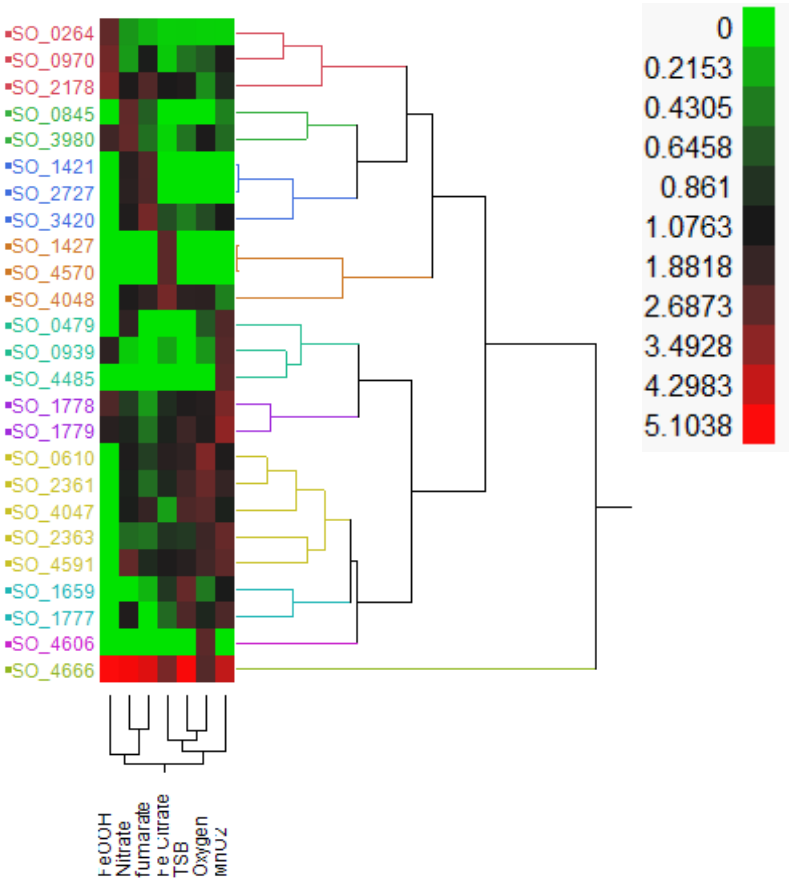


Figure 3.2. Hierarchical clustering of detected c-type cytochromes in *S.oneidensis* MR-1 cultures grown under various electron-accepting conditions. Hierarchical clustering and the heatmap of protein abundances were generated using JMP genomics 6.0 (SAS, Cary, NC). Hierarchical clustering was conducted using Ward algorithm. Each row was scaled so that the variance of each row is 1. The green-black-red scale on the top-right indicates scaled nSpCs (normalized spectral counts) for all *c*-type cytochromes.

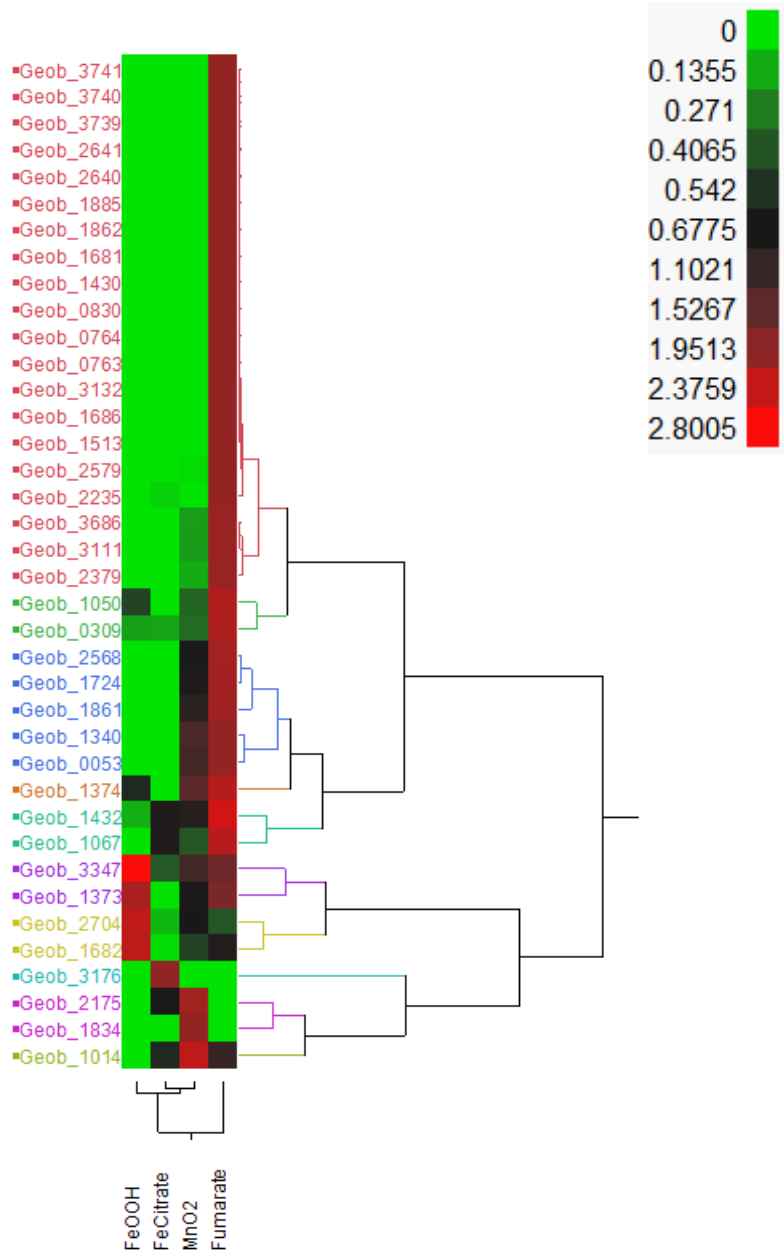


Figure 3.3 Hierarchical clustering of detected c-type cytochromes in *G.daltonii* FRC-32 cultures grown under various electron-accepting conditions. Hierarchical clustering and the heatmap of protein abundances were generated using JMP genomics 6.0 (SAS, Cary, NC). Hierarchical clustering was conducted using Ward algorithm. Each row was scaled so that the variance of each row is 1. The green-black-red scale on the top-right indicates scaled nSpCs (normalized spectral counts) for all *c*-type cytochromes.

2CP-C cultures: Adeh_0409, Adeh_0656, Adeh_1172 (CcoO), Adeh_2902 (NrfA), Adeh_2907, and Adeh_2963 (Table 3.2, Figure3.1). Strain MR-1 had seven "core" *c*-type cytochromes: SO_0264 (ScyA), SO_0970 (FccA), SO_1778 (MtrC/OmcB), SO_1779 (OmcA), SO_2178 (CcpA), SO_3980 (NrfA), and SO_4666 (CytcB) (Table 3.2, Figure 3.2), among which, the abundance of FccA (SO_0970) is 10-100 fold higher in magnitude than the other "core" cytochromes under all the growth conditions tested (Table 3.2). Three cytochrome *c* family proteins (Geob_0309, Geob_2704, and Geob_3347) and two hypothetical proteins (Geob_1432, and Geob_1682) were identified as "core" cytochromes in strain FRC-32 cultures (Table 3.2, Figure 3.3).

In strain 2CP-C, the "core" cytochrome Adeh_2902 encodes a pentaheme cytochrome *c* nitrite reductase (NrfA), which is the enzyme that directly reduces nitrite to ammonia. Although it is unclear whether NrfA contributes to the electron transfer chain, interestingly, in the other Proteobacteria such as *Wolinella succinogenes* and *Desulfovibrio vulgaris*, NrfA is involved in electron transfer from menaquinones to nitrite [186, 187]. The NrfA in strain MR-1 (SO_3980) is also detected under all growth conditions (Table 3.2). SO_3980 is predicted to contain four CXXCH heme-binding motifs, but a fifth atypical motif (CXXCK) has been identified [188], suggesting that both Adeh_2902 and SO_3980 contain five heme-binding motifs. Under nitrate amended growth condition, SO_3980 of strain MR-1 shows a very high abundance of detection compared to the other conditions (Table 3.2). However, the abundance of Adeh_2902 of strain 2CP-C in nitrate grown cells is the lowest among all the growth conditions tested (Table 3.2).

The *c*-type cytochrome Adeh_2963 of strain 2CP-C shares extremely high sequence similarity to the *c*-type cytochrome sulfite reductase SirA of strain MR-1 (encoded by SO_0479), based on Blastp search result (e-value: 0; bitscore: 645). SirA, previously designated MccA,

Table 3.2. The 'core' *c*-Type cytochromes of *A. dehalogenans* 2CP-C, *S. oneidensis* MR-1 and *G. daltonii* FRC-32.

Gene ID	GenBank annotation	Number of *c*-type cytochrome motifs predicted	Cellular localization predicted by Psort	Growth conditions						
				FeOOH	Fe citrate	MnO_2	Nitrate	Tryptic Soy Broth	Fumarate	Oxygen
***Anaeromyxobacter dehalogenans* strain 2CP-C**										
Adeh_0409	hypothetical protein	10 motifs	Extracellular	63	102	33	3	6	24	10
Adeh_0656	hypothetical protein	2 motifs	Extracellular	36	41	11	6	6	45	5
Adeh_1172	bifunctional cbb_3-type cytochrome c oxidase subunit II/cytochrome *c*	2 motifs	Unknown	6	68	90	72	50	49	30
Adeh_2902	respiratory nitrite reductase	5 motifs	Periplasmic	344	117	67	20	41	45	31
Adeh_2907	hypothetical protein	10 motifs	Unknown	34	165	114	18	194	288	34
Adeh_2963	cytochrome *c*	7 motifs	Periplasmic	147	269	18	9	10	189	18
***Shewanella oneidensis* strain MR-1**										
SO_0264	scyA, cytochrome *c*	1 motifs	Periplasmic	879	24	20	99	25	56	22
SO_0970	FccA, fumarate reductase flavoprotein subunit precursor	4 motifs	Periplasmic	5373	170	2175	505	843	1818	1088
SO_1778	omcB/mtrC, decaheme cytochrome *c*	10 motifs	Unknown	119	45	158	39	66	15	73
SO_1779	omcA, decaheme cytochrome *c*	10 motifs	Unknown	126	81	291	79	167	40	114
SO_2178	ccpA, cytochrome c551 peroxidase	2 motifs	Periplasmic	175	62	48	67	69	131	18

Table 3.2. Continued

Gene ID	GenBank annotation	Number of *c*-type cytochrome motifs predicted	Cellular localization predicted by Psort	Growth conditions						
				FeOOH	Fe citrate	MnO_2	Nitrate	Tryptic Soy Broth	Fumarate	Oxygen
SO_3980	nrfA, cytochrome c552 nitrite reductase	4 motifs	Periplasmic	406	11	101	542	92	96	234
SO_4666	cytcB, cytochrome *c*	2 motifs	Periplasmic	95	60	79	92	93	87	47
***Geobacter daltonii* strain FRC-32**										
Geob_3347	cytochrome *c* family protein	21 motifs	Unknown	116	16	52	N/A	N/A	69	N/A
Geob_2704	cytochrome *c* family protein	21 motifs	Unknown	45	2	13	N/A	N/A	8	N/A
Geob_1682	hypothetical protein	9 motifs	Unknown	164	1	33	N/A	N/A	59	N/A
Geob_1432	hypothetical protein	4 motifs	Unknown	57	366	410	N/A	N/A	1140	N/A
Geob_0309	cytochrome *c* family protein	5 motifs	Unknown	8	7	15	N/A	N/A	101	N/A

catalyzes dissimilatory sulfite reduction in strain MR-1 [189]. SirA (SO_0479) is an octaheme *c*-type cytochrome containing seven CXXCH motifs and one atypical heme-binding motif $CX_{15}CH$ [190]. Our prediction suggests that there are seven CXXCH motifs in Adeh_2963. But similar to SO_0479, an atypical motif $CX_{15}CH$ is also present in Adeh_2963, which makes Adeh_2963 an octaheme *c*-type cytochrome. The protein sequences of both Adeh_2963 and SO_0479 contain an iron-binding motif (CXXCG) similar to rubredoxin, which is also an bacterial electron transfer component [191]. SO_0479 is only detected when MnO_2, nitrate, or oxygen is provided as electron acceptor (Table 3.3). However, Adeh_2963 is detected under all growth conditions in strain 2CP-C cells and its abundance is very low in MnO_2, nitrate, or oxygen amended growth conditions compared to that in Fe Citrate, fumarate, or FeOOH conditions (Table 3.2). Adeh_2963 may be required for iron reduction.

Three "core" *c*-type cytochromes of strain 2CP-C, Adeh_0409, Adeh_0656, and Adeh_2907, are annotated as "hypothetical proteins" in GenBank (Table 3.2). These proteins are consistently expressed in 2CP-C cells grown under all conditions, albeit detected at different abundances, suggesting that these putative *c*-type cytochromes are important for the electron transfer machinery of strain 2CP-C. Adeh_0409 and Adeh_2907 are both decaheme *c*-type cytochromes. Blastp results showed that Adeh_2907 shares very high sequence similarity with Adeh_0409 with 0 e-value, 1013 bitscore, and 68% sequence identity. Adeh_0409 is predicted to be an extracellular *c*-type cytochrome (Table 3.4). Both Adeh_2907 and Adeh_0409 share sequence similarity with many hypothetical proteins in other organisms such as other *Anaeromyxobacter* spp*., Denitrovibrio acetiphilus*, other *Geobacter* spp*.*, other *Shewanella* spp*.*, and *Rhodoferax ferrireducens*, which suggests that Adeh_2907 and

Table 3.3. *c*-type cytochrome identification for *Shewanalla oneidensis* MR-1 grown with various electron acceptors.

Locus	Annotation	Gene Name	Motif	COG	Localization_psort	FeOOH	Fe Citrate	MnO_2	Nitrate	TSB	fumarate	Oxygen
SO_0264	cytochrome c	scyA	1	C	Periplasmic	879.5	24.0	20.0	99.1	24.8	56.1	21.5
SO_0479	cytochrome c, putative	SO_0479	7	-	Periplasmic	-	-	20.9	14.7	-	-	5.4
SO_0610	ubiquinol-cytochrome c reductase, cytochrome c1	petC	1	C	Cytoplasmic Membrane	-	34.8	29.3	28.9	39.0	17.8	75.0
SO_0714	monoheme cytochrome c	SO_0714	1	C	Periplasmic	-	-	-	-	-	-	-
SO_0716	monoheme cytochrome c, putative	SO_0716	1	-	Unknown	-	-	-	-	-	-	-
SO_0717	monoheme cytochrome c	SO_0717	1	C	Periplasmic	-	-	-	-	-	-	-
SO_0845	cytochrome c-type protein NapB	napB	2	C	Periplasmic	-	-	15.2	102.2	-	21.7	-
SO_0939	cytochrome c, putative	SO_0939	2	-	Unknown	68.9	10.2	104.5	3.7	-	-	12.6
SO_0970	fumarate reductase flavoprotein subunit precursor	FccA	4	C	Periplasmic	5372.5	170.0	2174.6	505.4	842.7	1818.4	1088.0
SO_1233	tetraheme cytochrome c	torC	5	C	Cytoplasmic Membrane	-	-	-	-	-	-	-
SO_1413	tetraheme cytochrome c, putative	SO_1413	4	-	Periplasmic	-	-	-	-	-	-	-
SO_1421	fumarate reductase flavoprotein subunit	ifcA-1	4	C	Periplasmic	-	-	-	4.2	-	6.8	-
SO_1427	decaheme cytochrome c	SO_1427	10	-	Periplasmic	-	10.2	-	-	-	-	-
SO_1659	decaheme cytochrome c	SO_1659	10	-	Periplasmic	-	4.0	5.0	-	13.6	0.9	2.2
SO_1777	decaheme cytochrome c MtrA	mtrA	10	-	Unknown	-	4.3	19.0	10.5	19.3	-	7.7
SO_1778	decaheme cytochrome c	omcB/mtrC	10	-	Unknown	119.1	44.7	158.2	39.2	65.6	14.8	72.5
SO_1779	decaheme cytochrome c	omcA	10	-	Unknown	125.9	81.0	290.8	79.3	167.4	39.7	114.3
SO_1780	decaheme cytochrome c MtrF	mtrF	10	-	Unknown	-	-	-	-	-	-	-
SO_1782	decaheme cytochrome c MtrD	mtrD	10	-	Periplasmic	-	-	-	-	-	-	-
SO_2178	cytochrome c551 peroxidase	ccpA	2	P	Periplasmic	175.2	61.7	48.1	66.7	69.4	130.6	18.5
SO_2361	cytochrome c oxidase, cbb3-type, subunit III	ccoP	2	C	Cytoplasmic Membrane	-	26.0	50.3	27.6	58.8	14.0	79.9
SO_2363	cytochrome c oxidase, cbb3-type, subunit II	ccoO	1	C	Cytoplasmic	-	25.8	85.9	15.7	24.6	14.4	61.4
SO_2727	cytochrome c3	cctA	4	-	Periplasmic	-	-	-	16.7	-	25.5	-

Table 3.3. Continued

Locus	Annotation	Gene Name	Motif	COG	Localization_psort	FeOOH	Fe Citrate	MnO_2	Nitrate	TSB	fumarate	Oxygen
SO_3056	tetraheme cytochrome c	SO_3056	4	-	Periplasmic	-	-	-	-	-	-	-
SO_3300	cytochrome c	SO_3300	4	-	Unknown	-	-	-	-	-	-	-
SO_3420	cytochrome c'	SO_3420	1	C	Periplasmic	-	113.1	189.0	215.8	70.5	510.1	115.1
SO_3623	tetraheme cytochrome c	SO_3623	4	-	Periplasmic	-	-	-	-	-	-	-
SO_3980	cytochrome c552 nitrite reductase	nrfA	4	P	Periplasmic	405.6	11.0	101.4	541.6	91.7	95.8	233.7
SO_4047	cytochrome c family protein	SO_4047	2	C	Periplasmic	-	6.2	23.1	23.6	54.8	42.8	59.5
SO_4048	cytochrome c family protein	SO_4048	2	C	Periplasmic	-	53.5	7.1	21.9	28.2	29.7	28.0
SO_4142	cytochrome c family protein	SO_4142	1	-	Unknown	-	-	-	-	-	-	-
SO_4144	cytochrome c, putative	SO_4144	8	-	Periplasmic	-	-	-	-	-	-	-
SO_4360	decaheme cytochrome c	SO_4360	10	-	Periplasmic	-	-	-	-	-	-	-
SO_4484	cytochrome c-type protein Shp	SO_4484	1	-	Unknown	-	-	-	-	-	-	-
SO_4485	diheme cytochrome c	SO_4485	2	-	Unknown	-	-	17.9	-	-	-	-
SO_4570	hypothetical protein	SO_4570	1	-	Unknown	-	24.0	-	-	-	-	-
SO_4572	cytochrome c, putative	SO_4572	3	-	Unknown	-	-	-	-	-	-	-
SO_4591	tetraheme cytochrome c	cymA	4	C	Cytoplasmic Membrane	-	21.1	45.2	46.8	25.4	15.5	35.8
SO_4606	cytochrome c oxidase, subunit II	SO_4606	2	C	Cytoplasmic Membrane	-	-	-	-	-	-	2.6
SO_4666	cytochrome c	cytcB	2	C	Periplasmic	95.3	59.5	79.5	91.9	93.3	86.6	47.1

Cellualr localizations are predicted by Psort 3.0

Numbers in cells are normalized spectral counts (nSpC).

Red: Core *c*-type cytochromes

Yellow: Undetected *c*-type cytochromes

Orange: Detected under all but FeOOH

Green: Unique *c*-type cytochromes

Purple: Detected under two conditions

Blue: Detected under all but one condition

Table 3.4. *c*-type cytochrome identification for *Anaeromyxobacter dehalogenans* 2CP-C grown with various electron acceptors.

Chromosome Locus ID	GenBank Annotation	Predicted C-type cytochrome motifs	Cellular Localization	COG	FeOOH	Fe Citrate	MnO_2	Nitrate	TSB	Fumarate	Oxygen
Adeh_0029	Fibronectin, type III	9 motifs	Unknown	none	-	14.5	4.5	0.4	3.3	0.3	-
Adeh_0108	hypothetical protein	12 motifs	Periplasmic	none	-	-	-	-	-	-	-
Adeh_0109	hypothetical protein	10 motifs	Cytoplasmic Membrane	C	-	0.4	-	2.5	4.3	-	4.0
Adeh_0215	Multihaem cytochrome	7 motifs	Periplasmic	none	-	-	-	-	-	-	-
Adeh_0318	split soret cytochrome c precursor	2 motifs	Periplasmic	none	78.0	116.6	10.9	3.0	3.3	25.7	-
Adeh_0409	hypothetical protein	10 motifs	Extracellular	none	62.9	102.2	32.7	3.2	5.5	23.9	10.3
Adeh_0656	hypothetical protein	2 motifs	Extracellular	none	36.0	40.9	10.9	5.7	5.7	44.6	5.1
Adeh_0728	hypothetical protein	3 motifs	Periplasmic	S	-	-	-	3.0	6.7	4.8	-
Adeh_0795	quinol:cytochrome c oxidoreductase pentaheme cytochrome subunit	5 motifs	Unknown	none	-	-	-	-	-	-	3.8
Adeh_0799	quinol:cytochrome c oxidoreductase monoheme COG2010, CccA	1 motif	Periplasmic	none	-	-	-	-	-	-	34.3
Adeh_0803	cytochrome c oxidase, subunit II 1 motif	1 motif	Cytoplasmic Membrane	C	-	-	-	6.9	8.8	-	94.8
Adeh_0909	hypothetical protein	4 motifs	Periplasmic	none	-	-	-	10.0	-	-	-
Adeh_0910	NrfA Nitrite reductase (cytochrome; ammonia-forming)	5 motifs	Periplasmic	P	-	-	-	3.0	-	-	-
Adeh_0911	NrfH cytochrome c-type protein	4 motifs	Unknown	C	-	-	-	-	-	-	-
Adeh_0918	hypothetical protein	1 motif	Periplasmic	none	-	-	-	-	5.1	23.8	-
Adeh_0972	cytochrome c family protein	8 motifs	Periplasmic	none	-	14.1	55.3	5.7	-	-	6.5
Adeh_1172	cytochrome C oxidase, mono-heme subunit/FixO	2 motifs	Unknown	C	6.3	68.2	90.0	71.8	49.9	49.3	30.0
Adeh_1278	cytochrome c, class I	3 motifs	Periplasmic	C	-	-	74.5	-	-	-	-
Adeh_1422	hypothetical protein	4 motifs	Unknown	none	-	-	-	18.7	21.9	16.6	-
Adeh_1425	cytochrome c, class I	1 motif	Periplasmic	C	-	6.3	14.5	322.6	262.6	209.6	106.2
Adeh_1504	hypothetical protein putative cyt. c peroxidase precursor	3 motifs	Unknown	none	-	-	-	-	-	-	-

Table 3.4. Continued

Chromosome Locus ID	GenBank Annotation	Predicted C-type cytochrome motifs	Cellular Localization	COG	FeOOH	Fe Citrate	MnO_2	Nitrate	TSB	Fumarate	Oxygen
Adeh_1659	cytochrome c family protein	6 motifs	Cytoplasmic Membrane	none	-	6.5	-	-	-	4.8	-
Adeh_1696	cytochrome c3	3 motifs	Periplasmic	none	-	-	-	15.7	-	-	-
Adeh_1697	cytochrome c, class III	3 motifs	Unknown	none	-	-	-	30.8	98.3	285.4	-
Adeh_1719	hypothetical protein	1 motif	Unknown	none	-	10.4	31.7	6.6	11.3	9.4	12.6
Adeh_1764	hypothetical protein	8 motifs	Periplasmic	none	-	1.1	3.2	150.1	1.2	20.3	13.6
Adeh_1891	hypothetical protein	3 motifs	Unknown	none	-	-	-	2.9	-	5.5	-
Adeh_2002	cytochrome c family protein	9 motifs	Periplasmic	none	-	8.6	-	1.8	16.4	22.5	-
Adeh_2004	hypothetical protein	3 motifs	Periplasmic	none	-	9.4	-	-	-	8.4	-
Adeh_2097	hypothetical protein	7 motifs	Periplasmic	none	-	-	-	-	-	-	-
Adeh_2169	hypothetical protein preceeds Nitrate reductase alpha subunit	1 motifs	Unknown	none	-	-	7.0	8.7	-	3.1	-
Adeh_2216	hypothetical protein	10 motifs	Extracellular	none	-	62.2	57.8	68.4	42.8	15.6	45.4
Adeh_2269	hypothetical protein	3 motifs	Unknown	none	-	-	-	-	-	-	-
Adeh_2275	cyt c oxidase, subunitII	1 motif	Cytoplasmic Membrane	C	-	1.5	-	12.3	12.6	9.8	31.6
Adeh_2277	cytochrome c, class I	1 motif	Periplasmic	none	-	-	-	55.4	29.3	-	164.1
Adeh_2280	monoheme cytochrome SoxX (sulfur oxidation)	1 motifs	Unknown	none	10.1	-	21.5	59.0	77.4	51.1	285.3
Adeh_2281	diheme cytochrome SoxA (sulfur oxidation)	1 motifs	Unknown	none	-	-	-	25.9	31.6	16.9	141.4
Adeh_2285	sulfur dehydrogenase subunit SoxD cytochrome c, class I	2 motifs	Periplasmic	C	-	-	-	11.1	13.0	8.3	76.9
Adeh_2403	hypothetical protein preceeds Nitrous oxide reductase	5 motifs	Cytoplasmic	none	55.1	17.4	14.7	-	-	4.1	-
Adeh_2603	hypothetical protein	10 motifs	Unknown	none	-	5.6	6.5	1.3	4.7	-	6.2
Adeh_2621	hypothetical protein	2 motifs	Unknown	none	-	-	-	-	-	-	4.3
Adeh_2664	hypothetical protein	9 motifs	Periplasmic	none	-	-	-	-	-	-	-
Adeh_2665	cytochrome c family protein	10 motifs	Periplasmic	none	-	-	-	-	-	-	-

Table 3.4. Continued

Chromosome Locus ID	GenBank Annotation	Predicted C-type cytochrome motifs	Cellular Localization	COG	FeOOH	Fe Citrate	MnO_2	Nitrate	TSB	Fumarate	Oxygen
Adeh_2666	hypothetical protein	20 motifs	Periplasmic	none	-	-	-	-	-	-	-
Adeh_2682	hypothetical protein cytochrome c-554 homolog	4 motifs	Unknown	none	-	-	-	3.6	-	1.4	3.0
Adeh_2757	hypothetical protein	4 motifs	Unknown	C	-	-	-	-	-	-	10.2
Adeh_2815	cytochrome c like protein	13 motifs	Unknown	none	-	-	5.0	0.9	3.0	9.7	1.4
Adeh_2816	cytochrome c like protein	14 motifs	Periplasmic	none	-	5.1	10.7	6.9	4.3	6.0	0.4
Adeh_2902	NrfA Nitrite reductase (cytochrome; ammonia-forming) EC:1.7.2.2	5 motifs	Periplasmic	P	344.0	117.0	67.0	19.8	41.4	45.0	30.6
Adeh_2903	NrfH cytochrome c-type protein	4 motifs	Unknown	C	-	-	-	-	-	-	-
Adeh_2907	hypothetical protein	10 motifs	Unknown	none	34.0	164.9	113.7	17.6	194.3	288.3	34.4
Adeh_2927	hypothetical protein cytochrome c553 family protein in Methylococcus capsulatus	9 motifs	Unknown	none	6.4	2.9	18.3	6.9	-	-	1.6
Adeh_2940	hypothetical protein putative di-heme cytochrome c peroxidase	2 motifs	Unknown	P	-	-	-	-	-	-	-
Adeh_2963	cytochrome c, putative	7 motifs	Periplasmic	none	146.6	268.9	17.9	9.0	10.3	189.0	18.2
Adeh_2966	hypothetical protein	2 motifs	Unknown	none	-	-	-	2.7	-	-	70.1
Adeh_2967	putative lipoprotein	1 motif	Periplasmic	none	-	-	-	-	-	-	75.9
Adeh_2989	hypothetical protein	4 motifs	Periplasmic	none	-	46.0	69.2	66.0	74.5	87.5	-
Adeh_3065	cytochrome c family protein	6 motifs	Unknown	none	-	1.2	-	-	-	-	-
Adeh_3067	cytochrome c family protein	16 motifs	Periplasmic	none	-	-	-	-	-	0.6	-
Adeh_3068	cytochrome c family protein	26 motifs	Periplasmic	none	-	-	-	-	-	-	-
Adeh_3074	cytochrome c family protein	6 motifs	Extracellular	none	-	-	-	-	-	-	-
Adeh_3077	Multihaem cytochrome	40 motifs	Unknown	none	-	1.9	1.6	0.4	0.6	2.6	-
Adeh_3086	cytochrome c family protein	4 motifs	Unknown	none	-	-	-	-	-	-	-
Adeh_3090	hypothetical protein	8 motifs	Periplasmic	none	-	7.6	33.6	-	-	-	-
Adeh_3151	putative lipoprotein	2 motifs	Unknown	P	-	-	-	-	-	-	-
Adeh_3163	hypothetical protein	10 motifs	Extracellular	none	18.4	4.8	-	-	-	-	-

Table 3.4. Continued

Chromosome Locus ID	GenBank Annotation	Predicted C-type cytochrome motifs	Cellular Localization	COG	FeOOH	Fe Citrate	MnO_2	Nitrate	TSB	Fumarate	Oxygen
Adeh_3200	hypothetical protein	10 motifs	Unknown	none	-	-	-	-	-	-	-
Adeh_3392	hypothetical protein	12 motifs	Periplasmic	none	-	-	2.4	-	-	3.2	-
Adeh_3775	cytochrome c family protein	9 motifs	Periplasmic	none	-	-	-	-	-	-	1.4

Cellualr localizations are predicted by Psort 3.0

Numbers in cells are normalized spectral counts (nSpC).

Red: Core *c*-type cytochromes

Yellow: Undetected *c*-type cytochromes

Orange: Detected under all but FeOOH

Green: Unique *c*-type cytochromes

Purple: Detected under two conditions

Blue: Detected under all but one condition

Adeh_0409 may belong to a conserved uncharacterized protein family. Adeh_0656 shares sequence similarity with split-soret cytochrome *c* proteins in other organisms such as *Desulfovibrio desulfuricans, Archaeoglobus veneficus, Thermodesulfovibrio yellowstonii, Thermincola potens, Melioribacter roseus*, etc. These cytochromes are characteristic of possessing an unusual split at the Soret band in their visible absorption spectrum [192]. While the function of split-soret proteins is unclear, da Silva *et al.* proposed that the split-soret protein of *D. desulfuricans* is the electron input point for the electron transfer chain in the reduction of nitrate [193]. However, our prediction shows that Adeh_0656 is an extracellular *c*-type cytochrome, which is more likely a terminal reductase than an electron input point.

Adeh_1172 is annotated as the subunit II of the bifunctional cbb_3-type cytochrome *c* oxidase, which belongs to the heme-copper oxidase (HCO) superfamily. In many organisms, the protein complex of cbb_3-type cytochrome *c* oxidase is encoded by the *ccoNOPQ* operon [194, 195]. The genes encoding for the subunit II and III (or subunit O and P) are named *ccoO* and *ccoP*. Typically, CcoO and CcoP contain one and two heme-binding sites, respectively. However, Adeh_1172 (CcoO) contains two CXXCH motifs. There is no other protein of *A. dehalogenans* annotated as other subunit of cbb_3-type HCO (e.g. CcoN, CcoP, etc.). However, the monoheme cytochrome Adeh_0799 contains a cbb_3 subunit III (CcoP) domain, and was only detected under oxygen amended condition (Figure 3.1, Table 3.4). We speculate that under low oxygen conditions, Adeh_1172 and Adeh_0799 form a protein complex which functions similarly as a cbb_3-type HCO for oxygen reduction and proton pumping, while under anoxic condition, Adeh_1172 works as electron transfer intermediate which receives and/or delivers electrons from/to other cytochromes in the electron transfer chain. In strain MR-1, we have also

detected the expression of the *ccoNOPQ* operon. The CcoO (SO_2363) and CcoP (SO_2361) in strain MR-1 were expressed under all growth conditions except for FeOOH (Table 3.3). We suspect its absence is due to the poor overall cellular growth in FeOOH amended culture. CcoN (SO_2364) was only detected under oxygen-amended condition (data not shown).

3.3.3 Unique *c*-Type Cytochromes

In contrast to the "core" *c*-type cytochromes, some *c*-type cytochromes were only detected under one or a few growth conditions. Twenty *c*-type cytochromes of strain 2CP-C were identified as unique cytochromes (Table 3.5, Figure 3.1). Adeh_0909, Adeh_0910 (NrfA), and Adeh_1696 were unique to nitrate grown cultures (Table 3.5, Figure 3.1). Six *c*-type cytochromes of strain 2CP-C were unique to growth under oxic conditions, among which, Adeh_0799 and Adeh_2967 (lipoprotein) were detected at a normalized spectral count (nSpC) of 34 and 76, respectively, which were very high levels compared to other unique *c*-type cytochromes (Table 3.5, Figure 3.1). Adeh_1278 was only expressed when MnO_2 was provided as electron acceptor (nSpC: 74) (Table 3.5, Figure 1). Likewise, unique *c*-type cytochromes were also revealed for strain MR-1 and FRC-32 with various expression levels (Table 3.3 and 3.6, Figure 3.2 and 3.3).

Adeh_0909 and Adeh_0910 are both predicted *c*-type cytochromes that were detected only in nitrate-grown *A. dehalogenans* cells (Table 3.5, Figure 3.1). Adeh_0909 is a protein with unknown function, while Adeh_0910 encodes for nitrite reductase (NrfA). Adeh_0910 is not identical to the other copy of NrfA encoded by Adeh_2902 (44% sequence identity), which was detected as a "core" cytochrome. The observation of Adeh_0910 only in nitrate grown cells

Table 3.5. The unique *c*-type cytochromes of *A. dehalogenans* 2CP-C expressed only in one growth condition

Gene ID	GeneBank Annotation	Motifs	Cellular Localization	COG	Electron Acceptor	nSpC
Adeh_0795	hypothetical protein	5 motifs	Unknown	none	oxygen	4
Adeh_0799	hypothetical protein	1 motif	Periplasmic	none	oxygen	34
Adeh_0909	hypothetical protein	4 motifs	Periplasmic	none	nitrate	10
Adeh_0910	respiratory nitrite reductase	5 motifs	Periplasmic	P	nitrate	3
Adeh_1278	cytochrome *c*, class I	3 motifs	Periplasmic	C	MnO_2	74
Adeh_1696	cytochrome c_3	3 motifs	Periplasmic	none	nitrate	16
Adeh_2621	hypothetical protein	2 motifs	Unknown	none	oxygen	4
Adeh_2757	hypothetical protein	4 motifs	Unknown	C	oxygen	10
Adeh_2967	lipoprotein	1 motif	Periplasmic	none	oxygen	76
Adeh_3065	NHL repeat-containing protein	6 motifs	Unknown	none	Fe citrate	1
Adeh_3067	cytochrome *c* family protein	16 motifs	Periplasmic	none	fumarate	1
Adeh_3775	cytochrome *c* family protein	9 motifs	Periplasmic	none	oxygen	1

nSpC: normalized spectral counts.

Table 3.6. The unique *c*-type cytochromes of *G. daltonii* FRC-32 expressed only in one growth condition

Gene ID	GenBank Annotation	Motifs	Cellular Localization Predicted by Psort	Electron Acceptor	nSpC
Geob_3176	hypothetical protein	7	Unknown	Fe citrate	1
Geob_1834	cytochrome c family protein	37	Unknown	MnO_2	1
Geob_3741	cytochrome c family protein	44	Unknown	fumarate	1
Geob_3740	cytochrome c family protein	35	Unknown	fumarate	2
Geob_3739	cytochrome c family protein	53	Unknown	fumarate	1
Geob_3132	cytochrome c family protein	36	Unknown	fumarate	0.46
Geob_2641	multiheme cytochrome	8	Unknown	fumarate	13
Geob_2640	hypothetical protein	2	Unknown	fumarate	7
Geob_1885	hypothetical protein	6	Unknown	fumarate	5
Geob_1862	hypothetical protein	10	Unknown	fumarate	5
Geob_1686	hypothetical protein	11	Unknown	fumarate	14
Geob_1681	hypothetical protein	8	Periplasmic	fumarate	7
Geob_1513	cytochrome c family protein	36	Unknown	fumarate	1
Geob_1430	cytochrome c family protein	8	Periplasmic	fumarate	7
Geob_0830	hypothetical protein	12	Periplasmic	fumarate	25
Geob_0764	hypothetical protein	6	Unknown	fumarate	4
Geob_0763	hypothetical protein	6	Unknown	fumarate	19

nSpC: normalized spectral counts.

suggests a role as a terminal nitrite reductase rather than a member of a more extensive electron transfer chain. Adeh_0909 and Adeh_0910 are located on the same operon, together with three other genes: Adeh_0907 (hypothetical protein), Adeh_0908 (NapA), and Adeh_0911 (NrfH). Adeh_0907 was also only detected in cells grown with nitrate (data not shown), Adeh_0908 was detected when grown with nitrate, ferric iron (FeOOH, Fe citrate), or MnO_2 (data not shown), and Adeh_0911 was not detected under any growth condition (Table 3.5). It is likely that Adeh_0907- 0910 are co-expressed when nitrate is the sole electron acceptor and are all involved in nitrite reduction to ammonium. When the electron acceptor is a metal, such as Fe(III) and Mn(IV), only NapA (Adeh_0908) was expressed.

Adeh_1278 contains three heme-binding motifs and is detected only when MnO_2 is the electron acceptor. This cytochrome *c* shares sequence similarity (e-value of 4e-39) to SO_4047, a SoxA-like diheme cytochrome *c*. Previous work indicated that SO_4047 is important for aerobic growth and Cr(VI) reduction [196]. Proteomic measurements showed that SO_4047 was expressed under all but FeOOH condition, and its expression level was highest under the low oxic condition (Table 3.3). SO_4047 is located in the same operon as SO_4048, which is also expressed under all but FeOOH condition (Table 3.3). This agrees with previous finding that the expression of SO_4047 and SO_4048 are not impacted by different environmental electron acceptors [179], although the specific functions of SO_4047 and SO_4048 remain unknown. There are two other predicted *c*-type cytochromes of strain 2CP-C annotated as SoxA (Adeh_2281) and SoxX (Adeh_2280). SoxA is detected under all growth conditions except metal growth conditions (Fe citrate, FeOOH, MnO_2) and SoxX is detected under all but Fe Citrate condition (Figure 3.1, Table 3.4). SoxA and

SoxX are postulated to form a complex for the binding of substrate to another carrier protein during sulfur oxidation [197]. Our result suggests that SoxA and SoxX of strain 2CP-C may also be involved in electron transfer in an independent manner, since they are not always co-expressed.

3.3.4 Non-detected *c*-Type Cytochromes

In strain 2CP-C cells, the 16 undetected *c*-type cytochromes include eight proteins annotated as hypothetical proteins, five annotated as cytochrome *c*, one as lipoprotein, and two as NrfH nitrite reductases (Figure 3.1, Table 3.7). The protein sequences of all 16 undetected *c*-type cytochromes of *A. dehalogenans* were inspected for their detectability by mass spectrometry. All of the undetected *c*-type cytochromes should yield tryptic peptides upon proteolysis with appropriate length and hydrophobicity for mass spectrometry detection. Thus, if these were expressed, they should have been detectable under our experimental conditions; however, we did not detect these proteins under any of the seven growth conditions (in any of the triplicates). Possible reasons could be that the expression levels of these proteins were too low for detection, or the proteins were not expressed under these growth conditions or in planktonic cells. For example, we did not detect MtrD and MtrF in *S. oneidensis* cells (Table 3.3), which agrees with previous findings that the expression of MtrD and MtrF is more favorable in cells forming biofilms rather than planktonic cells [198]. It is possible that some of the undetected *c*-type cytochromes have unique functions in biofilm cells and thus were not detected in our experiments. Another possible reason could be that the expression of some undetected proteins is growth phase dependent. Since all the cells were

Table 3.7. The undetected *c*-Type cytochromes for *A. dehalogenans* 2CP-C

Locus ID	Annotation	Motif	Cellular Localization	COG
Adeh_0108	hypothetical protein	12 motifs	Periplasmic	none
Adeh_0215	multiheme cytochrome	7 motifs	Periplasmic	none
Adeh_0911	respiratory nitrite reductase specific menaquinol--cytochrome-c reductase (NrfH)	4 motifs	Unknown	C
Adeh_1504	hypothetical protein	3 motifs	Unknown	none
Adeh_2097	hypothetical protein	7 motifs	Periplasmic	none
Adeh_2269	hypothetical protein	3 motifs	Unknown	none
Adeh_2664	hypothetical protein	9 motifs	Periplasmic	none
Adeh_2665	cytochrome c family protein	10 motifs	Periplasmic	none
Adeh_2666	hypothetical protein	20 motifs	Periplasmic	none
Adeh_2903	respiratory nitrite reductase specific menaquinol--cytochrome-c reductase (NrfH)	4 motifs	Unknown	C
Adeh_2940	hypothetical protein	2 motifs	Unknown	P
Adeh_3068	cytochrome c family protein	26 motifs	Periplasmic	none
Adeh_3074	cytochrome c family protein	6 motifs	Extracellular	none
Adeh_3086	cytochrome c family protein	4 motifs	Unknown	none
Adeh_3151	lipoprotein	2 motifs	Unknown	P
Adeh_3200	hypothetical protein	10 motifs	Unknown	none

harvested in late exponential / early stationary phase in our experiments, some c-type cytochromes may have been expressed earlier on and have been degraded by the time cells reached harvest time. Additionally, only seven growth conditions were tested in this study, which does not include all possible electron acceptor conditions these organisms can experience.

NrfH is believed to interact with NrfA during electron transfer. It is unclear why both copies of NrfH (Adeh_0911 and Adeh_2903) are not detected (Table 3.7), considering both copies of NrfA are detected. However, NrfH is not present in the genome of *S. oneidensis*; instead, the 'NrfA function' is replaced by CymA (SO_4591) [196]. It is likely that strain 2CP-C also possesses an atypical Nrf system similar to that of strain MR-1. Thus, both copies of NrfH in strain 2CP-C may not be expressed, and other proteins may replace their function.

