From Functional Metagenomics to Unique Synthetic Expression Strategies in Iron-Reducing Bacteria

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Tanhia Denys Gonzalez

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Daniel R. Bond and Michael J. Sadowsky

May 2012

Dedication

This dissertation is dedicated to my parents Carlos and Leticia, my brother and sister, Carlos and Ileana, and my husband, David.

Abstract

Cellulose and hemicellulose are renewable sources of fermentable sugars. The use of fermentable sugars for the production of alternative energy sources (i.e. ethanol, butanol, etc.) is an attractive solution to alleviate the shortage and high prices of petroleum. Cellulases and hemicellulases are the two groups of glycosyl hydrolases responsible for breaking down the polysaccharide component of biomass into their respective sugar moieties. The enzymatic hydrolysis of cellulose and hemicellulose has relied on enzymes originally produced by culturable organisms. This thesis describes the use of metagenomics coupled to high-throughput screening techniques to identify glycosyl hydrolases originally encoded by uncultured organisms. The findings of this thesis include the identification and biochemical characterization of a unique endoglucanase. Besides catalyzing the hydrolysis of soluble and insoluble cellulosic substrates, this endoglucanase exhibited a domain architecture that has not been previously reported in the literature. This thesis also describes two different strategies to engineer the surface of (Fe^{+3})-reducing bacteria. These expression systems are a valuable tool for studying the cellular respiration of *Geobacter* and *Shewanella.* Furthermore, they have practical applications in the area of whole-cell biocatalysis in microbial fuel cells. The first strategy involved using an autodisplay system to engineer the cell envelope of *Geobacter* and *Shewanella*. The autodisplay system translocated a functional β-galactosidase enzyme to the cell envelope of *G. sulfurreducens* and *S. oneidensis*. Furthermore, this system proved to be an effective tool for catalyzing reactions in electrochemical cells using biofilms of *G. sulfurreducens* cells. The second strategy exploited the use of in-frame fusions with the *c*-type cytochrome OmcZ to translocate a recombinant protein to the outer membrane and extracellular matrix of *Geobacter sulfurreducens*. This is the first time that the *c*-type cytochrome OmcZ has been used to engineer biofilms of *Geobacter sulfurreducens*.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1. Introduction

1. Introduction

1.1 Thesis overview

This thesis work is focused on the application of existent and novel biotechnological techniques to explore different aspects of microbial biocatalysis. This work has expanded our knowledge in the area of biocatalysis and it provides a genetic platform to answer fundamental questions in the microbial physiology of metal reducing bacteria.

Chapter 1

Chapter 1 is a comprehensive literature review about the three areas of research used in this thesis. The first part of the chapter describes the history, principles and applications of metagenomics. The second part of the chapter provides an overview about dissimilatory iron reducing bacteria and the challenges related to understanding the electron transfer mechanisms used by these microorganisms. A brief review of the model organisms used to study iron (Fe^{+3}) reduction is also provided. The last part of this chapter provides an overview of surface display systems and proposes using one of these systems as a tool for understanding fundamental questions regarding the cellular respiration of iron reducing bacteria.

Chapter 2

Chapter 2 describes a functional high-throughput screening of a soil metagenome for the identification of unique enzymes involved in cellulose and hemicellulose degradation. This chapter describes the identification of functional glycosyl hydrolases and the biochemical characterization of a unique endoglucanase.

Chapter 3

Chapter 3 describes a strategy to implement an existent autodisplay system for the expression of proteins on the surface of *E. coli*, *Shewanella oneidensis* and *Geobacter sulfurreducens*. Besides its relevant applications for studying extracellular electron transfer mechanisms in iron (Fe^{+3}) bacteria, this system represents a valuable tool for microbial fuel cell biocatalysis.

Chapter 4

Chapter 4 describes a novel strategy to engineer the surface of *Geobacter sulfurreducens*. This strategy exploits the use of *c*-type cytochrome OmcZ to translocate a recombinant protein to the outer membrane and extracellular matrix of *Geobacter sulfurreducens*. The surface of *Geobacter* cells can be exploited to catalyze redox reactions in microbial fuel cells.

Chapter 5

Chapter 5 summarizes the findings of this research and it discusses their possible implications in area of biotechnology. It also provides conclusive remarks and future directions.

2. Overview of Metagenomics

For over 200 years, the use of microscopy and culturing techniques has allowed microbiologists to describe a portion of the microbial population. These techniques were critical for creating the basis of our knowledge in modern microbiology. Additionally, these advances were crucial for the development of antibiotics and food products. Today, powerful techniques including metagenomics have expanded our knowledge of microbial diversity, as well as, some of the ecological roles of uncultured organisms (46).

2.1 History of Metagenomics; From Cultured to Uncultured Microorganisms

Some of the most significant breakthroughs made in the area of microbiology have relied on the development of technologies that allow scientists to take a closer look into the microbial world. This trend started with the pioneering work of Antonie van Leeuwenhoek in 1663. His powerful microscopy techniques gave the first glimpse of microbial diversity and allowed other microbiologists to make similar observations. In 1876, Ferdinand Cohn first used microscopy to classify bacterial cells, plants and algae (Figure 1). The pour plate technique developed by Koch in 1881, allowed microbiologists to isolate microorganisms in pure cultures. This technique was used to classify cultured from uncultured organisms. Those microorganisms that could not be grown using standard culturing techniques were considered uncultured microorganisms.

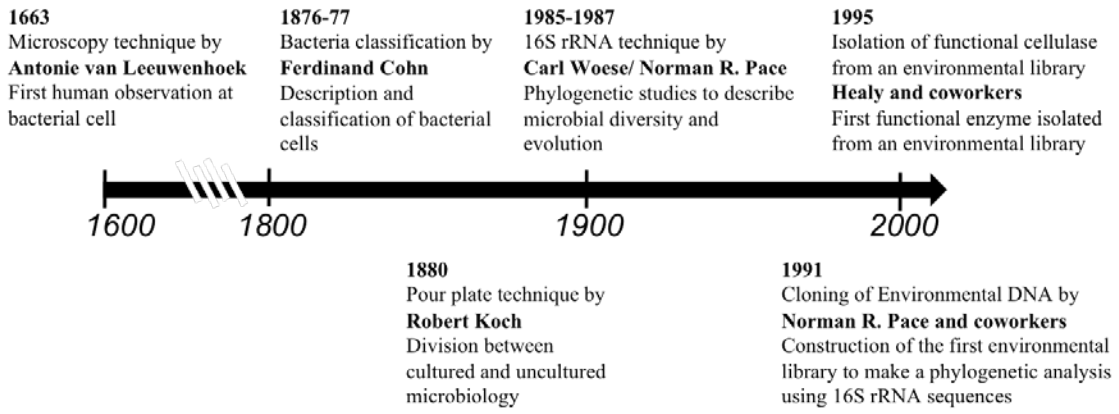


Figure 1. Historical milestones in microbiology and metagenomics

The studies focused on cultured organisms created the basis of scientific knowledge in modern microbiology, as well as, microbial physiology and genetics. However, there were discrepancies in the microbial populations observed through microscopy and plate count techniques. The largest discrepancies were found in soil and water. It was estimated than less than 1% of the cells observed in soil could be cultured under standard culturing techniques (166).

Carl Woose redefined bacterial evolution in 1985 by proposing the use of rRNA gene sequences for the identification of new species and predictions of phylogenetic relationships between them (182). The rRNA sequencing technique coupled with the advances in PCR technologies have allowed microbiologists to survey the microbial diversity of complex environments and to make better assessments about the fraction of organisms that remain unknown.

In 1991, Pace and coworkers constructed the first environmental library to study microbial diversity in soil samples (143). In a different study, Halley and coworkers identified a functional cellulase in a metagenomic library made from samples taken at an anaerobic digester (48). These studies marked the beginning of the metagenomic era.

2.2 Metagenomics; The Principle

Metagenomic studies access the total amount of genetic information contained in an environmental sample. This method provides the genetic information of uncultured microorganisms by overcoming the limitations imposed by standard culturing techniques. Metagenomics has been proven to be a powerful technique to study microbial diversity (167). Furthermore, this technique has accelerated the discovery of novel enzymes by accessing the biocatalytic capabilities of uncultured microorganisms. Currently, there are a wide variety of enzymes derived from soil and marine metagenomes including esterases, lipases, amylases, cellulases, etc. (30, 66, 94, 158).

Some of the first large metagenomic projects studied biofilms from an acid mine drainage and seawater samples from the Sargasso Sea (168, 173). Over the last 10 years, metagenomics has been used to explore the microbial diversity of a wide variety of environments ranging from soil to insect guts (29, 179).

The construction of a metagenomic library begins with the isolation of the total DNA from a particular sample (i.e. soil, water, etc.). Subsequently, the extracted DNA is

ligated to a suitable cloning vector and the resulting constructs are used to transform host cells (Figure 2). The construction of metagenomic libraries is not trivial since challenges can be found in isolating large DNA fragments, finding suitable cloning vectors and selecting expression hosts (29, 135). Additionally, metagenomics relies on sequence-based or functional-based screening assays to identify novel enzymes (29).

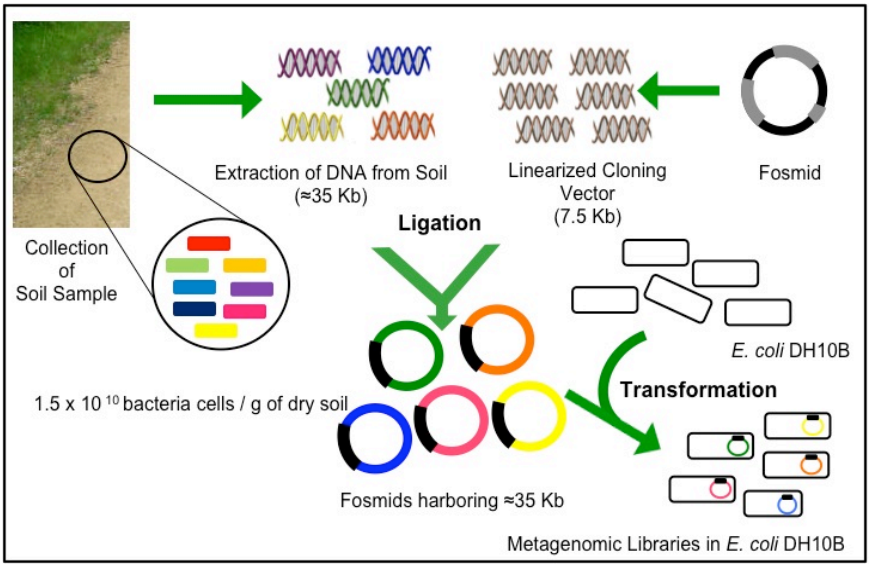


Figure 2. Construction of metagenomic libraries

2.3 High-Throughput Screening Analysis

Functional screening approaches rely on sensitive assays and the heterologous host expression. During this particular approach, genes have to be transcribed and the resulting RNA-fragments have to be translated into proteins. Furthermore, in some cases the resulting proteins have to be secreted in order to be identified by the screening assays. This approach does not rely on the homology to specific nucleotide sequences. Therefore,

one of the advantages of using functional screening strategies is the opportunity to identify novel genes.

The sequence-based screening approach relies on DNA sequencing from either full genomes or specific DNA-fragments. The identification of specific genes is obtained through nucleotide sequence comparisons, which limit the possibility of finding novel and functional genes. During the past few years, this approach has become popular due to the increasing number of low-cost and high-throughput DNA sequencing platforms. The sequence-based approach has been particularly successful during phylogenetic studies, since the resulting information can be used to infer the distribution and function of microbial populations along with the evolutionary relationships among them.

On the other hand, metagenomics coupled with high-throughput functional screenings has shown to identify novel and functional biocatalysts while also providing access to large amounts of genomic information (70, 87, 131, 158). Recent metagenomic studies have identified novel cellulases and hemicellulases in complex environments such as soil, cow rumen and insect guts (15, 34, 52, 70, 131). These studies represent only the beginning of the exploratory efforts, since the majority of the environments remain unexplored. The first hypothesis of this work was: 1) functional metagenomics could identify cellulases and hemicellulases from uncultured organisms living in soils that had been exposed to biomass for long periods of time.

2.4 Overview of Bioconversion of Cellulose and Hemicellulose

Cellulose and hemicellulose are renewable sources of fermentable sugars. The use of fermentable sugars for the production of alternative energy sources (i.e. ethanol, butanol, etc.) is an attractive solution to alleviate the shortage and high prices of petroleum. However, the bioconversion of these recalcitrant polysaccharides remains unfeasible due to the lack of efficient pretreatment technologies specialized on biomass conversion (106, 185).

The methods used for biomass conversion are usually grouped into thermochemical and biochemical methods (23). Thermochemical methods are not selective and they use high temperatures, high pressure and generate large amounts of waste products (23, 119). Biochemical methods are an attractive alternative for biomass conversion because they combine milder chemical pretreatments with highly selective enzymatic hydrolysis. However, one of the major challenges of enzymatic hydrolysis is the cost and availability of large quantities of highly specific enzymes to accomplish the biomass conversion (54). Therefore, there is a great interest in discovering novel and more effective enzymes to optimize the conversion of these recalcitrant substrates.

2.5 Biomass Composition

Cellulose and hemicellulose are the two major polysaccharides found in lignocellulose, agricultural residues and energy crops (Table 1). Cellulose comprises between 40% and 50% of the total dry weight of wood while hemicellulose represents between 25% and

35% of the total weight. Cellulose is a recalcitrant biopolymer constituted of unbranched chains of glucose monomers linked together by a β-1,4-glycosidic bonds. This biopolymer can be found in a crystalline and amorphous state. Crystalline cellulose is composed of highly organized cellulose chains inaccessible to enzymatic hydrolysis. In contrast, the disorganized chains found in amorphous cellulose are more accessible to enzymatic hydrolysis than its crystalline counterpart (113, 132).