3.3.5 Heme-binding Motifs of *c*-Type Cytochromes

Most of the predicted *c*-type cytochromes contained more than one heme-binding motif per protein (Figure 3.4). *c*-Type cytochromes of strain 2CP-C and FRC-32 were predicted to contain as many as 40 and 53 motifs, respectively, whereas in *S. oneidensis*, the highest number of heme-binding motifs for a *c*-type cytochrome was 10 (Figure 3.4). On average, the cytochromes in strain 2CP-C, MR-1, and FRC-32 contain 6, 4, and 12 CXXCH motifs (Figure 3.4d).

Monoheme *c*-type cytochromes constituted 25% of all the predicted *c*-type cytochromes of strain MR-1, which was the highest percentage among all strains (Figure 3.4). However, the measured abundances of monoheme *c*-type cytochromes of strain MR-1 culture were not as

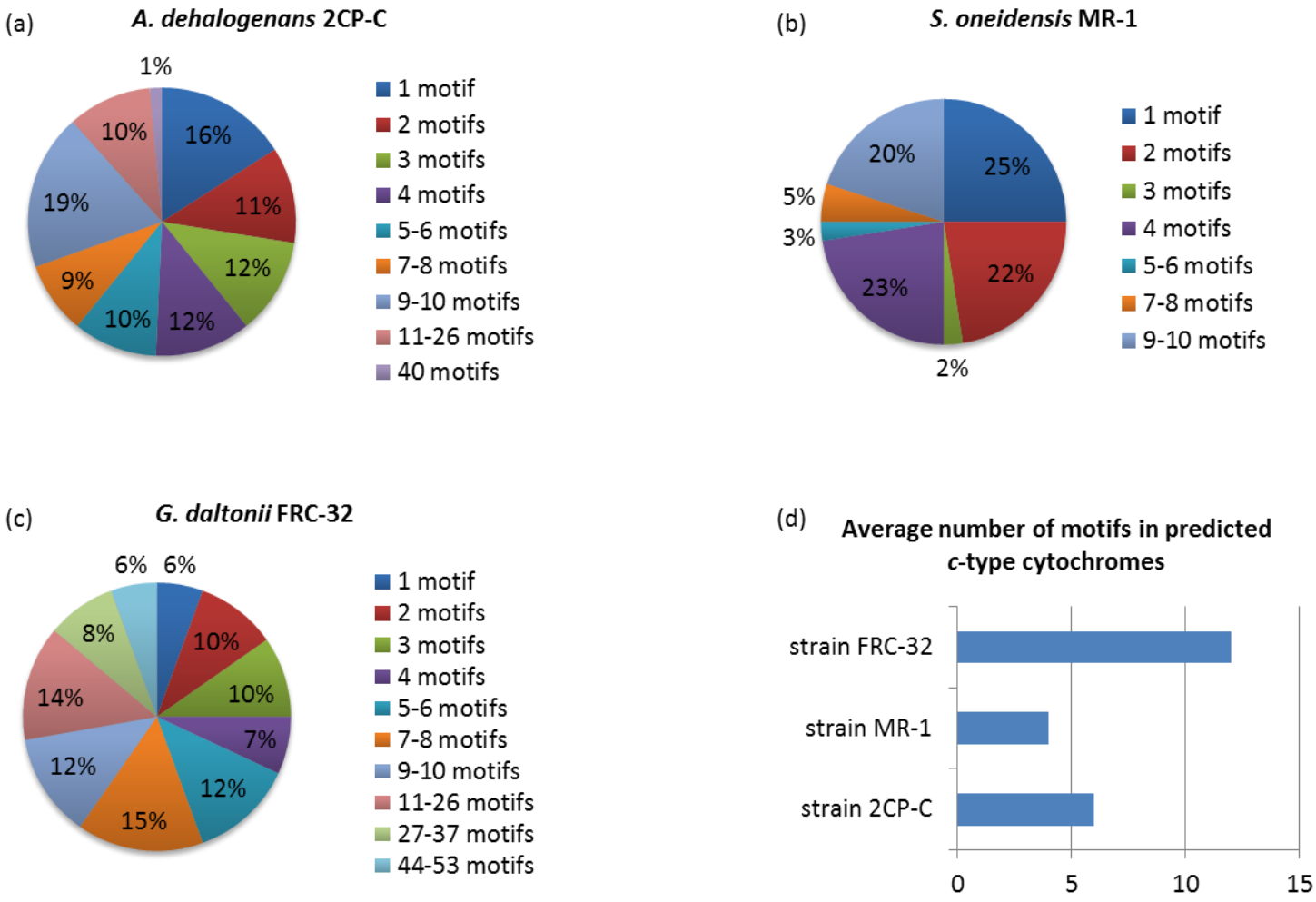


Figure 3.4. Heme-binding motifs of predicted c-type cytochromes of *Anaeromyxobacter dehalogenans* 2CP-C (a), *Shewanella oneidensis* MR-1 (b), and *Geobacter daltonii* FRC-32 (c). Average number of motifs in predicted *c*-type cytochromes (d) were calculated by multiply the number of motif with its corresponding percentage, and sum all the multiplications for each organism.

high as their counterparts in strain 2CP-C culture, especially in the oxygen grown cells (Figure 3.5). In strain 2CP-C cells grown with ferric iron (FeOOH, Fe Citrate) or MnO_2, monoheme *c*-type cytochromes were less abundant compared to other growth conditions (nitrate, TSB, or fumarate) (Figure 3.5). This abundance difference of monoheme *c*-type cytochromes in metal and non-metal growth conditions was not observed for strain MR-1 and FRC-32 (Figure 3.5).

Among all the predicted multiple heme-containing *c*-type cytochromes, tetraheme cytochromes were the largest proportion in strain MR-1 (23%), while in strain 2CP-C and MR-1, only 12% and 7% of the predicted *c*-type cytochromes had four predicted heme-binding motifs, respectively (Figure 3.4). However, the percentage of tetraheme *c*-type cytochromes expression in strain FRC-32 was similar to strain MR-1*,* as shown by cumulative normalized spectral counts (nSpC) (Figure 3.5). Both strain MR-1 and FRC-32 cultures were dominated by tetraheme *c*-type cytochromes under all growth conditions, while in strain 2CP-C tetraheme *c*-type cytochromes were only measured in low abundances under a few growth conditions (e.g. FeOOH, Fe Citrate, and oxygen) (Figure 3.5, Table 3.4).

In strain MR-1, only 28% predicted *c*-type cytochromes contained more than 4 heme-binding motifs, whereas 49% and 67% predicted *c*-type cytochromes with more than 4 heme-binding motifs were identified in strain 2CP-C and FRC-32, respectively (Figure 3.4). Even though the percentage/distribution of predicted >4 heme-containing *c*-type cytochromes in strain 2CP-C was not the highest among all three populations, their detection level was the most abundant among all organisms, as shown by the cumulative nSpC under all growth conditions (Figure 3.5). Under the metal conditions (FeOOH, Fe Citrate, and MnO_2), about 50 – 80% of the cumulative nSpC of the detected *c*-type cytochromes contained more than 4 heme-

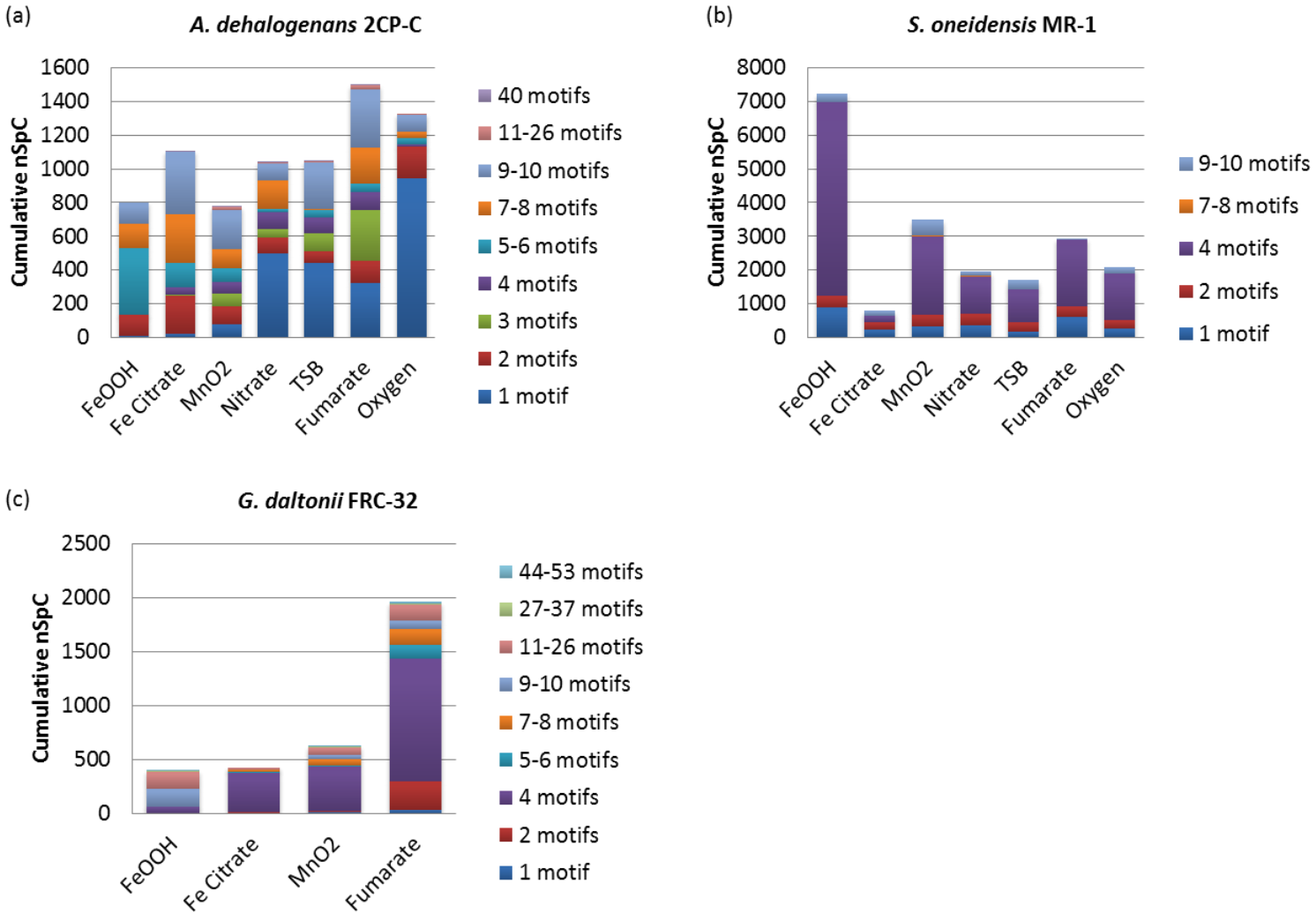


Figure 3.5. Abundances of detected *c*-type cytochromes grouped by the number of heme-binding motifs for (a.) *A. dehalogenans* 2CP-C, (b.) *S. oneidensis* MR-1 and (c.) *G. daltonii* FRC-32. Abundances are represented by cumulative nSpC of detected *c*-type cytochromes. nSpc: normalized spectral count.

binding motifs (Figure 3.5). Cytochromes with 5-6 heme-binding motifs were not detected under any growth condition for strain MR-1; however, their expression was detected in FeOOH and Fe Citrate grown strain 2CP-C cells (Figure 3.5). We detected 7-10 heme-containing cytochromes at high abundances in all the growth conditions of strain 2CP-C but at very low abundances in strain MR-1 and FRC-32 (Figure 3.5). *c*-Type cytochromes of high heme content (>10) were only present in strain 2CP-C and FRC-32 cells (Figure 3.4). Cytochromes with 11-26 heme-binding motifs were detected in very low abundance in strain 2CP-C cells but were significantly more abundant in fumarate grown FRC-32 (Figure 3.5).

Both strain 2CP-C and FRC-32 have *c*-type cytochromes with more than 25 heme-binding sites based on computational predictions (Figure 3.4). For example, the highest number of hemes per cytochrome in strain 2CP-C was predicted to be 40 (Adeh_3077), and in strain FRC-32 was 53 (Geob_3739). The LC-MS/MS measurement yielded definitive identification of the expression of Adeh_3077 under all but FeOOH and oxygen conditions (Figure 3.1, Figure 3.5), whereas Geob_3739 was only detected in fumarate grown culture (Table 3.6). There were 10 *c*-type cytochromes in strain FRC-32 containing more than 26 heme-binding motifs, only 1 of which was not detected (Geob_2170). The remaining 9 were detected under various conditions (Figure 3.5, Table 3.8).

3.3.6 Cellular Localization of *c*-Type Cytochromes

Computational prediction of cellular localization revealed that most of the putative *c*-type cytochromes of strain 2CP-C (45%), MR-1 (55%), and FRC-32 (26%) were periplasmic (Figure 3.6 left). While the *c*-type cytochromes of strain FRC-*32* were predicted to localize at all

Table 3.8. *c*-type cytochrome identification for *Geobacter daltonii* FRC-32 grown with various electron acceptors.

Gene Locus	Annotation	Heme-binding sites	Cellular Localization	COG	FeOOH	FeCitrate	MnO_2	Fumarate
Geob_3741	cytochrome C family protein	44	Unknown	-	-	-	-	0.9
Geob_3740	cytochrome C family protein	35	Unknown	-	-	-	-	2.1
Geob_3739	cytochrome C family protein	53	Unknown	-	-	-	-	1.3
Geob_3686	hypothetical protein	9	Cytoplasmic Membrane	-	-	-	1.0	11.5
Geob_3403	multiheme cytochrome	7	Unknown	-	-	-	-	-
Geob_3347	cytochrome C family protein	21	Unknown	-	116.5	16.3	52.4	69.4
Geob_3176	hypothetical protein	7	Unknown	-	-	0.9	-	-
Geob_3166	hypothetical protein	6	Extracellular	-	-	-	-	-
Geob_3160	cytochrome C family protein	26	Periplasmic	-	-	-	-	-
Geob_3159	cytochrome C family protein	16	Periplasmic	-	-	-	-	-
Geob_3155	hypothetical protein	6	Unknown	-	-	-	-	-
Geob_3132	cytochrome C family protein	36	Unknown	-	-	-	-	0.5
Geob_3111	formate-dependent nitrite reductase periplasmic cytochrome c552 subunit-like protein	7	Periplasmic	P	-	-	8.3	93.0
Geob_2912	cytochrome C family protein	10	Periplasmic	-	-	-	-	-
Geob_2704	cytochrome C family protein	21	Unknown	-	45.2	2.1	13.1	7.8
Geob_2641	multiheme cytochrome	8	Unknown	-	-	-	-	13.0
Geob_2640	hypothetical protein	2	Unknown	-	-	-	-	7.1
Geob_2579	Di-heme cytochrome c peroxidase	2	Periplasmic	P	-	-	1.3	155.2
Geob_2568	hypothetical protein	1	Unknown	-	-	-	12.0	35.3
Geob_2379	hypothetical protein	12	Unknown	-	-	-	1.8	26.9
Geob_2338	hypothetical protein	2	Unknown	-	-	-	-	-
Geob_2235	cytochrome c peroxidase	2	Periplasmic	P	-	1.2	-	57.1
Geob_2175	hypothetical protein	8	Unknown	-	-	2.1	6.7	-
Geob_2170	hypothetical protein	27	Periplasmic	-	-	-	-	-
Geob_1966	hypothetical protein	3	Unknown	-	-	-	-	-
Geob_1885	hypothetical protein	6	Unknown	-	-	-	-	4.8

Table 3.8. Continued

Gene Locus	Annotation	Heme-binding sites	Cellular Localization	COG	FeOOH	FeCitrate	MnO_2	Fumarate
Geob_1862	hypothetical protein	10	Unknown	-	-	-	-	4.7
Geob_1861	hypothetical protein	10	Extracellular	-	-	-	1.9	4.3
Geob_1834	cytochrome C family protein	37	Unknown	-	-	-	0.9	-
Geob_1813	cytochrome c3	3	Unknown	-	-	-	-	-
Geob_1740	hypothetical protein	4	Periplasmic	-	-	-	-	-
Geob_1737	hypothetical protein	7	Extracellular	-	-	-	-	-
Geob_1735	PKD domain-containing protein	7	Extracellular	-	-	-	-	-
Geob_1724	cytochrome c family protein	9	Periplasmic	-	-	-	1.0	2.9
Geob_1686	hypothetical protein	11	Unknown	-	-	-	-	14.3
Geob_1685	hypothetical protein	9	Unknown	-	-	-	-	-
Geob_1682	hypothetical protein	9	Unknown	-	163.6	1.1	33.1	59.1
Geob_1681	hypothetical protein	8	Periplasmic	-	-	-	-	6.7
Geob_1594	hypothetical protein	5	Periplasmic	-	-	-	-	-
Geob_1591	hypothetical protein	9	Unknown	-	-	-	-	-
Geob_1513	cytochrome C family protein	36	Unknown	-	-	-	-	1.4
Geob_1432	hypothetical protein	4	Unknown	-	56.8	365.7	410.5	1139.8
Geob_1430	cytochrome C family protein	8	Periplasmic	-	-	-	-	7.1
Geob_1429	cytochrome C family protein	11	Periplasmic	-	-	-	-	-
Geob_1426	hypothetical protein	3	Periplasmic	-	-	-	-	-
Geob_1401	hypothetical protein	3	Unknown	-	-	-	-	-
Geob_1374	cytochrome C family protein	45	Unknown	-	2.3	-	6.0	9.0
Geob_1373	cytochrome C family protein	27	Unknown	-	8.9	-	2.9	7.3
Geob_1359	hypothetical protein	2	Unknown	-	-	-	-	-
Geob_1348	hypothetical protein	4	Unknown	-	-	-	-	-
Geob_1344	hypothetical protein	10	Periplasmic	-	-	-	-	-
Geob_1343	hypothetical protein	12	Cytoplasmic	-	-	-	-	-
Geob_1340	hypothetical protein	5	Unknown	-	-	-	2.8	4.2

Table 3.8. Continued

Gene Locus	Annotation	Heme-binding sites	Cellular Localization	COG	FeOOH	FeCitrate	MnO_2	Fumarate
Geob_1067	hypothetical protein	2	Unknown	-	-	13.7	6.8	39.2
Geob_1050	cytochrome C family protein	44	Outer Membrane	-	1.8	-	1.3	8.7
Geob_1015	hypothetical protein	4	Unknown	-	-	-	-	-
Geob_1014	hypothetical protein	7	Cytoplasmic Membrane	-	-	11.2	45.0	20.5
Geob_0874	hypothetical protein	1	Cytoplasmic Membrane	-	-	-	-	-
Geob_0870	cytochrome C oxidase mono-heme subunit/FixO	2	Unknown	C	-	-	-	-
Geob_0868	hypothetical protein	5	Cytoplasmic Membrane	-	-	-	-	-
Geob_0830	hypothetical protein	12	Periplasmic	-	-	-	-	24.8
Geob_0764	hypothetical protein	6	Unknown	-	-	-	-	4.4
Geob_0763	hypothetical protein	6	Unknown	-	-	-	-	19.1
Geob_0619	hypothetical protein	3	Periplasmic	-	-	-	-	-
Geob_0616	hypothetical protein	3	Periplasmic	-	-	-	-	-
Geob_0517	hypothetical protein	8	Unknown	-	-	-	-	-
Geob_0311	cytochrome c3	3	Unknown	-	-	-	-	-
Geob_0309	cytochrome c family protein	5	Unknown	-	7.6	7.0	14.9	100.7
Geob_0301	hypothetical protein	1	Unknown	-	-	-	-	-
Geob_0122	cytochrome c class I	4	Periplasmic	-	-	-	-	-
Geob_0121	hypothetical protein	1	Unknown	-	-	-	-	-
Geob_0053	cytochrome C family protein	26	Outer Membrane	-	-	-	0.6	1.0

Cellualr localizations are predicted by Psort 3.0

Numbers in cells are normalized spectral counts (nSpC).

Red: Core *c*-type cytochromes

Yellow: Undetected *c*-type cytochromes

Orange: Detected under all but FeOOH

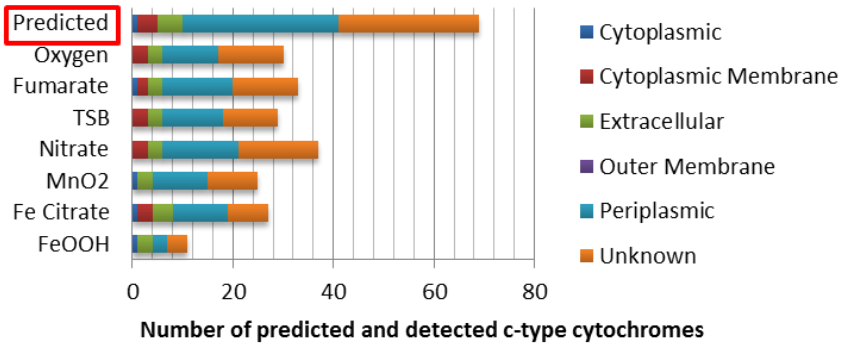
Green: Unique *c*-type cytochromes

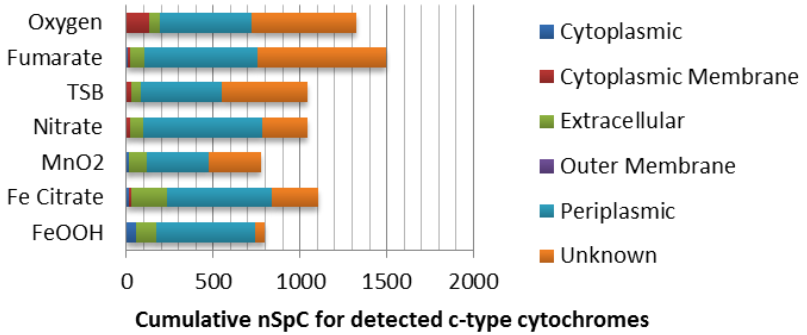
Purple: Detected under two conditions

Blue: Detected under all but one condition

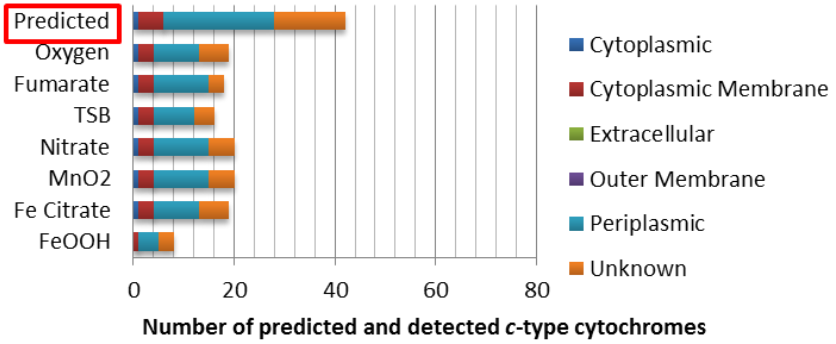
Figure 3.6. Cellular localization distribution for predicted and detected c-type cytochromes in *A. dehalogenans* 2CP-C, *S. oneidensis* MR-1 and *G. daltonii* FRC-32. Cellular localizations were predicted using Psortb v3.0. The left panel shows predicted localizations for all predicted and detected putative *c*-type cytochromes for all strains. The right panel shows the abundance of all the detected *c*-type cytochromes of all localizations. nSpC: normalized spectral count.

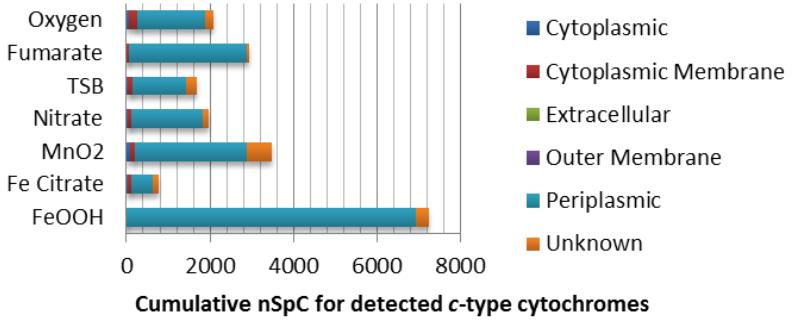
A. dehalogenans 2CP-C



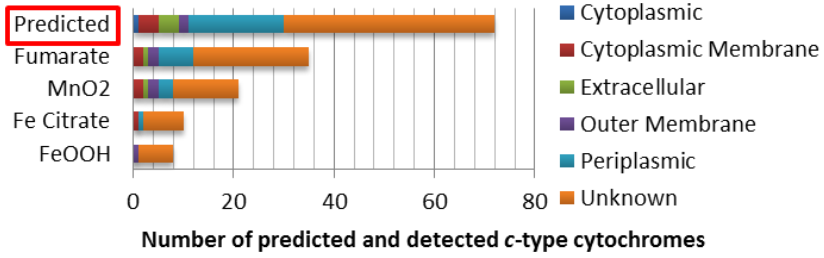


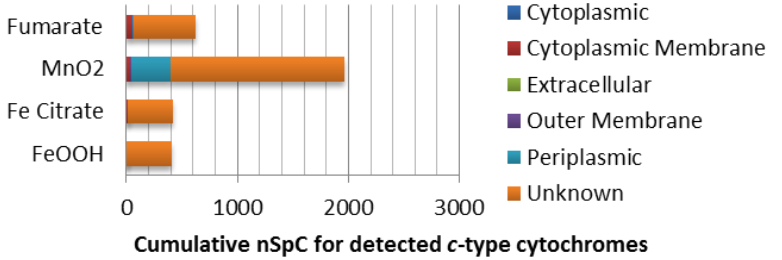
S. oneidensis MR-1





G. daltonii FRC-32





cellular localizations, no outer membrane *c*-type cytochrome were predicted for either strain 2CP-C or MR-1, and no extracellular *c*-type cytochrome were predicted for strain MR-1 (Figure 3.6 left). Among all the detected *c*-type cytochromes of all three cultures, the majority was predicted to be either periplasmic or with an unknown location (Figure 3.6 right).

There was one cytochrome for each bacteria predicted to be cytoplasmic (Figure 3.6 left, Table 3.3, 3.4 and 3.9), although previous study suggests that cytochromes mature outside of the cytoplasm [199, 200]. The one cytoplasmic cytochrome of strain MR-1, SO_2363 (CcoO), was detected under all growth conditions except for FeOOH, whereas the one of strain FRC-32, Geob_1343, was not detected under any growth condition (Table 3.3, 3.9). In strain 2CP-C cells, the cytoplasmic cytochrome Adeh_2403 was detected when FeOOH, Fe citrate, MnO_2, or fumarate was the electron acceptor, and its expression level under FeOOH condition (nSpC: 55) was about 3 to 11 fold of that under the other three conditions (Table 3.4).

The *c*-type cytochromes predicted to be associated with the cytoplasmic membrane were detected under almost all growth conditions for all three cultures (Figure 3.6). However, their abundances were very low, with the exception of oxygen grown strain 2CP-C cells, as shown by cumulative nSpC (Figure 3.6 right).

Strain 2CP-C and FRC-32 have similar numbers of detected extracellular cytochrome *c* (Figure 3.6 left). Only one extracellular *c*-type cytochrome of strain FRC-32 was found under MnO_2 and fumarate growth conditions at very low abundances with nSpC of 1.9 and 4.3, respectively (Figure 3.6 right, Table 3.8). In contrast, the extracellular *c*-type cytochromes of strain 2CP-C were detected at a relatively higher level with nSpCs ranging from 3.2 to 102.2 (Figure 3.6 right, Table 3.4).

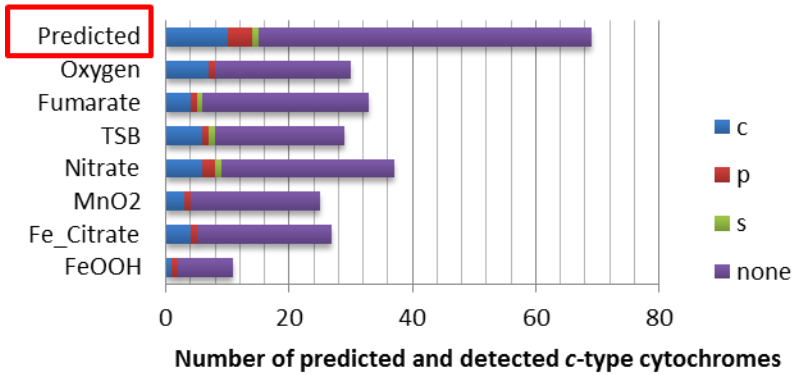
3.3.7 Clusters of Orthologous Groups (COG) Functional Categories of *c*-Type Cytochromes

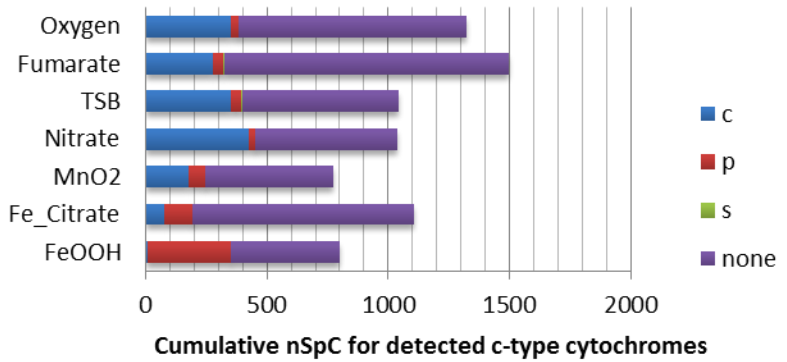
Most of the predicted *c*-type cytochromes of these bacteria could not be grouped into any functional category (none) (Figure 3.7 left). The detection of these *c*-type cytochromes dominated the cumulative nSpC of all *c*-type cytochromes in strain 2CP-C and FRC-32 cultures under all growth conditions (Figure 3.7 right). However, their abundance level in strain MR-1 was not as dominant (Figure 3.7 right).

The remainder of the *c*-type cytochromes of all three strains mainly fell into two COG functional categories: energy production and conversion (C), and inorganic ion transport and metabolism (P) (Figure 3.7). More "C" category *c*-type cytochromes were predicted and detected for *S. oneidensis* with higher abundances compared to that of strain 2CP-C and FRC-32 across all growth conditions (Figure 3.7). A few *c*-type cytochromes of each strain were grouped into the inorganic ion transport and metabolism category (P) (Figure 3.7 left). Overall, their expression levels were not as high as that of the "C" category *c*-type cytochromes or the "none" category *c*-type cytochromes. However, the "P" category *c*-type cytochromes are detected at high abundance in strain 2CP-C cells grown with FeOOH and strain FRC-32 cells grown with MnO_2 (Figure 3.7 right). One *c*-type cytochrome (Adeh_0728) of strain 2CP-C was grouped into the unknown function category (S) (Figure 3.7 left, Table 3.4). The "S" category means proteins are clustered into an orthologous group but the function for that group is poorly characterized.

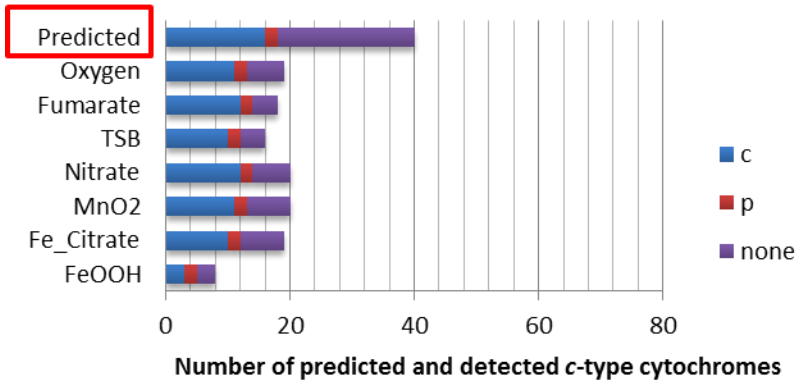
Figure 3.7. COG functional category distribution of predicted and detected *c*-type cytochromes in *A. dehalogenans* 2CP-C, *S. oneidensis* MR-1 and *G. daltonii* FRC-32 under various growth conditions. The left panel shows the COG functional categories of all predicted and detected *c*-type cytochromes. The right panel shows the abundance of detected *c*-type cytochromes in each COG category. Each alphabet represents one COG functional category. c (Energy production and conversion), p (Inorganic ion transport and metabolism), s (Unknown Function). nSpC: normalized spectral count.

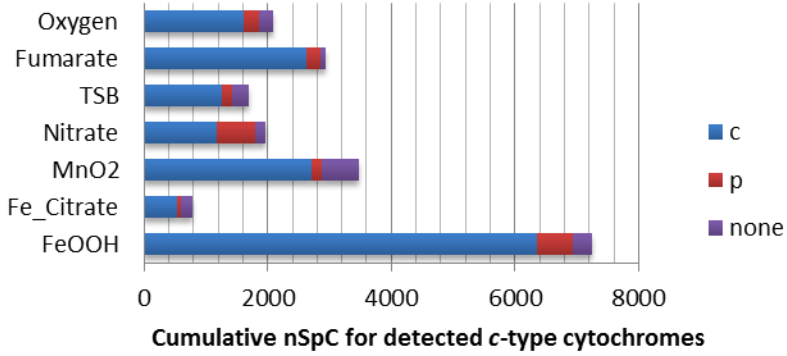
***A. dehalogenans* 2CP-C**



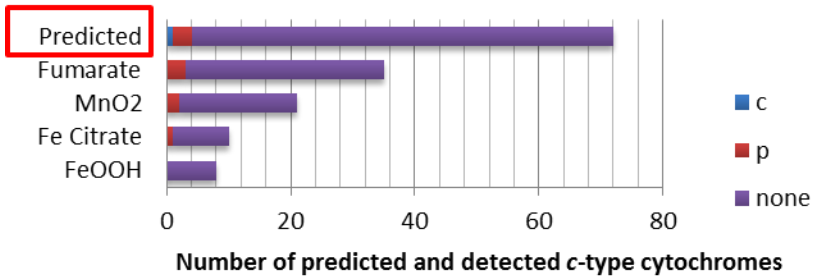


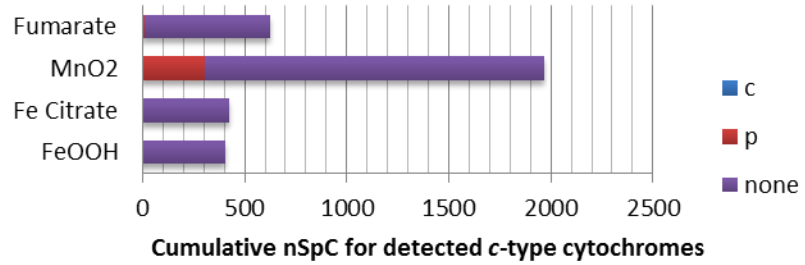
***S. oneidensis* MR-1**





***G. daltonii* FRC-32**





3.4 Discussion

The current models of electron transfer in *S. oneidensis* and *G. sulfurreducens* are constructed from information obtained from deletion mutants and heterologous expression studies under selective growth conditions [99, 102, 107]. In this study, cytochrome expression evaluated a wide range of additional electron acceptors, including solid metals (FeOOH, MnO_2), soluble metal (Fe citrate), and non-metal substrates (fumarate, nitrate, oxygen). In *S. oneidensis*, all the electron transfer models propose the inner membrane *c*-type cytochrome CymA as the initial electron carrier in the electron transfer chain, which receives electrons from the quinone pool and transfers them to periplasmic cytochromes [99]. During ferric iron reduction, the outer membrane protein complex containing *c*-type cytochrome OmcA and OmcB delivers electron to ferric iron [102, 201]. OmcA and OmcB were both core cytochromes expressed under all growth conditions tested, and CymA could tentatively be classified as a "core" *c*-type cytochrome since it was detected under all growth condition except in FeOOH-grown cultures (Table 3.3). Our results not only confirmed the role of OmcA, OmcA, and CymA in ferric iron reduction (Fe citrate), but also indicated their roles in electron transfer to other electron acceptors including nitrate, fumarate, oxygen, and manganese oxide. In strain MR-1, many other "core" *c*-type cytochromes (e.g. FccA, NrfA, etc.) were identified, but were not examined in detail in the scope of this study.

Based on the *c*-type cytochrome identification and cellular localizations predicted by PSORT, a special grouping of the *c*-type cytochromes network was constructed for strain 2CP-C (Table 3.9), which spatially segregates *c*-type cytochromes into cellular localizations. Clear commonalities and differences between growth conditions can be easily delineated by

highlighting the cytochromes detected (nitrate and fumarate growth shown as examples in Table 3.9 (a) and (b), respectively). With this table, one can begin to draw connections between *c*-type cytochromes and relate them to different electron acceptors. Future research could take this diagram as a starting point to build the electron transfer model for strain 2CP-C. Caution must be exercised with PSORT predictions because, in a few cases, the predictions did not agree with previously reported findings. For example, PSORT was unable to determine the localization of OmcA and OmcB which are outer membrane proteins [102, 110]. Experimental validation of cellular localization for most *c*-type cytochromes has not been accomplished, demonstrating the need for detailed protein characterization. Despite the minor imperfection, PSORT provides a generalized overview of all the predicted cellular localizations for all *c*-type cytochromes, which is useful for global *c*-type cytochrome proteomics analysis.

The majority of the predicted *c*-type cytochromes of strain 2CP-C and strain FRC-32 were annotated as "hypothetical proteins" in GenBank (Table 3.4 and 3.8). Although these proteins have conserved motifs or domains that permit them to be classified as *c*-type cytochromes, detailed functional information has been lacking. The unambiguous detection, identification, and abundance measurements with the MS approach verified that the "hypothetical proteins" were expressed and functional. Furthermore, many of these detected "hypothetical" putative *c*-type cytochromes were linked to specific growth conditions. Thus, in addition to information about the identities and relative abundances of these *c*-type cytochromes, this MS proteomic approach also provides information about under what experimental conditions these proteins appear. Obviously, further research needs to be conducted to more fully explore their roles in electron transfer and detail their contributions for

Table 3.9. Putative *c*-type cytochromes of *A. dehalogenans* 2CP-C grouped by cellular localizations predicted by Psort. Detected proteins in fumarate and nitrate cultures are highlighted in purple and blue, respectively, in table (a) and (b).

(a)

Cellular Localization	Putative *c*-Type Cytochrome				
Extracellular	Adeh_0409	Adeh_0656	Adeh_2216	Adeh_3074	Adeh_3163
Outer Membrane					
Periplasm	Adeh_0108	Adeh_0215	Adeh_0318	Adeh_0728	Adeh_0799
	Adeh_0909	Adeh_0910	Adeh_0918	Adeh_0972	Adeh_1278
	Adeh_1425	Adeh_1696	Adeh_1764	Adeh_2002	Adeh_2004
	Adeh_2097	Adeh_2277	Adeh_2285	Adeh_2664	Adeh_2665
	Adeh_2666	Adeh_2816	Adeh_2902	Adeh_2963	Adeh_2967
	Adeh_2989	Adeh_3067	Adeh_3068	Adeh_3090	Adeh_3392
	Adeh_3775				
Cytoplasmic Membrane	Adeh_0109	Adeh_0803	Adeh_1659	Adeh_2275	
Cytoplasm	Adeh_2403				

Table 3.9. Continued

(b)

Cellular Localization	Putative *c*-Type Cytochrome				
Extracellular	Adeh_0409	Adeh_0656	Adeh_2216	Adeh_3074	Adeh_3163
Outer Membrane					
Periplasm	Adeh_0108	Adeh_0215	Adeh_0318	Adeh_0728	Adeh_0799
	Adeh_0909	Adeh_0910	Adeh_0918	Adeh_0972	Adeh_1278
	Adeh_1425	Adeh_1696	Adeh_1764	Adeh_2002	Adeh_2004
	Adeh_2097	Adeh_2277	Adeh_2285	Adeh_2664	Adeh_2665
	Adeh_2666	Adeh_2816	Adeh_2902	Adeh_2963	Adeh_2967
	Adeh_2989	Adeh_3067	Adeh_3068	Adeh_3090	Adeh_3392
	Adeh_3775				
Cytoplasmic Membrane	Adeh_0109	Adeh_0803	Adeh_1659	Adeh_2275	
Cytoplasm	Adeh_2403				

Cellualr localizations are predicted by Psort 3.0

Red: Core *c*-type cytochromes

Green: Undetected *c*-type cytochromes

Orange: Detected under all but FeOOH condition

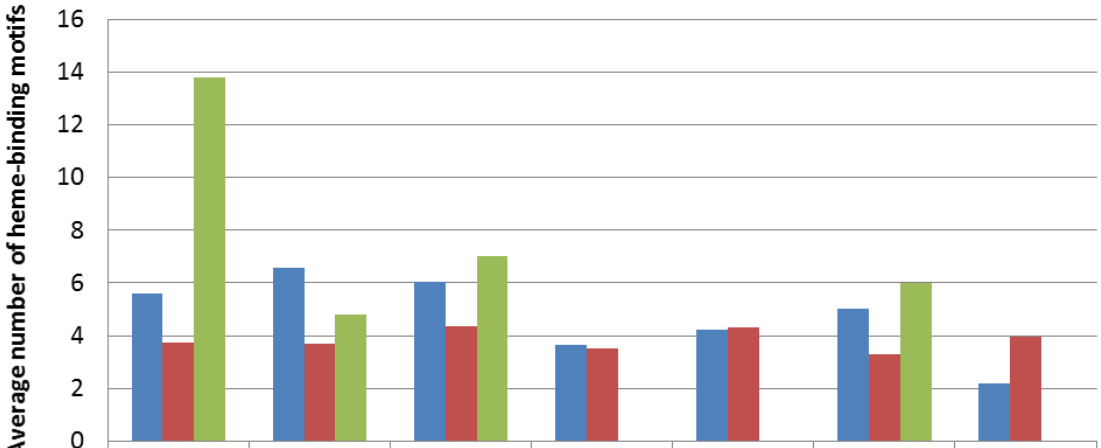
Purple: Detected with fumarate growth

Blue: Detected with nitrate growth

respiration of certain electron acceptors. An attempt to determine the orthologous *c*-type cytochromes in strain 2CP-C and FRC-32 was made by reciprocal Blastp search, which only identified a few orthologs. For example, SO_0479 in strain MR-1 and Adeh_2963 in strain 2CP-C were matched during reciprocal Blast with an e-value of 0. No orthologous proteins were found for the majority of putative *c*-type cytochromes in strain MR-1 and 2CP-C.

Of particular interest is the high number of heme motifs in some of the *c*-type cytochromes because such cytochromes in *Geobacter* species are hypothesized as 'iron lungs' [108, 109]. The iron lung theory proposes that the hemes in the periplasmic and outer-membrane *c*-type cytochromes provide a repertoire of heme groups with electron storage capacity [107-109], which is essential for the bacteria when there are limited electron acceptors in the environment. In this study, detected *c*-type cytochromes in FeOOH-grown FRC-32 cells contained more heme-binding motifs on average than grown with Fe citrate, MnO_2, or fumarate (Figure 3.5, Figure 3.8), supporting the 'iron lung' hypothesis. Higher electron storage capacity may be needed due to the low accessibility of the solid FeOOH, hence, *c*-type cytochromes with higher heme-content were expressed to provide electron-accepting capacity to temporarily store electrons before the FRC-32 cells can contact the solid electron acceptor. It is unclear why MnO_2, being another solid electron acceptor, did not show similar effect on the *c*-type cytochrome expression as FeOOH, but this might indicate a more amenable electron transfer environment.

In this study, the expression profile of *c*-type cytochrome proteins for each strain was significantly different among growth with various electron acceptors (Figure 3.1-3.3). The presence of a *c*-type cytochrome and its relative abundance could be associated with the



	FeOOH	Fe citrate	MnO2	nitrate	TSB	fumarate	oxygen
strain 2CP-C	6	7	6	4	4	5	2
strain MR-1	4	4	4	4	4	3	4
strain FRC-32	14	5	7			6	

Figure 3.8. Average number of heme-binding motif for *c*-type cytochromes detected under various growth conditions. Average number of heme-binding motif expressed under certain growth condition was calculated based on the percentage of cumulative normalized spectral count, *P*(nSpC). The motif number was multiplied to its corresponding *P*(nSpC). The sum of all the multiplications for all motif numbers was the average number of heme-binding motif for cytochromes detected under this growth condition.

available electron acceptor in the environment, and unique *c*-type cytochromes were only observed when certain electron acceptors were provided. Such unique *c*-type cytochromes should be informative as biomarkers to reflect the environmental redox conditions and responsive biological processes carried out by the microbes.

Although there are at least 40 or more predicted *c*-type cytochromes in the three bacterial strains discussed above, only a handful of these proteins had been functionally characterized and assigned specific roles in the electron transfer process. The proteomics approach outlined here provides a means of profiling large-scale *c*-type cytochrome expression simultaneously, providing both qualitative and abundance information about their expression changes in response to different growth conditions. The identification of specific cytochromes detected only under certain growth conditions suggests that these microbes have the ability to adjust their cytochrome expression for specific electron acceptors. Thus, the information provided here should enable the application of specific cytochromes as biomarkers of specific electron acceptor conditions in the growth environment. Due to the significant impact of microbes on the biogeochemistry of metals and minerals, *c*-type cytochrome biomarkers might provide a powerful toolset that can be used to assess and amend microbial activities *in-situ*. The results obtained in this large scale characterization study provide detailed information which would be difficult if not impossible to acquire elsewhere about which annotated or "functionally unknown" *c*-type cytochromes would be the best candidates to track in environmental microbiology studies.

3.5 Conclusions

The differential proteome expression of *Anaeromyxobacter dehalogenans* strain 2CP-C, *Geobacter daltonii* strain FRC-32, and *Shewanella oneidensis* strain MR-1 were characterized using a mass spectrometry - based proteomic approach. The analyses revealed a set of core *c*-type cytochromes expressed under all growth conditions tested, as well as *c*-type cytochromes unique to specific growth conditions, suggesting their potential utility as biomarkers to assess terminal electron-accepting processes (TEAPs). Many *c*-type cytochromes with multiple heme-binding motifs were detected, in particular for *G. daltonii* FRC-32 grown with the solid metal electron acceptor ferric oxyhydroxide, which supports the hypothesis that high heme content *c*-type cytochromes act as capacitors when electron acceptors are not readily available. A basic model for *A. dehalogenans* 2CP-C *c*-type cytochrome network was constructed to link cytochrome detection with corresponding TEAP. These findings suggest that a MS-based proteome approach is powerful for linking *c*-type cytochrome expression profiles with specific TEAPs, and provides significant insight into the fine-tunable control of bacterial electron transfer processes for different electron acceptors.

Chapter 4

MS-based Proteomic Characterization of *Anaeromyxobacter dehalogenans* Strain 2CP-C Reveals Elevated Energy Metabolism in Growth with Metal Electron Acceptors

The text and figures presented in this chapter will be published as the following journal article:

> Xiaoxin Liu, Silke Nissen, Karuna Chourey, Frank Löffler, Robert Hettich. "MS-Based Proteomic Characterization of *Anaeromyxobacter dehalogenans* 2CP-C Reveals Elevated Energy Metabolism in Growth with Metal Electron Acceptors". Submitting to *Journal of Proteome Research* 2014.

Xiaoxin Liu contributed to the proteome sample preparation, mass spectrometry experiments performing, biological data analysis and visualizations, and manuscript writing.

4.1 Introduction

Anaeromyxobacter dehalogenans strain 2CP-C is a facultative gram-negative *Myxobacteria* isolated from soil [87]. Unlike other members in the *Myxobacteria* family which are considered obligate aerobes, *A. dehalogenans* is capable of survival in anaerobic and microaerophilic environments [87, 202]. Studies of anaerobic growth of strain 2CP-C revealed tremendous respiration versatility, as evidenced by the ability to utilize a vast range of electron donors (e.g. acetate, hydrogen, pyruvate, lactate, succinate, formate) and electron acceptors (e.g. nitrate, fumarate, halogenated phenols, ferric iron, etc.) [87, 173, 203]. In particular, the dissimilatory metabolic reduction of metals and radionuclides by 2CP-C raised interest in implementing *Anaeromyxobacter* species for *in situ* bioremediation for soil and sediments with

heavy metal or radionuclide contamination [88, 204, 205]. Indeed, field studies of uranium contaminated sites confirmed the bioremediation contribution of *A. dehalogenans* populations [90, 91, 172, 206, 207]. The increase of *Anaeromyxobacter* abundance in uranium reduction sediment after oxygen intrusion distinguished *Anaeromyxobacter* from other *in situ* dissimilatory uranium reducing bacteria (e.g. *Geobacter*), and suggested the need for persistent presence to maintain the stability of immobilized uranium by competing with uranium reoxidation [208].

Despite the respiration versatility and profound environmental impact of *A. dehalogenans* strain 2CP-C, the physiology and metabolic basis that account for the electron donor/acceptor utilization is largely unknown. Complete genome sequencing of strain 2CP-C revealed genes responsible for a wide metabolism range [111]. For example, the identification of 68 *c*-type cytochrome coding genes reflected the respiration versatility, and the aerobic growth ability was supported by genetic evidence of oxygen-respiration and detoxification system. The genome provided metabolic potential information which is carried out by functional units – proteins. In order to understand how 2CP-C utilizes its genetic content for functional activities during growth under various redox conditions, proteomic characterization is desirable, in that proteins are the actual metabolic units that support the overall cellular-level activities. Thus, global proteome profiling is in need to provide insight into the functional machineries for bacterial physiology.

For comprehensive proteomic characterizations, mass spectrometry (MS) is currently the state-of-the-art technic and has achieve tremendous success for the proteomics studies on

numerous biological systems [35, 183]. A previous study has characterized the proteome of strain 2CP-C using two-dimensional gel electrophoresis (2DE) and MALDI-TOF mass spectrometry [209]. Growth with two electron acceptors (fumarate and ferric citrate) were compared in this study, and the results identified around 13% of predicted proteins in the strain 2CP-C proteome. Although 2DE coupled with MALDI-TOF has been a widely used technique for proteomics interrogation, the resolution and accuracy are not sufficient to resolve the complexity of a microbial proteome. In order to obtain more comprehensive and deeper proteome measurement for strain 2CP-C, liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) in this study. In addition, a wide range of electron acceptors are tested ranging from different forms of metals to non-metal organic and inorganic compounds.

4.2 Experimental Section

Bacterial Strains and Culture Conditions. In this chapter, a total of eight growth conditions were tested for *Anaeromyxobacter dehalogenans* strain 2CP-C, including FeOOH, Fe citrate, MnO_2, fumarate, nitrate, tryptic soy broth, oxygen, and N_2O. Detailed methods are provided in Chapter 2.

Cell Lysis and Protein Extraction. The cell lysis and protein extraction methods are described in Chapter 2.

LC-MS/MS. Peptides were analyzed by two-dimensional liquid chromatography (Ultimate HPLC System, LC Packings, a division of Dionex, San Francisco, CA) coupled online

with a linear ion trap mass spectrometer (LTQ XL, ThermoFisher Scientific, San Jose, CA), according to established protocol[132]. A total of 53 runs were collected, which comprised biological triplicates for all conditions tested. A 12-step 24-h Multidimensional Protein Identification Technology (MuDPIT) was used to analyze peptides as described [132, 139]. The LTQ XL was operated in a data dependent mode and one full MS scan was followed by five MS/MS scans. The m/z isolation width was set to 3 m/z and the dynamic exclusion repeat was set at 1 with a duration time of 60 sec.

Database Searching and Data Normalization. Experimental MS/MS spectra were searched against FASTA protein databases containing all predicted open reading frames of *A. dehalogenans* 2CP-C obtained from Joint Genome Institute (JGI, http://genome.jgi.doe.gov/, downloaded in July, 2011) and common contaminants such as trypsin, keratin, etc. A decoy database consisting of reversed sequences from the corresponding proteome database was also appended to the FASTA protein database in order to calculate False Discovery Rate (FDR) at the protein level. The SEQUEST search algorithm was used for peptide identification. Identified peptides were filtered and assembled into proteins by DTASelect [43]. The DeltCN filter was set to 0.08. The cross-correlation score filtering process started with conservative criteria as following: XCorr: +1 = 1.8, +2 = 2.5, +3 = 3.5. The XCorr criteria were slightly altered to adjust FDR to ~1% for each sample. At least two peptides were required per protein identification, and at least one peptide had to be unique to that protein. Raw spectral counts were normalized by protein lengths and the total spectral counts using Normalized Spectral Abundance Factor (NSAF) as previously described [184]. Normalized spectral counts are then uniformly multiplied by a factor of 100,000 for better readability and comparability to raw spectral counts. Adjusted

normalized spectral count is denoted as nSpC throughout this paper. The nSpCs for biological triplicates were averaged for each growth condition. The nSpCs reflect abundances of detected proteins. Larger nSpC value represents higher abundance, and vice versa.

Hierarchical clustering and ANOVA analysis: In order to achieve quantifiable proteins for statistical analysis, a prevalence value (PV) was determined for each detected protein as previously described [210]. Since the PVs are derived from nSpCs of detected proteins, they reflect the abundances as well as reproducibility for detected protein. The PV for each protein was ranked from the largest to smallest. Substantive proteins representing top 95% of the total PVs were retained for statistical analysis. After proteins are filtered by their PVs, normalized data was log10-transformed for hierarchical clustering and ANOVA analysis. Hierarchical clustering was performed by JMP genomics v.6.0 (SAS Institute) using Ward clustering algorithm. The following settings: “scale row”, “center row” were applied to make the variance of each row to be 1, and center the mean of each row to 0. Each row contains transformed abundances for one protein in different growth conditions. ANOVA was conducted in JMP Genomics v.6.0 (SAS Institute), to compare protein abundances across all conditions with Benjamini & Hochberg FDR correction (alpha = 0.05).

Pathway analysis: Metabolic maps were generated by KEGG pathway mapping software iPATH 2.0 [211].

Cellular Localization Prediction: Cellular localizations were predicted based on protein sequences using PSORTb.3.0 [185].

4.3 Results and Discussion

4.3.1 Characterization of the Pan-proteome of *A. dehalogenans* 2CP-C

Mass spectrometry-based proteomics measurements were obtained for biological triplicates of *A. dehalogenans* 2CP-C cells grown with different electron acceptors. Approximately 2,000 proteins were identified for most of the growth conditions (Fe citrate, MnO_2, nitrate, tryptic soy broth, fumarate, and oxygen - Figure 4.1). When FeOOH was provided as the sole electron acceptor, the 2CP-C cell growth was significantly slower than all other tested growth conditions, as evidenced by the longer time to reach late exponential phase for harvesting. The recalcitrant growth with FeOOH as electron acceptor appeared to yield fewer proteins being expressed and a lower complexity protein mixture, which yielded the least protein identifications (Figure 4.1). Although a different protein extraction strategy was applied on FeOOH grown cultures, the detected proteins were found to be distributed among all cellular localizations and clusters of orthologous group (COG) categories with highly comparable percentages among different growth conditions (Figure 4.2, 4.3), suggesting comprehensive sampling coverage and adequate proteome measurement of that sample.

In total, 2,846 proteins were identified across all growth conditions, representing the pan-proteome, which corresponded to 65% of all the predicted open reading frames in 2CP-C

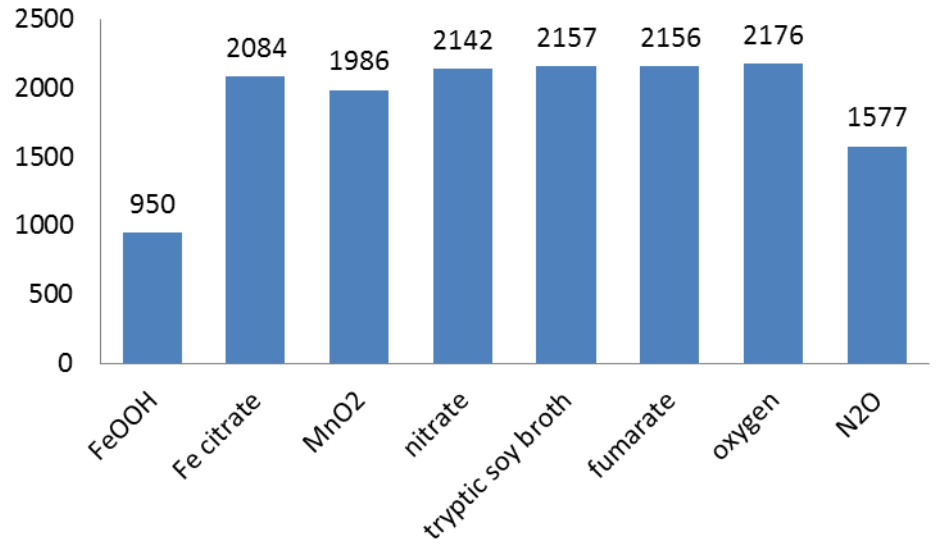


Figure 4.1. Proteome identifications for *A. dehalogenans* 2CP-C grown under various conditions. Total identified protein numbers are pooled for biological triplicates of each growth condition.

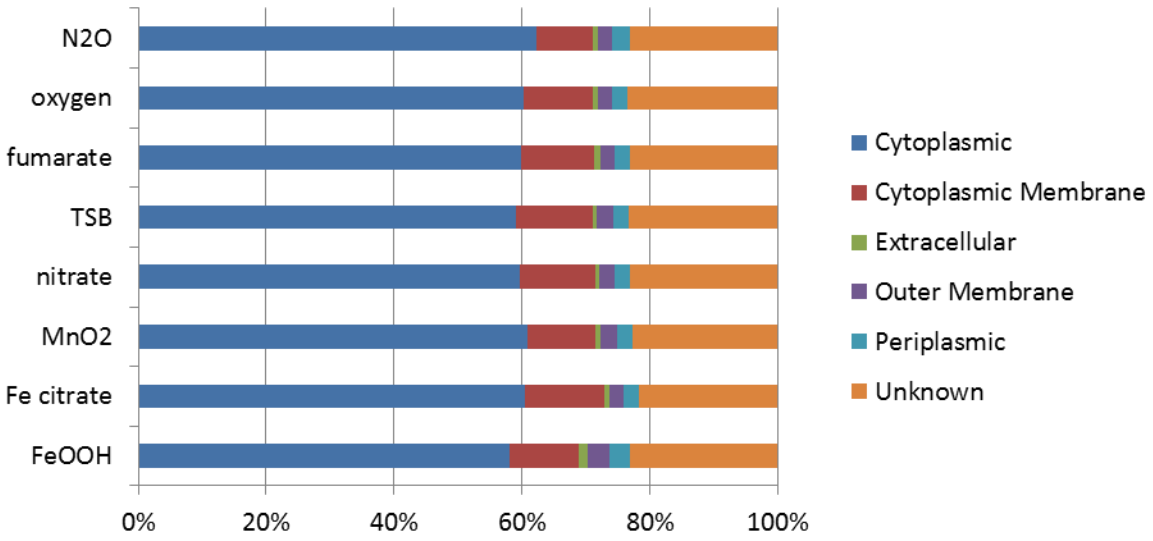


Figure 4.2. Cellular localization distribution of *A. dehalogenans* 2CP-C proteome under growth under different conditions.

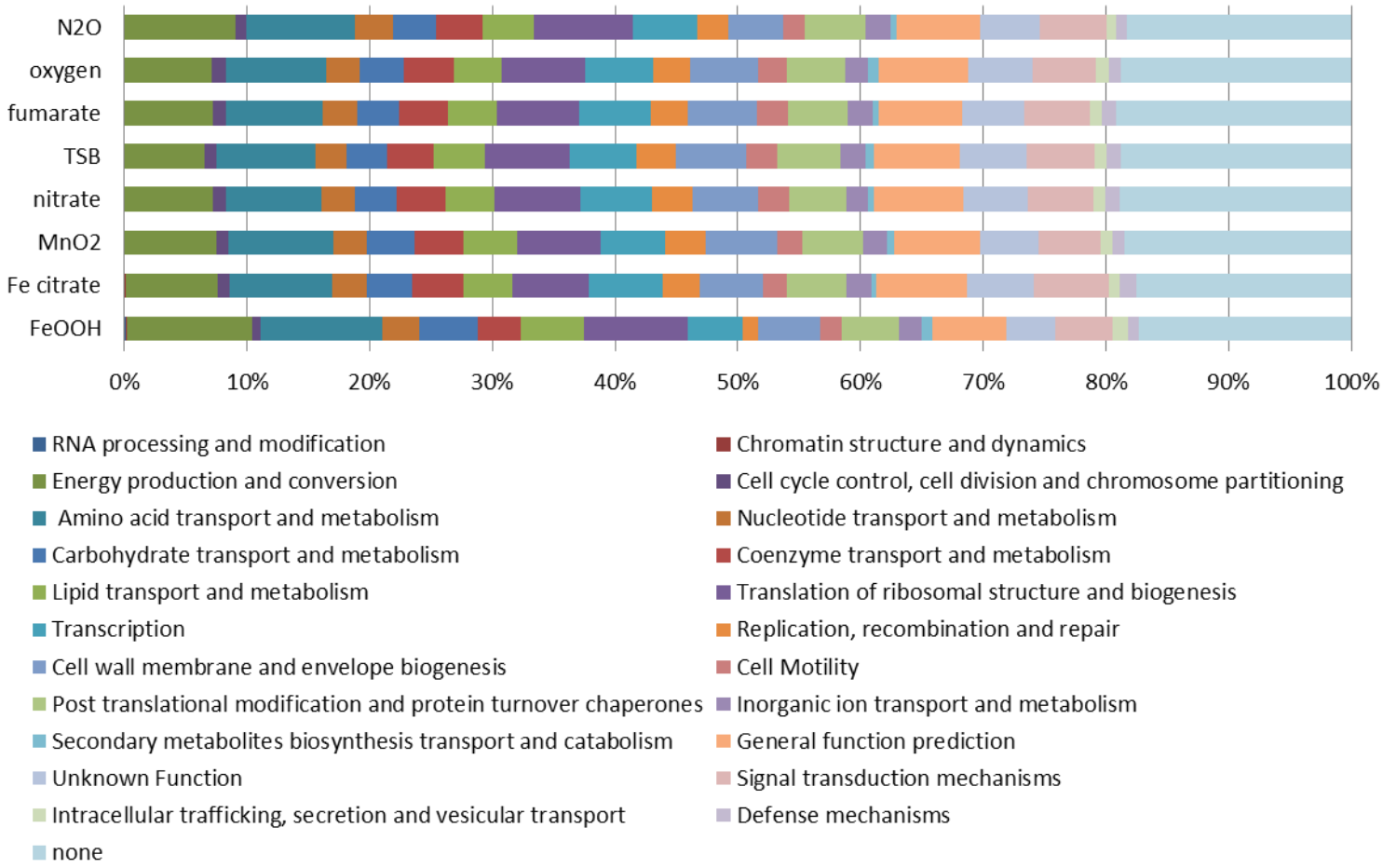


Figure 4.3. COG functional category distribution of *A. dehalogenans* 2CP-C proteome under growth under different conditions.

genome. Compared to the number of proteins identified in each growth conditions (around 2,000, as mentioned above), the pan-proteome exhibited only a slight increase in protein numbers, indicating a substantial overlap in the protein inventory between different growth conditions. The identified proteins distributed between all cellular localizations based on computational prediction (Figure 4.4). About 80% of the predicted cytoplasmic, periplasmic, outer membrane, and extracellular proteins were detected, whereas fewer predicted cytoplasmic membrane proteins (40%) and proteins with unknown localization (60%) were detected (Figure 4.4). Detected proteins belonged to all functional categories in the clusters of orthologous groups (COGs) classification database, with roughly even distributions (Figure 4.5). Proteins involved in energy metabolism, amino acid metabolism, translation, and transcription were readily identified, as might be expected since these are representatives of the core metabolism. Proteins that could not be classified into any COG functional categories ("none" category) were mostly annotated as "hypothetical proteins" in GenBank, and occupied the highest percentage of detected proteome (Figure 4.5). The "hypothetical protein" annotations were inferred from genome sequence analysis [212], but no experimental evidence was available for the expression of "hypothetical protein". The non-targeted LC-MS/MS proteome characterization provided evidence for the actual expression, identities, abundances, and growth condition dependence of many of these "hypothetical proteins".

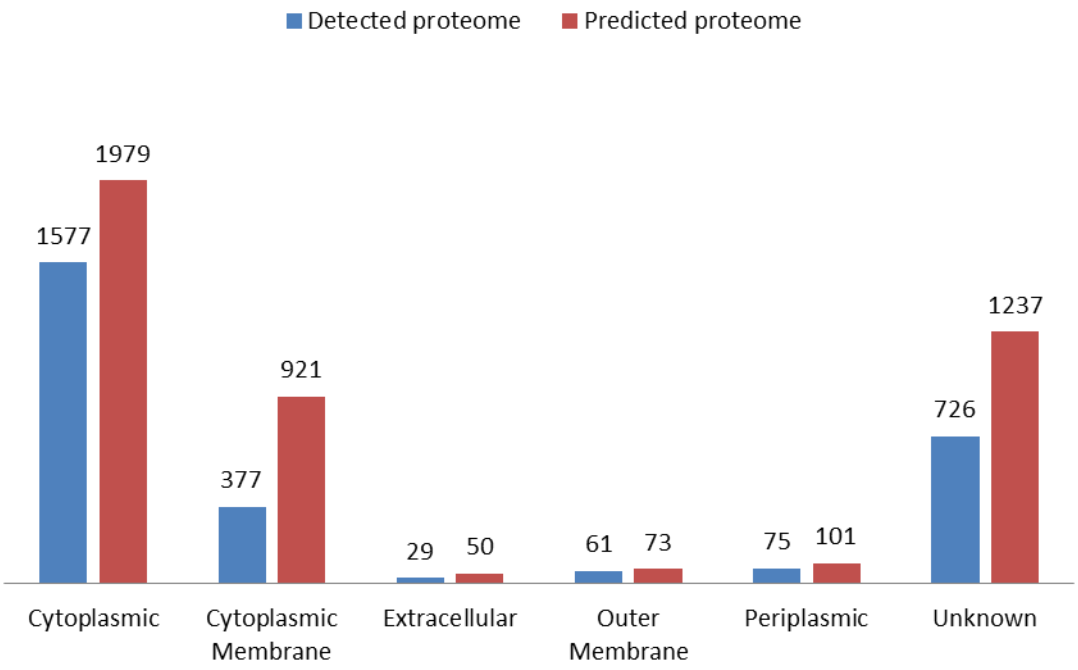


Figure 4.4. Cellular localization distribution of predicted (red) and detected (blue) proteomes of *A. dehalogenans* 2CP-C. Detected proteome is represented by pooled protein identifications of all replicates and all growth conditions.

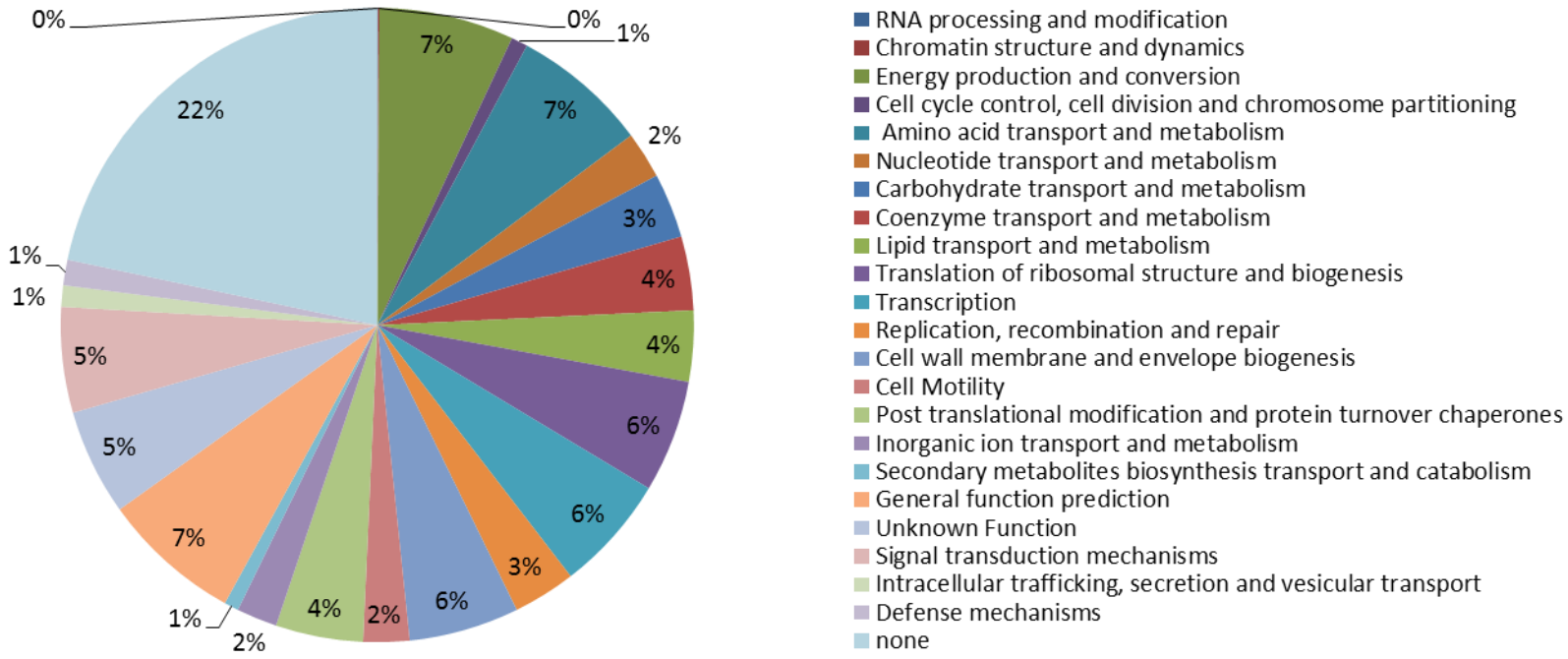


Figure 4.5. COG functional category distribution of detected proteome of *A. dehalogenans* 2CP-C. Detected proteome is represented by pooled protein identifications of all replicates and all growth conditions.

4.3.2 The "Core" Proteome of Strain 2CP-C

Combining proteome measurements for all 2CP-C cultures revealed the "core" proteome, which consisted of 710 proteins shared among growth with all tested electron acceptors. The abundances of the core proteome represented a vast range as indicated by the normalized spectral counts (nSpCs) (scaling from 10e-1 to 10e4). When sorted by the COG functional categories, the core proteome distributed almost all COG classifications with consistent abundances in each category across different growth conditions (Figure 4.6). Roughly 35-50% of the core proteome represent functions related to protein translation and metabolism, and energy production (Figure 4.6), which is probably not surprising in that the core proteome is essential to sustain the basic operational metabolism for cell survival.

To generate a more specific functional overview of the core proteome, iPath2.0 was used to map all the core proteins onto predicted metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Figure 4.7). The functional annotation provided by KEGG is not complete; only 1806 out of 4361 proteins in the entire predicted 2CP-C proteome have been assigned functions in KEGG database. Nonetheless, KEGG provides a more high-resolution functional view as compared to the broad COG functional classifications. Out of the 1806 proteins whose KEGG functional annotations were available, 1383 were detected across various growth conditions, and were mapped to metabolic pathways using iPath2.0 (Figure 4.7a). In the core proteome, 441 of the 710 proteins have assigned functions in the KEGG database, and were mapped to metabolic pathways involved in the metabolism of lipid, carbohydrate, amino acid, nucleotides, etc. (Figure 4.7b). Although the core proteome is only

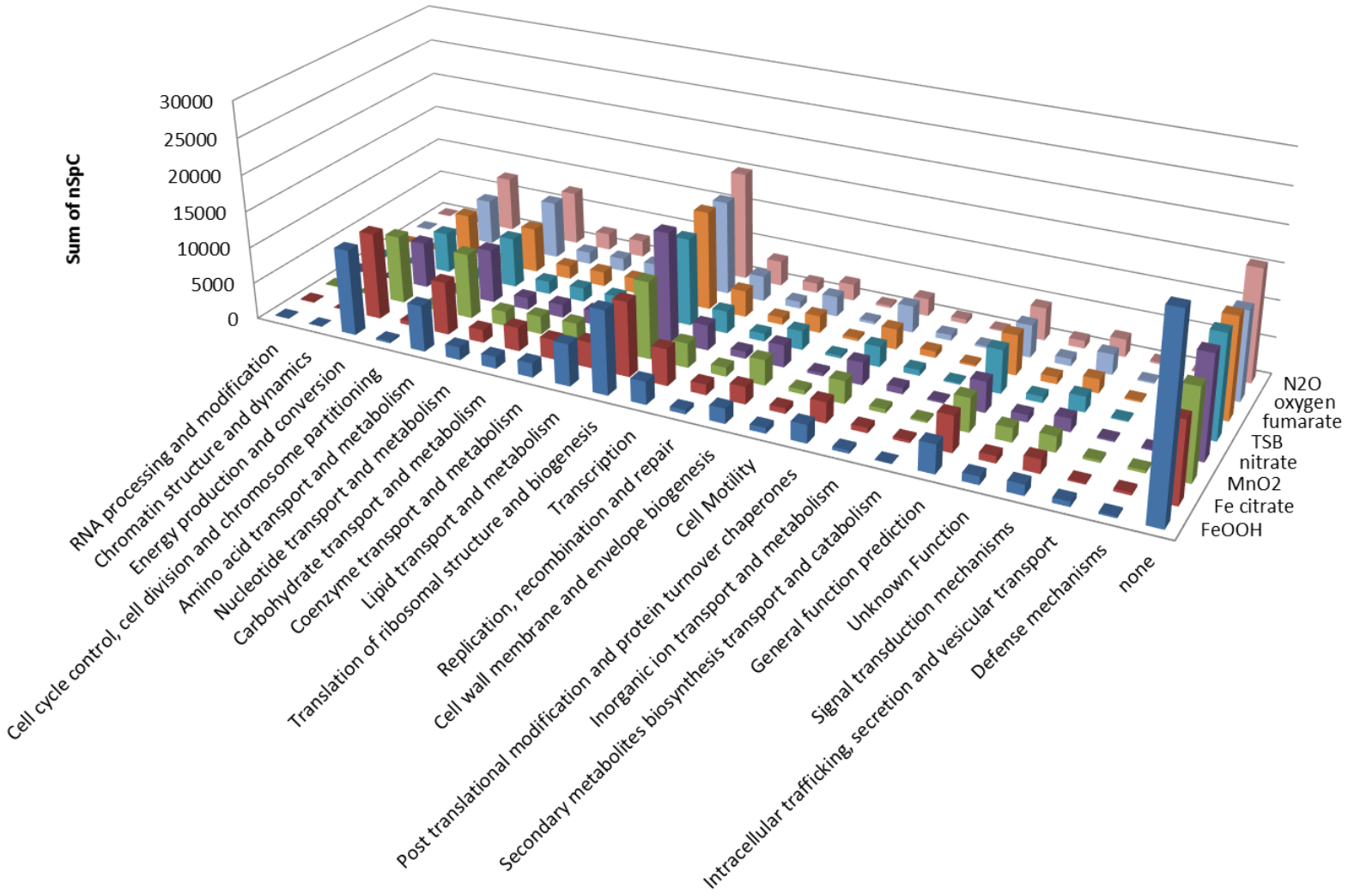
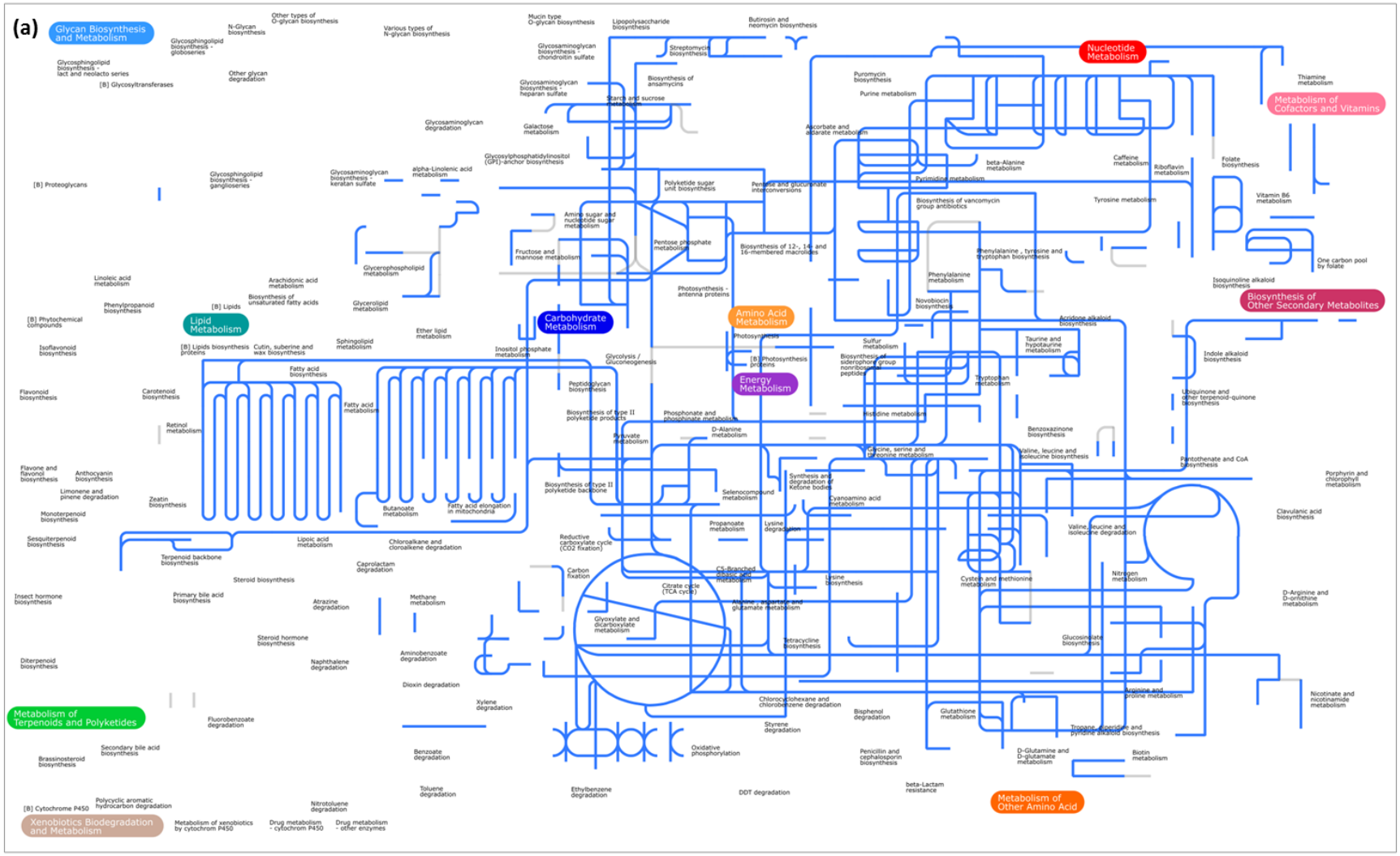
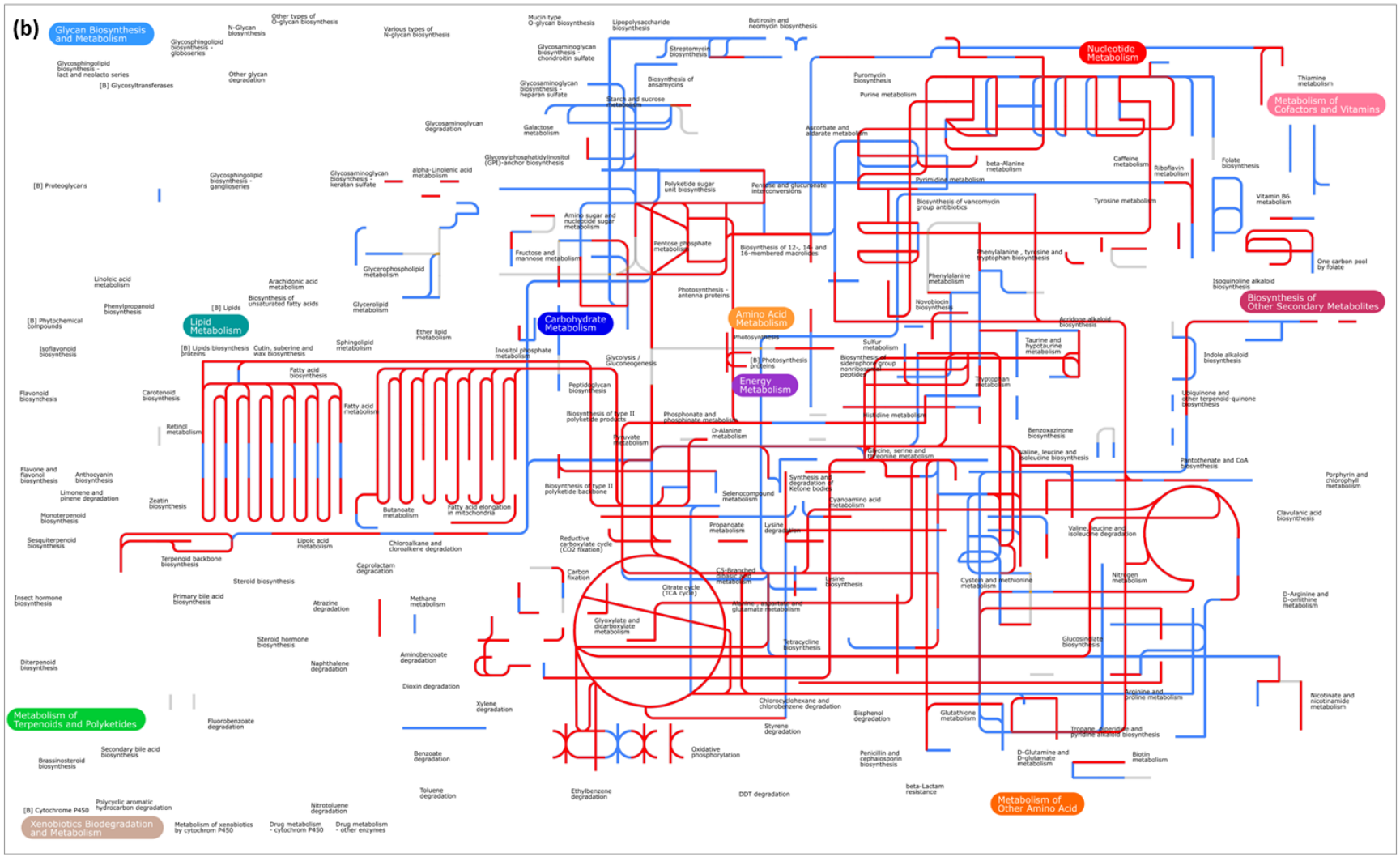


Figure 4.6. Quantitative distribution of *A. dehalogenans* 2CP-C core proteome by COG functional categories.

Figure 4.7. Metabolic pathways of detected *A. dehalogenans* 2CP-C proteome. The pan-proteome of 2CP-C represents all the detected proteins pooled across all growth conditions and all replicates. The core proteome represents proteins detected under all growth conditions. (a)The pan-proteome is mapped onto KEGG metabolic pathways and highlighted blue. (b)The core proteome represented pathways are highlighted in red.


(a)
Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid


(b)
Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid

16% of the predicted proteome of 2CP-C, it contains well annotated proteins participating in almost all metabolic processes. The most readily represented pathways by the core proteome participate in energy metabolism, and lipid metabolism (Figure 4.7b). The conservation of most metabolic pathways by the core proteome among tested growth conditions further reinforce the essentiality of the core proteome, and suggests that different electron acceptors or donors, by in large, do not affect the overall cellular processes.

4.3.3 The Unique Proteins of Strain 2CP-C

The proteins uniquely detected under specific electron acceptor growth conditions are likely responsible for specific /specialized biological processes related to the utilization of corresponding electron acceptor. When 2CP-C cells were grown with FeOOH or N_2O, relatively few unique proteins were identified (12 and 16, respectively). Oxygen and Fe citrate-growth yielded the most unique proteins (77 and 57, respectively), while growth under the remaining conditions yielded around 40 unique proteins each. The unique proteins of each condition were grouped by their COG functional classifications according to their abundances (Figure 4.8). Under all growth conditions, proteins that could not be classified into any category ("none category") had the highest representation. The MnO_2 grown cells generated the most abundant "energy production and conversion" protein detection. Proteins in the "post translational modification and protein turnover chaperones" category were most abundant in Fe citrate and tryptic soy broth grown cells. Unique proteins belonging to "inorganic ion transport and metabolism" and "signal transduction mechanisms" categories were both most abundant with

growth in tryptic soy broth. The “cell wall membrane and envelope biogenesis” proteins are most abundant with oxygen-growth (Figure 4.8).