Hemicellulose is commonly found associated through hydrogen bonds with cellulose. This polysaccharide is composed by an assortment of sugars including xylose, galactose, mannose, and arabinose. Xylan is the major hemillulosic component of hardwood while the major component of softwood is galactoglumannan (113, 132).

Table 1. Carbohydrate compositions of feedstocks

Biomass Type	Cellulose* (%)	Hemicellulose* (%)	References
Lignocellulose (Hardwoods and Softwoods)	40-50	25-35	(132)
Switchgrass	32	36	(63)
Corn Stover	35-37	21-28	(1)
Sweet Sorghum	26	20	(138)

*All percentages are based on dry weight basis

2.6 Enzymatic Hydrolysis of Cellulose and Hemicellulose

Cellulases are the group of enzymes that synergistically hydrolyze the β-1,4-glycosidic bonds of cellulose. They are classified by their mode of action into endoglucanases (EC 3.2.1.4), exoglucanses/cellobiohydrolases (EC 3.2.1.91) and glucosidades (EC 3.2.1.21). Endoglucanses hydrolyze random β-1,4-glycosidic bonds of cellulose producing reducing and non-reducing ends. Cellobiohydrolases hydrolyze reducing or non-reducing ends of cellulose chain releasing cellobiose. β-Glucosidases hydrolyze cellobiose into glucose monomers (188). Endoxylanases are the major group of hemicellulases and they specialize in the hydrolysis of β-1,4-xylopyranose residues of xylan (147).

2.7 Glycosyl Hydrolases Classification

Glycosyl hydrolases (GH) are classified based on their amino acid sequence and structure similarity by the Carbohydrate-Active Enzyme (CAZy) database. GH are usually found associated with a single or multiple carbohydrate binding modules (CBM) at their N- or C-terminal domain. In some cases, the presence of CBM has shown to improve glycosyl hydrolase activity by facilitating the binding to specific substrates and/or disrupting hydrogen bonds to improve substrate accessibility (12, 140).

Currently, there are more than 100 different families of structurally related glycosyl hydrolases and more than 50 different families of CBM. According to the CAZy database, there are at least 17 different families of endoglucanases (EC 3.2.1.4). Some of these families have more than 300 characterized enzymes (i.e. GH family 5), while others

have not been extensively studied (i.e GH family 44 and 74). GH family 10 and family 11 together have close to 400 characterized xylanases. GH family 43 has at least 50 characterized enzymes including xylosidases, arabinofuranosidases and xylanases.

3. Overview of Dissimilatory Iron (Fe^{+3}) Reduction

Cellular respiration is the fundamental process for generating biochemical energy to support cell growth. During microbial respiration, low potential donors are oxidized (i.e. sugars and organic acids) and high-potential acceptors (i.e. oxygen and Fe(III)) are reduced. This set of redox reactions generates a flow of electrons that results in the generation of biochemical energy. The flow of electrons moving through the cell membranes (electron transport chain) generates energy. The electron transport chain is facilitated by electron carrier molecules such as cytochromes and quinones. Microbial respiration is usually classified as aerobic or anaerobic. Aerobic respiration uses oxygen as a terminal electron acceptor while anaerobic uses a terminal acceptor other than oxygen. Alternative electron acceptors include nitrate (NO_3^-), sulfate (SO_4^{-2}) and iron (Fe^{+3}).

Iron is an abundant and widely distributed element on earth. This transition element commonly exists in +2 and +3 oxidation states. Under aerobic and circumneutral pH conditions, soluble ferrous iron (Fe^{+2}) is readily oxidized to form insoluble iron oxide (Fe^{+3}). Ferrous iron (Fe^{+2}) is only stable under anoxic or low pH conditions. The microbial reduction of iron oxide (Fe^{+3}) or iron respiration occurs in a wide range of anoxic environments including sediments and pristine aquifers. The microbial metabolism of iron reducers facilitates the cycling of metals and ultimately carbon in the environment (100, 102). Furthermore, the microbial reduction of metal oxides has practical applications in the area of bioremediation and electricity production (95, 96).

3.1 Introduction to Dissimilatory Metal Reducing Bacteria (DMRB)

Dissimilatory metal reducing bacteria are characterized by their ability to couple the oxidation of organic acids with the reduction of metal oxides such as iron (Fe^{+3}) and manganese (Mn^{+4}) (97, 150). Dissimilatory iron reduction was initially studied in fermentative iron (Fe^{+3}) reducers (98, 102). These organisms do not conserve energy from iron (Fe^{+3}) reduction and they transfer most of their electrons to fermentation products such as ethanol and hydrogen. In 1988, Lovley and Phillips identified the first organism that completely oxidized acetate to CO_2 using iron (Fe^{+3}) as a sole electron acceptor (103). This organism was called GS-15 and later named *Geobacter metallireducens* (99). Since then, a wide variety of dissimilatory iron reducers (Fe^{+3}) have been identified in the bacteria and archaea domains (100).

3.2 Electron Transfer Mechanisms of Dissimilatory Metal Reducing Bacteria

The cellular respiration of dissimilatory metal reducing bacteria relies on sophisticated electron transfer mechanisms to support the flow of electrons across membranes and into insoluble terminal electron acceptors. There have been four different strategies identified for extracellular electron transfer in biological systems. These strategies can be classified as direct or indirect electron transfer. During direct electron transfer, redox molecules have a direct contact with the terminal electron acceptors. In contrast, the indirect electron transfer mechanisms utilize cell filaments, shuttling molecules or metal chelators/siderophores to transfer electrons to the terminal electron acceptors.

The direct electron transfer mechanism is characteristic of *Geobacter* species. Previous biochemical studies have shown that iron reduction only occurs when *G. sulfurreducens* and *G. metallireducens* cells have direct contact with the surface iron oxide (Fe^{+3}) (91, 126). Furthermore, electrochemical studies have suggested that *Geobacter sulfurreducens* do not produce shuttling molecules when electrodes are used as terminal electron acceptors (10).

The indirect electron transfer assisted by shuttling molecules has been previously reported in *Shewanella oneidensis* and *Geothrix fermentans*. Biochemical and electrochemical techniques have demonstrated that *S. oneidensis* secretes flavin compounds that function as electron shuttles between the cell surface and the insoluble electron acceptors (107). The presence of shuttling molecules has also been suspected in *G. fermentas* cultures grown with either insoluble iron oxide (Fe^{+3}) or electrodes as terminal electron acceptors (11, 127).

3.3 Three-Electrode Electrochemical Cells

Electrochemical cells exploit the microbial metabolism as catalysts for the conversion of chemical energy to electricity. Besides their application for current production, electrochemical cells have been a powerful device for studying extracellular electron transfer mechanisms in *Geobacter* and *Shewanella* species. Some of these studies have been performed in three-electrode electrochemical cell (Figure 3).

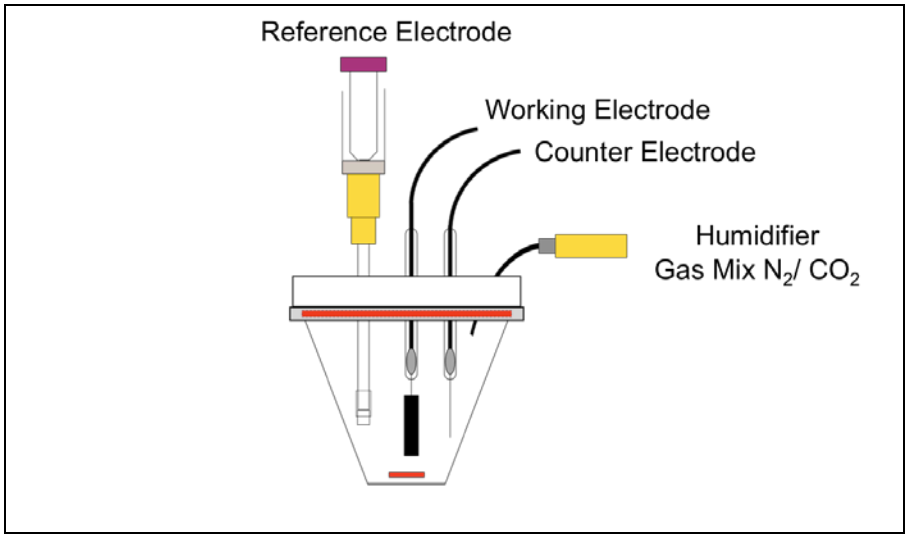


Figure 3. Three-electrode electrochemical cell (Figure extracted from Marsili et al. 2008)

Three-electrode electrochemical cells have a working electrode, a counter electrode and a reference electrode in a single anaerobic chamber. The working electrode functions as the terminal electron acceptor for bacteria. Poised electrodes are working electrodes that have been set to a potential that favors the occurrence of a particular redox reaction. The function of the reference electrode is to control the potential of the working electrode by using a potentiostat.

The counter electrode is an auxiliary electrode that supports the continuous flow of electrons from the working electrode. This particular set-up allows running electrochemical techniques with poised working electrode such as cyclic voltammetry (CV) and chronoamperometry (CA) (Figure 4).

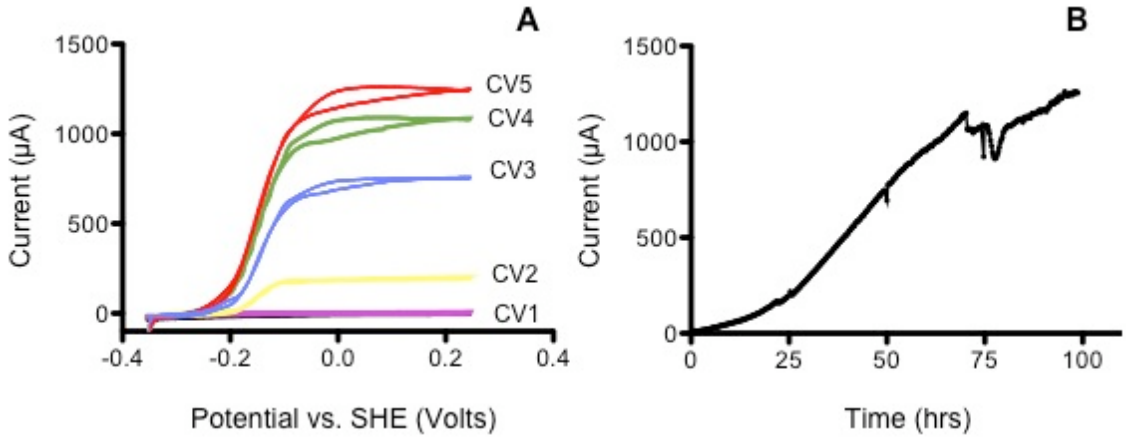


Figure 4. Cyclic voltammetry (A) and chronoamperometry (B) techniques applied to a biofilm of wild type *G. sulfurreducens* cells grown on a poised electrode. The cyclic voltammograms were taken every 24 hrs (CV1= 0 hrs, CV2= 24 hrs, CV3= 48 hrs, CV4= 72 hrs, CV5=96 hrs). Potential was calculated against a standard hydrogen electrode (SHE).

Cyclic voltammetry and chronoamperometry have previously been used to characterize the redox properties of proteins and microbial biofilms (4, 10, 108). During cyclic voltammetry, the amount of current produced by a biofilm is determined across a linear

range of potentials (Figure 4A). The range of potentials is typically scanned in both directions. One of the curves represents the flow of electrons being transferred into the electrode (reduction) and the reverse represents the electron flowing back into the protein (oxidation). The reversibility of the redox reaction can be determined by comparing the oxidation and reduction curves. Curves with similar midpoint potentials imply that redox reactions are reversible. During chronoamperometry (CA), the amount of current produced by a poised electrode is determined over time (Figure 4B). This curve reflects the amount of electrons transferred into the electrode over time.

3.4 *Geobacter*

Geobacter species were the first identified group of organisms to conserve energy from the complete oxidation of acetate using iron (Fe^{+3}) as a sole electron acceptor. This group δ-proteobacteria belongs to the *Geobacteraceae* family and they have been isolated from a wide range of anoxic environments including aquatic sediments, contaminated aquifers and deep subsurfaces (100). Other members of *Geobacteraceae* family include species of *Desulfuromonas* and *Pelobacter*. The study of members in the *Geobacteraceae* family has been favored over other iron reducers due to their relative abundance in places undergoing continuous iron (Fe^{+3}) reduction (53, 160). The availability of pure isolates and/or a genetic system have also favored the study of *Geobacter* species (27).

Along with acetate oxidation coupled to iron reduction (Fe^{+3}), *Geobacter* species have also been found to use alternative electron donors and acceptors to support cell growth.

For example, *G. metallireducens* has been found to oxidize additional organic acids (i.e. propionate and butyrate) aromatic compounds (i.e. toluene and phenol) and alcohols (i.e. ethanol) (99, 101). Alternative electron acceptors include other metals such as manganese (Mn^{+4}) and uranium (U^{+6}) (99).

Geobacter sulfurreducens has been one of the most extensively studied members of the *Geobacteraceae* family. This organism was isolated from sediments contaminated with hydrocarbons in Norman, Oklahoma (19). *G. sulfurreducens* couples the oxidation of acetate or hydrogen with the reduction of iron oxide (Fe^{+3}). The alternative electron acceptors used by *G. sulfurreducens* include ferric (Fe^{+3}) citrate, cobalt (Co^{+3}) and electrodes. *G. sulfurreducens* has been exploited in microbial fuel cells for harnessing electricity from sediments (96). Furthermore, biofilms of *G. sulfurreducens* cells grown on poised electrode surfaces have been a platform to study the electron transfer mechanism used by this organism.