In a few cases, proteins unique to different growth conditions shared the same functional annotation (Table 4.1), suggesting that these functions can be carried out by multiple different proteins, and that different environmental factors require specific proteins performing similar functions. The unique proteins under each growth condition were mapped to KEGG pathways using iPath2.0 to visualize the represented pathways in the metabolic network (Figure 4.9). Due to limited KEGG annotations, a total of 98 out of all unique proteins in all growth conditions had available KEGG information, and were mapped to 27 (redundant) metabolic pathways. For example, the pathways unique to MnO_2 belonged to nucleotide metabolism, and a few amino acid metabolic pathways were represented by unique proteins in tryptic soy broth grown cells (Figure 4.9). Proteins unique to nitrate growth did not match any metabolic pathways, but were involved in regulatory pathways for cell motility and signal transduction (data not shown). In addition to the overlapping functional annotations between unique proteins, the same functional annotations could also be shared with the core proteins or other non-unique proteins. Thus, in most cases, the unique proteins represented pathways were not exclusive to the corresponding growth condition. As a result, the pathways highlighted by unique proteins could also be represented in the core metabolic network (Figure 4.7 and 4.9). Only fumarate resulted in the expression of fatty acid biosynthesis and terpenoid backbone biosynthesis pathways that were not represented by proteins detected under other growth conditions (Figure 4.9).

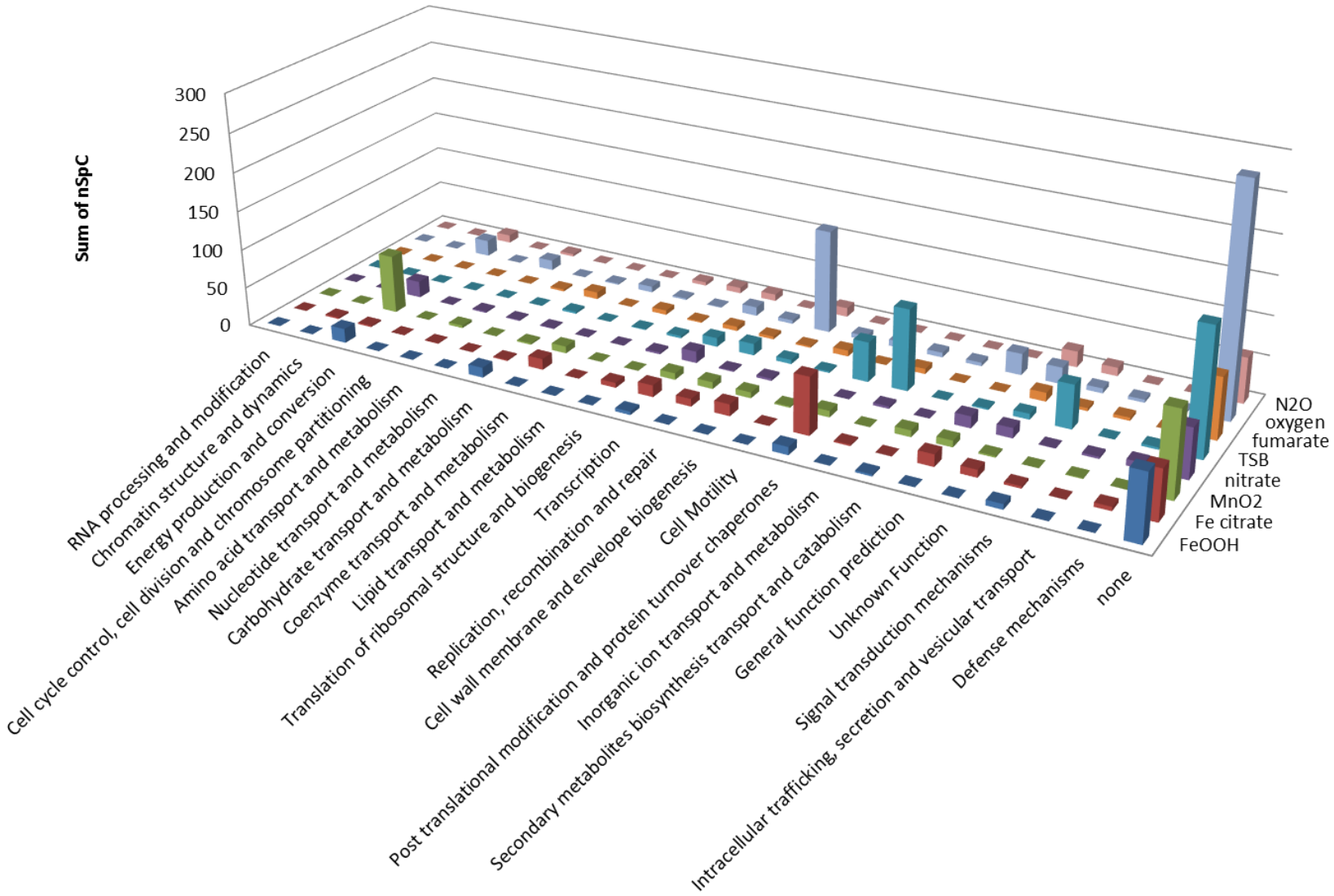


Figure 4.8. Quantitative distribution of detected *A. dehalogenans* 2CP-C proteins unique to certain growth conditions.

Table 4.1. Unique proteins from different growth conditions share overlapping functional annotations.

Unique Proteins	Growth with	KEGG Annotation	GenBank Annotation
Adeh_2542	FeOOH	K00773	queuine tRNA-ribosyltransferase
Adeh_1428	nitrate		
Adeh_0466	Fe citrate	K03088	sigma-24 (FecI-like)
Adeh_2378	MnO_2		
Adeh_1747	nitrate		
Adeh_0557	tryptic soy broth		
Adeh_2734	nitrate	K03408	CheW protein
Adeh_0602	oxygen		
Adeh_2057	tryptic soy broth	K06131	phospholipase D/Transphosphatidylase
Adeh_4242	N_2O		
Adeh_3016	Fe citrate	K07216	Hemerythrin-like, metal-binding protein
Adeh_0219	fumarate		

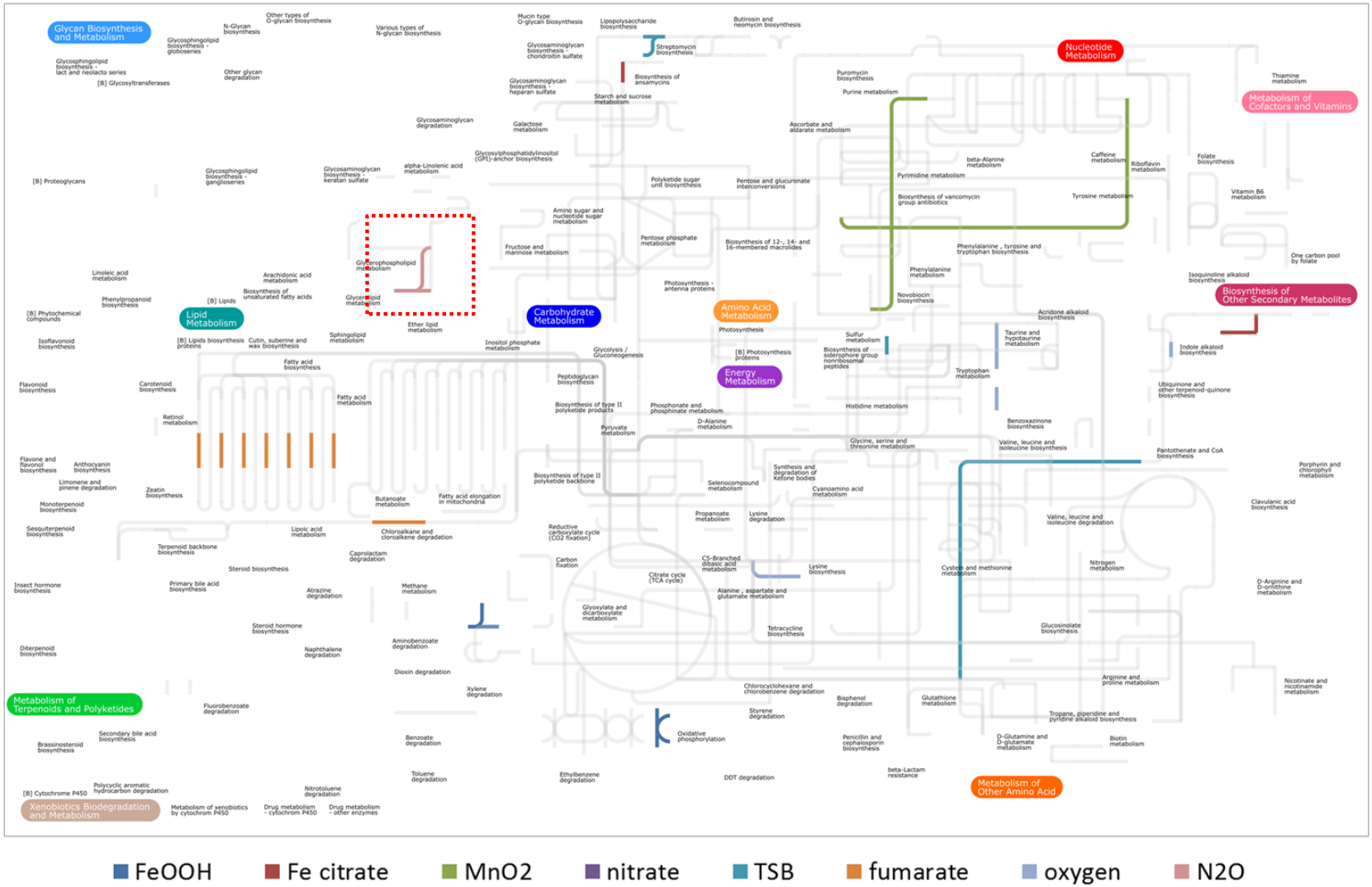


Figure 4.9. KEGG pathways represented by detected proteins of *A. dehalogenans* 2CP-C unique to certain growth condition. Pathways represented by unique proteins are highlighted in different colors by growth conditions. The proteins unique to nitrate do not map onto any of the metabolic pathways. The pathway in the dashed box is mapped by both unique proteins to tryptic soy broth and N_2O.

4.3.4 Quantitative Comparison of 2CP-C Proteome Expression Patterns

In order to *quantitatively* compare the proteomes of 2CP-C grown under different conditions, the proteomic data was filtered to remove low abundance proteins whose quantification is less robust due to the nature of peptide sampling [169]. The data was filtered by 95% prevalence value (PV) (see Experimental Section), resulting in 1554 quantifiable proteins representing ~55% of the total detection. The filtered data was analyzed using hierarchical clustering for overall quantitative distribution of protein expression across all tested growth conditions (Figure 4.10). The hierarchical clustering analysis generated a total of 10 protein clusters based on protein abundance patterns across growth conditions. The number of proteins within each cluster ranged from 59 (cluster 2) to 278 (cluster 3).

The protein clusters were dissected to reveal the quantitative contributions of each growth conditions (Figure 4.11a). The cumulative nSpC of each protein cluster was calculated for each growth condition, and the relative percentages reflected the growth condition representations in each cluster. Several protein clusters were most abundant in certain conditions. In particular, the protein abundances of cluster 1 and 10 had the highest contribution (~35%) by growth with FeOOH and oxygen, respectively (Figure 4.11a). Growth with Fe citrate, MnO_2, and N_2O contributed over 25% to the total protein abundances in cluster 4, 5, and 8, respectively (Figure 4.11a). Since each cluster contains detected proteins with similar abundance patterns across growth conditions, and many clusters (e.g. cluster 1, 10) represents proteins predominantly detected under certain condition, the functional attribute of proteins within each cluster could be linked to specific electron acceptors. To functionally

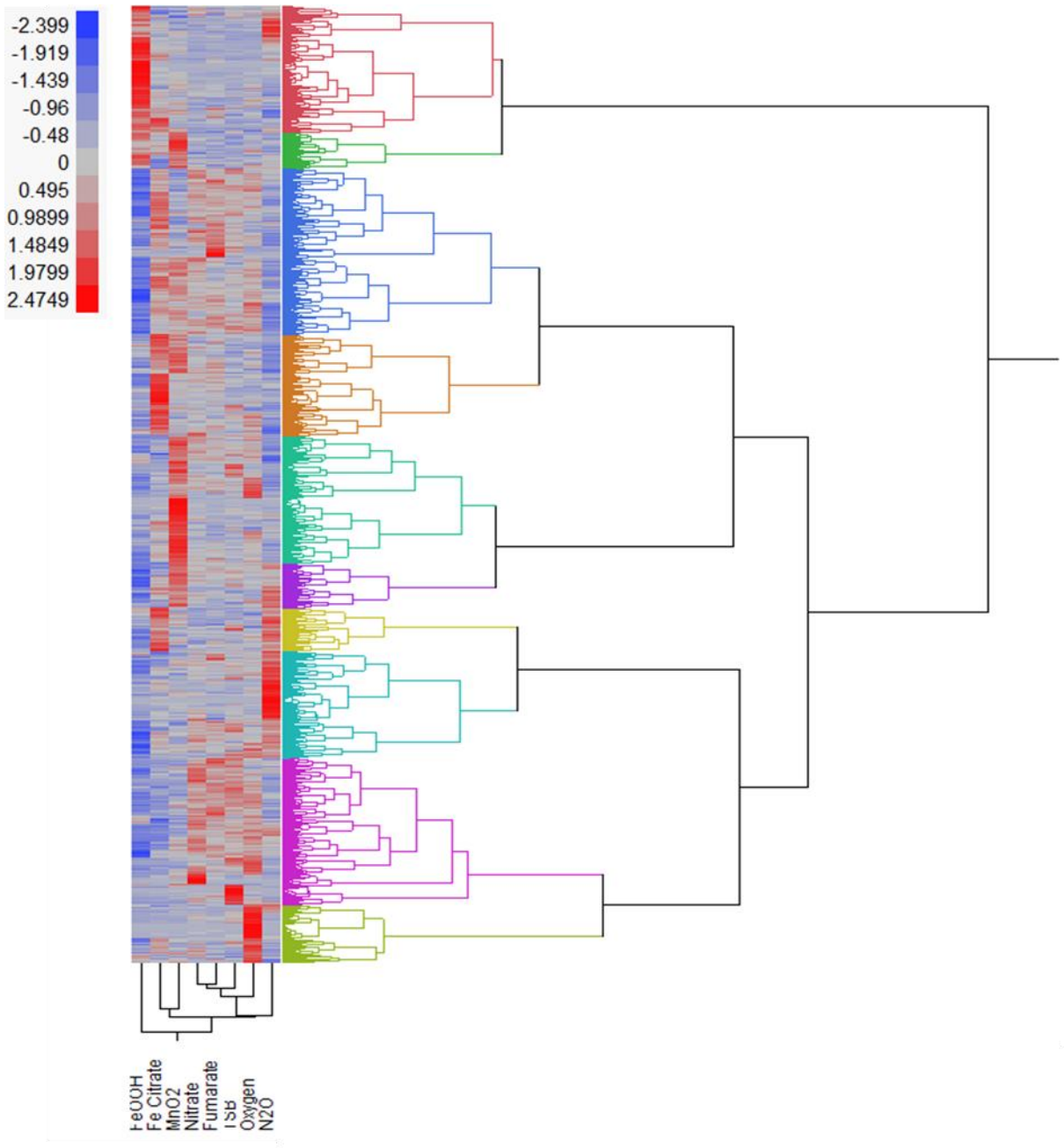


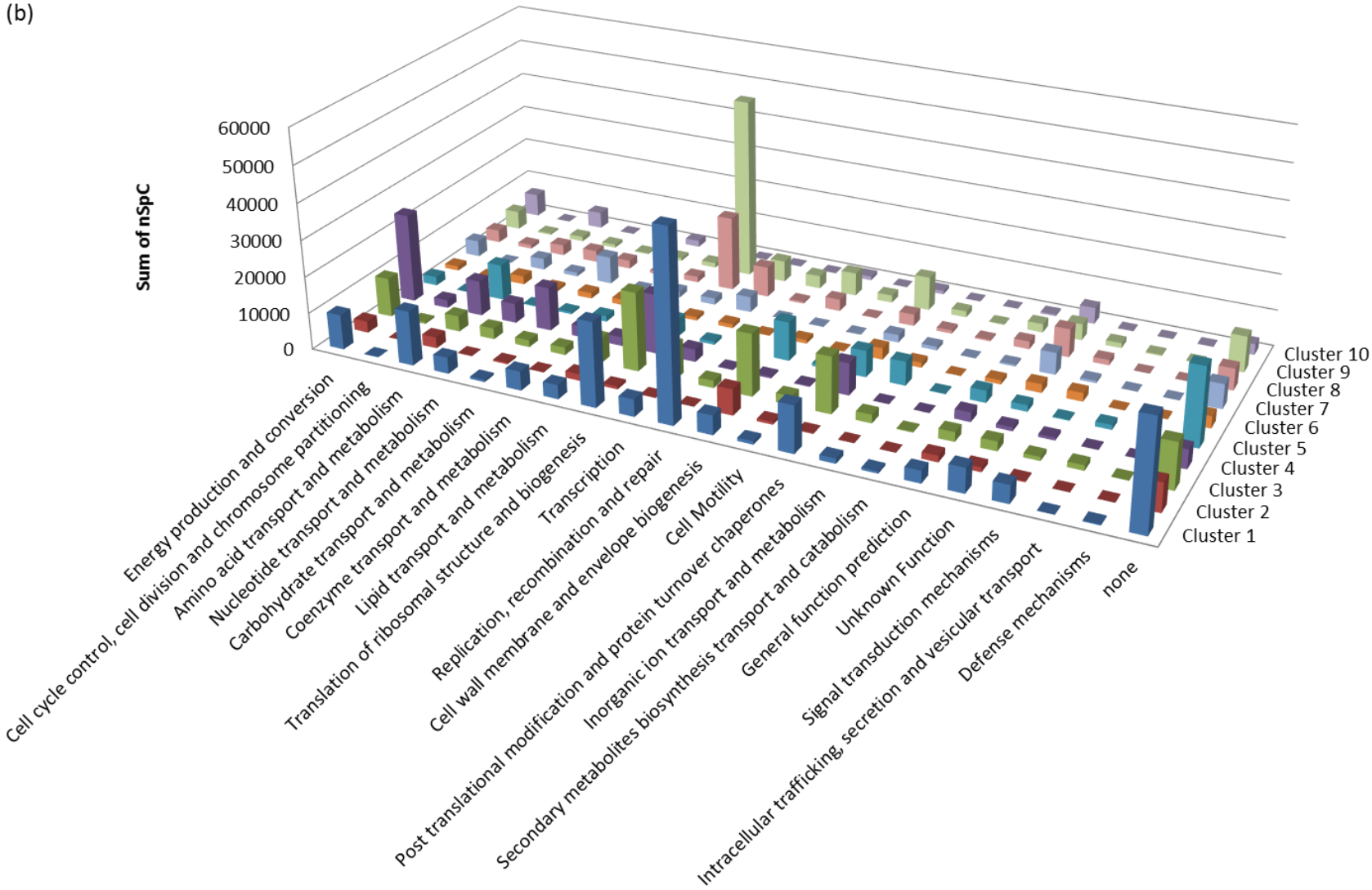
Figure 4.10. Hierarchical clustering of *A. dehalogenans* 2CP-C proteins and growth conditions. Quantifiable proteins were analyzed by hierarchical clustering to reveal the differential global proteome profiles. Protein abundances (nSpCs) were scaled so that the variance of each row is 1, and the mean of each row is centered to 0. The scaled and centered abundances ranged from -2.399 to +2.4749. Each hierarchical cluster contains detected proteins having similar abundance patterns across growth conditions. A total of 10 hierarchical clusters were generated, and denoted by different colors. From top to bottom are clusters 1 – 10. Different growth conditions were also clustered (bottom) based on the similarities of global proteome abundance patterns.

Figure 4.11. Quantitative distribution of *A. dehalogenans* 2CP-C protein hierarchical clusters by COG functional categories. (a)The quantitative contribution of proteins detected under each growth condition was determined for each hierarchical cluster. The quantitative contribution of each growth condition is represented by the percent share of the total nSpC in each protein cluster. The cells are color-coded (red:green::high:low) to visualize the degree of quantitative contributions. (b) Each hierarchical cluster contains detected proteins having similar abundance patterns across growth conditions. Proteins in each cluster (*z* axis) were classified into respective COG functional categories (*x* axis). The *y* axis represented sum of nSpCs of contributing proteins in each COG category in different clusters.

(a)

	FeOOH	Fe citrate	MnO_2	nitrate	tryptic soy broth	fumarate	oxygen	N_2O
cluster 1	35%	8%	7%	9%	9%	10%	7%	15%
cluster 2	21%	10%	23%	9%	8%	8%	10%	12%
cluster 3	3%	17%	11%	15%	15%	17%	13%	8%
cluster 4	8%	25%	15%	12%	10%	12%	11%	7%
cluster 5	6%	10%	27%	13%	12%	11%	14%	7%
cluster 6	4%	10%	20%	12%	12%	11%	12%	18%
cluster 7	8%	23%	11%	9%	10%	9%	10%	20%
cluster 8	4%	10%	9%	13%	13%	13%	12%	26%
cluster 9	4%	4%	8%	19%	20%	17%	17%	10%
cluster 10	7%	9%	9%	10%	10%	7%	36%	11%

(b)


Sum of nSpC
60000
50000
40000
30000
20000
10000
0
Energy production and conversion
Cell cycle control, cell division and chromosome partitioning
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Translation of ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Cell wall membrane and envelope biogenesis
Cell Motility
Post translational modification and protein turnover chaperones
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis transport and catabolism
General function prediction
Unknown Function
Signal transduction mechanisms
Intracellular trafficking, secretion and vesicular transport
Defense mechanisms
none
Cluster 10
Cluster 9
Cluster 8
Cluster 7
Cluster 6
Cluster 5
Cluster 4
Cluster 3
Cluster 2
Cluster 1

analyze the hierarchical clusters, the protein complement in each cluster was examined for their representing COG functional categories (Figure 4.11b). Each protein cluster demonstrated different abundance distributions across COG categories. For example, proteins predominantly detected with MnO_2 growth (cluster 4) mostly represented “energy production and conversion” and “translation of ribosomal structure and biogenesis”; whereas, for proteins most abundant with FeOOH growth (cluster 1), the highest functional representations were “replication, recombination and repair”, “translation of ribosomal structure and biogenesis”, and “none”. The “none” category refers to proteins that could not be classified into any COG functional category, and was abundant with FeOOH and MnO_2 growth (cluster 1 and 4, respectively) (Figure 4.11b). There were 52 and 22 proteins in the “none” category for cluster 1 and 4, respectively, most of which were annotated as “hypothetical protein” (40 and 17, respectively). Proteins are annotated as “hypothetical protein” because their corresponding genes do not have homologs with known functions [212]. Although caution needs to be exercised, a biological role could be inferred if the protein sequence contains conserved domains or have similarity to proteins with known function. In order to delineate the functions of detected “hypothetical proteins”, the protein sequences were queried against Pfam database for associated protein families [213]. Unfortunately, the most abundant hypothetical proteins in the “none” category of cluster 1 and 4 did not associate with any protein family in Pfam database and their function remained unclear.

In the hierarchical clustering, different growth conditions were also clustered based on the global protein abundance patterns (Figure 4.10). All the non-metal electron acceptors clustered away from the metal conditions, suggesting that the 2CP-C proteome responded to

metal and non-metal electron acceptors by controlling the protein expression profiles. Among the tested non-metal electron acceptors, nitrate and fumarate resulted in the most similar proteome abundance patterns which clustered closely together. N_2O was the most distant non-metal electron acceptor in the hierarchical clustering, growth with which generated the most distinct protein abundances pattern. To our surprise, the global protein abundance pattern with oxygen growth did not show significant distinction to the other anaerobic growth. This suggests that the oxygen in microaerophilic condition tested do not cause toxicity (e.g. reactive oxygen species (ROS)) to the cell, and the cell could utilize low concentration of oxygen as electron acceptor without additional biological processes (e.g. detoxification of ROS). The soluble metal Fe citrate and solid metal MnO_2 were grouped together, suggesting that the overall protein expression pattern of 2CP-C is similar in response to different forms of metal electron acceptors. However, growth with another solid metal FeOOH resulted in a protein abundance pattern distinct from all other tested growth conditions, which could be attributed by the recalcitrant growth with FeOOH as electron acceptor.

4.3.5 Quantitative Differential Analysis of 2CP-C Proteome Responding to Different Electron Acceptors

To further delineate the detailed differentiations of proteome expression between different growth conditions, the semi-quantitation power of label-free shotgun proteomics was exploited through statistical analysis. Quantifiable detected proteins passing the 95% PV filter, as mentioned above, were analyzed by one-way ANOVA test, which generated a total of 28

pairwise comparisons for all eight tested growth conditions. Out of the 1554 analyzed proteins, the abundances of 1100 displayed significant differential change across various growth conditions. The number of significantly changed proteins in each pairwise comparison ranged from 54 (between fumarate and N_2O) to 658 (between FeOOH and oxygen) (Figure 4.12). In general, most proteins (450 - 658) demonstrated significant abundance change when growth with FeOOH is compared to growth with another electron acceptor, consistent with the distinction of FeOOH in the hierarchical clustering result (see above).

Three growth conditions analyzed included different forms of metal electron acceptors (FeOOH, Fe citrate, and MnO_2), and five conditions with non-metal electron acceptors (nitrate, fumarate, tryptic soy broth, oxygen, and N_2O). The ANOVA analysis revealed 166 proteins whose abundance change is significant only when comparison is made between growth with metal and non-metal electron acceptors (see attachment “Supplementary Table 4.1”). In other words, these 166 proteins did not show significant abundance change in comparisons within the three metal electron acceptors, or within the five non-metal electron acceptors. Interestingly, out of the 166 differentially abundant proteins, 62 were consistently up-regulated in all metal electron accepter-growth, and 104 were consistently up-regulated in all non-metal electron acceptor-growth (see attachment “Supplementary Table 4.1”), suggesting that the effect of metal and non-metal electron acceptors on protein abundance is distinct and directional. The 62 proteins significantly more abundant with metal electron acceptor-growth mainly belong to the “energy production and conversion” and “amino acid transport and metabolism” functional categories in the COG database (Figure 4.13). The 104 proteins whose

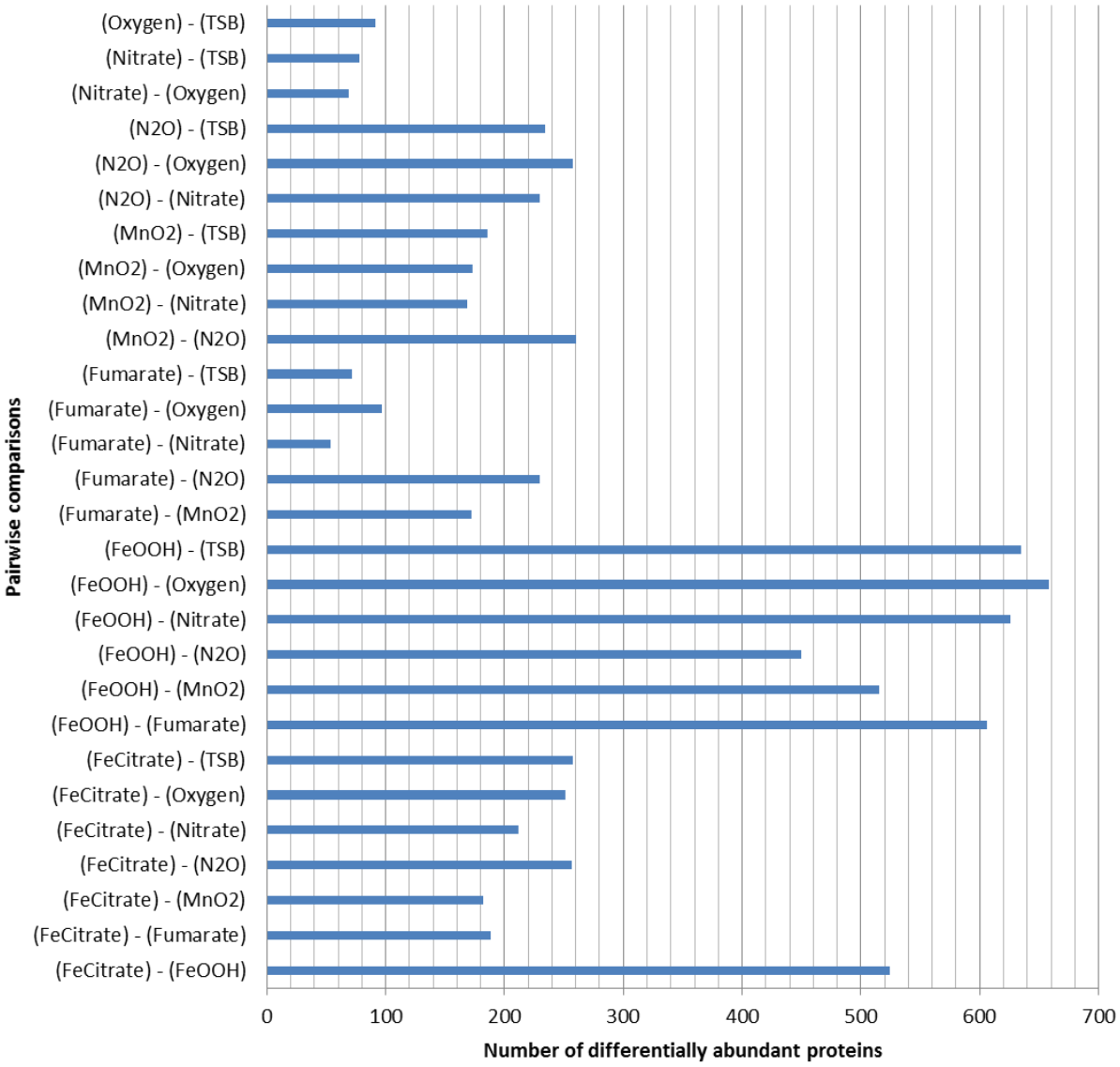
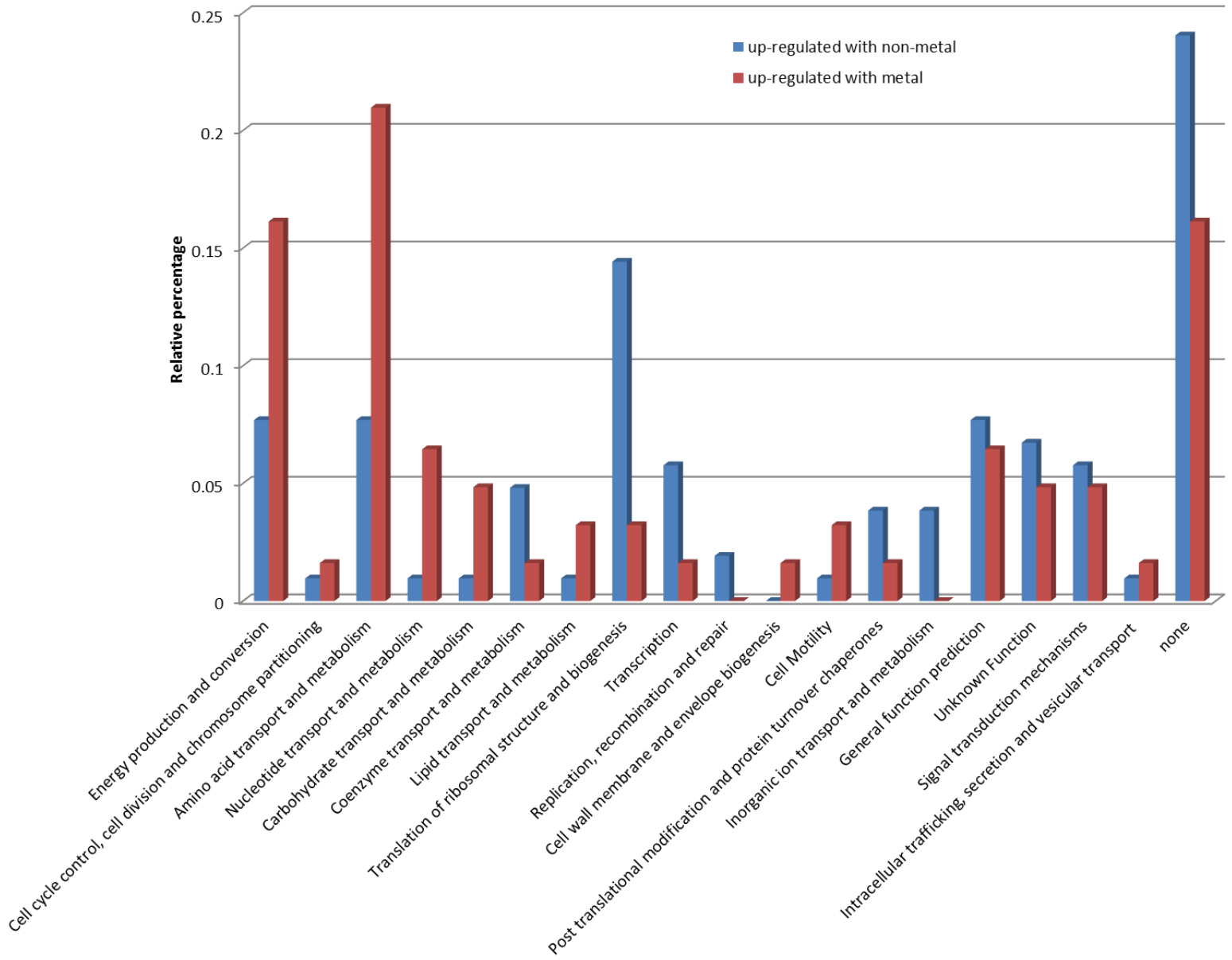


Figure 4.12. The number of differentially abundant proteins revealed by ANOVA test. The ANOVA test generated a total 28 pairwise comparisons (*y* axis) for all tested growth conditions. The number of proteins showing significant abundance differences (*x* axis) as determined by ANOVA test are plotted for each comparison.

Figure 4.13. COG functional category distribution for differentially abundant proteins. Proteins with differential abundances between metal and non-metal electron acceptor-growth are highlighted in red (more abundant with metal) or blue (more abundant with non-metal), and are plotted by the COG functional categories (*x* axis). The *y* axis represents the relative percentage of protein numbers in each functional category.

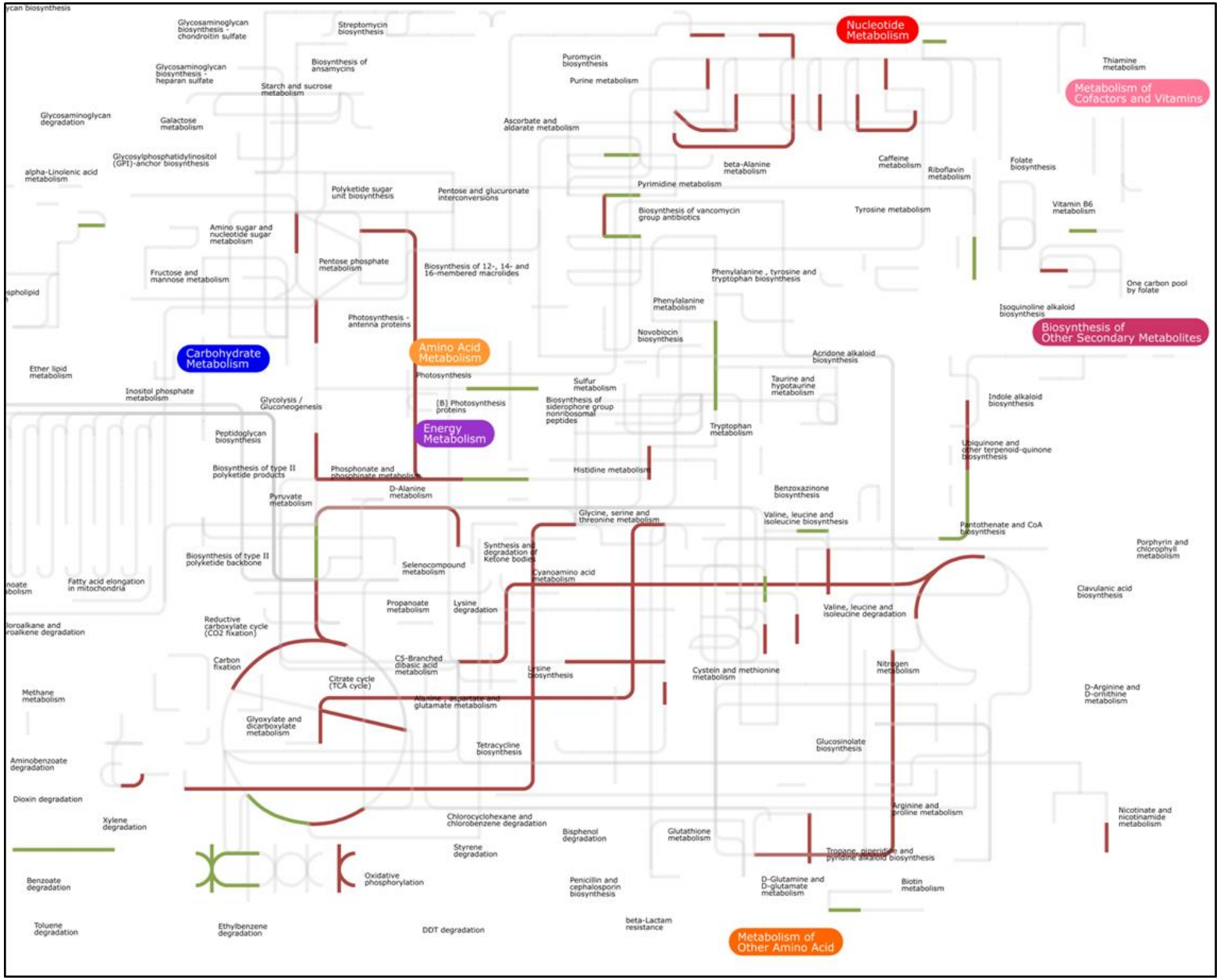

0.25
0.2
0.15
0.1
0.05
0
Relative percentage
up-regulated with non-metal
up-regulated with metal
Energy production and conversion
Cell cycle control, cell division and chromosome partitioning
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Translation of ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Cell wall membrane and envelope biogenesis
Cell Motility
Post translational modification and protein turnover chaperones
Inorganic ion transport and metabolism
General function prediction
Unknown Function
Signal transduction mechanisms
Intracellular trafficking, secretion and vesicular transport
none

abundances are significantly higher with non-metal electron acceptor-growth all showed more than 10 fold change (lg(metal)-lg(non-metal)<-1) (see attachment “Supplementary Table 4.1”), and “none” and “translation of ribosomal structure and biogenesis” were the highest represented functional categories (Figure 4.13).

To better visualize the functional differences of proteins significantly changed between growth with metal and non-metal electron acceptors, differentially abundant proteins were mapped to metabolic pathways in KEGG database using iPath 2.0. Out of the 166 proteins with differential changes between metal and non-metal electron acceptor-growths, 92 were annotated in the KEGG database and 42 of which were mapped to 60 (redundant) metabolic pathways (Figure 4.14). The metabolic pathways represented by proteins up-regulated in metal electron acceptor growth were highlighted in red, and no overlap was observed with pathways represented by proteins up-regulated in non-metal electron acceptor growth (highlighted in green) (Figure 4.12). Proteins up-regulated in metal electron acceptor-growth mapped to metabolic pathways participating in TCA cycle, metabolism of amino acid, nucleotide and carbohydrate (Figure 4.12), whereas proteins up-regulated in non-metal electron acceptor growth mostly involved in regulatory pathways for translation and cell motility.

Metabolic pathways mapping clearly indicated elevated expression of energy production pathways in growth with metal electron acceptors. As shown in Figure 4.14, major components in TCA cycle (KEGG pathway KO00020) were significantly up-regulated, contributed by increased abundances of 2-oxoglutarate ferredoxin oxidoreductase subunit beta [EC:1.2.7.3], malate dehydrogenase [EC:1.1.1.37], and citrate synthase [EC:2.3.3.1]. In addition,

Figure 4.14. Metabolic pathways represented by differentially expressed proteins with metal and non-metal electron acceptor-growth. Proteins exhibiting significant abundance change determined by ANOVA analysis are mapped onto KEGG metabolic pathways. The pathways corresponding to up-regulated proteins in metal electron acceptor-growth are highlighted in red, and up-regulated pathways in non-metal electron acceptors are highlighted in green.


Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Metabolism of Other Amino Acid
Glycosaminoglycan biosynthesis - chondroitin sulfate
Streptomycin biosynthesis
Biosynthesis of ansamycins
Puromycin biosynthesis
Purine metabolism
Thiamine metabolism
Glycosaminoglycan biosynthesis - heparan sulfate
Starch and sucrose metabolism
Glycosaminoglycan degradation
Galactose metabolism
Ascorbate and aldarate metabolism
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
alpha-Linolenic acid metabolism
beta-Alanine metabolism
Caffeine metabolism
Riboflavin metabolism
Folate biosynthesis
Polyketide sugar unit biosynthesis
Pentose and glucuronate interconversions
Pyrimidine metabolism
Biosynthesis of vancomycin group antibiotics
Tyrosine metabolism
Vitamin B6 metabolism
Amino sugar and nucleotide sugar metabolism
Fructose and mannose metabolism
Pentose phosphate metabolism
Biosynthesis of 12-, 14- and 16-membered macrolides
Phenylalanine , tyrosine and tryptophan biosynthesis
One carbon pool by folate
Photosynthesis - antenna proteins
Phenylalanine metabolism
Isoquinoline alkaloid biosynthesis
Novobiocin biosynthesis
Ether lipid metabolism
Acridone alkaloid biosynthesis
Photosynthesis
Sulfur metabolism
Taurine and hypotaurine metabolism
Inositol phosphate metabolism
Glycolysis / Gluconeogenesis
[B] Photosynthesis proteins
Biosynthesis of siderophore group nonribosomal peptides
Indole alkaloid biosynthesis
Peptidoglycan biosynthesis
Tryptophan metabolism
Ubiquinone and other terpenoid-quinone biosynthesis
Biosynthesis of type II polyketide products
Phosphonate and phosphinate metabolism
Histidine metabolism
D-Alanine metabolism
Benzoxazinone biosynthesis
Pyruvate metabolism
Glycine, serine and threonine metabolism
Valine, leucine and isoleucine biosynthesis
Pantothenate and CoA biosynthesis
Porphyrin and chlorophyll metabolism
Biosynthesis of type II polyketide backbone
Synthesis and degradation of ketone bodies
Selenocompound metabolism
Cyanoamino acid metabolism
Fatty acid elongation in mitochondria
Clavulanic acid biosynthesis
Propanoate metabolism
Lysine degradation
Valine, leucine and isoleucine degradation
Reductive carboxylate cycle (CO2 fixation)
Carbon fixation
C5-Branched dibasic acid metabolism
Lysine biosynthesis
Cysteine and methionine metabolism
Nitrogen metabolism
Citrate cycle (TCA cycle)
D-Arginine and D-ornithine metabolism
Methane metabolism
Alanine , aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
Tetracycline biosynthesis
Glucosinolate biosynthesis
Aminobenzoate degradation
Dioxin degradation
Arginine and proline metabolism
Nicotinate and nicotinamide metabolism
Xylene degradation
Chlorocyclohexane and chlorobenzene degradation
Bisphenol degradation
Glutathione metabolism
Styrene degradation
Tropane, piperidine and pyridine alkaloid biosynthesis
Oxidative phosphorylation
Penicillin and cephalosporin biosynthesis
D-Glutamine and D-glutamate metabolism
Biotin metabolism
Benzoate degradation
Toluene degradation
Ethylbenzene degradation
DDT degradation
beta-Lactam resistance

other ATP producing processes, glycolysis (KEGG pathway KO00630) and oxidative phosphorylation (KEGG pathway KO00190), also exhibited significant up-regulation with metal electron acceptor-growth. Due to the up-regulation of many energy producing pathways, it was not surprising that porphyrin metabolism (KEGG pathway KO00860), a pathway related to heme production, was up-regulated in metal electron acceptor-growth as represented by increased abundance of glutamate-1-semialdehyde 2,1-aminomutase [EC:5.4.3.8]. Heme is an electron carrier component in many proteins involved in the electron transfer chain (e.g. cytochrome), which inextricably relates to energy production. The increased expression of proteins involved in energy-producing pathways either reflects an overall higher energy production, or indicates that the cells grown with metal electron acceptors underwent stress condition and were making efforts to produce more energy.