The mechanism used by *Geobacter sulfurreducens* to transfer electron across membranes and into insoluble electron acceptors are not fully understood. However, it has been demonstrated that *c*-type cytochromes and other redox active proteins play a significant role during the reduction of insoluble substrates. The genome of *Geobacter sulfurreducens* contained more than 100 putative *c*-type cytochromes (33).

c-type cytochromes are redox active proteins that function as electron carriers. The prosthetic group of a cytochrome is called heme. In contrast to other cytochromes, the heme groups in *c*-type cytochromes are covalently bound via thioether bonds to the cysteine residues of the protein. Cytochromes exploit the reversible oxidation state of the metal iron ion ($Fe^{+2} \leftrightarrow Fe^{+3}$) in heme groups to transport electron across several angstroms (Å) in biological structures.

The role of *c*-type cytochromes during reduction of insoluble electron acceptors has been studied in *G. sulfurreducens*. For example, PpcA is a *c*-type cytochrome that functions as an intermediate electron carrier during reduction of insoluble iron (Fe^{+3}) and other insoluble electron acceptors (92). Based on its function, it has been predicted that PpcA is localized in the periplasm of *G. sulfurreducens* cells. Δ*ppcA* mutants have shown to be deficient in the reduction of insoluble electron acceptors while reducing soluble electron acceptors at wild type rates (92).

The expression of outer membrane *c*-type cytochromes has also shown to be essential for extracellular electron transfer. For example, *G. sulfurreducens* Δ*omcB*, Δ*omcS* and Δ*omcZ* mutants exhibited a defect reducing insoluble electron acceptors but not a defect reducing soluble acceptors. OmcB is an 87-kDa dodeca-heme *c*-type cytochrome localized in the outer membrane of *G. sulfurreducens* (133). The function of OmcB has been predicted to be as terminal iron (Fe^{+3}) reductase (82).

Other outer membrane proteins have shown to be directly or indirectly involved with the electron transfer mechanisms. For example, the expression of OmpB, a multicopper oxidase localized on the cell surface of *G. sulfurreducens,* has been shown to be required for reducing insoluble iron (Fe^{+3}) (114, 133). OmpJ, an abundant outer membrane porin, has shown to be essential for proper localization of *c*-type cytochromes in the outer membrane and ultimately the ability to reduce insoluble electron acceptors (2).

The study of extracellular electron transfer mechanisms in *Geobacter* species remains challenging. Even though the genetic system of *G. sulfurreducens* has allowed studying the phenotypes of mutants lacking the expression of cytochromes and other proteins, these systems seem to be dependently regulated. For example, the expression of MacA, a periplasmic *c*-type cytochrome of *G. sulfurreducens*, was originally shown to be essential for reduction of insoluble iron (18). Later studies suggested that Δ*macA* mutants could recover the iron reduction (Fe^{+3}) phenotype with *in trans* expression of the *omcB* gene. This study demonstrated the lack of transcription of the *omcB* gene and lack of OmcB protein in Δ*macA* mutants (68). However, it remains unknown what is the transcriptional regulatory mechanism for the expression of these two proteins. Other regulatory mechanisms have been observed between *c*-type cytochromes. For example, Westernblot and Northern blot analysis identified that *G. sulfurreducens* Δ*omcG* and Δ*omcH* mutants posttranscriptionally regulated the expression of OmcB (67).

Besides the challenge of transferring the electron across membranes and into insoluble electron acceptors, *G. sulfurreducens* cells have to find mechanisms to transfer electrons between cells and across long distances. Studying these mechanisms has been essential for understanding how biofilms of *G. sulfurreducens* cells are capable of respiring on anodic surfaces. Some of the few elucidated mechanisms also rely on *c*-type cytochromes. For example, OmcS is a predicted 50-kDa hexa-heme *c*-type cytochrome associated along the pili of *G. sulfurreducens* (83, 115). It has been speculated that the presence of OmcS on pili facilitates long-range electron transfer from pili to insoluble electron acceptors (115). However, there is not direct evidence supporting that particular theory other than the spatial localization of OmcS along the pili.

There is a recent theory proposing that the extracellular electron transfer mechanisms observed in current-producing biofilms are supported by tunneling electrons through redox active centers (161). This theory contradicts the model suggesting the presence of conductive biological filaments (105). The electron exchange theory agrees with the fact that the presence of outer membrane *c*-type cytochrome OmcZ is essential for the formation of high density current-producing biofilms on anodes (125).

The studies leading to the identification of OmcZ were based first on the observation that cyclic voltammograms of a biofilms composed of *G. sulfurreducens* Δ*omcZ* mutant cells showed significant differences to those produced by wild type and other *c*-type cytochrome *G. sulfurreducens* mutants (Δ*omcB,* Δ*omcE* and Δ*omcST*). These studies

suggested that Δ*omcZ* mutant cells grown on an electrode surface imposed a greater resistance than wild type cells (134). Furthermore, later studies observed that the transcription of the *omcZ* gene increased 5.7-fold when *G. sulfurreducens* cells were grown using an anode as terminal electron acceptor (125). Considering the importance of OmcZ in current producing biofilms of *G. sulfurreducens*, there have been strong efforts on characterizing this protein further. Immunogold labeling studies performed on anodic biofilms of *G. sulfurreducens* cells determined that OmcZ was localized in the extracellular matrix formed throughout the biofilm (56, 136).

OmcZ is a predicted octaheme *c*-type cytochrome comprised by 473 amino acids. Biochemical studies have identified the presence of two protein populations of OmcZ on the outer membrane and extracellular matrix of *G. sulfurreducens* cells harvested from current-producing anodic biofilms (56, 125). These two OmcZ populations have predicted molecular weighs of 30-kDa and 50-kDa. N-terminal sequences have indicated that 30-kDa protein is a truncated version of the full-length 50-kDa OmcZ (57).

The 30-kDa OmcZ form is composed of 282 amino acids including a signal peptide and eight predicted heme c-binding motifs. Since antibodies have only been raised for the 30-kDa form of OmcZ, the localization of the remaining 191 amino acids has only been identified in the 50-kDa OmcZ form. The processing site of OmcZ is found between residues 328 and 329 (57).

The sequence of the remaining 191 amino acids is predicted to contain a glycosyl hydrolase family domain from family 13 (GH13). More than 20 different functions have been reported for proteins harboring a GH13 domain including amylases (EC 3.2.1.1), pullanases (EC 3.2.1.41), and cyclomaltodextrin glucanotransferases (EC 2.4.1.19).

Even though the 30-kDa form has shown to be more abundant than the 50-kDa form of OmcZ in the outer membrane of wild type *G. sulfurreducens*, this protein distribution has only been confirmed with *G. sulfurreducens* cells grown on acetate-fumarate conditions. Since the *omcZ* gene has shown to be differentially transcribed in cells harvested from anodic biofilms compared to cells grown on fumarate conditions, further studies need to be conducted to determine which OmcZ form is the most abundant in anodic biofilms of *G. sulfurreducens* cells.

3.5 *Shewanella*

Dissimilatory iron (Fe^{+3}) reduction has been intensively studied in *Shewanella*. This group of facultative anaerobes can readily be isolated from marine and freshwater environments. The carbon requirements of *Shewanella* are different than those of *Geobacter*. *Shewanella* partially oxidize lactate and pyruvate to acetate. Besides oxygen, *Shewanella* species are capable of using a wide range of electron acceptors including organic acids (i.e. fumarate), insoluble metals (i.e. iron (Fe^{+3}) and manganese (Mn^{+4})) and sulfur compounds (i.e. sulfur and dimethyl sulfoxide) (124). The ability of using many different kinds of electron acceptors is one of the most unique features of

Shewanella species. Furthermore, *Shewanella* species have shown to have a different mechanism for reducing insoluble metals than other metal reducers.

Shewanella oneidensis MR-1 has been the model organism for studying extracellular respiration in *Shewanella* species. Previous reports have shown the complexity of the transmembrane and extracellular pathways utilized by *S. oneidensis* for reducing insoluble electron acceptors (28, 150, 151). This respiratory pathway is also known as metal respiratory system (Mtr) and it facilitates the transfer of electrons across the cell envelope and into soluble and insoluble electron acceptors.

The electron conduit formed by multi-heme *c*-type cytochromes (CymA, MtrA, and MtrC) and an integral outer membrane protein (MtrB) has shown to be essential for supporting the of flux electrons from quinone pools, across membranes and into soluble and insoluble electron acceptors (28). MtrF, a recently crystallized decaheme *c*-type cytochrome, is an example of the highly sophisticated redox-active structures localized in the cell surface of *Shewanella* to facilitate extracellular electron transfer between cells and into insoluble electron acceptors (24). The size and spatial organization of the heme groups in MtrF are suspected to contribute to the direct electron transfer mechanisms of *S. oneidensis* (24).

Electron shuttling is the main extracellular respiration pathway identified in *S. oneidensis.* Biochemical and electrochemical studies have shown that *S. oneidensis*

secretes flavin compounds as an electron shuttling molecule (107, 177). This shuttling mechanism facilitates the transport of electrons beyond cell membranes and to insoluble electron acceptors (17).

3.6 Relevance of Studying Electron Transfer Mechanisms in *Geobacter* and *Shewanella* species

The electron transfer mechanisms used by *Geobacter* and *Shewanella* denote fundamental differences in their cellular respiration. The Mtr pathway in *Shewanella* illustrates a more definitive pathway for moving electron across membranes. Meanwhile, the electron transfer mechanisms in *Geobacter* still remain unknown.

Considering the importance of the cell surface of *G. sulfurreducens* and *S. oneidensis,* it is critical to develop genetic tools for understanding the fundamentals of extracellular electron transfer mechanisms in DMRB. Currently, there are not available expression systems to engineer the membrane and extracellular compartment of *G. sulfurreducens* and *S. oneidensis*. Thus, the second hypothesis of this work was: 2) surface display expression systems could be used to engineer the surface of *G. sulfurreducens* and *S. oneidensis*. Furthermore, there is a potential application of the redox active space created throughout the microbial biofilm in redox biotechnology. The next section of this chapter reviews the history and principles of current surface display systems.

4. Overview of Current Surface Display Systems in Gram-Negative Bacteria

4.1 Surface Display Principle

Cell surface display is a biological process in which proteins or short peptides are transported across the biological membranes and exposed to the cell surface. The use of recombinant DNA technologies have allowed using this naturally occurring process to develop biological tools for displaying heterologous proteins at the surface of phages, bacteria and yeast (21, 120, 154). For over twenty years, cell surface display systems have been extensively studied and optimized for the expression of heterologous proteins particularly in bacteria (170). Surface display in bacteria has been of special interest because of its multiple applications in the area of biocatalysis, bioremediation, biosensors, drug delivery, and others (84, 184).

The principle of the surface display systems involves using a membrane-associated protein, which functions as a carrier, fused to a passenger protein to be anchored on the surface of the host. For simplicity, the surface display systems have been classified by the nature of the carrier protein.

Some of the most common surface display systems utilize outer membrane proteins, surface appendages, lipoproteins or autotransporters as carrier proteins (142, 170). In gram-negative bacteria, the passenger protein has to be translocated from the cytoplasm to the periplasmic space and then it has to be secreted throughout the outer membrane and into the cell surface. The performance of the surface display systems can be

influenced by the secretion system used by the carrier protein, its abundance and by the size of the passenger protein.

4.2 Cell Surface Display Systems

The history of surface display systems began in 1985, when George Smith developed a bacteriophage displaying a fragment of EcoRI endonuclease through a fusion with the coat protein pIII. His pioneering work resulted in the development and commercialization of the technology known today as phage display (154, 155). Even though this system was limited to the display of short peptides, the idea of targeting peptides to particular biological compartments using protein fusions was a breakthrough. Since then, an extensive number of protein carriers have been investigated including outer membrane proteins, lipoproteins, surface appendages and autotransporters (170).

4.3 Outer Membrane Proteins

Freudl and co-workers (1986) and Charbit and co-workers (1986) reported the use of two bacterial outer membrane proteins, OmpA and LamB, respectively, to display short peptides on the surface of *E. coli* (21, 40). This was the first time that bacterial outer membrane proteins were used as a carrier to display peptides on the cell surface.

The systems based on OmpA have shown success in displaying short peptides and small proteins on the surface of bacteria. However, they are hindered by the size of the passenger protein and constraints imposed by the abundance of these proteins in the

bacterial membranes. Stathopoulos and coworkers (1996) reported the inability of OmpA to display alkaline phosphatase (PhoA), a protein with a molecular weight of 90-kDa, on the surface of *E. coli* (157). The LamB system has also been used to display other proteins and peptides in bacteria (16, 162). Other outer membrane proteins have been used to display proteins at the surface of bacteria including OmpX, OmpC, OprR and most recently PgsA (74, 123). The PgsA is a transmembrane protein from *Bacillus subtilis* that has shown to display proteins of up to 78-kDa on the surface of *E. coli* (123).

4.4 Lipoproteins

Lipoproteins have been another class of carrier proteins used in surface display systems. These proteins are characterized for being covalently bound to the lipid moieties of the bacterial outer membrane. The study of lipoproteins in bacteria started in 1969 when Braun and coworkers identified an abundant peptidoglycan associated lipoprotein (PAL) in the outer membrane of *E. coli* (14). In 1992, Francisco and coworkers optimized the OmpA surface display system by exchanging the native secretion signal of OmpA for the secretion signal of Braun's lipoprotein (38). Since then, this system has been used extensively in *E. coli* to display proteins such as β-lactamase, cellulase, cellulose and chitin binding domains, and a green fluorescent protein (38, 39, 148, 178).

The ice nucleation protein (INP) is an alternative lipoprotein carrier used for cell surface display. INP is encoded by plant pathogenic bacteria including *Pseudomonas, Erwinia* and *Xanthomonas*. The INP is responsible for catalyzing the ice formation of supercooled

water (183). The ice nucleation activity is associated with the outer membrane of the bacteria (90). The formation of ice crystals promotes the frost damage of the plant tissue thus facilitating the infection of the pathogenic bacteria (89). INP is composed by three distinctive domains. The N-terminal domain is partially hydrophobic and it remains attached to the surface via glycosyl-phosphatidylinositol anchoring. The central domain contains a series of repetitive elements and it has been found to be responsible for the ice nucleating activity. The C-terminal domain is highly hydrophilic and it has been employed as a template for fusion of heterologous proteins (65). Full-length and truncated versions of the INP have been used to display proteins such as a carboxymethylcellulase, a levansucrase, an organophosphorous hydrolase and a green fluorescent protein (61, 62, 86, 152). In comparison to previous outer membrane-based surface displays system, the INP is not hindered by the size of the passenger protein nor by the membrane stability (86, 186).

4.5 Surface Appendages

Surface appendages in bacteria such as flagella, fimbriae and pili have been an alternative cell surface display system for short peptides (73, 181). Majander and coworkers (2005) demostrated that FliD and FliC, two major proteins with an important role during flagellar polymerization and assembly, could be used to display short peptides at the surface of *E. coli* (104, 164).

4.6 Autotransporters

The autodisplay system is the most popular mechanism exploited to display proteins on the surface of bacteria. This system is based on using autotransporter proteins to facilitate the transport of heterologous proteins across bacterial membranes. The secretion of autotransporter proteins is characteristic of the model type V secretion system.

Autotransporters contain three distinct domains. The N-terminal domain contains the leader sequence responsible for signaling the protein translocation across the inner membrane. The passenger domain encodes the protein to be displayed on the cell surface. The helper domain forms a pore throughout the outer membrane and facilitates the secretion of the passenger protein across the outer membrane.

Autotransporters contain a secretion-dependent signal sequence (sec) at their N-terminal domain. However, it remains unknown if there are additional secretory pathways or accessory proteins involved during the translocation of the autotransporters to the cell surface.

The immunoglobulin A1 protease (IgA1) of *Neisseria gonorrhoeae* was discovered in 1984 and it was the first autotransporter protein to be described and characterized (45, 72). The IgA1 is synthesized as a 1500 amino acid preprotein. This autotransporter is composed by four domains including a protease domain flanked by a leader peptide and an α-domain ($IgA1_{\alpha}$) followed by a carrier β-domain ($IgA1_{\beta}$) (72). The translocation unit

$IgA1_{\beta}$, has been used to display proteins including single chain antibody fragments (scFv) and metal binding proteins (169, 172).

The adhesin involved in diffused adherence (AIDA-I) of enteropathogenic *E. coli* (EPEC) was discovered in 1989 (9). However, it was not until 1995 that it was classified as an autotransporter due to its structural similarity with the IgA1 protein. The AIDA-I is composed of a 78-kDa passenger protein and a 45-kDa carrier protein. Jochem and coworkers (1997) developed an autodisplay system using the 45-kDa carrier domain of AIDA-I to display a 13-kDa cholera subunit (110, 111). Since then, this system has shown to display a wide variety of proteins including a β-lactamase, a sorbitol dehydrogenase and an organophosphorus hydrolase (60, 81, 85). The AIDA-I autotransporter has been used extensively because it is produced natively in *E. coli.* Furthermore, the AIDA-I system has shown to display proteins of higher molecular weight than other carrier proteins (59, 170).

In order to develop tools for understanding the electron transfer mechanisms of *S. oneidensis* and *G. sulfurreducens*, Chapter 3 and Chapter 4 describes two independent strategies for engineering the surface of these organisms. The first approach exploits the use of the existent autodisplay system based on the AIDA-I autransporter. The second approach describes a unique strategy based on outer membrane *c*-type cytochromes.

Chapter 2. Identification and Characterization of a Unique Glycosyl Hydrolase from the Soil Metagenome of Minnesota

Metagenomics has shown to be a powerful technique for uncovering the genomes of uncultivated microorganisms and their potential biocatalytic capabilities. In this study, metagenomic DNA from soil was used to construct four genomic libraries. Soil samples were collected at different Minnesota locations continuously exposed to biomass. The metagenomic libraries were comprised of 139,000 fosmid clones in *E. coli.* The clones had an average insert size of 35-Kb and together they contained approximately 4.8-Gbp of DNA. Clones were screened for enzymatic activity by using colorimetric assays based on insoluble substrates, including cellulose (A-HE), xylan (A-XYL) or arabinoxylan (A-ARA) cross-linked with an azo dye and the soluble substrates carboxymethyl cellulose (CMC), 4-methylumbelliferyl-β-D-cellobioside (MUC) and 4-methylumbelliferyll-β-xylopyranoside (MUX). Forty-five clones hydrolyzed one or more of the cellulosic substrates. Four clones were selected for further studies since they exhibited hydrolytic activity towards insoluble substrates or had unique phenotypes. The DNA-inserts of selected clones were fully sequenced and the genes responsible of putative activity were identified through random transposon mutagenesis. While some glycosyl hydrolases were related to enzymes previously identified using similar metagenomic approaches, a hydrolase, designated CelK25, showed a unique domain architecture and its activity was comparable to an endoglucanase expressed by *Clostridum thermocellum*. The CelK25 was a 90-kDa glycosyl hydrolase and was comprised of a carbohydrate binding domain from family 9 (CBM9) fused to an accessory Immunoglobulin (Ig)-like domain and a catalytic glycosyl hydrolase domain from family 9 (GH9). The CBM9 has been exclusively found in bacterial xylanases, and this is the first report of its association with

a GH9 domain. The CelK25 hydrolyzed soluble (CMC, MUC and pNPC) and insoluble (A-HE) cellulosic substrates, and it exhibited the highest specific activity at pH 5 and at 40°C. The specific activity of CelK25 was 389 (U/μmol of protein) and the presence of 10mM $CaCl_2$ did not affect the hydrolytic activity. A truncated protein (CelK25-GH9), devoid of CBM9, was used to determine the influence of this domain in the hydrolytic activity of CelK25. The absence of CBM9 resulted in a loss of 40% of hydrolytic activity towards the soluble substrates, while the hydrolysis of the insoluble substrate increased 90% compared to the full-length CelK25. Results of these studies indicate that metagenomic analyses, along with functional screening approaches are powerful tools to identify novel unique cellulases with novel domain architectures that may be encoded by unculturable microorganisms.

1. Introduction

The use of cellulosic biomass for the production of alternative energy sources (i.e. ethanol, butanol, etc.) remains unfeasible due to the lack of efficient pretreatment technologies specialized on biomass conversion (106, 185). In 2007, 23% of the total production cost of bioethanol was attributed to chemical pretreatment and biochemical hydrolysis of biomass (44, 54). Therefore, there is a great opportunity to reduce the cost of bioethanol production by improving the conversion of biomass to its respective monosaccharide moieties (54).

Cellulose and hemicellulose are the two major polysaccharides found in lignocellulose, agricultural residues and energy crops. Cellulose is a recalcitrant biopolymer comprised of unbranched chains of glucose monomers linked together by a β-1,4-glycosidic bonds (132). Hemicellulose from hardwoods and grasses is mainly composed of xylan, while hemicellulose from softwoods is rich in glucomannans (51, 113, 132, 185).

Both thermochemical and biochemical methods are used for biomass conversion (23). Thermochemical methods are not selective and they use high temperatures, high pressure, and generate large amounts of waste products (23, 119). Biochemical methods are an attractive alternative for biomass conversion because they combine milder chemical pretreatments with highly selective enzymatic hydrolysis. One of the major challenges in enzymatic hydrolysis, however, is the cost and availability of large quantities of highly specific enzymes to accomplish biomass conversion (54).

Cellulases and hemicellulases are the major glycosyl hydrolases (GH) responsible for breaking-down the polysaccharide components of biomass. Cellulases are specialized in hydrolyzing the β-1,4-glycosidic bonds of cellulose chains. These enzymes are usually grouped according to their mode of action into endoglucanases, exoglucanases/ cellobiohydrolases and glucosidases (188). Endoxylanases are the most predominant group of hemicellulases and they are specialized in the hydrolysis of β-1,4-xylopyranose residues of xylan (147).

Over the last several years, improvements in biochemical hydrolysis have come especially from engineered enzymes produced by culturable organisms such as *Trichoderma reesei* and *Aspergillus nige*r (153, 188). Therefore, there is a strong interest in the discovery of novel and more effective enzymes from unculturable microorganisms to lower the cost of the biochemical pretreatment of cellulosic biomass (54).

The discovery of novel biocatalysts has been hindered by the lack of tools for culturing microorganisms in the laboratory, and our lack of understanding of the genomes of uncultivated microorganisms. Over the past decade, however, novel molecular techniques, together with more economical sequencing technologies, have facilitated access to the genomic information of unculturable microbial communities (55, 159). Besides providing information about microbial diversity, this genomic information carries unlimited possibilities for discovering novel biocatalysts.

Metagenomics approaches have been successfully employed to discover novel biocatalysts in complex environments ranging from soil to termite guts (135, 137, 158, 173, 179). The construction of metagenomic libraries is not trivial since challenges can be found in isolating large DNA fragments, finding suitable cloning vectors, and selecting expression hosts (29, 135). Additionally, metagenomics relies on sequence-based or functional-based screening assays to identify novel enzymes (29). In some cases, the exploratory capabilities of the sequence-based approach are limited by the sequences of known enzymes and the risk of finding incomplete and non-functional genes (87). The functional-based approach overcomes this limitation by relying on sensitive assays to identify only those clones expressing functional genes. Consequently, coupling metagenomics with high-throughput functional screening has the potential to identify novel and functional biocatalysts while also providing access to large amounts of genomic information (70, 87, 131, 158).

Recent metagenomic studies have identified novel cellulases and hemicellulases in complex environments such as soil, cow rumens and insect guts (15, 34, 52, 70, 131). These studies represent only the beginning of the exploratory efforts, since the majority of the environments remain unexplored. This chapter describes a strategy to identify and characterize unique cellulases from the soil metagenome.

2. Materials and Methods

2.1 Bacterial Strains and Plasmids

The bacterial strains and plasmids used in this study are described in Table 1. The *E. coli* DH10B, BL21 (DE3), and EPI300 were grown in Luria-Bertani (LB) medium at 37°C. The *E. coli* cells harboring the different modification of pEpiFOS-5 and pet27/bmod were grown in LB medium supplemented with 12 μg/ml of chloroamphenicol or 50 μg/ml of kanamycin, respectively.

2.2 Collection of Soil Samples and Construction of Metagenomic Libraries

Soil samples were collected at the SAPPI paper mill facility (Cloquet, MN), a tree processing facility (North Branch, MN), a prairie (Elk Lake, MN), and cornfields from UMore Park (Rosemount, MN). Four soil metagenomic libraries were constructed by Clemson University Genomics Institute (CUGI). The libraries were composed of approximately 139,000 clones. On average, each clone contained a 35-Kb DNA-fragment. Each DNA-fragment was ligated into a pEpiFos-5 fosmid (Epicenter, Madison, WI, USA) and cloned into *E. coli* DH10B. The clones were grown on LB agar plates supplemented with 12 μg/ml of chloroamphenicol and robotically-picked and arrayed in 384-well microplates.

Table 1. Strains and plasmids used during this study

Strain or Plasmid	Genotype or Description	Source
E. coli		
DH10B	*F⁻ endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ⁻*	Lab Stock
BL21 (DE3)	*F⁻ ompT hsdSB (rB- mB-) gal dcm (DE3)*	Lucigen
EPI300	*F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ⁻ rpsL (Str^R) nupG trfA dhfr*	Epicentre
pEpiFOS-5	7.5-Kb fosmid backbone. Cm^R.	Epicentre
pCelK	39.9-Kb DNA fragment extracted from soil collected at the paper mill facility and cloned into the pEpiFOS-5 fosmid.	This study
pCelK; Δ*celK*25::Tn5	1.9-Kb DNA fragment from transposon EZ-Tn5 <oriV/Kan-2> inserted at position 27,159 bp of pCelK fosmid.	This study
pCelP	37.5-Kb DNA fragment extracted from soil collected at UMore Park and cloned into the pEpiFOS-5 fosmid.	This study
pCelP; Δ*celP*16::Tn5 a	1.9-Kb DNA fragment from transposon EZ-Tn5 <oriV/Kan-2> inserted at position 28,596 bp of pCelP fosmid.	This study
pCelP; Δ*celP*16::Tn5 b	1.9-Kb DNA fragment from transposon EZ-Tn5 <oriV/Kan-2> inserted at position 28,724 bp of pCelP fosmid.	This study
pCelP; Δ*celP*16::Tn5 c	1.9-Kb DNA fragment from transposon EZ-Tn5 <oriV/Kan-2> inserted at position 28,903 bp of pCelP fosmid.	This study
pCelD	34.2-Kb DNA fragment extracted from soil collected at UMore Park and cloned into the pEpiFOS-5 fosmid.	This study
pCelD; Δ*celD*1::Tn5 a	1.9-Kb DNA fragment from transposon EZ-Tn5 <oriV/Kan-2> inserted at position 1 bp of pCelD fosmid.	This study
pCelD; Δ*celD*1::Tn5 b	1.9-Kb DNA fragment from transposon EZ-Tn5 <oriV/Kan-2> inserted at position 841 bp of pCelD fosmid.	This study
pXylP	33.2-Kb DNA fragment extracted from soil collected at UMore Park and cloned into the pEpiFOS-5 fosmid.	This study
pXylP: Δ*xylP*19::Tn5	1.9-Kb DNA fragment from transposon EZ-Tn5 <oriV/Kan-2> inserted at position 18, 823 bp of pXylP fosmid.	This study
pet27b/mod	5.3-Kb Expression vector backbone. Km^R.	[34]
pCelK25	2.5-Kb PCR fragment amplified from fosmid pCelK and cloned into the *Nco*I-*Xho*I site of pet27b/mod.	This study
pCelK25-GH9	1.7-Kb PCR fragment amplified from fosmid pCelK and clone into the *Nco*I-*Xho*I site of pet27b/mod.	This study
pCelP16	2.5-Kb PCR fragment amplified from fosmid pCelP and cloned into the *Nco*I-*Xho*I site of pet27b/mod.	This study
pCelD1	3.6-Kb PCR fragment amplified from fosmid pCelD and cloned into the *Hind*III-*Nhe*I site of pet27b/mod.	This study
pXylP19	1-Kb PCR fragment amplified from fosmid pXylP and cloned into the *Sac*I-*Xho*I site of pet27b/mod.	This study

2.3 Functional High-Throughput Screening of Metagenomic Libraries

Clones were screened for endoglucanase, exoglucanase, carboxymethyl cellulase (CMCase), xylanase and xylosidase activity using soluble and insoluble substrates. The endoglucanase and xylanase activity were assayed by incorporating 0.1% (w/v) of insoluble hydroxymethyl cellulose, 0.1% (w/v) xylan, or 0.1% (w/v) arabinoxylan cross-linked with azurine (Megazyme, Ireland) into LB agar plates. Clones were robotically inoculated onto the assay plates using Qbot robot (Molecular Devices, Sunnyvale, CA, USA) and they were incubated for 7 days at 27°C. Clones producing blue zones were isolated and their activities were confirmed. The CMCase activity was assayed by using the congo red overlay method (165). Briefly, LB agar plates were supplemented with 0.2% (w/v) soluble carboxymethyl cellulose (CMC) and chloroamphenicol (12 μg/ul). The clones were robotically inoculated onto the assay plates and incubated for 7 days at 27°C. The resulting plates were overlaid with an aqueous solution of 0.2% (w/v) congo red dye and destained with a 1M NaCl solution. Clones producing clearing zones were isolated and their CMCase activity was confirmed.