Amino acid and nucleotide metabolism also demonstrated up-regulation in metal electron acceptor-growth (Figure 4.12). Specifically, the following proteins responsible for amino acid synthesis all showed higher abundances with metal electron acceptor-growth: branched-chain amino acid aminotransferase [EC:2.6.1.42], shikimate dehydrogenase [EC:1.1.1.25], aspartate-semialdehyde dehydrogenase [EC:1.2.1.11], 3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54], argininosuccinate synthase [EC:6.3.4.5], carbamoyl-phosphate synthase large subunit [EC:6.3.5.5], 4-hydroxy-tetrahydrodipicolinate reductase [EC:1.17.1.8], and aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit A [EC:6.3.5.6 6.3.5.7]. For biosynthesis of purine and pyrimidine, increased abundances of purine-nucleoside phosphorylase [EC:2.4.2.1], ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1], and phosphoribosylaminoimidazole-succinocarboxamide synthase

[EC:6.3.2.6] in metal electron acceptor-growth was determined significant by ANOVA analysis and contributed to the elevation of nucleotide metabolism. The increased synthesis of amino acids and nucleotides could be a result of overall higher expression of energy-generating pathways in cells grown with metal electron acceptor.

The significantly increased proteins with non-metal electron acceptor-growth, on the other hand, only mapped to a few metabolic pathways for amino acid and nucleotide biosynthesis (e.g. KEGG pathway KO00400, KO00240), lipid metabolism (KEGG pathway KO00564), etc. However, most abundant proteins in non-metal electron acceptor-growth were ribosomal proteins involved in regulatory pathways for translation (KEGG pathway KO03010). Protein CheW was significantly more abundant in non-metal electron acceptor-growth, and was mapped to 10 elements in the regulatory pathways for chemotaxis (KO02030) and two-component system (KO02020).

4.4 Conclusions

Using the LC-MS/MS approach, deep measurement comprehensively characterized the proteome of the DMRB *A. dehalogenans* strain 2CP-C, and provided a global survey of the proteome-wide responses of 2CP-C to seven different electron acceptors. High proteome coverage was achieved, as represented by the pan-proteome. The results also revealed a core proteome sustaining major metabolic processes regardless of different electron-accepting environments. Quantitative analysis revealed that metal and non-metal electron acceptors have distinct effect on the proteome profiles of 2CP-C. Growth with metal electron acceptor

resulted in the expression of more protein participating in energy production compared to non-metal electron acceptor-growth.

Chapter 5

Proteome Characterizations for *Shewanella oneidensis* Strain MR-1 and *Geobacter daltonii* Strain FRC-32

5.1 Introduction to *Shewanella oneidensis* MR-1 and *Geobacter daltonii* FRC-32

Many species within the *Shewanella*, *Geobacter*, and *Anaeromyxobacter* genera are dissimilatory metal reducing bacteria (DMRB) with versatile respiratory abilities, as mentioned in Chapter 1. All three genera have been found to be the intrinsic microbial colonizers in many environments where metal reduction is a major biogeochemical process taking place. In an *in situ* sediment cap study, all three genera were abundant in relation to the total *Bacteria* population, with *Geobacter* spp. being the most dominant [214]. *Anaeromyxobacter* and *Shewanella* spp. displayed higher abundances in upper zones compared to the rest of the sediment cap possibly due to oxygen utilization [214]. In other field studies of uranium contaminated sediments, *Anaeromyxobacter* spp. and *Geobacter* spp. were proven to be most predominant microorganisms related to *in situ* uranium bioremediation [90, 172]. Despite their co-presence in the environment and commonalities of respiratory versatility, the similarities or differences of the overall metabolic activities of *Shewanella* spp., *Geobacter* spp., and *Anaeromyxobacter* spp. are largely unknown. In order to gain insight into the molecular level functional activities of *Shewanella*, *Geobacter*, and *Anaeromyxobacter*, the proteome characterizations for *Shewanella oneidensis* strain MR-1 and *Geobacter daltonii* strain FRC-32 are detailed in this chapter, and are compared to the proteome of *Anaeromyxobacter*

dehalogenans strain 2CP-C (presented in Chapter 4) to reveal the similarities or differences between these three DMRBs.

As briefly introduced in Chapter 1, *S. oneidensis* strain MR-1 and *G. daltonii* strain FRC-32 are both important DMRBs. While *S. oneidensis* strain MR-1 belongs to the gamma subdivision of Proteobacteria, *G. daltonii* strain FRC-32 is phylogenetically closer related to *A. dehalogenans* strain 2CP-C, both of which are members in the delta subdivision of Proteobacteria. Strain FRC-32 is an obligate anaerobe, and strain MR-1 and 2CP-C are microaerophiles able to grow in environment with low level of oxygen. All three bacterial strains have demonstrated versatile respiratory capabilities for the reduction of ferric iron, manganese, uranium, etc. based on physiological characterizations [87, 88, 215-217]. Using mass spectrometry-based proteomics approach, details about the proteome expression and representing metabolic pathways in response to different electron acceptors are presented and contrasted for *S. oneidensis* strain MR-1, *G. daltonii* strain FRC-32, and *A. dehalogenans* strain 2CP-C.

5.2 Experimental Section

Bacterial culture growth, sample collection, protein extraction and digestion, peptide separation through two-dimensional liquid chromatography, and following protein identification using tandem mass spectrometry on a linear ion trap (Thermo Scientific, Inc.) are described in detail in Chapter 2 and 3.

5.3 Results and Discussion

5.3.1 Characterization of the Pan-proteomes of *S. oneidensis* MR-1 and *G. daltonii* FRC-32

The proteomic characterization for *S. oneidensis* strain MR-1 was achieved for growth with seven conditions, including FeOOH, Fe citrate, MnO_2, nitrate, tryptic soy broth, fumarate, and oxygen. *G. daltonii* strain FRC-32 is an obligate anaerobe, and could only grow with FeOOH, Fe citrate, MnO_2, and fumarate, out the seven tested growth conditions. The total number of proteins identified under most growth conditions for strain MR-1 was around 1,600 (Fe citrate, MnO_2, nitrate, fumarate, and oxygen – Figure 5.1). With tryptic soy broth-growth (fumarate provided as electron acceptor), strain MR-1 yielded slightly lower number of protein identification (1270) (Figure 5.1). Compared to the minimum salt medium, the tryptic soy broth medium contains a complex mixture of undefined nutrients, which may be directly utilized by the cells to support growth without initiating nutrient-synthesizing pathways that are necessary to grow with minimal-nutrients. For strain FRC-32, growth with fumarate generated the highest number (1543) of total protein identification among all tested growth conditions (Figure 5.2). The least number of total protein identification was achieved when FeOOH was used as electron acceptor for both strain MR-1 and FRC-32, 296 and 773, respectively, which represent approximately 19% and 50% of proteins identified in other conditions, respectively (Figure 5.1 and 5.2). Similar to *A. dehalogenans* strain 2CP-C (discussed in Chapter 4), growth of strain MR-1 and FRC-32 was slowest with FeOOH supplied as the sole electron acceptor, which led to less variation of proteins being expressed and measured. The protein extraction method used for FeOOH-grown cultures was different from other growth conditions; however, it should not

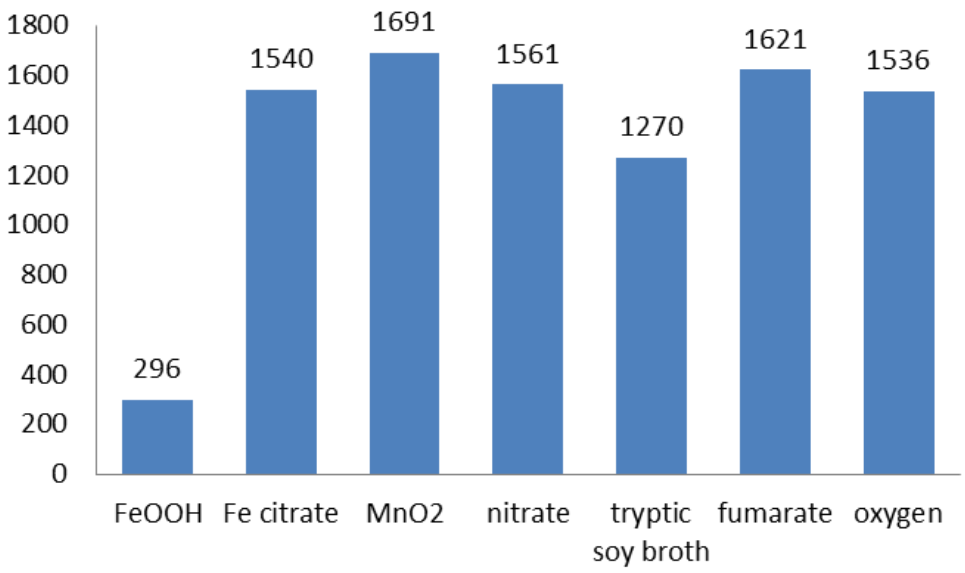


Figure 5.1. Proteome identifications for *S. oneidensis* MR-1 grown under various conditions. Total identified protein numbers are pooled for biological triplicates of each growth condition.

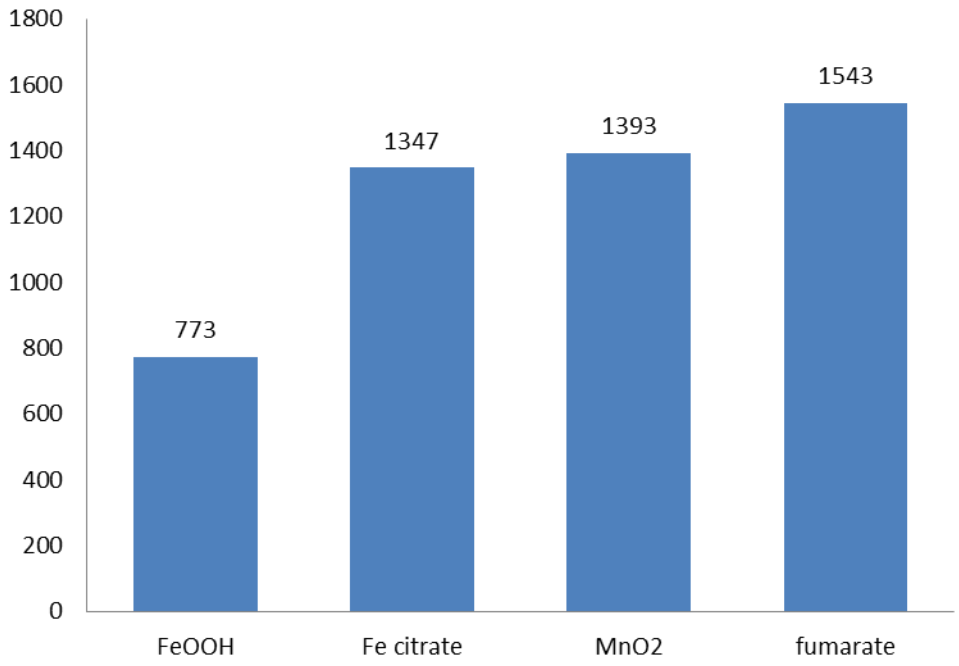


Figure 5.2. Proteome identifications for *G. daltonii* FRC-32 grown under various conditions. Total identified protein numbers are pooled for biological triplicates of each growth condition.

affect the total proteome identification, as discussed in Chapter 2 and 3. In addition, if the proteome expression with FeOOH-growth was comparable to other conditions, sample preparation bias would have resulted in similar protein identification impairments to both strain MR-1 and FRC-32 (unlike the observation of respective 19% and 50% of protein identifications in FeOOH-growth compared to other conditions). Thus, the protein identification results of strain MR-1 and FRC-32 further testify that the low protein identification numbers with FeOOH growth was not due to the different performance of proteome extraction methods.

The pan-proteome for each strain represents the all identified proteins pooled across all growth conditions. Combining the proteome identifications from all tested growth conditions, strain MR-1 yielded 2256 non-redundant proteins (the pan-proteome), representing 50% of the total predicted open reading frames (ORFs) in strain MR-1 proteome. Similarly, the identified pan-proteome of strain FRC-32 had 1931 non-redundant proteins, also corresponding to 50% of all possible ORFs in the predicted proteome database. Similar to *A. dehalogenans* strain 2CP-C (shown in Chapter 4), the pan-proteomes of strain MR-1 and FRC-32 both have only a slight increase in protein numbers, as compared to proteomes identified in most individual growth conditions. The substantial proteome overlap between different growth conditions suggest that the entire proteomes of strain MR-1 and FRC-32 are not significantly distinct when grown with different electron acceptors.

The pan-proteomes of strain MR-1 and FRC-32 spanned all cellular localizations predicted by PSORTb.3.0 [185]. About 60-75% of the predicted cytoplasmic, periplasmic, outer membrane, and extracellular proteins were detected for both strain MR-1 and FRC-32, whereas

fewer percentages (roughly 30-40%) of predicted outer membrane proteins and proteins with unknown cellular localizations were detected (Figure 5.3 and 5.4). The cellular localization distributions for proteins identified under each growth conditions also showed complete coverage of proteins localized in all cellular compartments (Figure 5.5 and 5.6). For strain MR-1 grown with FeOOH, even though the total detected proteome is only 19% of that in other growth conditions (Figure 5.1), the percentages of detected proteins in each cellular localizations were highly comparable to all other growth conditions (Figure 5.5). And similar localization distribution for strain FRC-32 proteome was also observed for all tested growth conditions (Figure 5.6), suggesting comprehensive protein samplings as well as adequate proteome coverages were achieved for both strains grown with all tested electron acceptors.

Detected proteins comprised almost all functional classifications in the clusters of orthologous groups (COG) database [218] (Figure 5.7 and 5.8), except a histone deacetylase super family protein of strain MR-1 (SO_1815) classified as “chromatin structure and dynamics”, and an RNA 3’-phosphate cyclase protein of strain FRC-32 (Geob_0963) belonging to the “RNA processing and modification” category, were not detected under any growth condition. For both strain MR-1 and FRC-32, most detected proteins (17% and 20%, respectively) in the pan-proteomes could not be classified into any functional categories and were designated as “none” in the COG database (Figure 5.7 and 5.8). Compared to *A. dehalogenans* strain 2CP-C (Chapter 4) whose pan-proteome contained 22% “none” category proteins, the MR-1 proteome is slightly more completely functionally classified and annotated. Typically, most proteins in the “none” categories are annotated as hypothetical proteins because of the lack of conserved domain or

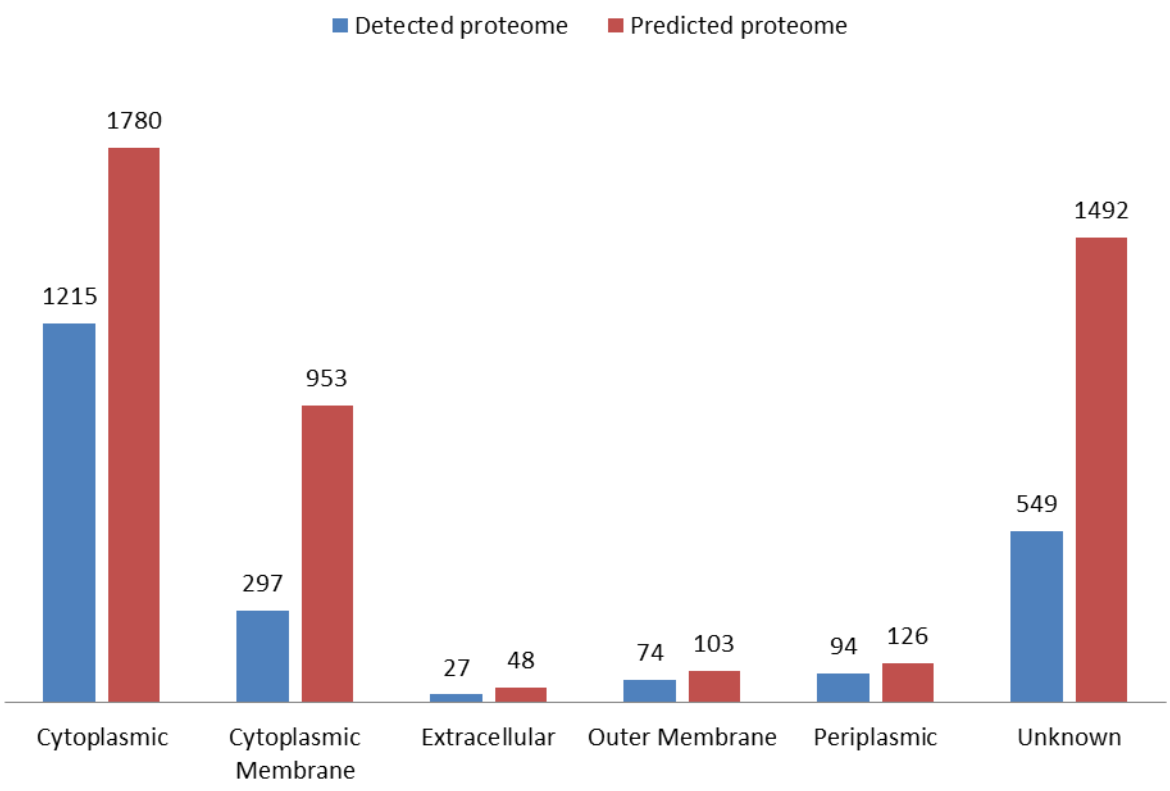


Figure 5.3. Cellular localization distribution of predicted (red) and detected (blue) proteome of *S. oneidensis* MR-1. Detected proteome is represented by pooled protein identifications of all replicates and all growth conditions.

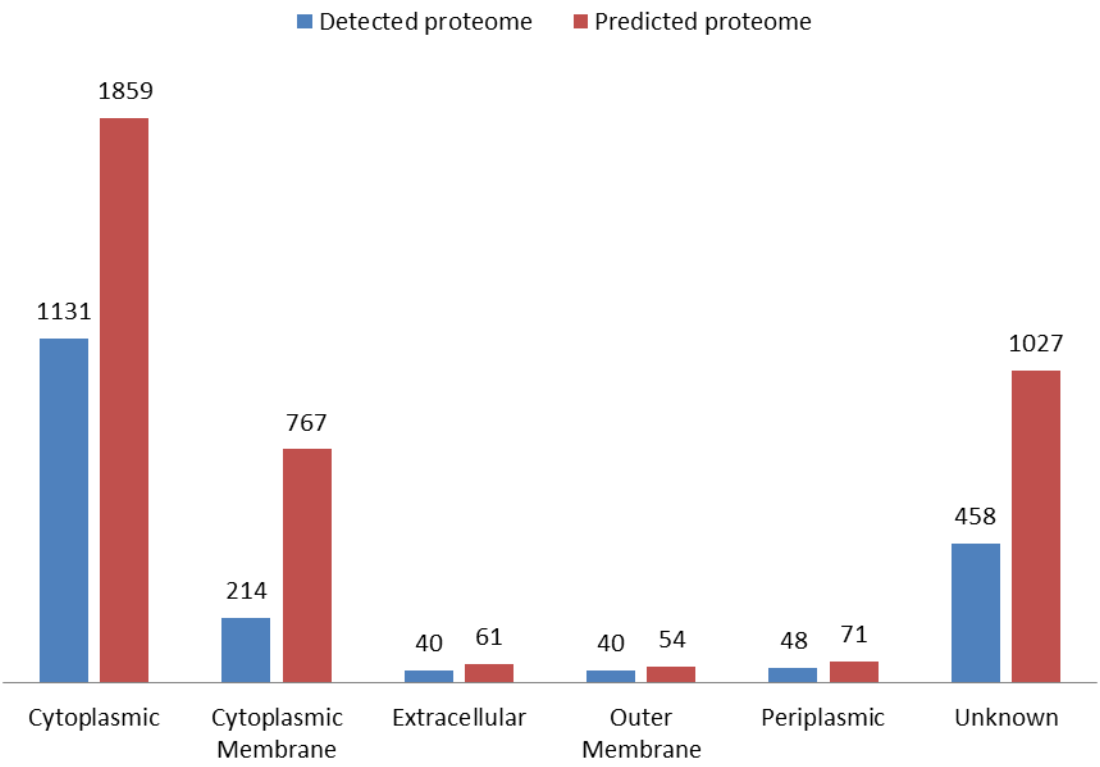


Figure 5.4. Cellular localization distribution of predicted (red) and detected (blue) proteome of *G. daltonii* FRC-32. Detected proteome is represented by pooled protein identifications of all replicates and all growth conditions.

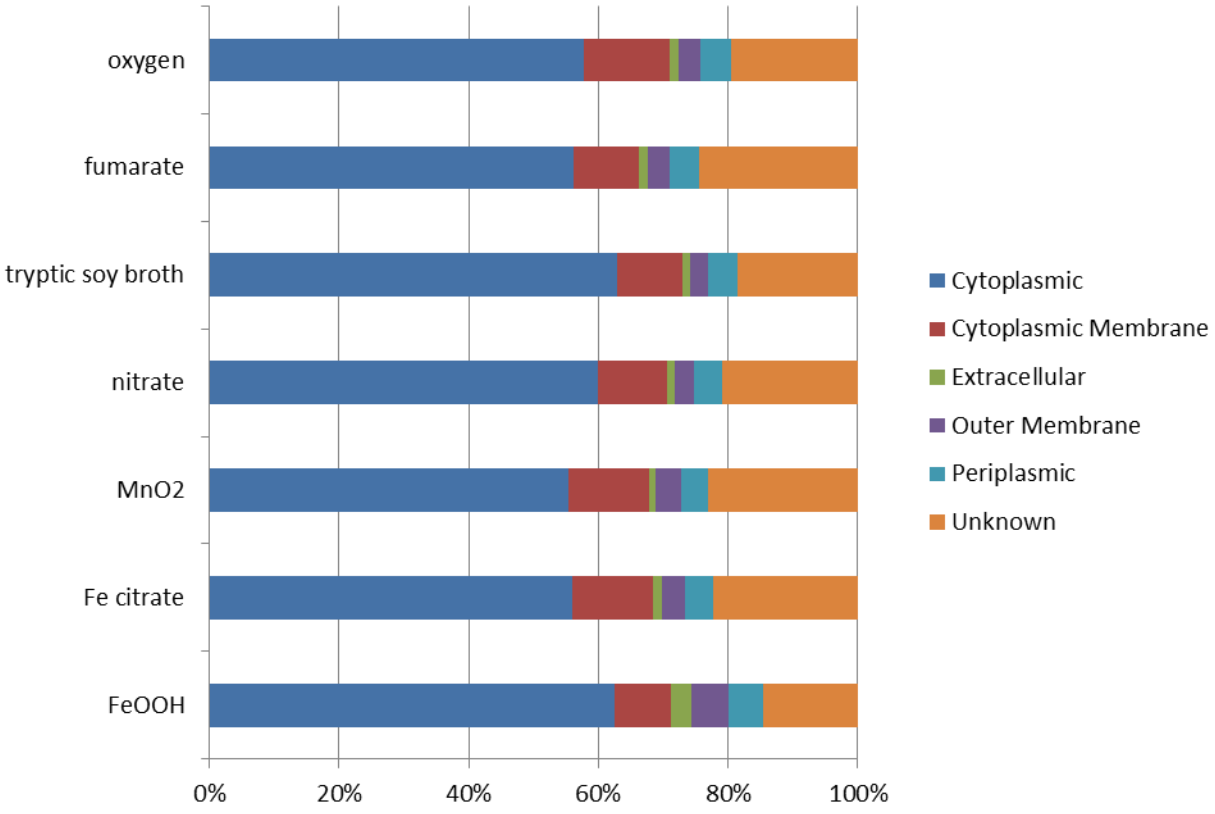


Figure 5.5. Cellular localization distribution of *S. oneidensis* MR-1 proteome under growth under different conditions.

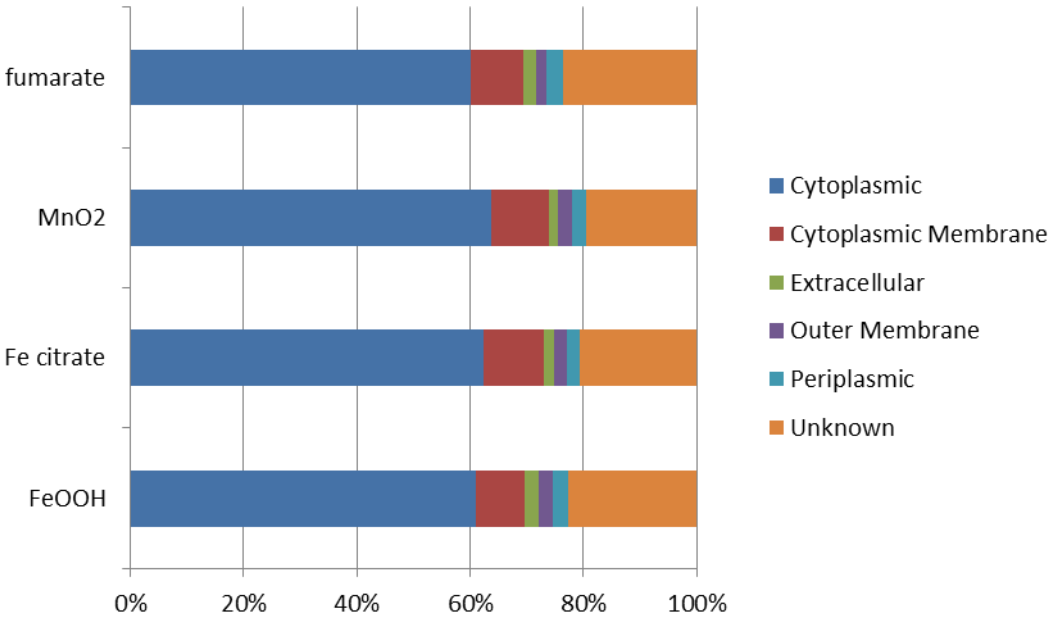


Figure 5.6. Cellular localization distribution of *G. daltonii* FRC-32 proteome under growth under different conditions.

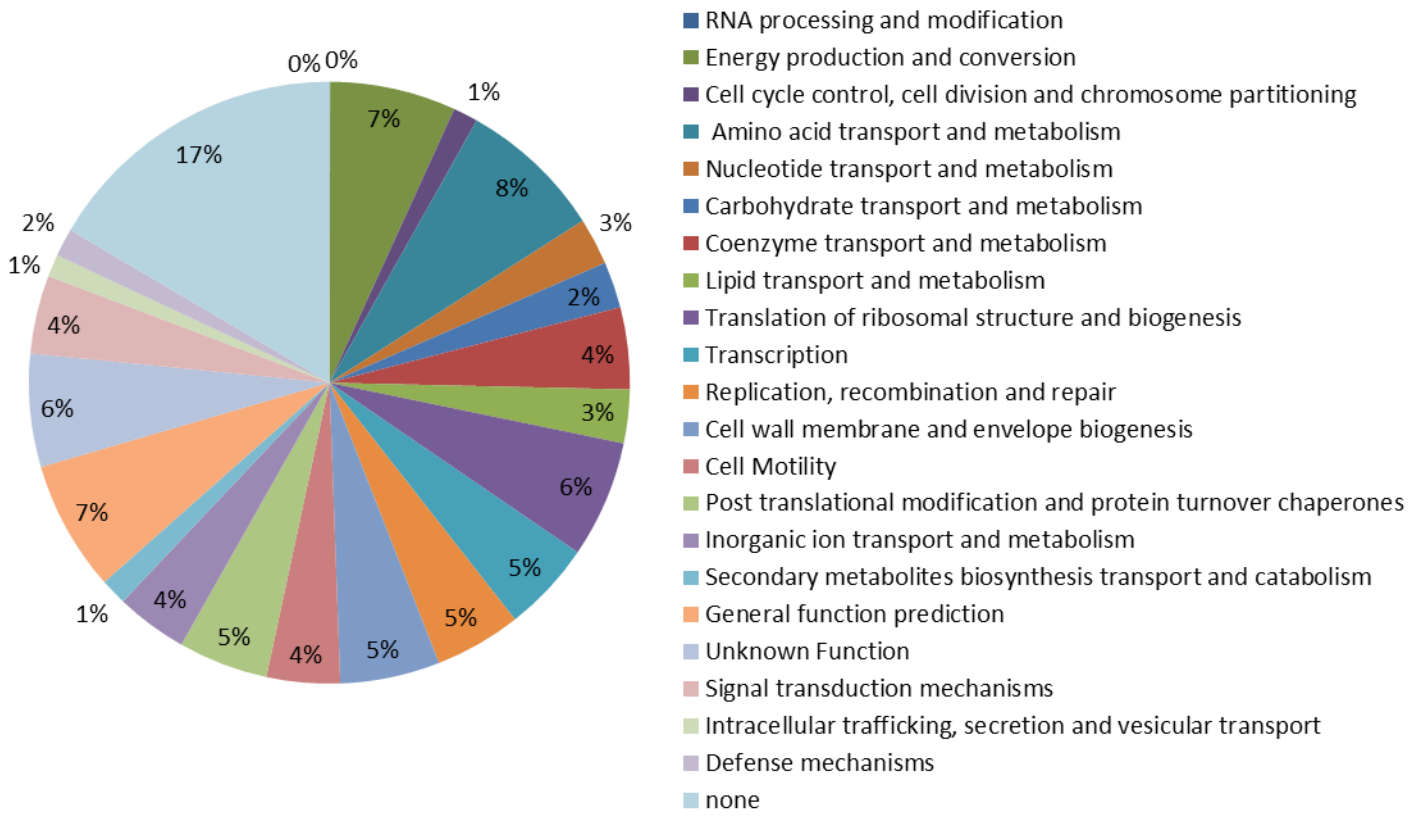


Figure 5.7. COG functional category distribution of detected proteome of *S. oneidensis* MR-1. Detected proteome is represented by pooled protein identifications of all replicates and all growth conditions.

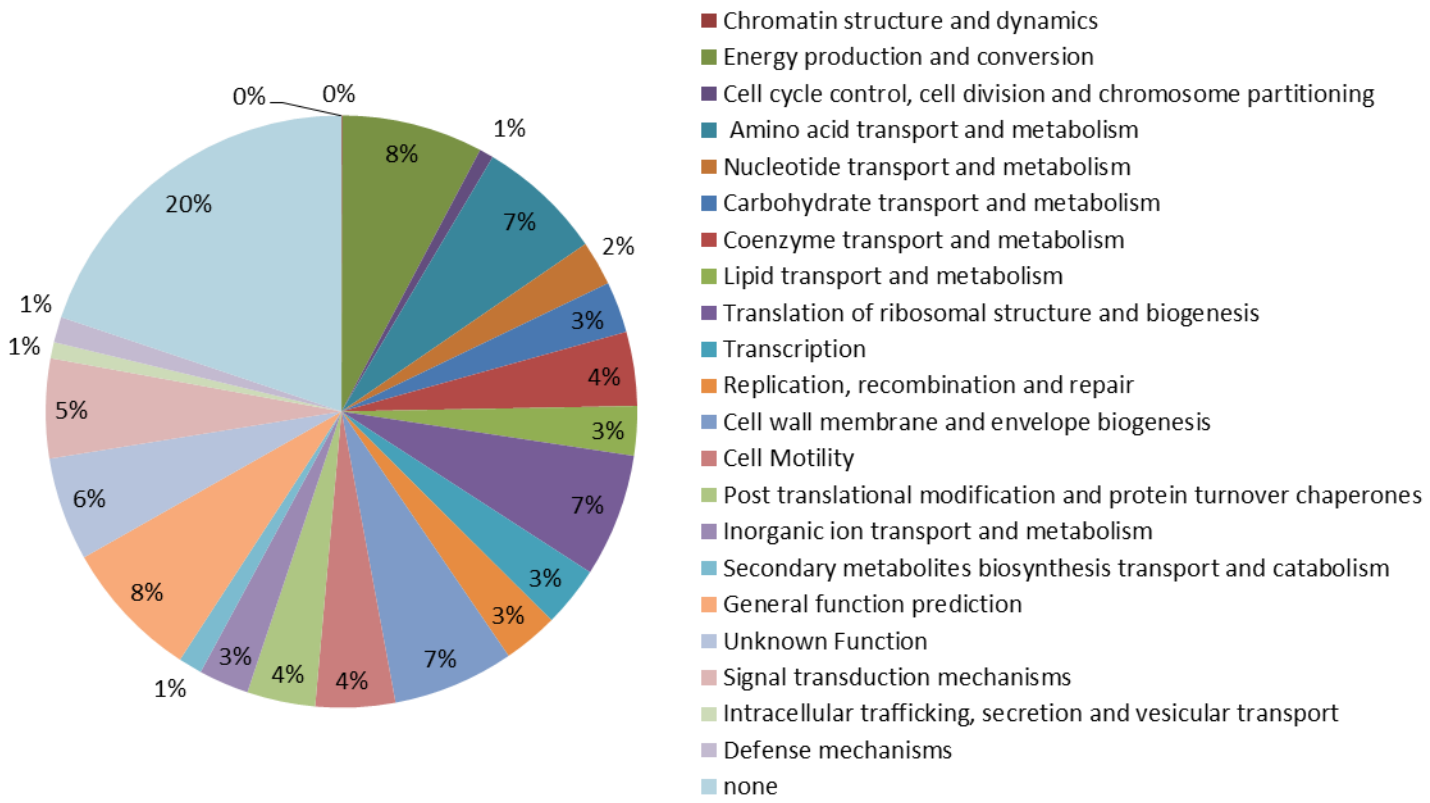


Figure 5.8. COG functional category distribution of detected proteome of *G. daltonii* FRC-32. Detected proteome is represented by pooled protein identifications of all replicates and all growth conditions.

motif and limited sequence similarity to functionally characterized proteins in the database. With the comprehensive proteome measurements, the expression of “none” category proteins were validated, and could be linked to certain tested growth conditions (Figure 5.9 and 5.10). The COG functional category distributions of proteins detected under each growth conditions have comparable patterns for strain MR-1, except with FeOOH-growth (Figure 5.9) which had higher percentages of proteins in the “energy production and conversion” and “translation of ribosomal structure and biogenesis” categories. For strain FRC-32, the COG distributions are variable among all tested growth conditions (Figure 5.10), suggesting that each tested electron acceptor has distinct impact on expressed proteins in each functional classification. Overall, the percent share of total detected proteins in each COG category is roughly consistent between different organisms (strain MR-1, FRC-32, and 2CP-C) (Figure 5.9 and 5.10, and Chapter 4).

To achieve a more in-depth functional representation of the pan-proteomes of strain MR-1 and FRC-32, all detected proteins across all biological replicates under all tested growth conditions were pooled and mapped onto metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [219] using the pathway mapping tool iPath2.0 [211] (Figure 5.11 and 5.12). Due to incomplete annotations, only 1478 proteins out of the detected pan-proteome (2256 proteins) of strain MR-1, and 1075 proteins out of the pan-proteome (1931 proteins) of strain FRC-32 were functionally annotated by KEGG. For those detected protein with available functional information, their representing metabolic pathways were highlighted in blue in Figure 5.11 and 5.12 for strain MR-1 and FRC-32, respectively. Both pan-proteomes showed high coverage of almost all the explored pathways for each organism, ranging from energy metabolism, metabolism of amino acid, nucleotide, lipid, to the

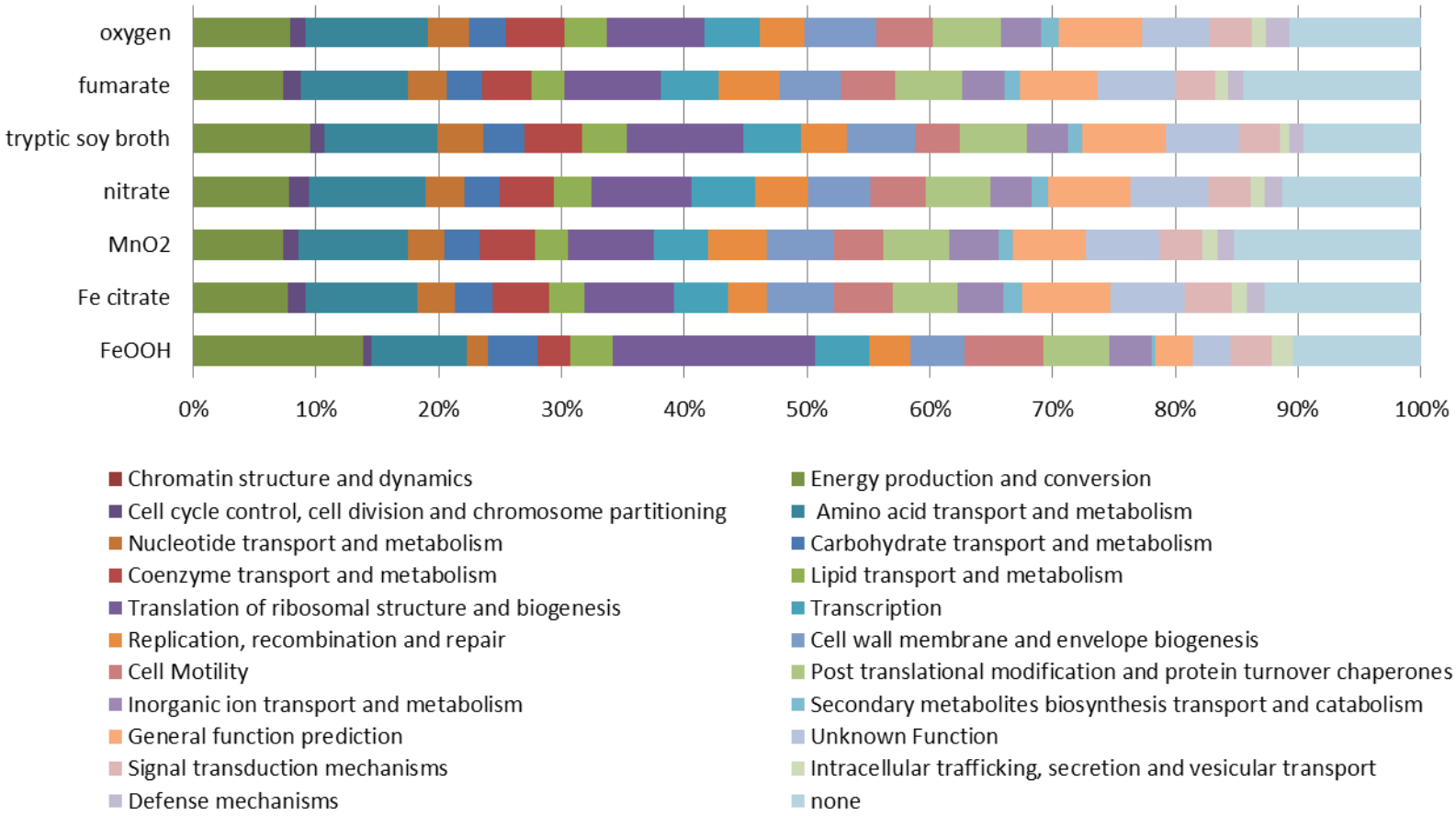


Figure 5.9. COG functional category distribution of *S. oneidensis* MR-1 proteome under growth under different conditions.

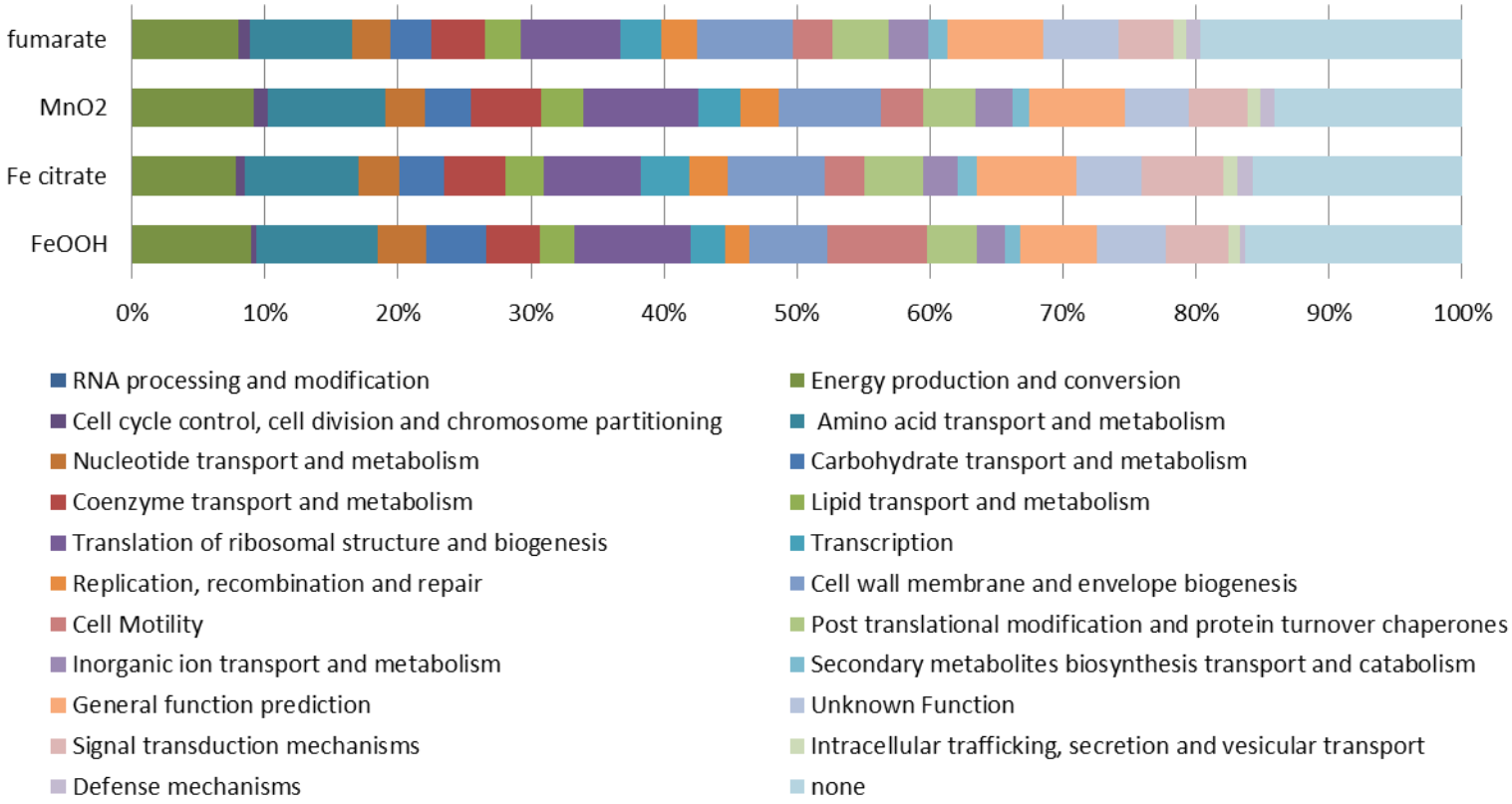


Figure 5.10. COG functional category distribution of *G. daltonii* FRC-32 proteome under growth under different conditions.