The exoglucanase and xylosidase activities were assayed using fluorogenic substrates (171). Briefly, clones were grown in 384-well plates containing LB medium supplemented with 50 μg/ml 4-methylumbelliferyl-β-D-cellobioside (MUC) or 4-methylumbelliferyl-β-D-xylosidase (MUX) and chloroamphenicol (12 μg/μl). Plates were incubated for 7 days at 26°C and exposed to UV light. Clones showing fluorescence were

isolated and their activities were confirmed. The CMC, Congo red dye, MUC, and MUX were all purchased from Sigma (Sigma-Aldrich, St. Louis, MO, USA).

2.4 *In vitro* Transposon Mutagenesis

The genes responsible for the putative activity were identified by using *in vitro* transposon mutagenesis. The fosmid DNA of selected clones was extracted by using FosmidMAX DNA Purification Kit (Epicenter, Madison, WI, USA). *In vitro* transposon mutagenesis was performed using the EZ-Tn5 <oriV/Kan-2> Insertion Kit (Epicenter, Madison, WI, USA) following the manufacturer's recommendations. A transposon library was created for each selected clone. The resulting clones were analyzed using the screening assays described above (Section 3.3). Clones lacking the original phenotype were selected and their fosmids were extracted by using the plasmid mini purification kit (Qiagen, Valencia, CA, USA). The insertion sites were determined by sequencing fosmids with a forward primer (5'-gccaacgactacgcactagccaac-3') that annealed to the transposon.

2.5 Fosmid Pyrosequencing

Fosmids from selected clones were fully sequenced in both directions using Roche 454 technology (454 Life Sciences, Branford, CT, USA) at the University of Minnesota Biomedical Genomic Center. Pyrosequencing reads were assembled using the GS *De Novo* Assembler and assembled sequences had an average depth of coverage of 31. The open reading frames (ORF) were annotated by using the RAST system (5). The

sequences were deposited in Genbank under accession numbers (pCelK [JQ822235], pCelP [JQ822236], pCelD [JQ822237], pXylP [JQ822238 and JQ822239]). Additionally, each predicted open reading frame was independently compared to sequences in the Genbank database and its predicted function was described (Supplementary Information; Table 6-9).

2.6 Subcloning and Overexpression of Putative Enzymes

The genes identified during transposon mutagenesis were PCR amplified using the primers listed in Table 2. The resulting fragments were ligated into a pet27b/mod expression vector and transformed into *E. coli* BL21 (DE3) (Table 1). To perform qualitative enzymatic assays, the transformed cells were grown overnight in LB medium supplemented with kanamycin (50 μg/ml) and 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). The resulting cells were sonicated and the lysates were spotted onto assay plates, similar to those used for the screening assays (Section 3.3). Hydrolytic activity was detected as previously described (Section 3.3). Transformed cells harboring pCelK25 plasmids were grown in LB medium supplemented with kanamycin (50 μg/ml) to an OD_{600} of 0.6 at 37°C, with continuous shaking. To induce protein expression, 1 mM IPTG was added to the LB medium and the cell cultures were incubated for an additional 16 hr at 16°C, with continuous shaking. The resulting cells were collected by centrifugation and were resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole, pH 8, 0.5 mM PMSF and DNAse I crystals). The cell suspensions were sonicated and intact cells were removed by centrifugation at 13,000

x *g* for 20 min. Additionally, the resulting supernatant was centrifuged at 100,000 x *g* for 1 hr to remove cell membranes. The clarified lysate was incubated with pre-equilibrated Ni-NTA (Qiagen, Valencia, CA, USA) slurry for 1 hr at 4°C in a tube rotator. Enzymes were purified from the affinity chromatography matrix (Ni-NTA) by using the manufacturer's recommendations. Further protein purification was performed by ÄKTA$_{FPLC}$, (GE Healthcare, Piscataway, NJ, USA) using a Superdex 200 gel filtration column (GE Healthcare, Piscataway, NJ, USA).

Table 2. Primers used during this study

Gene	Sequence of Forward (F) and Reverse (R) Primers (5'- 3')	Restriction Sites
*celK*25	F: gcacatatgtccatggatgtgtctcatctgaaacgcgcgctgtttctggtgttgg R: ctcgagtcagtgatggtgatggtgatgttgcgccccgttaaactgct	NcoI XhoI
*celK*25-GH9	F: gcaccatggatgcttctatcaggcaggcagcgccaaccacc R: gcatcactcgagtcagtgatggtgatggtgatg	NcoI XhoI
*celP*16	F: gcacatatgtccatggatcatcgcaccaaccgcatcatctatc R: gctctcgagtcagtgatggtgatggtgatgcccgaccaccagcagcgtga	NcoI XhoI
*celD*1	F: gcaaagcttggttactggcataccagcggcgcaaagatcc R: gctgctagctcagtgatggtgatggtgatgtggaactggcgggaaggcat	HindIII NheI
*xylP*27	F: gcagagctctatgaacaacgcacccgcacacccgacg R: tccctcgagtcagtgatggtgatggtgatggctggcgttgcaggagacgg	SacI XhoI

2.7 Protein Determination and Analysis

The Bradford Reagent (Sigma-Aldrich, St. Louis, MO, USA) was used to determine the protein concentration of samples. Protein detection was performed by sodium dodecyl sulfate-polyacrylamide electrophoresis (SDS-PAGE) in 7.5% Tris-HCl gels under non-

reducing conditions (80). Samples containing 15 μg of protein were boiled at 95°C for 5 min in SDS loading buffer, and subsequently loaded onto gels. Proteins were visualized by staining gels with coomassie brilliant blue.

2.8 Characterization of Enzymes

The CMCase activity was determined using the dinitrosalicylic (DNS) reducing sugar method as described by Miller (1959) and adapted by Mandels and coworkers (1976) (42, 116). Briefly, 0.1 ml of an enzyme dilution was incubated with 0.4 ml of CMC (2% w/v) dissolved in citrate buffer (50 mM, pH 5). Samples were incubated for 30 min at 40°C in a water bath and the reactions were stopped by adding 1.5 ml of DNS reagent. The resulting samples were boiled for 5 min in a water bath and cooled in an ice bath. Samples were diluted with 10 ml of H_2O and 200 μl of each sample was placed in a microplate reader to determine the absorbance at 540 nm (Abs_{540}). The enzyme dilution releasing 0.5 mg/ml of glucose was used for further studies. A unit of CMCase activity was defined as micromoles of glucose equivalents released per minute per micromole of protein (μmoles of glucose eq. min^{-1} μmol^{-1} protein). The optimal pH was determined by incubating an enzyme dilution with carboxymethyl cellulose (2% w/v) dissolved in 50 mM citrate buffers (pH 3-5), 100 mM phosphate buffers (pH 6-8) and 50 mM Tris buffer (pH 9). The optimal temperature was determined by incubating the pre-equilibrated reactions in heating blocks with temperatures ranging from 10°C to 60°C. Birchwood xylan and oat spelt were prepared and tested using the same procedure as CMC. A unit of xylanase activity was defined as micromoles of xylose equivalents released per minute

per micromole of protein (μmoles of xylose eq. min^{-1} μmol^{-1} protein). The exoglucanase activity was determined by using a modified version of the Deshpande and coworkers (1982) method (32). Briefly, 90 μl of 2 mM p-nitrophenyl-β-D-cellobioside (pNPC) dissolved in citrate buffer (50 mM, pH 5) was incubated with 10 μl of enzyme dilutions. The reactions were incubated for 30 min, and every 5 minutes, the reactions were quenched with the addition of 150 μl of 2 M Na_2CO_3. The absorbance was measured at 410 nm (Abs_{410}) in a microplate reader. The experimental extinction coefficient of o-nitrophenol was found to be 3.06 Abs_{410}/mM and it was used to determine the final exocellulase activity. The AZCELase activity was determined by incubating an enzyme dilution with a suspension of A-HE in citrate buffer (50 mM, pH 5) at 40°C for 60 min. The resulting samples were filtrated to remove non-hydrolyzed substrate and the absorbance at 590 nm (Abs_{590}) was determined using a microplate reader. A unit of AZCELase activity was defined as the Abs_{590} per minute. The specific AZCELase activity was expressed as Abs_{590} min^{-1} μmol^{-1} protein.

2.9 Mass Spectrometry

Enzyme CelK25 was analyzed by mass spectroscopy at the University of Minnesota Center for Mass Spectrometry and Proteomics. The protein was excised from a pre-casted SDS-PAGE (BioRad Laboratories, Hercules, CA, USA) gels, followed by trypsin digestion as previously described (146). The resulting protein was subjected to tandem mass spectrometry on a Thermo Finnigan LTQ. The analysis of the MS/MS data and the protein identification was performed using Sequest and X! Tandem using a library

composed of bacterial genomes. Resulting peptides were confirmed with Scaffold (v. Scaffold_01_05_14).

2.10 Phylogenetic Analysis of Glycosyl Hydrolases

The protein sequences encoded by genes identified by transposon mutagenesis were subjected to phylogenetic analysis. A set of similar proteins was obtained for each of the selected proteins from PSI Blast. Maximum likelihood phylogenetic trees (PhyML aLRT) were generated with the Phylogeny.fr platform using JTT as a substitution matrix (http://phylogeny.lirmm.fr/phylo_cgi/index.cgi). Sequences were aligned by MUSCLE (v3.7) and curated by G-Blocks (v0.91b). Phylogenetic trees were formatted with FigTree (v1.3.1).

3. Results

3.1 Functional High-Throughput Screening of Metagenomic Libraries

The metagenomic libraries were robotically screened for glycosyl hydrolase activity using a series of functional, high-throughput, assays. Forty-five clones were grouped into eight different phenotypes and were identified to have putative glycosyl hydrolase activity in at least one of the screening substrates (Table 3).

Table 3. Clones with putative glycosyl hydrolase activity identified during the functional screening

No. of Clones	Soluble Substrates			Insoluble Substrates		
	MUC[a]	MUX[b]	CMC[c]	A-HE[d]	A-XYL[e]	A-ARA[f]
1*	+	-	+	+	-	-
1*	-	-	+	+	-	-
1*	-	-	+	-	-	-
1*	-	-	-	-	+	+
2	+	+	+	-	-	-
6	+	-	-	-	-	-
16	-	+	-	-	-	-
17	+	+	-	-	-	-

*Clones selected for further studies. [a]MUC; 4-methylumbelliferyl-β-D-cellobioside ;[b]MUX; 4-methylumbelliferyl-β-D-xylosidase; [c]CMC; carboxymethyl cellulose; [d]A-HE; hydroxymethyl cellulose cross-linked with azurine; [e]A-XYL; xylan (from oats) cross-linked with azurine ; [f]A-ARA; arabinoxylan (from wheat) cross-linked with azurine.

The most predominant phenotype of clones hydrolyzed MUC and MUX substrates, which is indicative of putative cellobiohydrolase activity and xylosidase activity, respectively (130, 171). Ninety-four percent of the identified clones exhibited putative hydrolase activity exclusively with soluble substrates (MUC, MUX and CMC) and the rate of finding these clones in the libraries was approximately 1:3000. During this study, three independent clones exhibited glycosyl hydrolase activity towards insoluble substrates (A-HE, A-XYL or A-ARA), making the rate of finding a clone with these particular phenotypes at 1:46,000.

3.2 Identification of Genes Responsible for Glycosyl Hydrolase Activity in Selected Clones

Genes responsible for the putative hydrolase activity were identified using *in vitro* random transposon mutagenesis. A library of transposon mutants was created for each of the four selected clones. The resulting libraries were screened under the same conditions used during the initial screening assays (Table 4). Transposon mutants lacking the hydrolytic phenotype were selected for further studies, and their insertion sites were determined by DNA sequencing (Table 4, Figure 1). Additionally, four genes identified during transposon mutagenesis were cloned independently in expression vectors and over-expressed in *E. coli* BL21 (DE3) as previously described (Table 1). The transformed cells were tested for glycosyl hydrolase activity under the same conditions used during the initial screening assays. The activity results confirmed the functionality and the hydrolytic activity of the expressed proteins (Table 4).