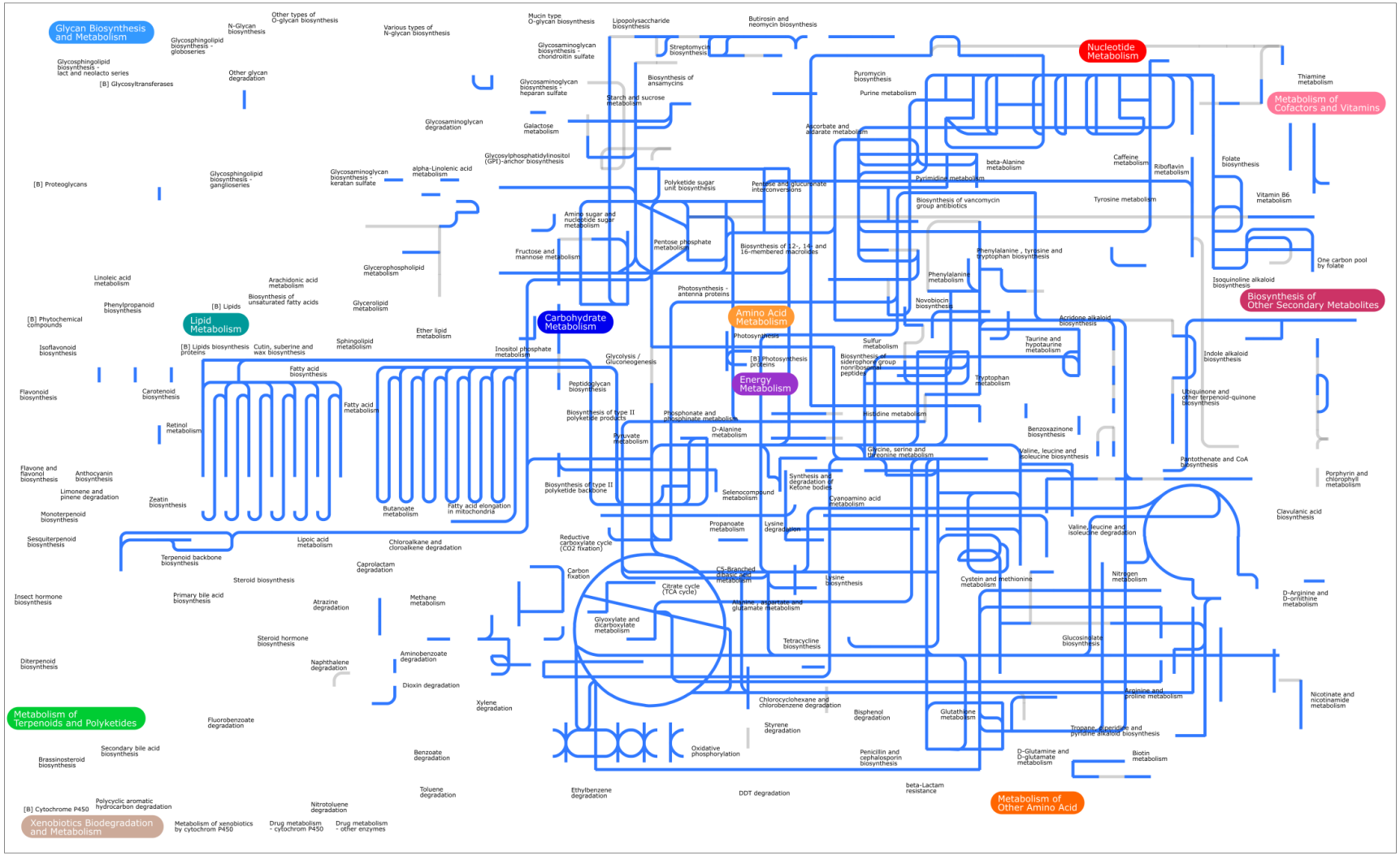


Figure 5.11. Metabolic pathways of detected *S. oneidensis* MR-1 pan-proteome. The pan-proteome of MR-1 consists detected proteins pooled across all growth conditions and all replicates, and is mapped onto KEGG metabolic pathways and highlighted blue.

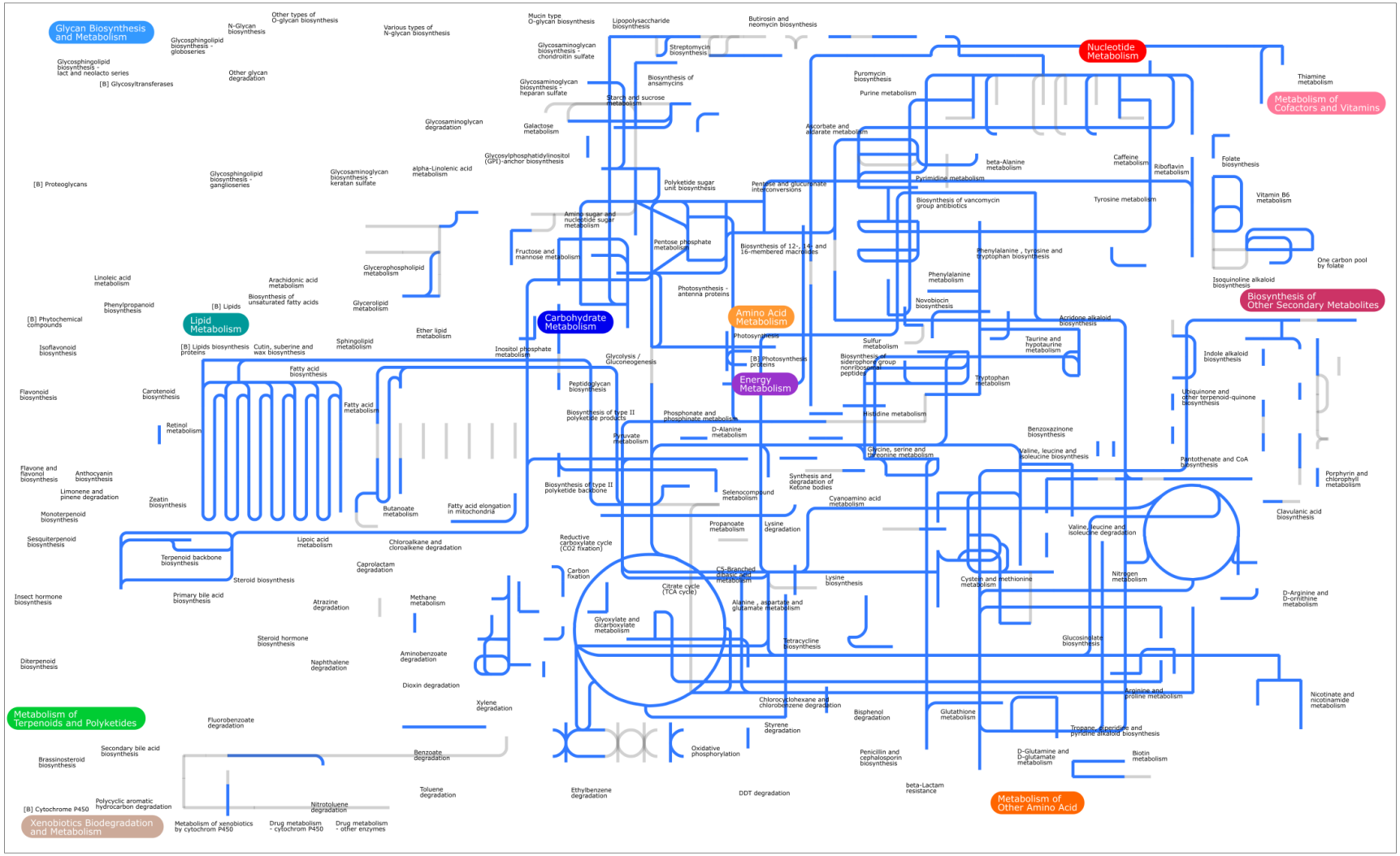


Figure 5.12. Metabolic pathways of detected *G. daltonii* FRC-32 pan-proteome. The pan-proteome of FRC-32 consists detected proteins pooled across all growth conditions and all replicates, and is mapped onto KEGG metabolic pathways and highlighted blue.

biosynthesis of secondary metabolites (Figure 5.11 and 5.12). The metabolic maps of the pan-proteomes reflect the cellular activities that are required for cell growth.

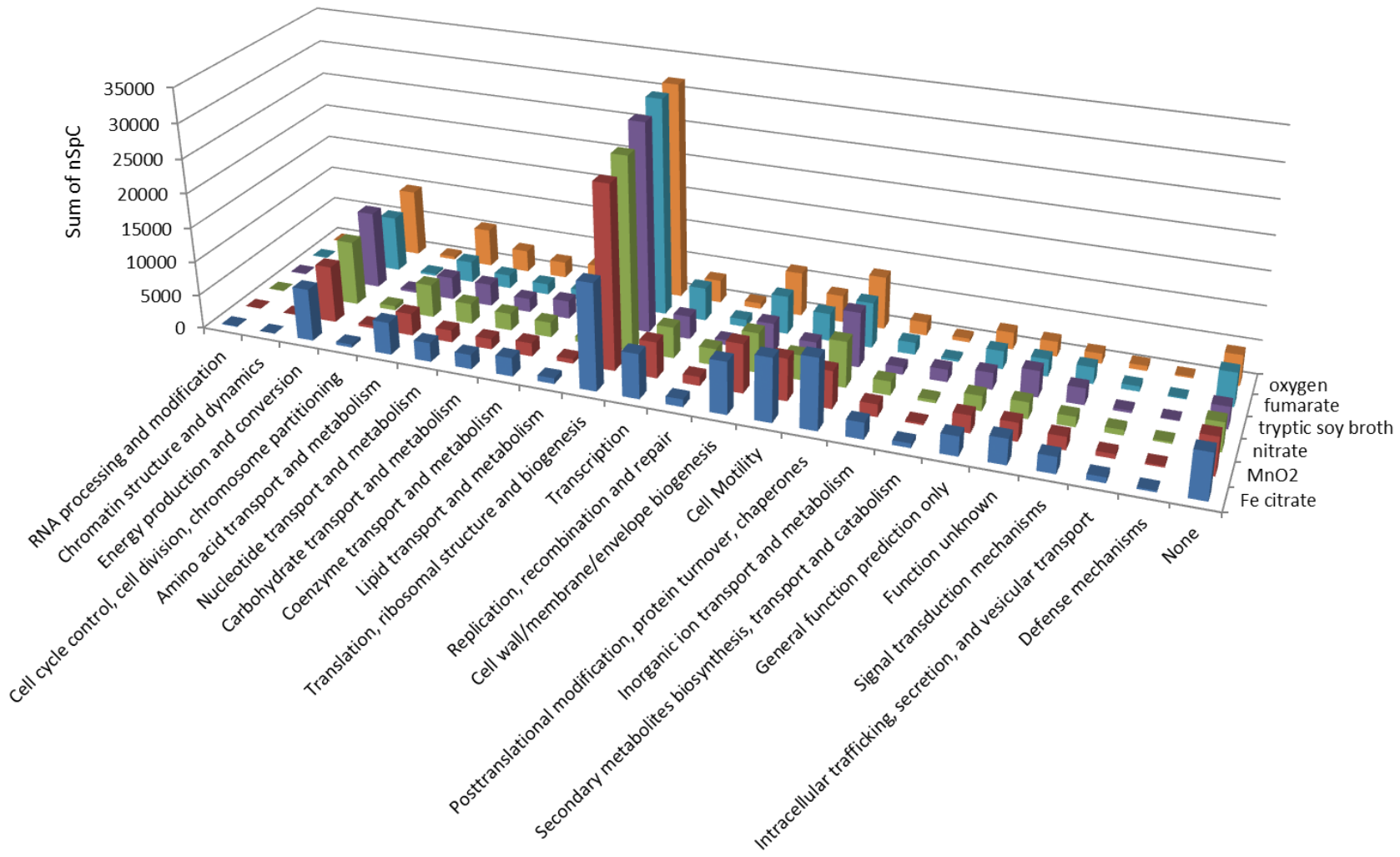
5.3.2 The "Common" and "Core" Proteomes of *S. oneidensis* MR-1 and *G. daltonii* FRC-32

The "common" proteome represents proteins detected under all tested growth conditions for each organism. For strain MR-1, a measured overlapping proteome of 260 proteins was shared across all tested electron acceptors. In comparison, the measured common proteome for strain FRC-32 contained 594 proteins, representing 31% of the pan-proteome. The relatively larger common proteome in strain FRC-32 is probably due to the fewer growth conditions tested, resulting in a larger overlap between detected proteomes under each condition, and more protein identifications with FeOOH growth. The common proteome of strain MR-1 appears to be the smallest among all three bacterial strains, but it is likely to be heavily skewed by the low protein identification with FeOOH growth. Thus, in order to represent the true biological core proteome, FeOOH condition should be omitted. The "core" proteome represents all the proteins detected under all tested growth conditions with FeOOH condition omitted. When FeOOH condition is omitted from the analysis here, the detected core proteomes of strains MR-1 and FRC-32 contain around 922 and 1,022 proteins, respectively, representing 41% and 53% of the pan-proteomes, respectively.

In order to probe the functional perspective of the common and core proteomes of strain MR-1 and FRC-32, the quantitative distributions by COG functional categories of common and core proteomes in each growth conditions are presented in Figure 5.13 and 5.14,

Figure 5.13. Quantitative distribution of *S. oneidensis* MR-1 common proteome and core proteome by COG functional categories. (a) The common proteome contains identified proteins overlapping all tested growth conditions. (b) The core proteome represents detected proteins overlapping all conditions with FeOOH condition omitted.

(a)


Sum of nSpC
35000
30000
25000
20000
15000
10000
5000
0
RNA processing and modification
Chromatin structure and dynamics
Energy production and conversion
Cell cycle control, cell division, chromosome partitioning
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Translation, ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Cell wall/membrane/envelope biogenesis
Cell Motility
Posttranslational modification, protein turnover, chaperones
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport and catabolism
General function prediction only
Function unknown
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
Defense mechanisms
None
oxygen
fumarate
tryptic soy broth
nitrate
MnO2
Fe citrate

(b)

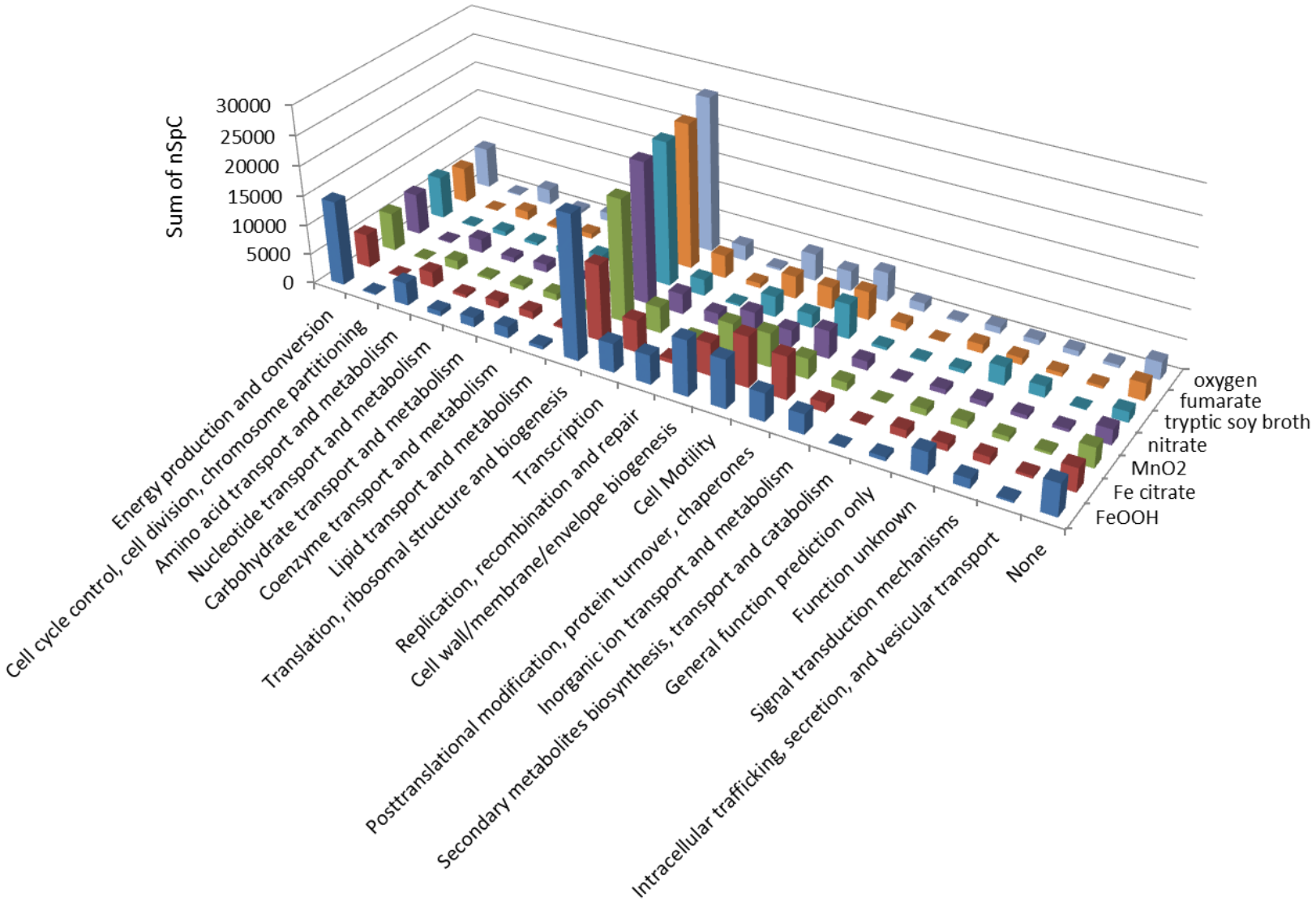
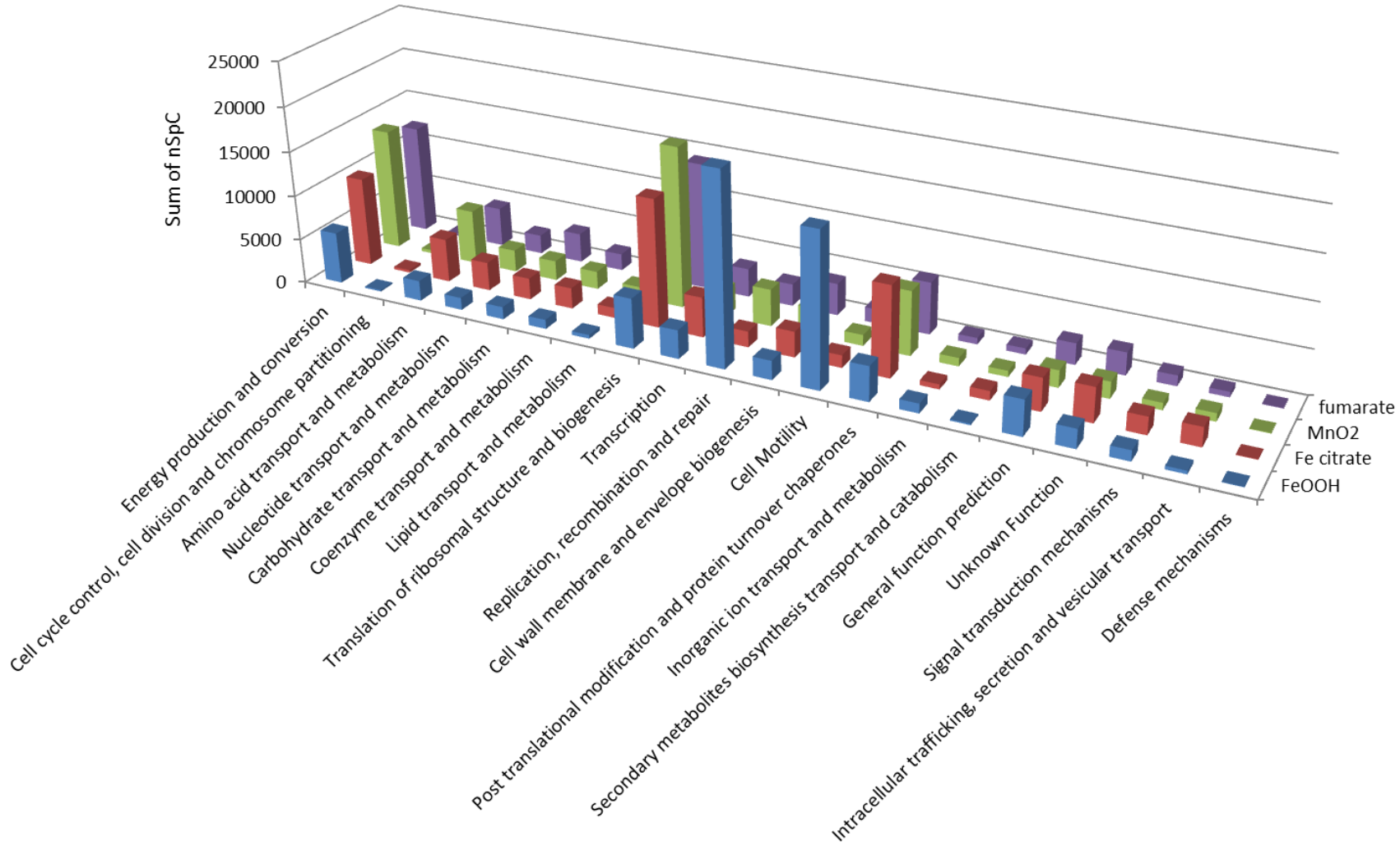
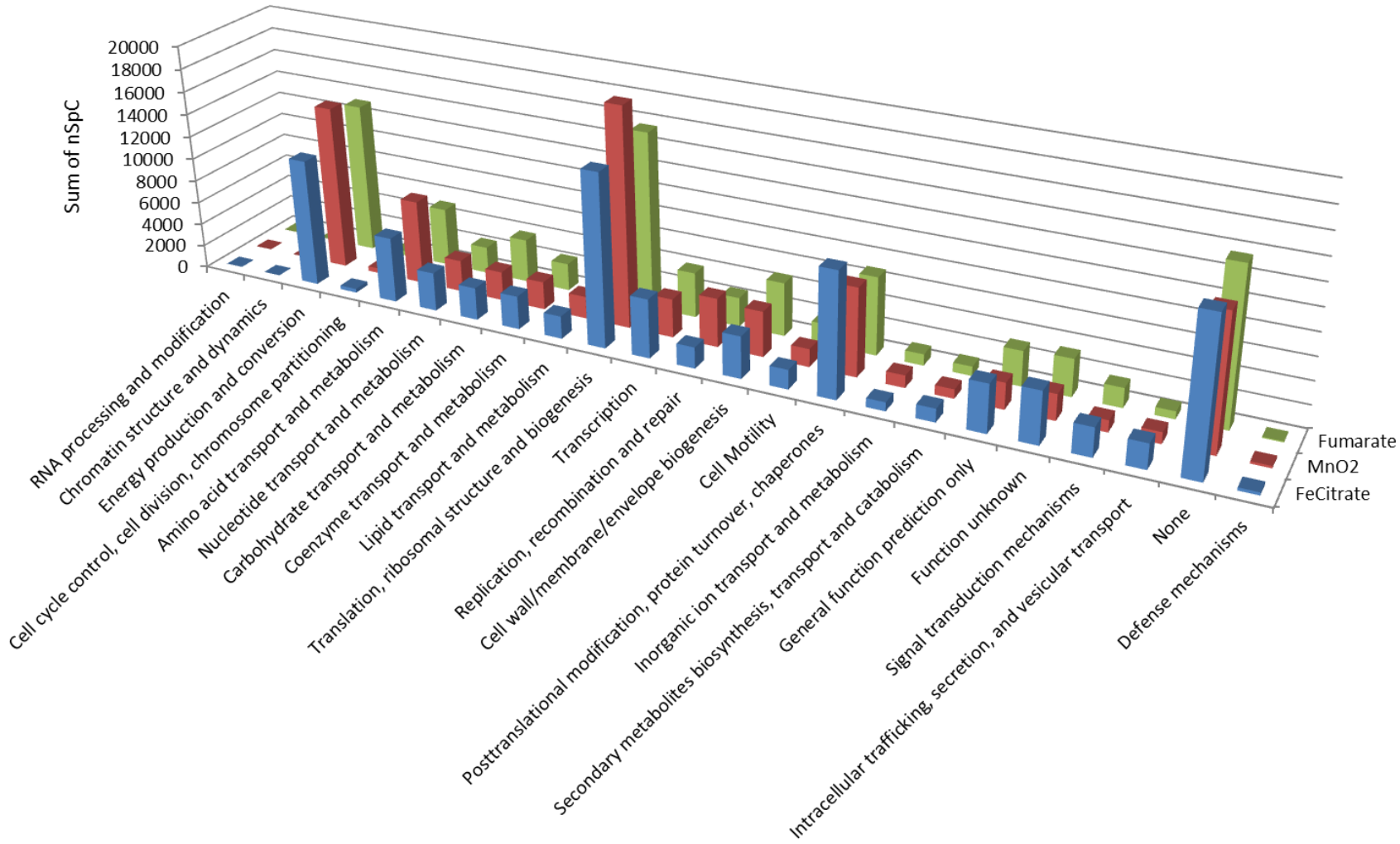

Sum of nSpC
30000
25000
20000
15000
10000
5000
0
Energy production and conversion
Cell cycle control, cell division, chromosome partitioning
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Translation, ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Cell wall/membrane/envelope biogenesis
Cell Motility
Posttranslational modification, protein turnover, chaperones
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport and catabolism
General function prediction only
Function unknown
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
None
oxygen
fumarate
tryptic soy broth
nitrate
MnO2
Fe citrate
FeOOH

Figure 5.14. Quantitative distribution of *G. daltonii* FRC-32 common proteome and core proteome by COG functional categories. (a) The common proteome contains identified proteins overlapping all tested growth conditions. (b) The core proteome represents detected proteins overlapping all conditions with FeOOH condition omitted.

(a)


Sum of nSpC
25000
20000
15000
10000
5000
0
Energy production and conversion
Cell cycle control, cell division and chromosome partitioning
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Translation of ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Cell wall membrane and envelope biogenesis
Cell Motility
Post translational modification and protein turnover chaperones
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis transport and catabolism
General function prediction
Unknown Function
Signal transduction mechanisms
Intracellular trafficking, secretion and vesicular transport
Defense mechanisms
fumarate
MnO2
Fe citrate
FeOOH

(b)

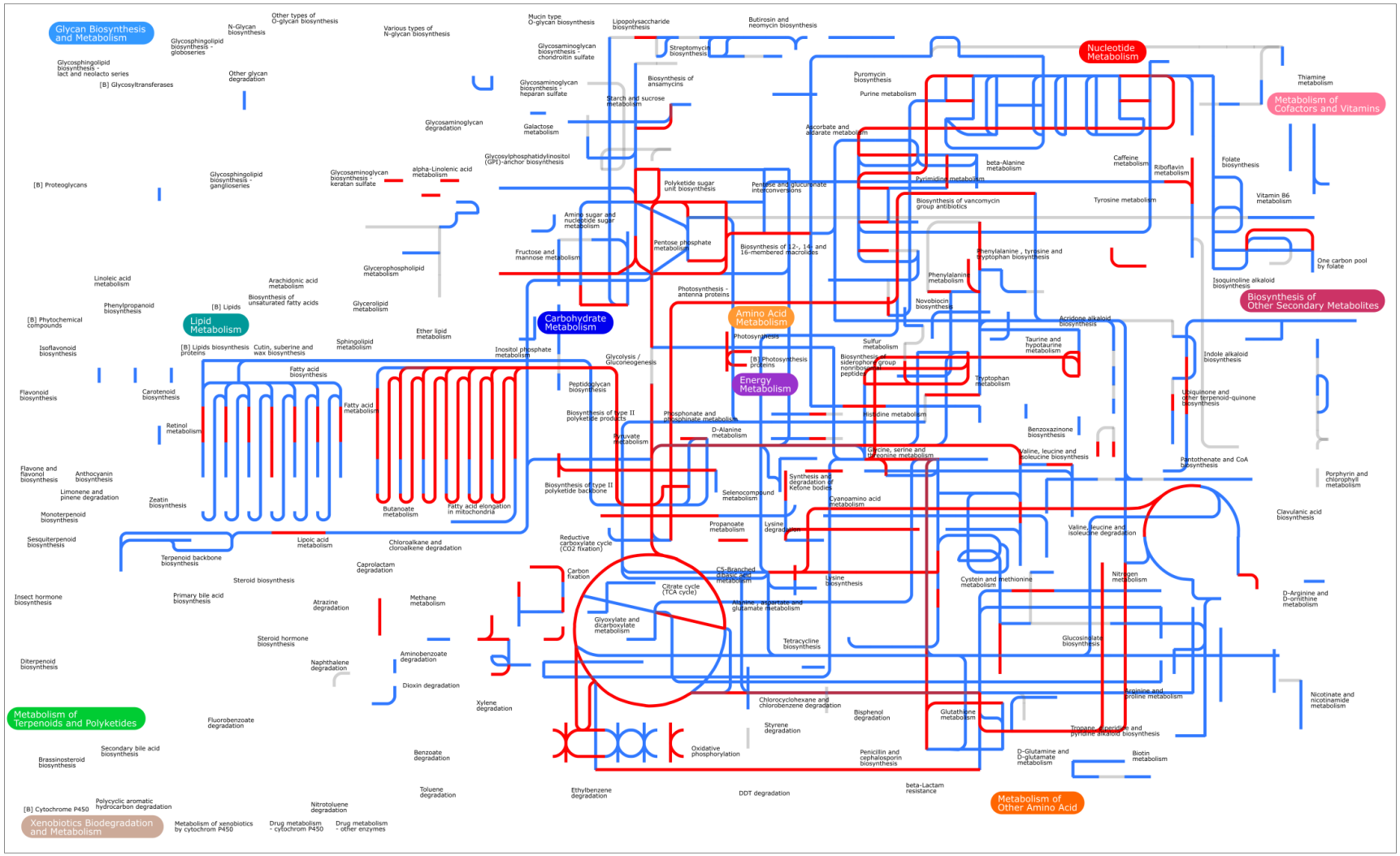

Sum of nSpC
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
RNA processing and modification
Chromatin structure and dynamics
Energy production and conversion
Cell cycle control, cell division, chromosome partitioning
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Translation, ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Cell wall/membrane/envelope biogenesis
Cell Motility
Posttranslational modification, protein turnover, chaperones
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport and catabolism
General function prediction only
Function unknown
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
None
Defense mechanisms
Fumarate
MnO2
FeCitrate

respectively. In general, the abundances of common and core proteins in each functional category are roughly consistent between different growth conditions, with proteins in the "translation, ribosomal structure and biogenesis" category having the highest abundances for both organisms (Figure 5.13 and 5.14). For MR-1 grown with FeOOH, proteins related to energy production and conversion is roughly twice as abundant as with other growth conditions (Figure 5.13a), which could be either a skewed result from the overall low protein identification, or that the cells are attempting to generate more energy with FeOOH growth. In contrast to strains MR-1 and 2CP-C (Chapter 4), the common proteome of strain FRC-32 grown with FeOOH had the most distinct quantitative COG distribution, in which proteins related to replication, recombination and repair were the most abundant, suggesting more active cell proliferation or repair with FeOOH growth for strain FRC-32 (Figure 5.14a). The second highest abundant protein functional category of FeOOH grown FRC-32 core proteome was "cell motility" (Figure 5.14a), which is in agreement with previous studies showing that *Geobacter* spp. could generate type IV pili structure during electron transfer to solid-phase iron electron acceptor [78, 174, 220]. The cell motility proteins did not show elevated abundance during growth with MnO_2 (Figure 5.14), another solid metal electron acceptor, suggesting that electron transfer through pili may be unique to solid-phase iron.

A more detailed and specific functional overview of the core proteomes was generated by mapping detected common and core proteins onto metabolic pathways in KEGG database using iPath2.0 (Figure 5.15 and 5.16). For strain MR-1 and FRC-32, functional annotations were available for 213 and 404 proteins in the common proteomes, respectively, and for 745 and 676

Figure 5.15. Metabolic pathways of detected *S. oneidensis* MR-1 common and core proteomes. (a) Proteins detected under all growth conditions contribute to the common proteome, and the representing metabolic pathways are highlighted in red. All detected non-common proteins are highlighted in blue. (b) When FeOOH condition is omitted, the detected core proteome mapped to more metabolic pathways than the common proteome did, and are highlighted in red. All detected non-core proteins are highlighted in blue.

(a)


Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid
N-Glycan biosynthesis
Other types of O-glycan biosynthesis
Various types of N-glycan biosynthesis
Mucin type O-glycan biosynthesis
Lipopolysaccharide biosynthesis
Butirosin and neomycin biosynthesis
Streptomycin biosynthesis
Glycosphingolipid biosynthesis - globoseries
Glycosphingolipid biosynthesis - lact and neolacto series
[B] Glycosyltransferases
Other glycan degradation
Glycosaminoglycan biosynthesis - chondroitin sulfate
Glycosaminoglycan biosynthesis - heparan sulfate
Biosynthesis of ansamycins
Starch and sucrose metabolism
Glycosaminoglycan degradation
Galactose metabolism
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
Puromycin biosynthesis
Purine metabolism
Ascorbate and aldarate metabolism
Thiamine metabolism
[B] Proteoglycans
Glycosphingolipid biosynthesis - ganglioseries
Glycosaminoglycan biosynthesis - keratan sulfate
alpha-Linolenic acid metabolism
Polyketide sugar unit biosynthesis
Pentose and glucuronate interconversions
beta-Alanine metabolism
Caffeine metabolism
Riboflavin metabolism
Folate biosynthesis
Pyrimidine metabolism
Vitamin B6 metabolism
Amino sugar and nucleotide sugar metabolism
Biosynthesis of vancomycin group antibiotics
Tyrosine metabolism
Pentose phosphate metabolism
Fructose and mannose metabolism
Biosynthesis of 12-, 14- and 16-membered macrolides
Phenylalanine , tyrosine and tryptophan biosynthesis
One carbon pool by folate
Glycerophospholipid metabolism
Linoleic acid metabolism
Arachidonic acid metabolism
Phenylalanine metabolism
Isoquinoline alkaloid biosynthesis
Biosynthesis of unsaturated fatty acids
Glycerolipid metabolism
Photosynthesis - antenna proteins
Novobiocin biosynthesis
Phenylpropanoid biosynthesis
[B] Lipids
[B] Phytochemical compounds
Acridone alkaloid biosynthesis
Ether lipid metabolism
Sphingolipid metabolism
Photosynthesis
Sulfur metabolism
Taurine and hypotaurine metabolism
Isoflavonoid biosynthesis
[B] Lipids biosynthesis proteins
Cutin, suberine and wax biosynthesis
Inositol phosphate metabolism
Glycolysis / Gluconeogenesis
[B] Photosynthesis proteins
Biosynthesis of siderophore group nonribosomal peptides
Indole alkaloid biosynthesis
Fatty acid biosynthesis
Peptidoglycan biosynthesis
Tryptophan metabolism
Flavonoid biosynthesis
Carotenoid biosynthesis
Fatty acid metabolism
Biosynthesis of type II polyketide products
Phosphonate and phosphinate metabolism
Histidine metabolism
Ubiquinone and other terpenoid-quinone biosynthesis
Retinol metabolism
D-Alanine metabolism
Pyruvate metabolism
Benzoxazinone biosynthesis
Glycine, serine and threonine metabolism
Valine, leucine and isoleucine biosynthesis
Flavone and flavonol biosynthesis
Anthocyanin biosynthesis
Pantothenate and CoA biosynthesis
Porphyrin and chlorophyll metabolism
Limonene and pinene degradation
Biosynthesis of type II polyketide backbone
Synthesis and degradation of ketone bodies
Selenocompound metabolism
Cyanoamino acid metabolism
Zeatin biosynthesis
Butanoate metabolism
Fatty acid elongation in mitochondria
Clavulanic acid biosynthesis
Monoterpenoid biosynthesis
Propanoate metabolism
Lysine degradation
Valine, leucine and isoleucine degradation
Sesquiterpenoid biosynthesis
Lipoic acid metabolism
Chloroalkane and chloroalkene degradation
Reductive carboxylate cycle (CO2 fixation)
Terpenoid backbone biosynthesis
Caprolactam degradation
Carbon fixation
C5-Branched dibasic acid metabolism
Lysine biosynthesis
Cysteine and methionine metabolism
Nitrogen metabolism
Steroid biosynthesis
Citrate cycle (TCA cycle)
Insect hormone biosynthesis
Primary bile acid biosynthesis
Atrazine degradation
Methane metabolism
Alanine , aspartate and glutamate metabolism
D-Arginine and D-ornithine metabolism
Glyoxylate and dicarboxylate metabolism
Steroid hormone biosynthesis
Tetracycline biosynthesis
Glucosinolate biosynthesis
Diterpenoid biosynthesis
Aminobenzoate degradation
Naphthalene degradation
Dioxin degradation
Arginine and proline metabolism
Nicotinate and nicotinamide metabolism
Xylene degradation
Chlorocyclohexane and chlorobenzene degradation
Bisphenol degradation
Glutathione metabolism
Fluorobenzoate degradation
Styrene degradation
Tropane, piperidine and pyridine alkaloid biosynthesis
Secondary bile acid biosynthesis
Oxidative phosphorylation
Penicillin and cephalosporin biosynthesis
D-Glutamine and D-glutamate metabolism
Biotin metabolism
Brassinosteroid biosynthesis
Benzoate degradation
Toluene degradation
Ethylbenzene degradation
DDT degradation
beta-Lactam resistance
Polycyclic aromatic hydrocarbon degradation
[B] Cytochrome P450
Nitrotoluene degradation
Metabolism of xenobiotics by cytochrom P450
Drug metabolism - cytochrom P450
Drug metabolism - other enzymes

(b)