Table 4. Glycosyl hydrolase activity of clones harboring fosmids or expression plasmids

Fosmid Id or Protein	Soluble Substrates		Insoluble Substrates	
	MUC[a]	CMC[b]	A-HE[c]	A-XYL[d]
pCelK	+	+	+	-
pCelK;Δ*celK*25::Tn5	-	-	-	-
CelK25	+	+	+	-
pCelP	-	+	+	-
pCelP;Δ*celP*16::Tn5 a,b,c	-	-	-	-
CelP16	-	+	+	-
pCelD	-	+	-	-
pCelD; Δ*celD*1::Tn5 a,b	-	-	-	-
CelD1	-	+	-	-
pXylP	-	-	-	+
pXylP; Δ*xylP*19::Tn5	-	-	-	-
XylP19	-	-	-	+

[a]MUC; 4-methylumbelliferyl-β-D-cellobioside ; [b]CMC; carboxymethyl cellulose; [c]A-HE; hydroxymethyl cellulose cross-linked with azurine; [d]A-XYL; xylan cross-linked with azurine.

Fosmid pCelK. This fosmid harbored a DNA-fragment extracted from soil collected at the SAPPI paper mill facility (Table 1). The clone carrying the pCelK fosmid hydrolyzed soluble (MUC and CMC) and insoluble (A-HE) substrates (Table 4). The observed phenotype was indicative of putative cellobiohydrolase, CMCase, and endoglucanase activity.

The c*elK*25 gene was found responsible for the cellulolytic activity since a mutant harboring pCelK;Δ*celK*25::Tn5 lost the observed phenotype (Table 4, Figure 1).

Additionally, the cells overexpressing CelK25 protein exhibited the same substrate specificity as did, the clone harboring pCelK fosmid (Table 4). The *celK*25 gene was predicted to encode a glycosyl hydrolase protein with three distinctive domains. The catalytic domain belonged to the GH9 family and it was fused to an Ig-like domain and a CBM9 at the N-terminal end of the protein (Figure 1).

The GH9 domain of CelK25 was 61% similar to the catalytic domain of a characterized cellobiohydrolase from *C. thermocellum*, while the CBM9 shared 38% similarity to a sequence found in *Thermotoga maritima* (144). In addition, the N-terminal sequence of CelK25 was predicted to contain a signal peptide recognized by gram-positive or gram-negative bacteria for protein secretion.

The protein sequences with the highest similarity to the GH9 domain of CelK25 were compared through phylogenetic analysis (Supplementary Information; Figure 7). The resulting phylogenetic tree indicated the evolutionary relationship between the catalytic domain GH9 of CelK25 and a putative endoglucanase found in the γ-proteobacterium *Hahella chejuensis*. Gram-positive bacteria, including *Streptomyces*, *Micromonospora*, and *Thermobifida* were also found to have evolutionary related endoglucanases and cellobiohydrolases.

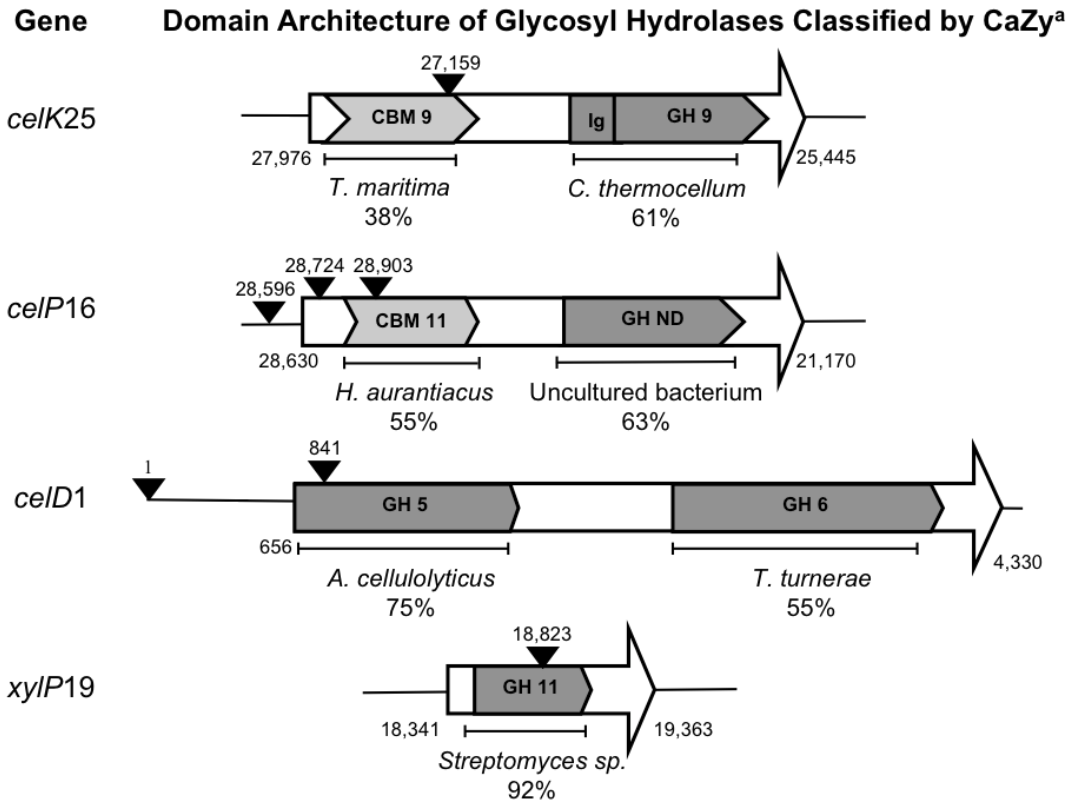


Figure 1. Genes responsible for glycosyl hydrolase activity identified with transposon mutagenesis. [a]CaZy, Carbohydrate-Active Enzymes Database; CBM, Carbohydrate Binding Modules; GH; Glycosyl Hydrolases*;* ▼ Transposon insertion sites; % , percentage similarity; The numbers at the extremes of each open reading frame indicate the exact gene location in the DNA fragment; *T*, *Thermotoga*; *C*, *Clostridium*; *H*, *Herpetosiphon*; *A*, *Acidothermus*; T, *Teredinibacter.*

Fosmid pCelP. Fosmid pCelP contained a DNA-fragment extracted from soil collected from the UMore Park corn fields (Table 1). This clone hydrolyzed insoluble (A-HE) and soluble (CMC) substrates, which was indicative of putative endoglucanase and CMCase activity, respectively (Table 4, Figure 1). *celP*16 gene was found to be responsible for the hydrolytic phenotype since three independent transposon mutants harboring pCelP;Δ*celP*16::Tn5 a, b, and c fosmids lost their hydrolytic phenotype (Table 4, Figure 1). Additionally, the clone expressing the CelP16 protein exhibited the same substrate specificity as did, the clone harboring the pCelP fosmid (Table 4).

Gene *celP*16 was predicted to encode a protein containing a non-classified catalytic GH domain associated with a truncated CBM11 domain (Figure 2). The GH domain of CelP16 was 63% similar to a bifunctional glucanase/xylanase enzyme recently isolated and characterized from an uncultured microorganism found in a rabbit cecum (35). The phylogenetic analysis of similar proteins indicated the evolutionary relationship between the catalytic domain of CelP16 and other endoglucanases isolated from uncultured bacteria found in soil and sediments (Supplementary Information; Figure 8). There was also an evolutionary relationship (70% similarity) between the GH domain of CelP16 and putative endoglucanases encoded by *Streptomyces*.

Fosmid pCelD. The fosmid pCelD also contained a DNA-fragment extracted from soil collected from the UMore Park corn fields (Table 1). This clone exclusively hydrolyzed CMC, which was indicative of CMCase activity (Table 4). The *celD*1 gene was found

responsible for the CMCase activity since two independent transposon mutants harboring pCelD;Δ*celD*1::Tn5 a and b lost their hydrolytic phenotype (Table 4, Figure 1). Additionally, the clone expressing the CelD1 protein exhibited the same substrate specificity and hydrolytic activity as did, the clone harboring the fosmid pCelD (Table 4).

Gene *celD*1 was predicted to encode a protein with dual catalytic domains (Figure 1). The catalytic domain associated at the N-terminus of the protein belonged to the GH5 family and was fused to a GH6 domain. The GH5 domain of CelD1 was found to be 75% similar to a protein found in *Acidothermus cellulolyticus*. The GH6 domain shared 55% similarity with a protein found in *Teredinibacter turnerae*. The catalytic domain GH5 was evolutionary-related to endoglucanases predicted to be encoded by actinobacteria including *Streptomyces* sp., *Catenulispora acidiphila,* and *Acidothermus cellulolyticus* (Supplementary Information; Figure 9-10).

Fosmid pXylP. Fosmid pXylP also contained a DNA-fragment extracted from soil collected from UMore Park (Table 1). This clone exhibited putative xylanase activity and hydrolyzed insoluble xylan (A-XYL) and arabinoxylan (A-ARA) (Table 4). Gene *xylP*19 was found to be responsible for the xylanase activity since a transposon mutant harboring pXylP;Δ*xylP*19::Tn5 fosmid was unable to hydrolyze insoluble xylan (A-XYL) and arabinoxylan (A-ARA) (Table 4, Figure 1). Additionally, a clone expressing XylP19 exhibited the same substrate specificity and hydrolytic activity as did, the clone harboring

the pXylP fosmid (Table 4). The XylP19 protein contained a GH11 domain that was 92% similar to a putative endoxylanase found in *Streptomyces* sp. (Figure 1).

3.3 Cloning and Overexpression of Recombinant Glycosyl Hydrolase CelK25

The protein CelK25 was selected for purification and biochemical characterization as it had the most unique domain architecture and hydrolyzed both soluble and insoluble substrates. The CelK25 was composed of 843 amino acids and had a predicted molecular weight of 91-kDa. The recombinant CelK25 protein was purified from a clarified cell lysate using a Ni-NTA column and size exclusion chromatography (FPLC). The over-expression of CelK25 was confirmed with SDS-PAGE (Figure 2). A protein band with an apparent molecular weight of 90-kDa was found and presumptively confirmed the presence of CelK25 (Figure 2).

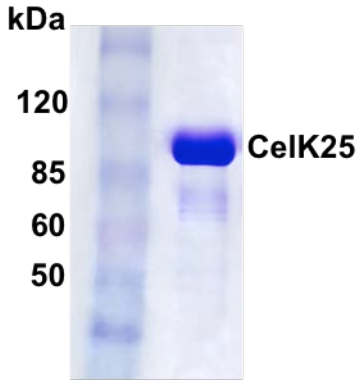


Figure 2. Purification of recombinant glycosyl hydrolase CelK25 identified in the soil metagenome and over-expressed in *E. coli*.

The identity of CelK25 was also confirmed by mass spectrometric analysis. Additionally, Western blot analysis, using anti-His-tag antibodies together with zymographs confirmed the presence of the affinity tag (His-Tag) and its functionality, respectively (data not shown).

3.4 Substrate Specificity, Optimum pH and Temperature Profile of CelK25

Biochemical studies were performed using the CelK25 protein (90-kDa). The optimal pH and temperature and the substrate specificity were determined using the reducing sugar assay and CMC as a substrate.

CelK25 exhibited the highest specific activity at pH 5 and around 50% of the activity was conserved between pH 6 and pH 7 (Figure 3A). The CMCase activity decreased dramatically below pH 5 and above pH 8. The optimal temperature for CelK25 was 40°C. However, CelK25 reactions mixtures incubated at temperatures ranging between 10°C to 50°C retained more than 50% of their hydrolytic activity (Figure 3B).

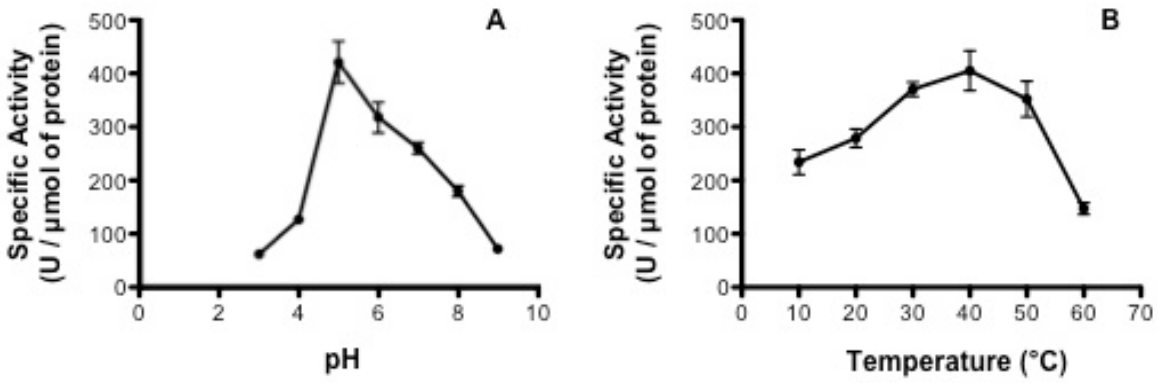


Figure 3. The influence of pH and temperature on glycosyl hydrolase activity of CelK25 determined by the DNS reducing sugar assay. **A.** Specific activity of CelK25 incubated at pH conditions ranging from 3 to 9. **B.** Specific activity of CelK25 incubated at temperatures ranging from 10°C to 60°C.

The specific CMCase activity of CelK25 was 389 Units/μmol protein and the exoglucanase activity of CelK25 was 113 Units/μmol protein. Even though the presence of CBM9 has been previously shown to bind oat spelt xylan, the CelK25 isolated in this study did not hydrolyze oat spelt xylan nor birchwood xylan (13) (Table 5).

Table 5. Substrate specificity of glycosyl hydrolase CelK25

Substrate	**Specific Activity (Units / μmol)**
CMC[a]	389 ± 55
CMC[a] +10 mM $CaCl_2$	400 ± 29
pNPC[b]	113 ± 12
A-HE[c]	20 ± 6
Oat Spelt Xylan	ND
Birchwood Xylan	ND

[a]CMC; Carboxymethyl cellulose, [b]pNPC; p-nitrophenyl-β-D-cellobioside, [c]A-HE; Hydroxymethyl cellulose cross-linked with azurine ND; Not detected.