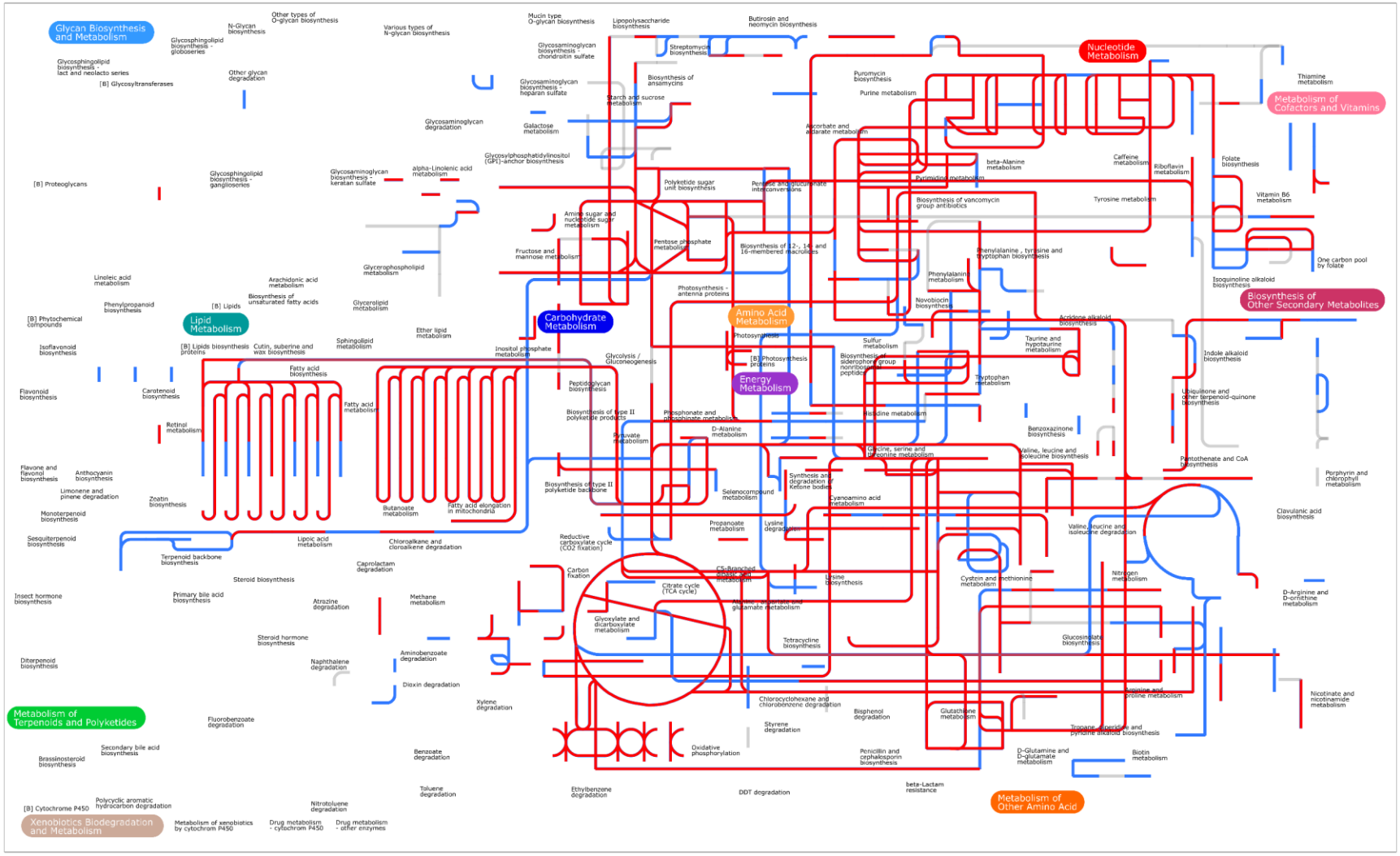
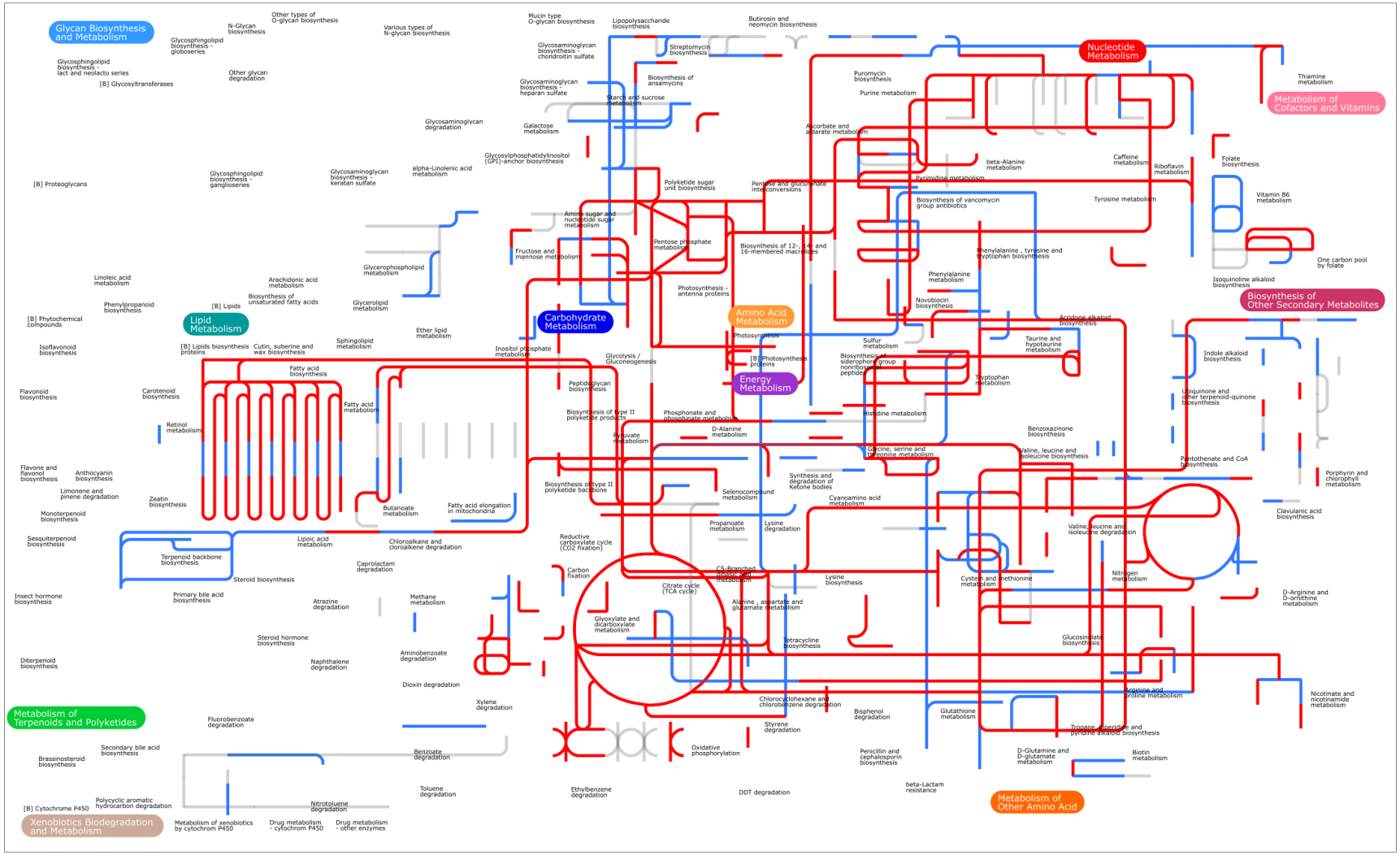
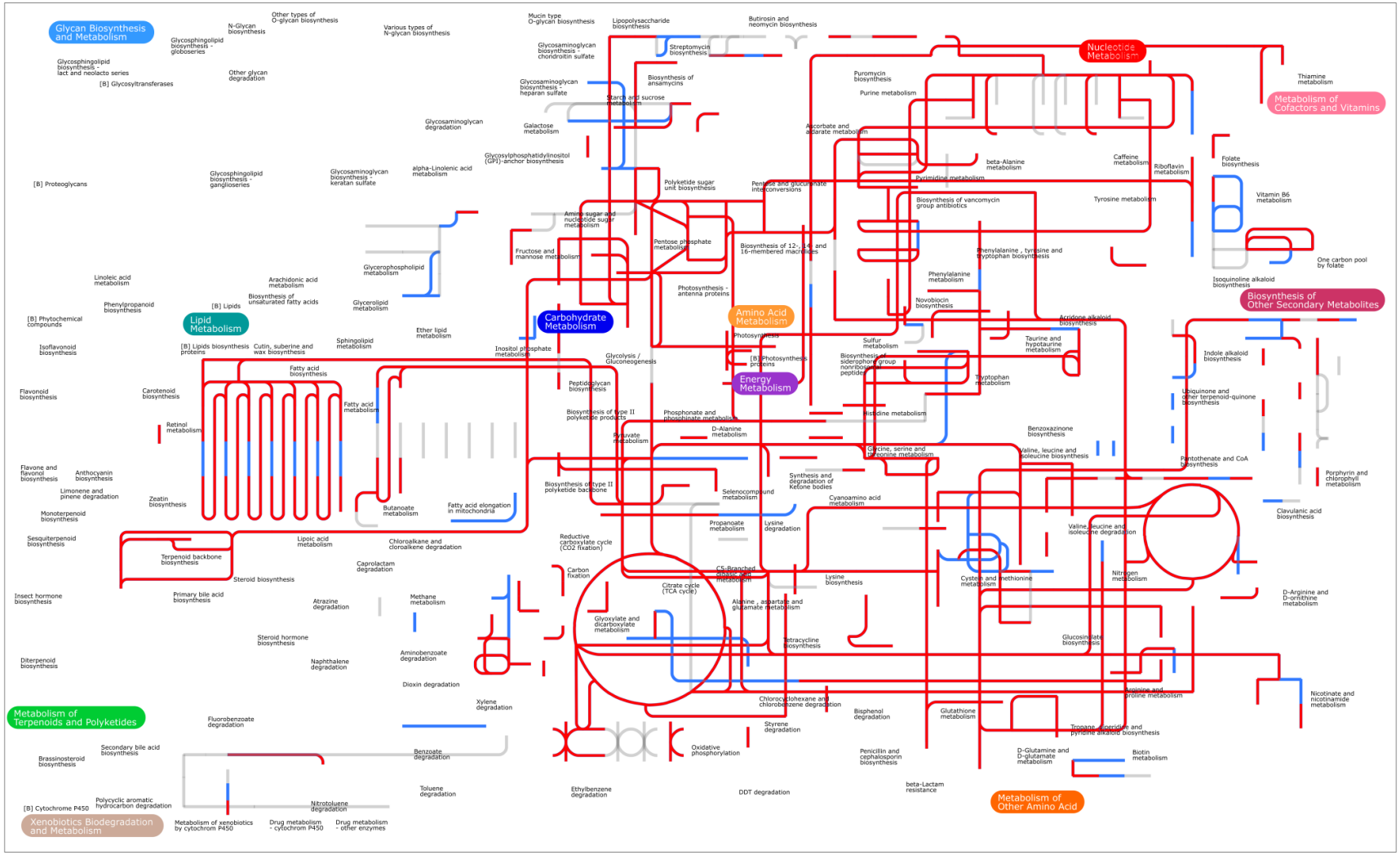

Glycan Biosynthesis and Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid
Glycosphingolipid biosynthesis - globoseries
Glycosphingolipid biosynthesis - lact and neolacto series
[B] Glycosyltransferases
N-Glycan biosynthesis
Other types of O-glycan biosynthesis
Various types of N-glycan biosynthesis
Other glycan degradation
Mucin type O-glycan biosynthesis
Lipopolysaccharide biosynthesis
Glycosaminoglycan biosynthesis - chondroitin sulfate
Glycosaminoglycan biosynthesis - heparan sulfate
Glycosaminoglycan degradation
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
[B] Proteoglycans
Glycosphingolipid biosynthesis - ganglioseries
Glycosaminoglycan biosynthesis - keratan sulfate
alpha-Linolenic acid metabolism
Linoleic acid metabolism
Arachidonic acid metabolism
Biosynthesis of unsaturated fatty acids
Phenylpropanoid biosynthesis
[B] Lipids
[B] Phytochemical compounds
Isoflavonoid biosynthesis
[B] Lipids biosynthesis proteins
Cutin, suberine and wax biosynthesis
Glycerophospholipid metabolism
Glycerolipid metabolism
Sphingolipid metabolism
Ether lipid metabolism
Fatty acid biosynthesis
Fatty acid metabolism
Flavonoid biosynthesis
Carotenoid biosynthesis
Retinol metabolism
Flavone and flavonol biosynthesis
Anthocyanin biosynthesis
Limonene and pinene degradation
Zeatin biosynthesis
Monoterpenoid biosynthesis
Sesquiterpenoid biosynthesis
Terpenoid backbone biosynthesis
Steroid biosynthesis
Insect hormone biosynthesis
Primary bile acid biosynthesis
Steroid hormone biosynthesis
Diterpenoid biosynthesis
Secondary bile acid biosynthesis
Brassinosteroid biosynthesis
Inositol phosphate metabolism
Glycolysis / Gluconeogenesis
Peptidoglycan biosynthesis
Biosynthesis of type II polyketide products
Pyruvate metabolism
Biosynthesis of type II polyketide backbone
Butanoate metabolism
Fatty acid elongation in mitochondria
Propanoate metabolism
Reductive carboxylate cycle (CO2 fixation)
Carbon fixation
Citrate cycle (TCA cycle)
Glyoxylate and dicarboxylate metabolism
Methane metabolism
Caprolactam degradation
Lipoic acid metabolism
Chloroalkane and chloroalkene degradation
Atrazine degradation
Aminobenzoate degradation
Naphthalene degradation
Dioxin degradation
Xylene degradation
Fluorobenzoate degradation
Benzoate degradation
Toluene degradation
Ethylbenzene degradation
Nitrotoluene degradation
Polycyclic aromatic hydrocarbon degradation
[B] Cytochrome P450
Metabolism of xenobiotics by cytochrom P450
Drug metabolism - cytochrom P450
Drug metabolism - other enzymes
Biosynthesis of ansamycins
Starch and sucrose metabolism
Galactose metabolism
Amino sugar and nucleotide sugar metabolism
Fructose and mannose metabolism
Pentose phosphate metabolism
Photosynthesis - antenna proteins
Photosynthesis
[B] Photosynthesis proteins
Oxidative phosphorylation
Selenocompound metabolism
Alanine, aspartate and glutamate metabolism
Cysteine and methionine metabolism
Lysine biosynthesis
Lysine degradation
Tetracycline biosynthesis
Chlorocyclohexane and chlorobenzene degradation
Styrene degradation
Bisphenol degradation
DDT degradation
Synthesis and degradation of Ketone bodies
Cyanoamino acid metabolism
C5-Branched dibasic acid metabolism
Valine, leucine and isoleucine biosynthesis
Glycine, serine and threonine metabolism
Histidine metabolism
D-Alanine metabolism
Phosphonate and phosphinate metabolism
Sulfur metabolism
Taurine and hypotaurine metabolism
Benzoxazinone biosynthesis
Novobiocin biosynthesis
Phenylalanine metabolism
Phenylalanine, tyrosine and tryptophan biosynthesis
Tryptophan metabolism
Biosynthesis of siderophore group nonribosomal peptides
Biosynthesis of 12-, 14- and 16-membered macrolides
Biosynthesis of vancomycin group antibiotics
Pentose and glucuronate interconversions
Polyketide sugar unit biosynthesis
Purine metabolism
Pyrimidine metabolism
Puromycin biosynthesis
Ascorbate and aldarate metabolism
beta-Alanine metabolism
Caffeine metabolism
Riboflavin metabolism
Folate biosynthesis
Tyrosine metabolism
Vitamin B6 metabolism
One carbon pool by folate
Thiamine metabolism
Isoquinoline alkaloid biosynthesis
Acridone alkaloid biosynthesis
Indole alkaloid biosynthesis
Ubiquinone and other terpenoid-quinone biosynthesis
Pantothenate and CoA biosynthesis
Porphyrin and chlorophyll metabolism
Clavulanic acid biosynthesis
Nitrogen metabolism
D-Arginine and D-ornithine metabolism
Arginine and proline metabolism
Glucosinolate biosynthesis
Nicotinate and nicotinamide metabolism
Tropane, piperidine and pyridine alkaloid biosynthesis
Glutathione metabolism
Penicillin and cephalosporin biosynthesis
D-Glutamine and D-glutamate metabolism
Biotin metabolism
beta-Lactam resistance
Butirosin and neomycin biosynthesis
Streptomycin biosynthesis

Figure 5.16. Metabolic pathways of detected *G. daltonii* FRC-32 common and core proteomes. (a) Proteins detected under all growth conditions contribute to the common proteome, and the representing metabolic pathways are highlighted in red. All detected non-common proteins are highlighted in blue. (b) When FeOOH condition is omitted, the detected core proteome mapped to more metabolic pathways than the common proteome did, and are highlighted in red. All detected non-core proteins are highlighted in blue.

(a)


Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid
Glycosphingolipid biosynthesis - globoseries
Glycosphingolipid biosynthesis - lact and neolacto series
[B] Glycosyltransferases
N-Glycan biosynthesis
Other types of O-glycan biosynthesis
Various types of N-glycan biosynthesis
Other glycan degradation
Mucin type O-glycan biosynthesis
Lipopolysaccharide biosynthesis
Glycosaminoglycan biosynthesis - chondroitin sulfate
Streptomycin biosynthesis
Butirosin and neomycin biosynthesis
Glycosaminoglycan biosynthesis - heparan sulfate
Biosynthesis of ansamycins
Starch and sucrose metabolism
Glycosaminoglycan degradation
Galactose metabolism
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
[B] Proteoglycans
Glycosphingolipid biosynthesis - ganglioseries
Glycosaminoglycan biosynthesis - keratan sulfate
alpha-Linolenic acid metabolism
Polyketide sugar unit biosynthesis
Amino sugar and nucleotide sugar metabolism
Fructose and mannose metabolism
Pentose phosphate metabolism
Glycerophospholipid metabolism
Linoleic acid metabolism
Arachidonic acid metabolism
Biosynthesis of unsaturated fatty acids
Glycerolipid metabolism
Phenylpropanoid biosynthesis
[B] Lipids
[B] Phytochemical compounds
Isoflavonoid biosynthesis
[B] Lipids biosynthesis proteins
Cutin, suberine and wax biosynthesis
Sphingolipid metabolism
Ether lipid metabolism
Inositol phosphate metabolism
Fatty acid biosynthesis
Fatty acid metabolism
Flavonoid biosynthesis
Carotenoid biosynthesis
Retinol metabolism
Flavone and flavonol biosynthesis
Anthocyanin biosynthesis
Limonene and pinene degradation
Zeatin biosynthesis
Monoterpenoid biosynthesis
Sesquiterpenoid biosynthesis
Terpenoid backbone biosynthesis
Steroid biosynthesis
Insect hormone biosynthesis
Primary bile acid biosynthesis
Steroid hormone biosynthesis
Diterpenoid biosynthesis
Secondary bile acid biosynthesis
Brassinosteroid biosynthesis
[B] Cytochrome P450
Polycyclic aromatic hydrocarbon degradation
Fluorobenzoate degradation
Benzoate degradation
Toluene degradation
Nitrotoluene degradation
Ethylbenzene degradation
Xylene degradation
Dioxin degradation
Aminobenzoate degradation
Naphthalene degradation
Atrazine degradation
Methane metabolism
Caprolactam degradation
Chloroalkane and chloroalkene degradation
Lipoic acid metabolism
Butanoate metabolism
Fatty acid elongation in mitochondria
Glycolysis / Gluconeogenesis
Peptidoglycan biosynthesis
Biosynthesis of type II polyketide products
Biosynthesis of type II polyketide backbone
Pyruvate metabolism
Reductive carboxylate cycle (CO2 fixation)
Carbon fixation
Citrate cycle (TCA cycle)
Glyoxylate and dicarboxylate metabolism
Propanoate metabolism
Lysine degradation
Synthesis and degradation of Ketone bodies
Selenocompound metabolism
Cyanoamino acid metabolism
Alanine , aspartate and glutamate metabolism
Lysine biosynthesis
Tetracycline biosynthesis
Oxidative phosphorylation
Chlorocyclohexane and chlorobenzene degradation
Styrene degradation
Bisphenol degradation
Penicillin and cephalosporin biosynthesis
beta-Lactam resistance
DDT degradation
Glutathione metabolism
D-Glutamine and D-glutamate metabolism
Cysteine and methionine metabolism
Valine, leucine and isoleucine degradation
Valine, leucine and isoleucine biosynthesis
Glycine, serine and threonine metabolism
D-Alanine metabolism
Phosphonate and phosphinate metabolism
Photosynthesis
[B] Photosynthesis proteins
Photosynthesis - antenna proteins
Sulfur metabolism
Biosynthesis of siderophore group nonribosomal peptides
Histidine metabolism
Tryptophan metabolism
Phenylalanine metabolism
Novobiocin biosynthesis
Biosynthesis of vancomycin group antibiotics
Pentose and glucuronate interconversions
Biosynthesis of 12-, 14- and 16-membered macrolides
Phenylalanine , tyrosine and tryptophan biosynthesis
Tyrosine metabolism
Taurine and hypotaurine metabolism
Acridone alkaloid biosynthesis
Benzoxazinone biosynthesis
Ubiquinone and other terpenoid-quinone biosynthesis
Pantothenate and CoA biosynthesis
Porphyrin and chlorophyll metabolism
Clavulanic acid biosynthesis
Nitrogen metabolism
D-Arginine and D-ornithine metabolism
Arginine and proline metabolism
Nicotinate and nicotinamide metabolism
Tropane, piperidine and pyridine alkaloid biosynthesis
Biotin metabolism
Glucosinolate biosynthesis
Indole alkaloid biosynthesis
Isoquinoline alkaloid biosynthesis
Vitamin B6 metabolism
One carbon pool by folate
Folate biosynthesis
Riboflavin metabolism
Caffeine metabolism
Thiamine metabolism
Pyrimidine metabolism
beta-Alanine metabolism
Purine metabolism
Puromycin biosynthesis
Ascorbate and aldarate metabolism
Metabolism of xenobiotics by cytochrom P450
Drug metabolism - cytochrom P450
Drug metabolism - other enzymes

(b)


Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid

proteins in the core proteomes, respectively. For strain MR-1, the common proteome represented pathways spread across all the functional modules, such as energy metabolism, lipid metabolism, etc. (Figure 5.15a). The majority of the pan-proteome covered pathways were not represented by the common proteome of strain MR-1 (Figure 5.15a). However, when FeOOH was omitted, the detected core proteome mapped onto almost all metabolic pathways and represented high coverage to the pan-proteome (Figure 5.15b). For strain FRC-32, the common proteome had much more overlap to almost all metabolic pathways represented by the pan-proteome (Figure 5.16a), which is similar to the common pathways of *A. dehalogenans* strain 2CP-C (shown in Chapter 4). The differences between the core and common proteomes of strain FRC-32 is not as big as that of strain MR-1. For strain FRC-32, both the core and common proteomes represented high coverage to the pan-proteome (Figure 5.16 a and b). Taken together, the core proteomes of strain MR-1 and FRC-32 conserve most of the metabolic pathways, which suggested that different electron acceptors did not impact the type of metabolic activities carried out in the living cell. In contrast, the common proteome of strain MR-1 represent much fewer metabolic pathways, which may due to the slow growth rate in FeOOH condition.

5.3.3 The Unique Proteins of *S. oneidensis* MR-1 and *G. daltonii* FRC-32

For both strains, many proteins were detected only under a specific growth condition. Such uniquely detected proteins could be functionally related to specialized biological activities corresponding to the utilization of certain electron acceptor. For strain MR-1, the number of

identified unique proteins was highest with MnO_2-growth, and lowest with FeOOH-growth (95 and 2, respectively). The other growth conditions (Fe citrate, nitrate, tryptic soy broth, fumarate, and oxygen) each resulted in the identification of approximately 60 unique proteins. For strain FRC-32, the unique proteins detected for growth with FeOOH, Fe citrate, MnO_2, and fumarate are 68, 94, 115, and 192, respectively. Overall, each tested electron acceptor stimulated the expression of unique set of proteins for each organism, and strain FRC-32 generated highest number of uniquely detected proteins under each condition, comparing to strain MR-1 and 2CP-C (shown in Chapter 4).

To functionally and quantitatively analyze the unique proteins, the abundance of unique proteins (represented by cumulative nSpC) detected in each growth condition was plotted against COG functional classifications (Figure 5.17 and 5.18 for strain MR-1 and FRC-32, respectively). For strain MR-1, the most abundant unique proteins detected with MnO_2-growth belonged to "inorganic ion transport and metabolism" (Figure 5.17), which suggest that MnO_2 reduction possibly requires a unique set of proteins participating in the electron transfer process. When grown with FeOOH, the most abundant unique proteins belonged to "lipid transport and metabolism" category for strain MR-1 (Figure 5.17), and belonged to "cell motility" category for strain FRC-32 (Figure 5.18). There were 26 unique proteins of strain FRC-32 in the "cell motility" category, among which, Geob_0503 was the most abundant (nSpC: 1555). The protein corresponding to Geob_0502 is annotated as "hypothetical protein" in the NCBI database, but the protein sequence matched to the "Flg_bb_rod" protein family in the Pfam database which corresponds to flagella basal body rod protein. Although Geob_0502 is currently uncharacterized, our results specifically linked this protein to growth with FeOOH,

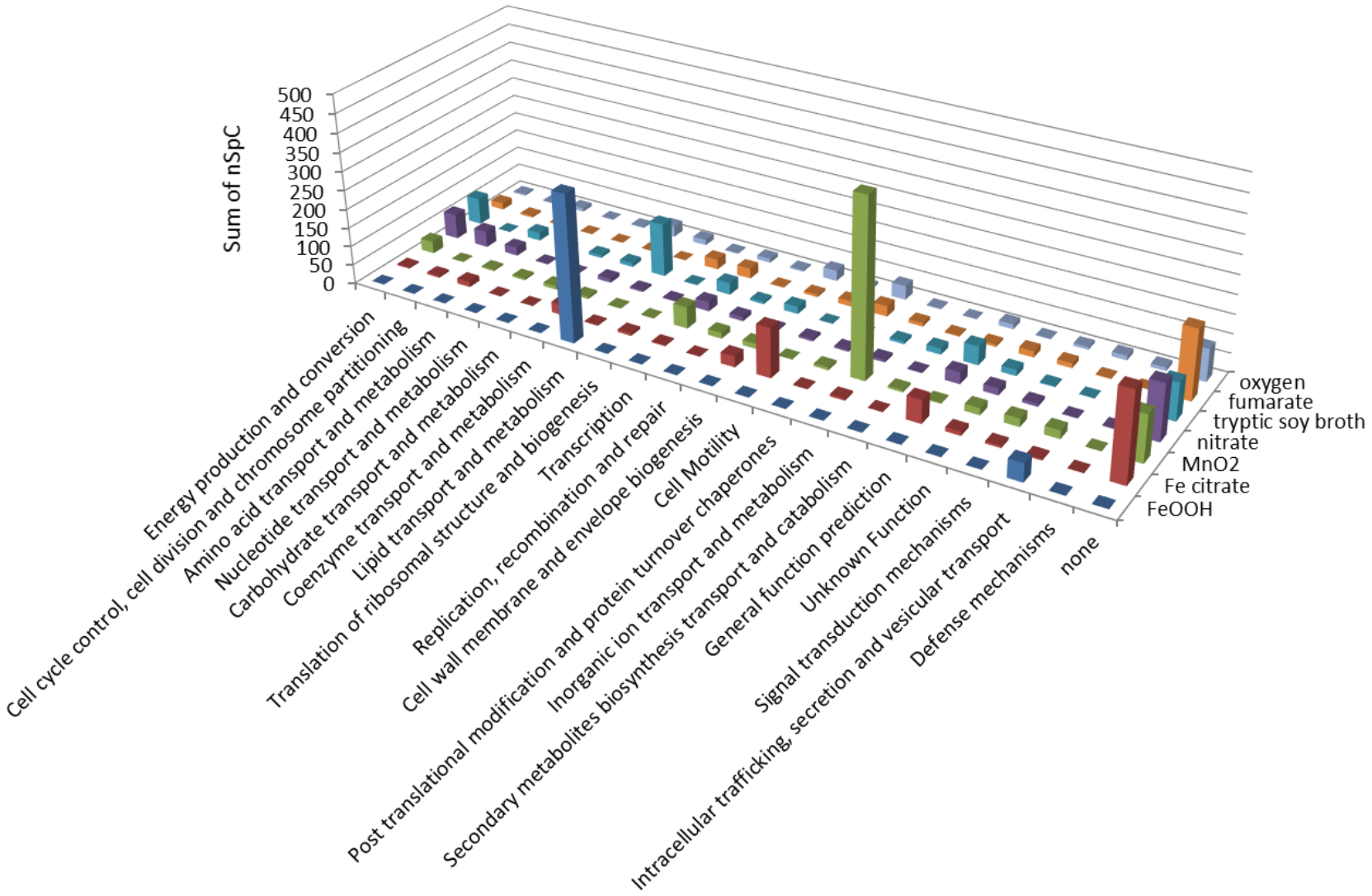


Figure 5.17. Quantitative distribution of detected *S. oneidensis* MR-1 proteins unique to certain growth conditions.

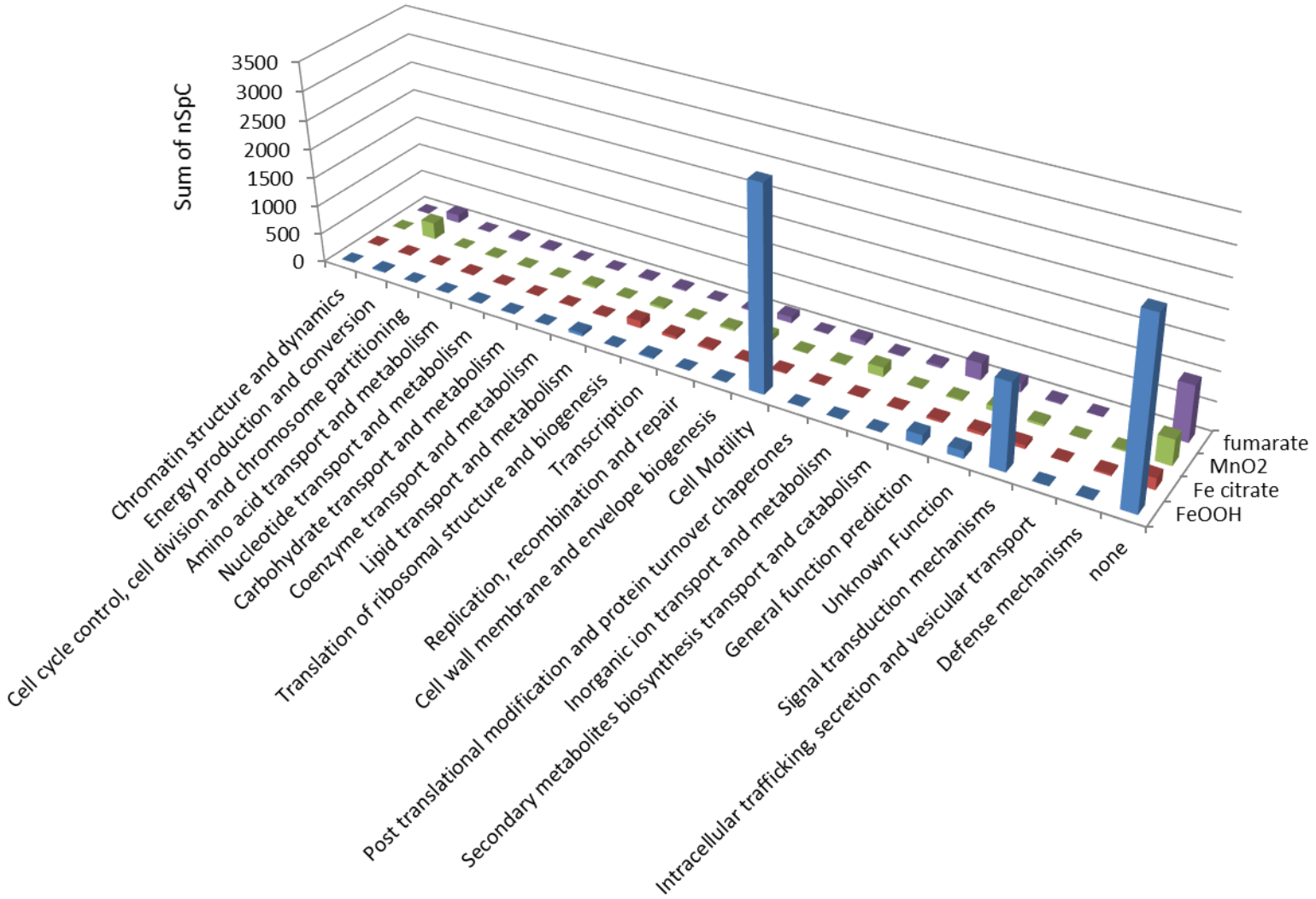


Figure 5.18. Quantitative distribution of detected *G. daltonii* FRC-32 proteins unique to certain growth conditions.

together with the Pfam information, suggesting that this protein plays important role in the conductive pili structure (mentioned above) during electron transfer to FeOOH. For both strain MR-1 and FRC-32, the abundances of functionally unclassified ("none" category) unique proteins are high in most growth condition (e.g. Fe citrate, fumarate for strain MR-1, and fumarate for strain FRC-32). The functional activities of these uncharacterized unique proteins could be important for the respiration of corresponding electron acceptors, and their unique detections are calling for further biochemical characterizations to reveal their biological functions.

5.4 Conclusions

In conclusion, mass spectrometry-based proteomics provided deep and comprehensive characterizations to the proteomes of *S. oneidensis* strain MR-1 and *G. daltonii* strain FRC-32 responding to a wide range of electron acceptors. The detected pan-proteome for each organism reflected almost all explored metabolic pathways. Comparing to *A. dehalogenans* strain 2CP-C, the common proteome of strain FRC-32 had comparable size and metabolic pathway representations, whereas for strain MR-1, poor growth with FeOOH as electron acceptor greatly distorted the direct comparison. When FeOOH condition is omitted, the core proteome for each strain contains around 1,000 proteins, covering almost all metabolic pathways represented by their corresponding pan-proteomes. Unique proteins were detected for each tested organisms, and their expression and possible functionalities were linked to specific growth conditions through proteomics measurements. For strain FRC-32 grown with FeOOH, both the core proteome and uniquely detected proteins reflected highly abundant

proteins relating to cell motility, which supports previous observations about conductive pili structure used for electron transfer to solid-phase iron. Taken together, our results provided systematic overview of the metabolic activities of different DMRB organisms, and revealed their proteomic reactions corresponding to the utilization of different electron acceptors.

Chapter 6

Concluding Remarks and Perspectives

6.1 Overview

The work presented in this dissertation demonstrated the use of a mass spectrometry-based proteomics approach to deepen our understanding of environmental microbiology. Dissimilatory metal reducing bacteria (DMRB) are a group of environmentally and ecologically important bacteria, whose functional activities greatly impact the planet earth. In the past, research efforts on DMRB has led to the complete genome sequencing of many individual strains, providing a great foundation for proteomics research to directly uncover the protein machineries. Relying on the available genome information, this dissertation achieved detailed proteomic characterizations for three dissimilatory metal reducing bacterial strains, *Anaeromyxobacter dehalogenans* strain 2CP-C, *Shewanella oneidensis* strain MR-1, and *Geobacter daltonii* strain FRC-32. Through high-performance proteomics measurement, this dissertation revealed the molecular-level phenotypes of DMRBs represented by protein signatures corresponding to the utilization of various electron acceptors for bacterial respiration. The metabolic activates of individual DMRB strains tested, in addition to the proteomic comparisons between different strains, provided insights into the overall understanding of DMRB molecular functions.

6.2 Conclusions for *c*-Type Cytochrome Study

When the research in this dissertation started about five years ago, our understanding about dissimilatory metal reducing bacteria was mainly limited to a broad and somewhat vague knowledge of their ability to use metal electron acceptors for energy preservation and survival. The efficiency of respiring different forms of metal compounds by various isolated DMRB strains, their preferences for different environmental electron acceptors, and their growth rates under different electron-accepting conditions were investigated previously, with the interest to characterize the microbial physiology and metal reducing properties. Although the available genomes of many DMRB strains have revealed their genetic makeup and functional potentials, there was still a knowledge gap for the metabolic activities and molecular mechanisms underlying the microbial activities of DMRBs. As mentioned in Chapter 1, for a few DMRB strains, such as *G. sulfurreducens* strain PCA and *S. oneidensis* strain MR-1, several specific genes encoding for *c*-type cytochromes had been extensively studied and had been linked to the respiration of certain electron acceptors through mutagenesis studies. However, the detailed molecular network that contributes to DMRB respiration was still unknown.

In Chapter 3, mass spectrometry-based proteomics profiled the expression of proteins responsible for the versatile respirational activities, revealing the members of *c*-type cytochrome proteins participating in the cellular metabolism under different growth conditions. This is the first time the entire repertoires of *c*-type cytochrome proteins for three different DMRB strains were monitored in detail. Instead of narrowing the research focus on a few specific proteins, the proteomics profiling broadened our view on the expression of all putative

c-type cytochromes in a global perspective. Previously, the majority of the putative *c*-type cytochromes were predicted based on genome information, but their actual expressions were not validated. Through mass spectrometry measurements, not only evidence for the actual expression of putative *c*-type cytochrome proteins was provided, but also a wide range of electron acceptors were linked to the presences and abundances of different sets of *c*-type cytochromes. The mass spectrometry detection, together with the localization prediction for putative *c*-type cytochromes, allowed us to propose a model for the electron transfer network of DMRBs. As demonstrated in Chapter 3, the spatial connections between detected *c*-type cytochromes under each individual growth conditions could be constructed, and the model for fumarate and nitrate grown *A. dehalogenan* strain 2CP-C were shown as examples. With the same approach, the *c*-type cytochrome networks for *S. oneidensis* strain MR-1, and *G. daltonii* strain FRC-32 could also be demonstrated in the same fashion.

Having large numbers of *c*-type cytochrome encoding genes is a common characteristic for all DMRBs, but it was not clear whether different *c*-type cytochrome in different strains have same or similar functional roles for microbial respiration. This dissertation compared the expression of different *c*-type cytochromes from different strains having protein sequence similarities. Although the majority of the *c*-type cytochromes in the three tested DMRB strains did not show sequence homology, extremely high sequence correlations between a few *c*-type cytochromes were discovered as shown in Chapter 3, calling for deeper investigations on the functions of conserved *c*-type cytochrome sequences. The expression patterns of homologous *c*-type cytochromes from different strains did not demonstrate similar response to different electron acceptors, which suggested different functional roles playing in different strains.

Although microbial organisms are microscopic, each organism is a fully developed biological system, and the *c*-type cytochrome profiling and comparisons between different DMRB strains demonstrated highly complex and controlled molecular networks for their respiratory activities. This dissertation demonstrated how to move the field of microbial *c*-type cytochrome research from the study of individual genes or proteins toward the global analysis of the all *c*-type cytochromes. This is by no means to say the studies on specific *c*-type cytochromes are unimportant; rather, the proteomics study serves as the groundwork for experimental design of future research to reveal the specific functions and interactions of individual *c*-type cytochromes. The complementary information on defined protein subsets and global proteome should enhance our knowledge about the molecular mechanisms for DMRB respiration.

6.3 Perspectives on *c*-Type Cytochrome Research

The expression of many predicted *c*-type cytochromes of three DMRB strains were characterized in this dissertation, but their presences and detected abundances under a specific electron-accepting condition are not sufficient to reveal their actual functions. Whether the detected *c*-type cytochrome participate in microbial electron transfer, and in what way they contribute to microbial respiration, are questions need to be answered through further investigations beyond the scope of this dissertation.

As mentioned, the few functionally or structurally characterized *c*-type cytochromes were only for a few extensively studied DMRB strains. Clearly, different bacterial genera, and

species are significantly different in terms of the composition of *c*-type cytochrome proteins, and the expression profiles corresponding to different growth conditions (as demonstrated in this dissertation). Knowing the mechanism of electron transfer process for a particular DMRB strain is not sufficient to infer the molecular mechanisms used in another strain. Due to the vast range of DMRB members, and their different specialties for respiring different electron acceptors, the *c*-type cytochromes of more bacterial species and strains in the DMRB family need to be studied with greater detail in order to understand their unique molecular basis for their respiratory activities. And more comparative studies between different *c*-type cytochromes from different DMRBs should provide deeper insight into the functionalities, utilizations, and evolution of *c*-type cytochrome proteins.

Another important discovery revealed by *c*-type cytochrome protein sequences is the presence of multiple heme-binding motifs in a single cytochrome protein. As shown in this dissertation, the highest number of heme-binding motifs per *c*-type cytochrome for *A. dehalogenan* strain 2CP-C is 40. Such *c*-type cytochromes with multiple heme-binding motifs could have large number of attached heme groups. But the actual occupancy of heme-binding motifs in these multi-heme proteins has not been validated to date. Although the proteomic measurements performed in this dissertation detected peptides belonging to *c*-type cytochromes with multiple heme-binding motifs, the detected peptides used for protein inference did not contain motifs. This was due to the proteome database reliance of the current peptide sequencing techniques. The peptide spectrum matching approach could only detect peptide sequences present in the *in silico* digested protein database. Peptides having the heme attachment would cause a mass shift to the parent peptide ion, and the fragmentation pattern

will also be altered because of the two covalent bonds formed between the heme-group and the two cysteine residues on the peptide. The current peptide sequence identification algorithms are able to search the protein database for known post-translational modifications on single amino acid residues, such as phosphorylation, acetylation, etc. Since heme-group attachment modifies two non-adjacent residues, the current informatics algorithms are unable to generate predicted fragmentation patterns for such peptides for spectrum matching. The fact that peptides with unoccupied heme-binding motif were not detected suggests the possibility that all motifs could be attached with hemes, and could be captured by mass spectrometer. There is a great need for improved search algorithms to identify heme-attached peptides in the MS measurement.

6.4 Conclusions on DMRB Proteome Study

Previous investigations on gene functional expression for microbial respiration were mostly focused on the transcript level. As mentioned in Chapter 1, gene activities of several DMRB strains represented by transcriptional products (mRNA) were monitored using microarray or RNA-seq techniques. Several proteomics studies have also been performed on a few DMRB strains, mostly using gel-based separation for proteome fractionation and mapping. This dissertation developed an unfractionated and untargeted mass spectrometry-based proteomics approach, which demonstrated protein identification with higher throughput, resolution, and accuracy, and achieved high proteome coverage compared to the 2DE-based approach. From sample preparation, protein digestion, peptide separation, to peptide

sequencing and proteome identification, careful considerations were taken to generate deep, comprehensive, and unbiased proteomic measurements for DMRBs.

In Chapter 4, the proteome characterization for the DMRB *A. dehalogenan* strain 2CP-C was conducted using the developed mass spectrometry workflow, and a wide range of electron acceptors were tested for their influence on the proteomic expression of strain 2CP-C. The direct comparisons between different forms of metal electron acceptors and non-metal electron acceptors are breakthroughs in the study of microbial respiration. With the proteomics characterization of strain 2CP-C under such a wide range of different electron-accepting conditions, the similarities and differences of microbial proteome expression and pathway expression toward the utilization of different electron acceptors for bacterial respiration were revealed. In this dissertation, the proteomics characterization suggested differential activities for energy metabolism in strain 2CP-C with metal electron acceptor-growth and non-metal electron acceptor-growth. This phenomenon was never demonstrated before through the physiological characterizations or transcriptomics analysis for strain 2CP-C.

The results in Chapter 4 revealed the core proteomes of strain 2CP-C, which represented essential proteins participating in all metabolic activities. Identifying proteins in the core proteome is important to understand the basic microbial protein machineries for survival, and provides detailed information for organism manipulations. In addition, uniquely expressed proteins were also detected by mass spectrometry. Through functional comparisons between unique proteins under different growth conditions, it was clear that the reduction of specific electron acceptor would require the activities of proteins with specific or specialized functions.

The linkage between unique proteins and their corresponding electron acceptor provides promising establishment of biomarker indicators for specific pathway activities, which could be exploited for microbial activities representation.

The developed mass spectrometry-based proteomics approach is widely applicable for various microbial organisms, and successful proteome characterization using this approach was demonstrated on two other DMRBs in Chapter 5. This provides the opportunity for complete proteome characterizations for more DMRB species, which is beneficial for the enhancement of existing knowledge about microbial molecular activities.

Through systematic comparisons of detected proteomes between different DMRB strains, this dissertation demonstrated the unique features for individual strains, such as the confirmation of conductive pili structure in *G. daltonii* strain FRC-32 when grown with solid iron electron acceptor. Previously, it was unclear whether the pili electron transfer mechanism is universal to the respiration of all solid metal electron acceptors. Our results provided evidence for the uniqueness of pili generation for solid iron-growth. Whether other DMRB species use similar electron transfer mechanisms via pili structure was also under debate. The proteomics characterization and comparisons for different DMRB made it clear that *Geobacter* was the only genus with the unique pili electron transfer mechanism out of the three tested organisms.

Direct proteome comparisons of different DMRB strains also revealed the similarities and differences of the essentiality of core proteomes. As shown in Chapter 5, the core proteome of *A. dehalogenan* strain 2CP-C, *S. oneidensis* strain MR-1 and *G. daltonii* strain FRC-32 represents almost all metabolic pathways to sustain energy conservation and growth. The

mass spectrometry analysis of DMRBs clearly showed that different strains have similar overall metabolic pathways, and in the same time, are unique in their own proteome activities to maintain individual lifestyles.

6.5 Perspectives on DMRB Proteomics Research

Over the course of this dissertation, mass spectrometry technology has improved rapidly. With many new instrumentations developed, such as LTQ-Orbitrap-Velos introduced in 2009, and LTQ-Orbitrap-Elite introduced in 2012, mass spectrometers with enhanced analytical figures of merit provide better qualities of proteome identification and quantification. The increased performances of newly developed instruments could be exploited for DMRB proteome characterizations with even higher coverage, better accuracy and resolution, higher throughput, and deeper measurements.

The instrumentation improvement also provides the capabilities of analyzing biological systems with higher complexities, such as environmental microbial communities. Since DMRB organisms naturally occur in various environmental settings from water, sediments, to soil, the proteomics characterization of the *in situ* DMRB metabolic activities are of tremendous importance to understand their functional roles in the ecosystem. To achieve accurate metaproteomics analysis on environmental DMRBs, the direct metagenome information for the same field sample would be highly desirable. Although artificially constructed genome database with concatenation of proteins from known DMRB species could provide information for metaproteomics analysis, the accuracy of protein identification, spectrum assignment and

protein quantification are confounded by the increased bacterial species, unpredictable abundances, uncharacterized and uncultivated strains, unknown organisms, and the interaction and communications between indigenous organisms. Currently, species differentiation and strain differentiation are tremendous challenges for metagenome studies, which also hinder the detailed differentiation of metaproteomics analysis. Ideally, strain-level information could be achieved through metagenomics and metaproteomics for environmental microbial communities, and the *in situ* metaproteomics results could be compared and contrasted to existing knowledge on individual strains obtained in laboratory conditions such as the results provided in this dissertation.

For field-based metaproteomics studies, protein extraction methods should be carefully chosen to maximize the protein recovery with unbiased efficiencies for different organisms present in the environmental sample. The main sample preparation used in this dissertation (SDS-TCA method) was developed for soil microbial proteome extraction, and have demonstrated success utilities for environmental samples. Metaproteomics studies using the SDS-TCA method described in Chapter 2 could be conveniently compared to results presented in this dissertation. However, cautions need to be exerted, in that the SDS-TCA approach was proven to be interfered by the presence of FeOOH in this dissertation. While FeOOH is a very common form of ferric iron mineral present in the environment, it is likely that the field sample containing FeOOH would be problematic for achieving comprehensive metaproteome extraction. Thus, continuous improvement for proteomics sample preparation should be conducted, in order to gain deeper insight and more complete information to the proteome analysis on either microbial isolates or environmental communities.

Overall, mass spectrometry-based proteomics is a fast developing field that is providing continuous proteomic understandings to environmental microbiology. With better technological improvements applied on more environmental microorganisms, the proteomics information will present groundbreaking insights into our knowledge for microbiology and life sciences.

References

1. Avery, O.T., C.M. MacLeod, and M. McCarty, *Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III.* The Journal of experimental medicine, 1944. **79**(2): p. 137-158.
2. Hershey, A.D. and M. Chase, *Independent functions of viral protein and nucleic acid in growth of bacteriophage.* The Journal of general physiology, 1952. **36**(1): p. 39-56.
3. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids.* Nature, 1953. **171**(4356): p. 737-738.
4. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-563.
5. Crick, F.H., *The origin of the genetic code.* Journal of molecular biology, 1968. **38**(3): p. 367-379.
6. Fleischmann, R.D., et al., *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* Science, 1995. **269**(5223): p. 496-512.
7. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* Journal of molecular biology, 1975. **94**(3): p. 441-448.
8. Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA.* Proceedings of the National Academy of Sciences, 1977. **74**(2): p. 560-564.
9. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the National Academy of Sciences, 1977. **74**(12): p. 5463-5467.
10. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-1351.
11. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
12. ANDERSON, M.A., et al., *A polymorphic DNA marker genetically linked to Huntington's disease.* Landmarks in Medical Genetics: Classic Papers with Commentaries, 2004. **51**: p. 153.
13. Riesenfeld, C.S., P.D. Schloss, and J. Handelsman, *Metagenomics: genomic analysis of microbial communities.* Annu. Rev. Genet., 2004. **38**: p. 525-552.
14. Handelsman, J., *Metagenomics: application of genomics to uncultured microorganisms.* Microbiology and Molecular Biology Reviews, 2004. **68**(4): p. 669-685.
15. Williamson, S.J., et al., *The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples.* PloS one, 2008. **3**(1): p. e1456.
16. Daniel, R., *The soil metagenome–a rich resource for the discovery of novel natural products.* Current opinion in biotechnology, 2004. **15**(3): p. 199-204.
17. Pernthaler, A., et al., *Diverse syntrophic partnerships from deep-sea methane vents revealed by direct cell capture and metagenomics.* Proceedings of the National Academy of Sciences, 2008. **105**(19): p. 7052-7057.
18. Gill, S.R., et al., *Metagenomic analysis of the human distal gut microbiome.* Science, 2006. **312**(5778): p. 1355-1359.
19. Lazarevic, V., et al., *Metagenomic study of the oral microbiota by Illumina high-throughput sequencing.* Journal of microbiological methods, 2009. **79**(3): p. 266-271.
20. DeRisi, J., et al., *Use of a cDNA microarray to analyse gene expression patterns in human cancer.* Nature genetics, 1996. **14**(4): p. 457-460.
21. Spellman, P.T., et al., *Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization.* Molecular biology of the cell, 1998. **9**(12): p. 3273-3297.
22. Zhou, J., *Microarrays for bacterial detection and microbial community analysis.* Current opinion in Microbiology, 2003. **6**(3): p. 288-294.
23. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature Reviews Genetics, 2009. **10**(1): p. 57-63.

24. Klose, J., *Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues.* Humangenetik, 1975. **26**(3): p. 231-243.
25. O'Farrell, P.H., *High resolution two-dimensional electrophoresis of proteins.* Journal of biological chemistry, 1975. **250**(10): p. 4007-4021.
26. Wilkins, M.R., et al., *From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and arnino acid analysis.* Nature Biotechnology, 1996. **14**(1): p. 61-65.
27. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
28. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Analytical chemistry, 1988. **60**(20): p. 2299-2301.
29. Gygi, S.P., B. Rist, and R. Aebersold, *Measuring gene expression by quantitative proteome analysis.* Current opinion in biotechnology, 2000. **11**(4): p. 396-401.
30. Peng, J. and S.P. Gygi, *Proteomics: the move to mixtures.* Journal of mass spectrometry, 2001. **36**(10): p. 1083-1091.
31. Abraham, P., et al., *Defining the boundaries and characterizing the landscape of functional genome expression in vascular tissues of Populus using shotgun proteomics.* Journal of proteome research, 2011. **11**(1): p. 449-460.
32. Chourey, K., et al., *Environmental proteomics reveals early microbial community responses to biostimulation at a uranium‐and nitrate‐contaminated site.* Proteomics, 2013. **13**(18-19): p. 2921-2930.
33. Nissen, S., et al., *Comparative c-type cytochrome expression analysis in Shewanella oneidensis strain MR-1 and Anaeromyxobacter dehalogenans strain 2CP-C grown with soluble and insoluble oxidized metal electron acceptors.* Biochemical Society Transactions, 2012. **40**(6).
34. Verberkmoes, N.C., et al., *Shotgun metaproteomics of the human distal gut microbiota.* The ISME journal, 2009. **3**(2): p. 179-189.
35. Young, J.C., et al., *Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in Streptococcus thermophilus.* PloS one, 2012. **7**(5): p. e38077.
36. Giannone, R.J., et al., *Proteomic characterization of cellular and molecular processes that enable the Nanoarchaeum equitans-Ignicoccus hospitalis relationship.* PloS one, 2011. **6**(8): p. e22942.
37. Abraham, P., et al., *Putting the pieces together: high-performance LC-MS/MS provides network-, pathway-, and protein-level perspectives in Populus.* Molecular & Cellular Proteomics, 2013. **12**(1): p. 106-119.
38. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: the apogee of the omics trilogy.* Nature reviews Molecular cell biology, 2012. **13**(4): p. 263-269.
39. Sugimoto, M., et al., *Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles.* Metabolomics, 2010. **6**(1): p. 78-95.
40. Bundy, J.G., M.P. Davey, and M.R. Viant, *Environmental metabolomics: a critical review and future perspectives.* Metabolomics, 2009. **5**(1): p. 3-21.
41. Compton, P.D., et al., *On the scalability and requirements of whole protein mass spectrometry.* Analytical chemistry, 2011. **83**(17): p. 6868-6874.
42. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.
43. Tabb, D.L., W.H. McDonald, and J.R. Yates, *DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics.* Journal of proteome research, 2002. **1**(1): p. 21-26.

44. Tabb, D.L., C.G. Fernando, and M.C. Chambers, *MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis.* Journal of proteome research, 2007. **6**(2): p. 654-661.
45. Ma, Z.-Q., et al., *IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering.* Journal of proteome research, 2009. **8**(8): p. 3872-3881.
46. Ram, R.J., et al., *Community proteomics of a natural microbial biofilm.* Science, 2005. **308**(5730): p. 1915-1920.
47. Denef, V.J., et al., *Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities.* Proceedings of the National Academy of Sciences, 2010. **107**(6): p. 2383-2390.
48. Mueller, R.S., et al., *Ecological distribution and population physiology defined by proteomics in a natural microbial community.* Molecular systems biology, 2010. **6**(1).
49. Benndorf, D. and U. Reichl, *Proteomics in environmental and technical microbiology.* Engineering in Life Sciences, 2014. **14**(1): p. 27-46.
50. Cowan, D.A., *The upper temperature for life–where do we draw the line?* Trends in microbiology, 2004. **12**(2): p. 58-60.
51. Baker-Austin, C. and M. Dopson, *Life in acid: pH homeostasis in acidophiles.* Trends in microbiology, 2007. **15**(4): p. 165-171.
52. Turnbaugh, P.J., et al., *The human microbiome project: exploring the microbial part of ourselves in a changing world.* Nature, 2007. **449**(7164): p. 804.
53. Madsen, E.L., *Environmental microbiology: from genomes to biogeochemistry*. 2011: John Wiley & Sons.
54. Paustian, T. and G. Roberts, *Through the Microscope: Adventures in Microbiology*. 2012, Textbook Consortia.
55. Chourey, K., et al., *Comparative temporal proteomics of a response regulator (SO2426)-deficient strain and wild-type Shewanella oneidensis MR-1 during chromate transformation.* Journal of proteome research, 2009. **8**(1): p. 59-71.
56. Brown, S.D., et al., *Molecular dynamics of the Shewanella oneidensis response to chromate stress.* Molecular & Cellular Proteomics, 2006. **5**(6): p. 1054-1071.
57. VerBerkmoes, N.C., et al., *Determination and Comparison of the Baseline Proteomes of the Versatile Microbe Rhodopseudomonas p alustris under Its Major Metabolic States.* Journal of proteome research, 2006. **5**(2): p. 287-298.
58. Kao, W.-C., et al., *Quantitative proteomic analysis of metabolic regulation by copper ions in Methylococcus capsulatus (Bath).* Journal of biological chemistry, 2004. **279**(49): p. 51554-51560.
59. Bergaust, L., et al., *Denitrification response patterns during the transition to anoxic respiration and posttranscriptional effects of suboptimal pH on nitrogen oxide reductase in Paracoccus denitrificans.* Applied and environmental microbiology, 2010. **76**(19): p. 6387-6396.
60. Lacerda, C.M. and K.F. Reardon, *Environmental proteomics: applications of proteome profiling in environmental microbiology and biotechnology.* Briefings in functional genomics & proteomics, 2009. **8**(1): p. 75-87.
61. Wilmes, P., et al., *Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal.* The ISME journal, 2008. **2**(8): p. 853-864.
62. Benndorf, D., et al., *Functional metaproteome analysis of protein extracts from contaminated soil and groundwater.* The ISME journal, 2007. **1**(3): p. 224-234.
63. Graham, D.E., et al., *Microbes in thawing permafrost: the unknown variable in the climate change equation.* The ISME journal, 2011. **6**(4): p. 709-712.

64. Fredrickson, J.K. and Y.A. Gorby, *Environmental processes mediated by iron-reducing bacteria.* Current opinion in biotechnology, 1996. **7**(3): p. 287-294.
65. Lonergan, D.J., et al., *Phylogenetic analysis of dissimilatory Fe (III)-reducing bacteria.* Journal of Bacteriology, 1996. **178**(8): p. 2402-2408.
66. Weber, K.A., L.A. Achenbach, and J.D. Coates, *Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction.* Nature Reviews Microbiology, 2006. **4**(10): p. 752-764.
67. Lovley, D.R., *Dissimilatory metal reduction.* Annual Reviews in Microbiology, 1993. **47**(1): p. 263-290.
68. Vargas, M., et al., *Microbiological evidence for Fe (III) reduction on early Earth.* Nature, 1998. **395**(6697): p. 65-67.
69. Cairns-Smith, A.G., A.J. Hall, and M.J. Russell, *Mineral theories of the origin of life and an iron sulfide example*, in *Marine Hydrothermal Systems and the Origin of Life*. 1992, Springer. p. 161-180.
70. Gold, T., *The deep, hot biosphere.* Proceedings of the National Academy of Sciences, 1992. **89**(13): p. 6045-6049.
71. Seckbach, J., *Origins: genesis, evolution and diversity of life*. Vol. 6. 2004: Springer.
72. Lovley, D.R. and E.J. Phillips, *Manganese inhibition of microbial iron reduction in anaerobic sediments.* Geomicrobiology Journal, 1988. **6**(3-4): p. 145-155.
73. Lovley, D.R. and E.J. Phillips, *Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese.* Applied and environmental microbiology, 1988. **54**(6): p. 1472-1480.
74. Lovley, D., *Dissimilatory Fe (III)-and Mn (IV)-reducing prokaryotes*, in *The prokaryotes*. 2006, Springer. p. 635-658.
75. Wall, J.D. and L.R. Krumholz, *Uranium reduction.* Annu. Rev. Microbiol., 2006. **60**: p. 149-166.
76. Rotaru, D.E.H., et al., *Geobacter: the microbe electric's physiology, ecology, and practical applications.* Advances in microbial physiology, 2011. **59**: p. 1.
77. Malvankar, N.S., et al., *Tunable metallic-like conductivity in microbial nanowire networks.* nature nanotechnology, 2011. **6**(9): p. 573-579.
78. Reguera, G., et al., *Extracellular electron transfer via microbial nanowires.* Nature, 2005. **435**(7045): p. 1098-1101.
79. Klimes, A., et al., *Production of pilus‐like filaments in Geobacter sulfurreducens in the absence of the type IV pilin protein PilA.* FEMS microbiology letters, 2010. **310**(1): p. 62-68.
80. Yi, H., et al., *Selection of a variant of< i> Geobacter sulfurreducens</i> with enhanced capacity for current production in microbial fuel cells.* Biosensors and Bioelectronics, 2009. **24**(12): p. 3498-3503.
81. MacDonell, M.T. and R.R. Colwell, *Phylogeny of the Vibrionaceae, and Recommendation for Two New Genera, Listonella and Shewanella.* Systematic and Applied Microbiology, 1985. **6**(2): p. 171-182.
82. MYERS, C.R. and K.H. NEALSON, *Bacterial Manganese Reduction and Growth with Manganese Oxide as the Sole Electron Acceptor.* Science, 1988. **240**(4857): p. 1319-1321.
83. Hau, H.H. and J.A. Gralnick, *Ecology and Biotechnology of the Genus Shewanella.* Annual Review of Microbiology, 2007. **61**(1): p. 237-258.
84. Hartshorne, R., et al., *Characterization of Shewanella oneidensis MtrC: a cell-surface decaheme cytochrome involved in respiratory electron transport to extracellular electron acceptors.* JBIC Journal of Biological Inorganic Chemistry, 2007. **12**(7): p. 1083-1094.
85. Marsili, E., et al., *Shewanella secretes flavins that mediate extracellular electron transfer.* Proceedings of the National Academy of Sciences, 2008. **105**(10): p. 3968-3973.

86. El-Naggar, M.Y., et al., *Electrical transport along bacterial nanowires from Shewanella oneidensis MR-1.* Proceedings of the National Academy of Sciences, 2010. **107**(42): p. 18127-18131.
87. Sanford, R.A., J.R. Cole, and J.M. Tiedje, *Characterization and description of Anaeromyxobacter dehalogenans gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium.* Applied and environmental microbiology, 2002. **68**(2): p. 893-900.
88. Wu, Q., R.A. Sanford, and F.E. Löffler, *Uranium (VI) reduction by Anaeromyxobacter dehalogenans strain 2CP-C.* Applied and environmental microbiology, 2006. **72**(5): p. 3608-3614.
89. Löffler, F.E., J.M. Tiedje, and R.A. Sanford, *Fraction of electrons consumed in electron acceptor reduction and hydrogen thresholds as indicators of halorespiratory physiology.* Applied and environmental microbiology, 1999. **65**(9): p. 4049-4056.
90. North, N.N., et al., *Change in bacterial community structure during in situ biostimulation of subsurface sediment cocontaminated with uranium and nitrate.* Applied and environmental microbiology, 2004. **70**(8): p. 4911-4920.
91. Petrie, L., et al., *Enumeration and characterization of iron (III)-reducing microbial communities from acidic subsurface sediments contaminated with uranium (VI).* Applied and environmental microbiology, 2003. **69**(12): p. 7467-7479.
92. He, Q. and R.A. Sanford, *Induction characteristics of reductive dehalogenation in theortho-halophenol-respiring bacterium, Anaeromyxobacter dehalogenans.* Biodegradation, 2002. **13**(5): p. 307-316.
93. Methe, B., et al., *Genome of Geobacter sulfurreducens: metal reduction in subsurface environments.* Science, 2003. **302**(5652): p. 1967-1969.
94. Kranz, R.G., et al., *Cytochrome c biogenesis: mechanisms for covalent modifications and trafficking of heme and for heme-iron redox control.* Microbiology and Molecular Biology Reviews, 2009. **73**(3): p. 510-528.
95. Moore, G. and G. Pettigrew, *Cytochromes c.* Evolutionary, Structural and Physicochemical Aspects. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, 1990.
96. Ow, Y.-L.P., et al., *Cytochrome c: functions beyond respiration.* Nature reviews Molecular cell biology, 2008. **9**(7): p. 532-542.
97. Ferguson, S.J., et al., *Cytochrome< i> c</i> assembly: A tale of ever increasing variation and mystery?* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 2008. **1777**(7): p. 980-984.
98. Richardson, D.J., *Bacterial respiration: a flexible process for a changing environment.* Microbiology, 2000. **146**(3): p. 551-571.
99. Shi, L., et al., *Respiration of metal (hydr) oxides by Shewanella and Geobacter: a key role for multihaem c‐type cytochromes.* Molecular microbiology, 2007. **65**(1): p. 12-20.
100. Myers, J.M. and C.R. Myers, *Role of the tetraheme cytochrome CymA in anaerobic electron transport in cells of Shewanella putrefaciens MR-1 with normal levels of menaquinone.* Journal of bacteriology, 2000. **182**(1): p. 67-75.
101. Zargar, K. and C.W. Saltikov, *Lysine-91 of the tetraheme c-type cytochrome CymA is essential for quinone interaction and arsenate respiration in Shewanella sp. strain ANA-3.* Archives of microbiology, 2009. **191**(11): p. 797-806.
102. Richardson, D.J., et al., *The 'porin‐cytochrome' model for microbe‐to‐mineral electron transfer.* Molecular microbiology, 2012. **85**(2): p. 201-212.
103. Gorby, Y.A., et al., *Electrically conductive bacterial nanowires produced by Shewanella oneidensis strain MR-1 and other microorganisms.* Proceedings of the National Academy of Sciences, 2006. **103**(30): p. 11358-11363.
104. Leang, C., M.V. Coppi, and D. Lovley, *OmcB, a c-type polyheme cytochrome, involved in Fe (III) reduction in Geobacter sulfurreducens.* Journal of Bacteriology, 2003. **185**(7): p. 2096-2103.

105. Mehta, T., et al., *Outer membrane c-type cytochromes required for Fe (III) and Mn (IV) oxide reduction in Geobacter sulfurreducens.* Applied and environmental microbiology, 2005. **71**(12): p. 8634-8641.
106. Afkar, E., et al., *A novel Geobacteraceae-specific outer membrane protein J (OmpJ) is essential for electron transport to Fe (III) and Mn (IV) oxides in Geobacter sulfurreducens.* BMC microbiology, 2005. **5**(1): p. 41.
107. Lovley, D., *Long-range electron transport to Fe (III) oxide via pili with metallic-like conductivity.* Biochemical Society Transactions, 2012. **40**(6): p. 1186-1190.
108. Esteve‐Núñez, A., et al., *Fluorescent properties of c‐type cytochromes reveal their potential role as an extracytoplasmic electron sink in Geobacter sulfurreducens.* Environmental microbiology, 2008. **10**(2): p. 497-505.
109. Lovley, D.R., *Extracellular electron transfer: wires, capacitors, iron lungs, and more.* Geobiology, 2008. **6**(3): p. 225-231.
110. Meyer, T.E., et al., *Identification of 42 possible cytochrome c genes in the Shewanella oneidensis genome and characterization of six soluble cytochromes.* Omics: a journal of integrative biology, 2004. **8**(1): p. 57-77.
111. Thomas, S.H., et al., *The mosaic genome of Anaeromyxobacter dehalogenans strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria.* PloS one, 2008. **3**(5): p. e2103.
112. Lin, W., M.V. Coppi, and D. Lovley, *Geobacter sulfurreducens can grow with oxygen as a terminal electron acceptor.* Applied and environmental microbiology, 2004. **70**(4): p. 2525-2528.
113. Heidelberg, J.F., et al., *Genome sequence of the dissimilatory metal ion–reducing bacterium Shewanella oneidensis.* Nature Biotechnology, 2002. **20**(11): p. 1118-1123.
114. Myers, J.M. and C.R. Myers, *Genetic complementation of an outer membrane cytochrome omcB mutant of Shewanella putrefaciens MR-1 requires omcB plus downstream DNA.* Applied and environmental microbiology, 2002. **68**(6): p. 2781-2793.
115. Coppi, M.V., et al., *Development of a Genetic System forGeobacter sulfurreducens.* Applied and environmental microbiology, 2001. **67**(7): p. 3180-3187.
116. Shelobolina, E.S., et al., *Importance of c-type cytochromes for U (VI) reduction by Geobacter sulfurreducens.* BMC microbiology, 2007. **7**(1): p. 16.
117. Bretschger, O., et al., *Current production and metal oxide reduction by Shewanella oneidensis MR-1 wild type and mutants.* Applied and environmental microbiology, 2007. **73**(21): p. 7003-7012.
118. Fredrickson, J.K. and M.F. Romine, *Genome-assisted analysis of dissimilatory metal-reducing bacteria.* Current opinion in biotechnology, 2005. **16**(3): p. 269-274.
119. Beliaev, A.S., et al., *Gene and protein expression profiles of Shewanella oneidensis during anaerobic growth with different electron acceptors.* Omics: a journal of integrative biology, 2002. **6**(1): p. 39-60.
120. Thompson, D.K., et al., *Transcriptional and proteomic analysis of a ferric uptake regulator (fur) mutant of Shewanella oneidensis: possible involvement of fur in energy metabolism, transcriptional regulation, and oxidative stress.* Applied and environmental microbiology, 2002. **68**(2): p. 881-892.
121. Wan, X.-F., et al., *Transcriptomic and proteomic characterization of the Fur modulon in the metal-reducing bacterium Shewanella oneidensis.* Journal of Bacteriology, 2004. **186**(24): p. 8385-8400.
122. Romine, M.F., et al., *Validation of Shewanella oneidensis MR-1 small proteins by AMT tag-based proteome analysis.* Omics: a journal of integrative biology, 2004. **8**(3): p. 239-254.

123. Elias, D.A., et al., *Global detection and characterization of hypothetical proteins in Shewanella oneidensis MR ‐1 using LC ‐MS based proteomics.* Proteomics, 2005. **5**(12): p. 3120-3130.
124. Kolker, E., et al., *Global profiling of Shewanella oneidensis MR-1: expression of hypothetical genes and improved functional annotations.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(6): p. 2099-2104.
125. Fang, R., et al., *Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach.* Molecular & Cellular Proteomics, 2006. **5**(4): p. 714-725.
126. Thompson, M.R., et al., *Dosage-dependent proteome response of Shewanella oneidensis MR-1 to acute chromate challenge.* Journal of proteome research, 2007. **6**(5): p. 1745-1757.
127. Ding, Y.-H.R., et al., *The proteome of dissimilatory metal-reducing microorganism< i> Geobacter sulfurreducens</i> under various growth conditions.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2006. **1764**(7): p. 1198-1206.
128. Ding, Y.-H.R., et al., *Proteome of< i> Geobacter sulfurreducens</i> grown with Fe (III) oxide or Fe (III) citrate as the electron acceptor.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2008. **1784**(12): p. 1935-1941.
129. Cole, J.R., B.Z. Fathepure, and J.M. Tiedje, *Tetrachloroethene and 3-chlorobenzoate dechlorination activities are co-induced inDesulfomonile tiedjei DCB-1.* Biodegradation, 1995. **6**(2): p. 167-172.
130. Loffler, F.E., R.A. Sanford, and J.M. Tiedje, *Initial Characterization of a Reductive Dehalogenase from Desulfitobacterium chlororespirans Co23.* Applied and Environmental Microbiology, 1996. **62**(10): p. 3809-3813.
131. Sanford, R.A., et al., *Characterization of Desulfitobacterium chlororespirans sp. nov., which grows by coupling the oxidation of lactate to the reductive dechlorination of 3-chloro-4-hydroxybenzoate.* Applied and environmental microbiology, 1996. **62**(10): p. 3800-3808.
132. Chourey, K., et al., *Direct cellular lysis/protein extraction protocol for soil metaproteomics.* Journal of proteome research, 2010. **9**(12): p. 6615-6622.
133. Wu, C.C. and J.R. Yates, *The application of mass spectrometry to membrane proteomics.* Nature Biotechnology, 2003. **21**(3): p. 262-267.
134. Lu, S., et al., *Insights into the structure and metabolic function of microbes that shape pelagic iron-rich aggregates ("iron snow").* Applied and environmental microbiology, 2013. **79**(14): p. 4272-4281.
135. Thompson, M.R., et al., *Experimental approach for deep proteome measurements from small-scale microbial biomass samples.* Analytical chemistry, 2008. **80**(24): p. 9517-9525.
136. Smith, P., et al., *Measurement of protein using bicinchoninic acid.* Analytical biochemistry, 1985. **150**(1): p. 76-85.
137. Issaq, H.J., et al., *Multidimensional separation of peptides for effective proteomic analysis.* Journal of Chromatography B, 2005. **817**(1): p. 35-47.
138. Issaq, H.J., *The role of separation science in proteomics research.* Electrophoresis, 2001. **22**(17): p. 3629.
139. Wolters, D.A., M.P. Washburn, and J.R. Yates, *An automated multidimensional protein identification technology for shotgun proteomics.* Analytical chemistry, 2001. **73**(23): p. 5683-5690.
140. Hillenkamp, F. and M. Karas, *[12] Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization.* Methods in enzymology, 1990. **193**: p. 280-295.
141. Banerjee, S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte.* International journal of analytical chemistry, 2012. **2012**.

142. Wilm, M. and M. Mann, *Analytical properties of the nanoelectrospray ion source.* Analytical chemistry, 1996. **68**(1): p. 1-8.
143. Karas, M., U. Bahr, and T. Dülcks, *Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine.* Fresenius' journal of analytical chemistry, 2000. **366**(6-7): p. 669-676.
144. Takats, Z., et al., *Electrosonic spray ionization. A gentle technique for generating folded proteins and protein complexes in the gas phase and for studying ion-molecule reactions at atmospheric pressure.* Analytical chemistry, 2004. **76**(14): p. 4050-4058.
145. Taylor, G., *Disintegration of water drops in an electric field.* Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 1964. **280**(1382): p. 383-397.
146. Kebarle, P. and L. Tang, *From ions in solution to ions in the gas phase-the mechanism of electrospray mass spectrometry.* Analytical chemistry, 1993. **65**(22): p. 972A-986A.
147. Fenn, J.B., *Ion formation from charged droplets: roles of geometry, energy, and time.* Journal of the American Society for Mass Spectrometry, 1993. **4**(7): p. 524-535.
148. Kebarle, P. and M. Peschke, *On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions.* Analytica Chimica Acta, 2000. **406**(1): p. 11-35.
149. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels.* Analytical chemistry, 1996. **68**(5): p. 850-858.
150. Rayleigh, L., *XX. On the equilibrium of liquid conducting masses charged with electricity.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1882. **14**(87): p. 184-186.
151. Schmelzeisen-Redeker, G., L. Bütfering, and F. Röllgen, *Desolvation of ions and molecules in thermospray mass spectrometry.* International Journal of Mass Spectrometry and Ion Processes, 1989. **90**(2): p. 139-150.
152. Nguyen, S. and J.B. Fenn, *Gas-phase ions of solute species from charged droplets of solutions.* Proceedings of the National Academy of Sciences, 2007. **104**(4): p. 1111-1117.
153. Dole, M., et al., *Molecular beams of macroions.* The Journal of Chemical Physics, 1968. **49**(5): p. 2240-2249.
154. Iribarne, J. and B. Thomson, *On the evaporation of small ions from charged droplets.* The Journal of Chemical Physics, 1976. **64**(6): p. 2287-2294.
155. Thomson, B. and J. Iribarne, *Field induced ion evaporation from liquid surfaces at atmospheric pressure.* The Journal of Chemical Physics, 1979. **71**(11): p. 4451-4463.
156. Schwartz, J.C., M.W. Senko, and J.E. Syka, *A two-dimensional quadrupole ion trap mass spectrometer.* Journal of the American Society for Mass Spectrometry, 2002. **13**(6): p. 659-669.
157. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* Journal of Mass Spectrometry, 2005. **40**(4): p. 430-443.
158. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Analytical chemistry, 2000. **72**(6): p. 1156-1162.
159. Shukla, A.K. and J.H. Futrell, *Tandem mass spectrometry: dissociation of ions by collisional activation.* Journal of Mass Spectrometry, 2000. **35**(9): p. 1069-1090.
160. Zubarev, R.A., N.L. Kelleher, and F.W. McLafferty, *Electron capture dissociation of multiply charged protein cations. A nonergodic process.* Journal of the American Chemical Society, 1998. **120**(13): p. 3265-3266.
161. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(26): p. 9528-9533.

162. Penn, S.G., et al., *The use of heated capillary dissociation and collision-induced dissociation to determine the strength of noncovalent bonding interactions in gas-phase peptide-cyclodextrin complexes.* Journal of the American Society for Mass Spectrometry, 1997. **8**(3): p. 244-252.
163. Zhang, Q., et al., *Analysis of non‐enzymatically glycated peptides: neutral‐loss‐triggered MS3 versus multi‐stage activation tandem mass spectrometry.* Rapid Communications in Mass Spectrometry, 2008. **22**(19): p. 3027-3034.
164. Guo, T., et al., *Hybridization of pulsed-Q dissociation and collision-activated dissociation in linear ion trap mass spectrometer for iTRAQ quantitation.* Journal of proteome research, 2008. **7**(11): p. 4831-4840.
165. McLuckey, S.A., *Principles of collisional activation in analytical mass spectrometry.* Journal of the American Society for Mass Spectrometry, 1992. **3**(6): p. 599-614.
166. Hunt, D.F., et al., *Sequence analysis of oligopeptides by secondary ion/collision activated dissociation mass spectrometry.* Analytical chemistry, 1981. **53**(11): p. 1704-1706.
167. Tabb, D.L., et al., *DBDigger: reorganized proteomic database identification that improves flexibility and speed.* Analytical chemistry, 2005. **77**(8): p. 2464-2474.
168. VerBerkmoes, N.C., et al., *Integrating "top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of Shewanella oneidensis.* Journal of proteome research, 2002. **1**(3): p. 239-252.
169. Liu, H., R.G. Sadygov, and J.R. Yates, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics.* Analytical chemistry, 2004. **76**(14): p. 4193-4201.
170. Zhang, B., et al., *Detecting differential and correlated protein expression in label-free shotgun proteomics.* Journal of proteome research, 2006. **5**(11): p. 2909-2918.
171. Paoletti, A.C., et al., *Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors.* Proceedings of the National Academy of Sciences, 2006. **103**(50): p. 18928-18933.
172. Van Nostrand, J.D., et al., *Dynamics of microbial community composition and function during in situ bioremediation of a uranium-contaminated aquifer.* Applied and Environmental Microbiology, 2011. **77**(11): p. 3860-3869.
173. He, Q. and R.A. Sanford, *Characterization of Fe (III) reduction by chlororespiring Anaeromxyobacter dehalogenans.* Applied and environmental microbiology, 2003. **69**(5): p. 2712-2718.
174. Childers, S.E., S. Ciufo, and D.R. Lovley, *Geobacter metallireducens accesses insoluble Fe (III) oxide by chemotaxis.* Nature, 2002. **416**(6882): p. 767-769.
175. Richter, K., M. Schicklberger, and J. Gescher, *Dissimilatory reduction of extracellular electron acceptors in anaerobic respiration.* Applied and Environmental Microbiology, 2012. **78**(4): p. 913-921.
176. Allen, J.W., et al., *C-type cytochromes: diverse structures and biogenesis systems pose evolutionary problems.* Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 2003. **358**(1429): p. 255-266.
177. Sophie, J.M., et al., *A functional description of CymA, an electron-transfer hub supporting anaerobic respiratory flexibility in Shewanella.* Biochemical Journal, 2012. **444**(3): p. 465-474.
178. Carmona-Martinez, A.A., et al., *Cyclic voltammetric analysis of the electron transfer of< i> Shewanella oneidensis</i> MR-1 and nanofilament and cytochrome knock-out mutants.* Bioelectrochemistry, 2011. **81**(2): p. 74-80.
179. Beliaev, A.S., et al., *Global transcriptome analysis of Shewanella oneidensis MR-1 exposed to different terminal electron acceptors.* Journal of bacteriology, 2005. **187**(20): p. 7138-7145.
180. Holmes, D.E., et al., *Transcriptome of Geobacter uraniireducens growing in uranium-contaminated subsurface sediments.* The ISME journal, 2008. **3**(2): p. 216-230.

181. Smith, J.A., D.R. Lovley, and P.-L. Tremblay, *Outer cell surface components essential for Fe (III) oxide reduction by Geobacter metallireducens.* Applied and Environmental Microbiology, 2013. **79**(3): p. 901-907.
182. Thompson, D.K., et al., *Proteomics reveals a core molecular response of Pseudomonas putida F1 to acute chromate challenge.* BMC genomics, 2010. **11**(1): p. 311.
183. Chourey, K., et al., *Global molecular and morphological effects of 24-hour chromium (VI) exposure on Shewanella oneidensis MR-1.* Applied and Environmental Microbiology, 2006. **72**(9): p. 6331-6344.
184. Zybailov, B., et al., *Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces c erevisiae.* Journal of proteome research, 2006. **5**(9): p. 2339-2347.
185. Nancy, Y.Y., et al., *PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes.* Bioinformatics, 2010. **26**(13): p. 1608-1615.
186. Kern, M. and J. Simon, *Production of recombinant multiheme cytochromes c in Wolinella succinogenes.* Methods Enzymol, 2011. **486**: p. 429-446.
187. Rodrigues, M.L., et al., *Quinol Oxidation by< i> c</i>-Type Cytochromes: Structural Characterization of the Menaquinol Binding Site of NrfHA.* Journal of molecular biology, 2008. **381**(2): p. 341-350.
188. Hartshorne, S., D. Richardson, and J. Simon, *Multiple haem lyase genes indicate substrate specificity in cytochrome c biogenesis.* Biochemical Society Transactions, 2006. **34**: p. 146-149.
189. Shirodkar, S., et al., *The octahaem SirA catalyses dissimilatory sulfite reduction in Shewanella oneidensis MR‐1.* Environmental microbiology, 2011. **13**(1): p. 108-115.
190. Jorg, S., et al., *Physiological function and catalytic versatility of bacterial multihaem cytochromes c involved in nitrogen and sulfur cycling.* Biochemical Society Transactions, 2011. **39**(6): p. 1864-1870.
191. Van Beilen, J.B., et al., *Rubredoxins involved in alkane oxidation.* Journal of bacteriology, 2002. **184**(6): p. 1722-1732.
192. Devreese, B., et al., *The Primary Structure of the Split‐Soret Cytochrome c from Desulfovibrio Desulfuricans ATCC 27774 Reveals an Unusual Type of Diheme Cytochrome C.* European Journal of Biochemistry, 1997. **248**(2): p. 445-451.
193. da Silva, S.M., I. Pacheco, and I.A.C. Pereira, *Electron transfer between periplasmic formate dehydrogenase and cytochromes c in Desulfovibrio desulfuricans ATCC 27774.* Journal of Biological Inorganic Chemistry, 2012: p. 1-8.
194. Pitcher, R.S., M.R. Cheesman, and N.J. Watmough, *Molecular and Spectroscopic Analysis of the Cytochromecbb 3 Oxidase from Pseudomonas stutzeri.* Journal of Biological Chemistry, 2002. **277**(35): p. 31474-31483.
195. Buschmann, S., et al., *The structure of cbb3 cytochrome oxidase provides insights into proton pumping.* Science, 2010. **329**(5989): p. 327-330.
196. Gao, H., et al., *Reduction of nitrate in Shewanella oneidensis depends on atypical NAP and NRF systems with NapB as a preferred electron transport protein from CymA to NapA.* The ISME journal, 2009. **3**(8): p. 966-976.
197. Friedrich, C.G., et al., *Prokaryotic sulfur oxidation.* Current opinion in microbiology, 2005. **8**(3): p. 253-259.
198. McLean, J.S., et al., *Oxygen‐dependent autoaggregation in Shewanella oneidensis MR‐1.* Environmental microbiology, 2008. **10**(7): p. 1861-1876.
199. Thöny‐Meyer, L., D. Ritz, and H. Hennecke, *Cytochrome c biogenesis in bacteria: a possible pathway begins to emerge.* Molecular microbiology, 1994. **12**(1): p. 1-9.

200. Kleingardner, J.G. and K.L. Bren, *Comparing substrate specificity between cytochrome c maturation and cytochrome c heme lyase systems for cytochrome c biogenesis.* Metallomics, 2011. **3**(4): p. 396-403.
201. Shi, L., et al., *The roles of outer membrane cytochromes of Shewanella and Geobacter in extracellular electron transfer.* Environmental microbiology reports, 2009. **1**(4): p. 220-227.
202. Dworkin, M., *Biology of the myxobacteria.* Annual Reviews in Microbiology, 1966. **20**(1): p. 75-106.
203. He, Q. and R.A. Sanford, *Acetate threshold concentrations suggest varying energy requirements during anaerobic respiration by Anaeromyxobacter dehalogenans.* Applied and Environmental Microbiology, 2004. **70**(11): p. 6940-6943.
204. Marshall, M.J., et al., *Electron donor‐dependent radionuclide reduction and nanoparticle formation by Anaeromyxobacter dehalogenans strain 2CP‐C.* Environmental microbiology, 2009. **11**(2): p. 534-543.
205. Sanford, R.A., et al., *Hexavalent uranium supports growth of Anaeromyxobacter dehalogenans and Geobacter spp. with lower than predicted biomass yields.* Environmental microbiology, 2007. **9**(11): p. 2885-2893.
206. Cardenas, E., et al., *Microbial communities in contaminated sediments, associated with bioremediation of uranium to submicromolar levels.* Applied and Environmental Microbiology, 2008. **74**(12): p. 3718-3729.
207. Xu, M., et al., *Responses of microbial community functional structures to pilot-scale uranium in situ bioremediation.* The ISME journal, 2010. **4**(8): p. 1060-1070.
208. Thomas, S.H., et al., *Unique ecophysiology among U (VI)-reducing bacteria as revealed by evaluation of oxygen metabolism in Anaeromyxobacter dehalogenans strain 2CP-C.* Applied and Environmental Microbiology, 2010. **76**(1): p. 176-183.
209. Chao, T.C., et al., *Comprehensive proteome profiling of the Fe (III)‐reducing myxobacterium Anaeromyxobacter dehalogenans 2CP‐C during growth with fumarate and ferric citrate.* Proteomics, 2010. **10**(8): p. 1673-1684.
210. Lochner, A., et al., *Label-free quantitative proteomics for the extremely thermophilic bacterium Caldicellulosiruptor obsidiansis reveal distinct abundance patterns upon growth on cellobiose, crystalline cellulose, and switchgrass.* Journal of proteome research, 2011. **10**(12): p. 5302-5314.
211. Yamada, T., et al., *iPath2. 0: interactive pathway explorer.* Nucleic acids research, 2011. **39**(suppl 2): p. W412-W415.
212. Sivashankari, S. and P. Shanmughavel, *Functional annotation of hypothetical proteins–A review.* Bioinformation, 2006. **1**(8): p. 335.
213. Punta, M., et al., *The Pfam protein families database.* Nucleic acids research, 2012. **40**(D1): p. D290-D301.
214. Himmelheber, D.W., et al., *Microbial colonization of an in situ sediment cap and correlation to stratified redox zones.* Environmental science & technology, 2008. **43**(1): p. 66-74.
215. Prakash, O., et al., *Geobacter daltonii sp. nov., an Fe (III)-and uranium (VI)-reducing bacterium isolated from a shallow subsurface exposed to mixed heavy metal and hydrocarbon contamination.* International journal of systematic and evolutionary microbiology, 2010. **60**(3): p. 546-553.
216. Nealson, K.H. and D. Saffarini, *Iron and manganese in anaerobic respiration: environmental significance, physiology, and regulation.* Annual Reviews in Microbiology, 1994. **48**(1): p. 311-343.
217. Myers, C.R. and K.H. Nealson, *Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor.* Science, 1988. **240**.

218. Natale, D.A., et al., *Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs).* Genome Biol, 2000. **1**(5): p. RESEARCH0009.
219. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic acids research, 2000. **28**(1): p. 27-30.
220. Richter, H., et al., *Cyclic voltammetry of biofilms of wild type and mutant Geobacter sulfurreducens on fuel cell anodes indicates possible roles of OmcB, OmcZ, type IV pili, and protons in extracellular electron transfer.* Energy & Environmental Science, 2009. **2**(5): p. 506-516.

Vita

Xiaoxin Liu was born and raised in Changsha, Hunan, China. She received her Bachelor of Science degree in Biological and Biomedical Sciences from Central South University, Changsha, China. In the year of 2009, she began her Ph.D study in the graduate school of Genome Science and Technology at the University of Tennessee, Knoxville. She expects to finish her dissertation work and receive a Ph.D degree in life sciences in August, 2014.