The CBM9 was also predicted to have three calcium binding sites (64, 129). Considering this, the effect of calcium ions in the hydrolytic activity of CelK25 was investigated. Results of these studies did not show a significant change in the hydrolytic activity of CelK25 in the presence of 10 mM $CaCl_2$ (Table 5).

3.5 Influence of CBM9 domain in the hydrolytic activity of CelK25

To further investigate the influence of the CBM9 domain on the hydrolytic activity of CelK25, a truncated protein devoid of CBM9 was expressed in *E. coli*. This protein, CelK25-GH9, was composed of 594 amino acids and has a predicted molecular weight of 64-kDa. The CelK25-GH9 was over-expressed and purified form *E. coli* cells, as previously described (Figure 4).

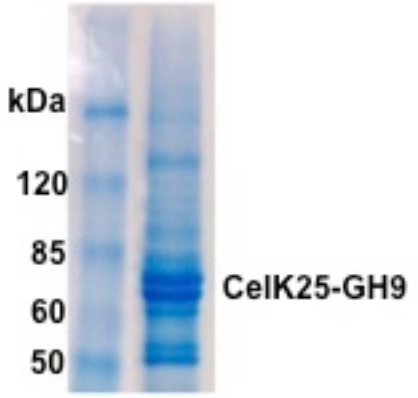


Figure 4. Protein fraction containing CelK25-GH9 devoid of carbohydrate binding module.

The SDS-PAGE revealed the presence of two predominant proteins bands with apparent molecular weights of 65-kDa and 70-kDa, respectively (Figure 4). The protein gel also indicated the presence of smaller proteins in this protein fraction suggesting the occurrence of additional proteolytic events. According to the amino acids sequence, CelK25-GH9 did not contain a signal peptide sequence for protein secretion. The second band may be the product of an additional processing event during protein expression or purification.

Activity assays indicated that the truncated CelK25-GH9 protein fraction lost 33% and 38%, respectively, of the specific CMCase and exoglucanase activity compared to the full length CelK25 (Figure 5). In contrast, CelK25-GH9 exhibited greater hydrolytic activity towards a commercial preparation of insoluble hydroxymethyl cellulose cross-linked with azurine (A-HE) than did the full-length CelK25.

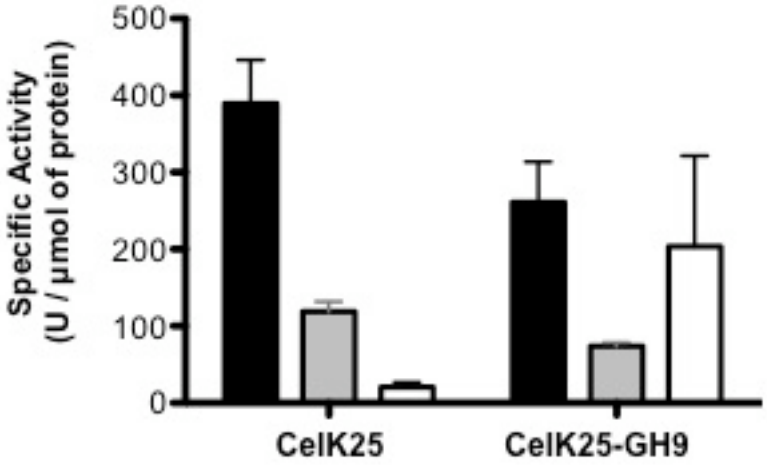


Figure 5. Specific activities of glycosyl hydrolases CelK25 and CelK25-GH9 incubated with ■ CMC, ▒ pNPC, and ▭ A-HE substrates. CMC; Carboxymethyl cellulose, pNPC; p-nitrophenyl-β-D-cellobioside, A-HE; Hydroxymethyl cellulose cross-linked with azurine.

4. Discussion

The present study demonstrated the potential of using a metagenomics approach, combined with a functional screening, to discover unique glycosyl hydrolases in soil metagenomes. Similar to what has been reported in other studies, environments continuously exposed to biomass contain DNA encoding unique glycosyl hydrolases (15, 70, 131, 176, 179). During this study, about 4.8-Gbp of DNA was screened for the identification of cellulases or hemicellulases, using both soluble and insoluble substrates. These assays identified a group of 45 clones exhibiting glycosyl hydrolases activity (Table 3).

While several studies have screened soil metagenomes for cellulases, only few have identified enzymes that hydrolyzed insoluble cellulosic substrates (34). The rate of finding a clone capable of hydrolyzing soluble cellulosic substrates in a metagenomic library made from soil has varied significantly. While some metagenomic studies reported 1:200 clones having hydrolytic activity, others have found this rate to be 10 to 300-fold lower (71, 79, 131, 175). The rate of finding a clone with hydrolytic activity towards insoluble cellulosic substrates is not well documented since most functional screening studies are performed using soluble substrates (131, 175). The challenge of hydrolyzing insoluble cellulosic substrates using biological systems is not trivial. It is well known that microorganisms with high cellulolytic activity, such as *Clostridium*

thermocellum, require specialized extracellular macrostructure to synergistically deconstruct insoluble cellulosic substrates (8, 145).

During this study, four clones were selected for further studies as they hydrolyzed insoluble substrates or exhibited unique phenotypes (Table 3). Sequencing and *in vitro* random transposon mutagenesis studies indicated that four genes were responsible for the observed phenotypes (Table 4). Expression experiments demonstrated that all four genes encoded functional glycosyl hydrolases (Table 4).

The sequence of CelK25 contained a CBM9 fused to an Ig-like domain and a GH9 domain. To our knowledge, this domain architecture has not been previously described. Bayer and coworkers (2006) classified GH9 into four different subfamilies Themes A, B, C, and D (Figure 6). The Ig-like domain is typically found associated with GH9 in Theme C and D subfamilies (7). While members of the GH9 family are commonly found associated with CBM from family 3 and 4, CBM9 has been exclusively found associated with bacteria xylanases, such as GH10 and GH11 (13, 20, 129, 144). The association of GH9 with a CBM9 in a single protein appears to be unique (Figure 6). The domain architecture of CelK25 showed for the first time a GH9 associated with a CBM9 in a single protein (Figure 6).

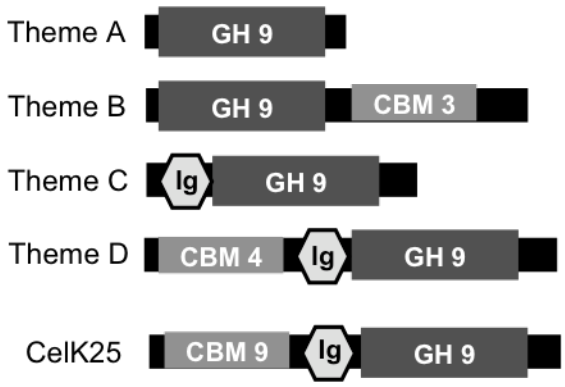


Figure 6. Schematic representation of domains architectures of GH9 subfamilies proposed by Baker and coworkers (2006) compared to CelK25.

The activity assays demonstrated that CelK25 hydrolyzed soluble (CMC and MUC) and insoluble cellulosic (A-HE) substrates (Table 4). The specific CMCase activity of CelK25 was similar to previously characterized GH9 proteins (118). However, the optimal pH of other GH9 proteins has varied quite significantly. While *Thermobifida halorotorans* produces a GH9 with optimum pH of 8, *Clostridium thermocellum* produces a GH9 with optimum pH of 5.5 (43, 187). The optimal temperature of GH9 proteins has also varied even within the same organism. For example, CelT and CelI are two major GH9 proteins produced by *C. thermocellum* and their optimum temperatures are 70°C and 60°C, respectively (43, 78).

In contrast to other studies, activity assays performed with CelK25-GH9 indicated that the GH9 domain of CelK25 by itself had the ability of hydrolyzing soluble and insoluble cellulosic substrates (Figure 4) (22, 43). When CBM3 domains were removed from a GH9 isolated from *Paenibacillus barcinonensis*, 95% of the specific activity was lost (22). The absence of CBM3 domains has also decreased the activity of CelI, a GH9 enzyme expressed by *Clostridium thermocellum*, towards insoluble substrates (43).

The binding specificities of CBM9 have been identified and characterized in a xylanase from *T. maritima* (13). In those studies, CBM9 was found to bind reducing ends of amorphous and crystalline cellulose, cellobiose, and oat spelt. In contrast, binding to CMC and hydroxyethyl cellulose was not observed. Furthermore, the influence of CBM9 in the glycosyl hydrolase activity of GH9 proteins has not been previously documented.

Our results, and those of others, suggest that the presence of CBM9 plays an important role during hydrolysis of soluble substrates. However, the same effect was not observed in the hydrolysis of insoluble substrates. The binding affinity of CBM9 for cellobiose may support the observed increase in the cellobiohydrolase activity in CelK25. However, this may not support the increase in CMCase activity since binding affinity of CBM9 to CMC has not been reported.

The presence of CBM in cellulases has been shown to produce different effects in the specific activity of glycosyl hydrolases. For example, CelT, a GH9 enzyme encoded by

C. thermocellum, is devoid of carbohydrate binding modules and it exhibits comparable hydrolytic activity to CelI, a GH9 harboring multiple CBM3 domains (43, 78).

In summary, this study demonstrated the identification of unique glycosyl hydrolases using the combination of metagenomics and functional screening approaches. Unique glycosyl hydrolase phenotypes were observed specially in clones harboring DNA extracted from soil collected at UMore Park corn fields. These results demonstrated that environments with continuous exposure to biomass contain uncultured organisms encoding unique glycosyl hydrolases.

5. Supplementary Information

5.1 Sequence Analysis of Selected Clones

The fosmids of the four selected clones were fully sequenced and the resulting data was deposited at GenBank. The fosmids had an average DNA insert size of 35-Kb and together they were predicted to encode more than 100 open reading frames. Additionally, each predicted open reading frame was independently searched against the GenBank database and its predicted function was described (Table 6-9).

Analysis of pCelK fosmid. According to sequencing results, fosmid pCelK carried a 39.9-Kb DNA-fragment and it was predicted to encode 37 open reading frames (Table 6). According to the sequence analysis, pCelK contained at least six genes encoding proteins involved in the biosynthesis or degradation of polysaccharides (Table 6). For example, the sequences of CelK7 and CelK10 are 49% and 73% similar to glycosyl transferases from family 1 and 2 encoded by *Sorangium cellulosum* and *Desulfovibrio magneticus*, respectively. Glycosyl transferases (GT) are responsible for catalyzing the formation of glycosydic bonds between sugar moieties and at least 94 different families have been described (20).

The protein sequences of CelK9 and CelK11 are 51% and 56% similar to a galactose oxidase and sorbosone dehydrogenase found in *Frankia alni* and *Herpetosiphon aurantiacus*, respectively. Galactose oxidase selectively catalyzes the oxidation of galactose moieties. Galactose can be found in hemicellulosic polysaccharides including

galactoglucomannan (47, 113). On the other hand, L-sorbose dehydrogenase catalyzes the oxidation of L-sorbose to produce L-sorbosone. L-sorbose has been found to induce transcription of cellulase genes in *Trichoderma reesei* (128). The sequence of pCelK was also found to contain a gene predicted to encode a glycosyl hydrolase from family 9 (*celK*25). The catalytic GH9 domain of CelK25 was found to be 61% similar to a cellobiohydrolase from *C. thermocellum* (144). Additional protein sequences similar to CelK25 were retrieved from the IMG/M. For example, a sequence found in the soil of Minnesota farms (2001321671) had 66% similarity with CelK25. An additional sequence found in a poplar biomass decaying microbial community (2020759778) showed 39% similarity with CelK25.

pCelK fosmid also contained putative oxidoreductases. One of these included a sequence (CelK17) 41% similar to a putative *c*-type cytochrome found in *Methylotenera* sp. followed by a sequenced (CelK18) 73% similar to a multicooper oxidase found in *Shpaerobacter thermophilus*. The amino acid sequences of another five open reading frames had high similarity to proteins originally encoded by phylogenetically related bacteria such as *Sphaerobacter thermophilus*, *Roseiflexus* sp. and *Chloroflexus aggregans*.

Analysis of pCelP fosmid. According to sequencing results, pCelP fosmid harbored a 37.5-Kb DNA-fragment. It was predicted to encode 18 independent open reading frames (Table 7). The pCelP fosmid contained a four-gene cluster predicted to encode proteins

involved in the synthesis and degradation of polysaccharides. This cluster was composed of three putative glycosyl transferases (GT) (CelP13-15) from family 39 (GT39) followed by a glycosyl hydrolase (CelP16) from an unknown family. The GT39 are mannosyltransferases involved in the O-glycosylation of proteins found in the cell walls. Although GT39 are mainly found in eukaryotic organisms, recently, the role of mannosyltransferases during glycosylation has been described in prokaryotes including *Mycobacterium* sp., *Coryneobacterium* sp. and *Streptomyces* sp. (93, 180). The encoded GT were between 38% and 61% similar to sequences found in different species of *Roseiflexus*. The catalytic GH domain of CelP16 was 63% similar to a bifunctional glucanase-xylanase recently characterized from the soil metagenome (122). Furthermore, a sequence (2001276605) retrieved from IMG/M was found in soil from Minnesota farms and it shared 57% similarity with CelP16.

The gene *celP*10 was predicted to encode a protein 65% similar to a putative pullulanase from *Streptosporangium roseum*. Pullanases are members of the GH13 family known to catalyze the hydrolysis of 1,6-α-D-glycsosidic bonds from pullulan, amylopectin and glycogen (69). The protein sequence of CelP10 contained dual α-amylase (GH13) domains fused to a CBM41 and a CBM48. The predicted function of 25% of the genes in the pCelP fosmid remained hypothetical.

Analysis of pCelD fosmid. Sequencing results indicated that the pCelD fosmid harbored a 34.2-Kb DNA-fragment and it was predicted to encode 28 open reading frames (Table

8). The function of 35% of the predicted open reading frames of pCelD remained hypothetical.

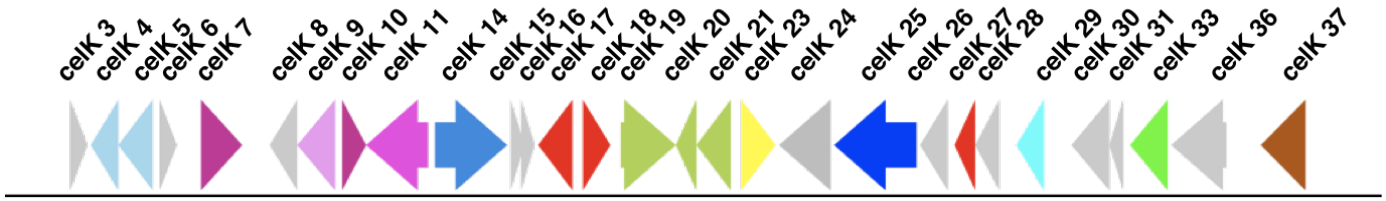
Sequencing results revealed the presence of at least three putative glycosyl hydrolases including an endoglucanse (CelD1), a xylanase (CelD26) and an arabinofuranosidase (CelD27). The sequence of CelD1 had a GH5 fused to a CBM3 and a GH6. The GH5 domain was found to be 74% similar to a sequence found in *Ktedonobacter racemifer*. The GH6 together with the CBM3 was found to be 50% similar to a putative cellobiohydrolase from *Streptosporangium roseum*. The sequence of CelD26 was 59% similar to a putative xylanase protein found in *Streptomyce*s sp., while the sequence of CelD27 was 49% similar to a putative arabinofuranosidase found in *Paenibacillus* sp. Xylanases and arabinofuranosidases are hemicellulases that hydrolyze xylose and arabinose moieties from xylan (147). The CelD27 sequence was preceded by two sequences predicted to be involved in galactose metabolism. This included a sequence (CelD24) 61% similar to a putative galactokinase found in *Roseiflexus castenholzii* and a sequence (CelD25) 71% similar to a galactose-1-phosphate uridylyltransferase (EC 2.7.7.12) found in *Rhodothermus marinus*.

Analysis of pXylP fosmid. According to sequencing results, pXylP fosmid carried a 40.7-Kb DNA-fragment and it was predicted to encode 30 open reading frames (Table 9). While the other three fosmids had GC contents ranging from 55% to 60%, pXylP had the

highest GC content with 70%. Furthermore, 45% of the sequences found in pXylP had protein homologs in different *Streptomyces* species (Figure 11).

Sequencing results indicated the presence of two proteins involved in polysaccharide degradation including a sequence (XylP19) 92% similar to a putative xylanase (GH11) found in *Streptomyces* sp. The putative xylanase was followed by a sequence (XylP20) 89% similar to a putative acetylxylan esterase found in *S. griseoflavus.* Overall, 36% of the predicted sequences in pXylP had hypothetical functions.

Table 6. Predicted open reading frames in fosmid pCelK



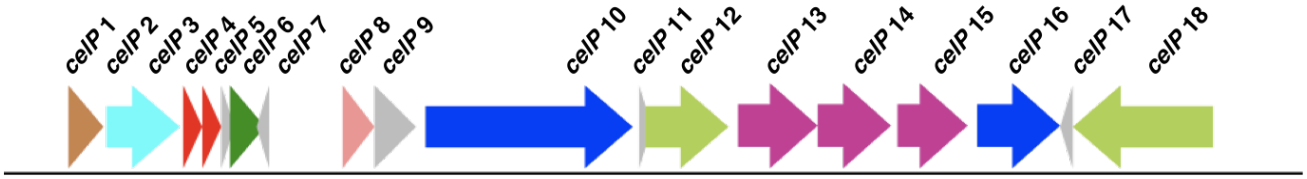
Fosmid pCelK (39920 bp)

Feature ID	Start	Stop	Length (bp)	Putative Function Prediction (Blast)	Organism	Identity (Similarity)
celK 1	342	506	165	Hypothetical protein	No similarity found	
celK 2	1437	2033	597	Hypothetical protein	No similarity found	
celK 3	2124	2702	579	Hypothetical protein	No similarity found	
celK 4	3626	2787	840	ABC-2 type transporter	*Exiguobacterium* sp.	44(62)
celK 5	4644	3676	969	Daunorubicin resistance ABC transporter ATPase subunit	*Exiguobacterium*	55 (71)
celK 6	4908	5417	510	Hypothetical protein	*Micromonas pusilla*	37 (52)
celK 7	6181	7419	1239	Glycosyltransferase-like protein	*Sorangium cellulosum*	29(49)
celK 8	9045	8278	768	Hypothetical protein	*Ktedonobacter racemifer*	53 (67)
celK 9	10258	9125	1134	Galactose oxidase	*Frankia alni*	38 (51)
celK 10	10429	11166	738	Glycosyltransferase	*Desulfovibrio magneticus*	53 (73)
celK 11	13046	11223	1824	Glucose/sorbosone	*Herpetosiphon aurantiacus*	36 (55)

				dehydrogenase-like		
*cel*K *12*	13588	13292	297	Hypothetical protein	*Acinetobacter* sp.	42 (54)
*cel*K *13*	13840	13676	165	Hypothetical protein	*Acinetobacter* sp.	53 (64)
*cel*K *14*	13833	15545	1713	Carbohydrate-binding protein	*Phytophthora infestas*	35 (43)
*cel*K *15*	15636	15863	228	Conserved hypothetical protein	*Ricinus communis*	36 (56)
*cel*K *16*	15879	16343	465	Hypothetical protein	*Microcoleus chthonoplastes*	50 (70)
*cel*K *17*	17451	16423	1029	Cytochrome c	*Methylotenera sp.*	28 (41)
*cel*K *18*	17704	18666	963	Multicopper oxidase type 3	*Sphaerobacter thermophilus*	59 (73)
*cel*K *19*	18855	20660	1806	Diguanylate cyclase with PAS/PAC sensor	*Desulforudis audaxviator*	36 (56)
*cel*K *20*	21288	20647	642	Dual specificity protein phosphatase	*Roseiflexus castenholzii*	42 (56)
*cel*K *21*	22306	21275	1032	Diacylglycerol kinase catalytic region	*Chloroflexus aggregans*	35 (51)
*cel*K *22*	22628	22440	189	Hypothetical protein	No similarity found	
*cel*K *23*	22638	23678	1041	Periplasmic trypsin-like serine protease	Idiomarina baltica	34 (46)
*cel*K *24*	25337	23832	1506	Tetratricopeptide repeat domain protein	*Microcoleus chthonoplastes*	36 (61)
*cel*K *25*	27976	25445	2532	Cellulase	*Sorangium cellulosum*	48 (62)
*cel*K *26*	28948	28151	798	Hypothetical protein	*Chromobacterium violaceum*	54 (63)
*cel*K *27*	29809	29183	627	Electron transport protein SCO1/SenC	*Roseiflexus sp.*	35 (58)
*cel*K *28*	30570	29827	744	Hypothetical protein	*Roseiflexus castenholzii*	30 (57)

celK 29	30896	31057	162	Hypothetical protein	No similarity found	
celK 30	31869	31045	825	UbiE/COQ5 methyltransferase [Rhodopseudomonas palustris HaA2	*Rhodopseudomonas palustris*	51 (68)
celK 31	33900	32761	1140	Hypothetical protein	uncultured methanogenic archaeon	37 (57)
celK 32	34308	33904	405	Hypothetical protein	*Clostridium leptum*	27 (43)
celK 33	35674	34478	1197	Glutamine-scyllo-inositol transaminase	*Natrialba magadii*	36 (56)
celK 34	37488	35761	1728	Hypothetical protein	*Thermotoga lettingae*	45 (63)
celK 35	37937	37791	147	Hypothetical protein	*Parabacteroides merdae*	59 (81)
celK 36	38241	38369	129	Hypothetical protein	No similarity found	
celK 37	39834	38554	1281	DNA repair protein RadA	*Sphaerobacter thermophilus*	58 (76)

Table 7. Predicted open reading frames in fosmid pCelP

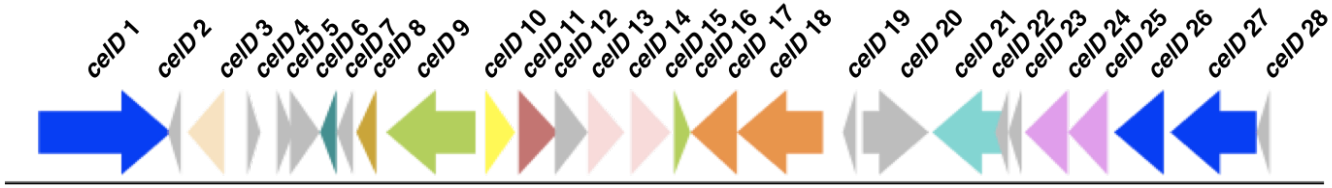


Fosmid pCelP (37537 bp)

6	Start	Stop	Length (bp)	Putative Function	Organism	Identity (Similarity)
*cel*P 1	421	1572	1152	Major facilitator superfamily	*Solibacter usitatus*	52 (68)
*cel*P 2	1658	3904	2247	Nitrate reductase	*Chryseobacterium gleum*	57 (73)
*cel*P 3	3964	4581	618	4Fe-4S ferredoxin, iron-sulfur binding domain protein	*Chthoniobacter flavus*	67 (78)
*cel*P 4	4618	5163	546	Rieske (2Fe-2S) domain protein	*Candidatus Solibacter usitatus*	43 (57)
*cel*P 5	5160	5384	225	Hypothetical protein	*Chthoniobacter flavus*	38(63)
*cel*P 6	5368	6456	1089	Putative C4-dicarboxylate transporter	*Geobacillus thermodenitrificans*	61 (73)
*cel*P 7	6681	6538	144	Hypothetical protein	No similarity found	
*cel*P 8	8954	9964	1011	Inosine/uridine-preferring nucleoside hydrolase	*Roseiflexus* sp.	77 (86)
*cel*P 9	9974	11167	1194	Hypothetical protein	*Rhodobacter capsulatus*	29 (51)
*cel*P 10	11558	17950	6393	Type II secretory pathway pullulanase PulA and related glycosidase-like protein	*Streptosporangium roseum*	51 (64)
*cel*P 11	18123	18275	153	Hypothetical protein	No similarity found	

*cel*P 12	18366	20897	2532	Sensory box histidine kinase/response regulator	*Geobacter sulfurreducens*	37 (56)
*cel*P 13	21172	23613	2442	Glycosyl transferase family 39	*Roseiflexus castenholzii*	27 (38)
*cel*P 14	23688	25979	2292	Glycosyl transferase, family 39	*Roseiflexus sp.*	49 (60)
*cel*P 15	26126	28267	2142	glycosyl transferase family 39	*Roseiflexus castenholzii*	49 (61)
*cel*P 16	28630	31170	2541	Endoglucanase	*Amycolatopsis mediterranei*	58 (70)
*cel*P 17	31531	31313	219	Hypothetical protein RoseRS_2661	*Roseiflexus* sp.	73 (81)
*cel*P 18	35922	31534	4389	Adenylyl cyclase class-3/4/guanylyl cyclase	*Roseiflexus* sp.	56 (70)

Table 8. Predicted open reading frames in fosmid pCelD

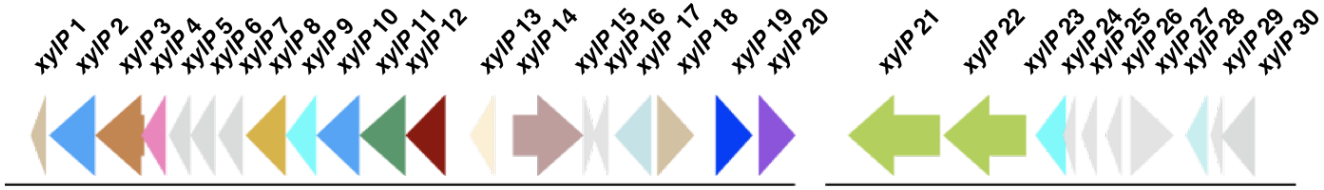


Fosmid pCelD (34284 bp)

Feature ID	Start	Stop	Length (bp)	Putative Function (Blast)	Organism	Identity (Similarity)
*cel*D 1	656	4330	3675	Glycoside hydrolase family 5	*Ktedonobacter racemifer*	65 (74)
*cel*D 2	4538	4425	114	Hypothetical protein	No similarity found	
*cel*D 3	5693	4698	996	Aldo/keto reductase	*Geobacter bemidjiensis*	74 (85)
*cel*D 4	6360	6512	153	Hypothetical protein	No similarity found	
*cel*D 5	7172	7627	456	Hypothetical protein	*Tsukamurella paurometabola*	45 (63)
*cel*D 6	7578	8381	804	Hypothetical protein	*Rhodococcus opacus*	40 (55)
*cel*D 7	8755	8369	387	Glyoxalase/bleomycin resistance protein/dioxygenase superfamily protein	*Pseudoalteromonas atlantica*	69 (82)
*cel*D 8	9247	8801	447	Conserved protein of unknown function	*Pseudoalteromonas haloplanktis*	40 (62)
*cel*D 9	9912	9316	597	Bifunctional deaminase-reductase domain protein	*Ktedonobacter racemifer*	73 (86)
*cel*D 10	12592	10175	2418	PAS/PAC sensor protein	*Candidatus Desulforudis audaxviator*	53 (69)
*cel*D 11	12846	13694	849	Peptidase S1 and S6, chymotrypsin/Hap	*Polaromonas* sp.	42 (61)

*cel*D 12	13790	14800	1011	ATPase associated with various cellular activities	*Polaromonas sp.*	70 (83)
*cel*D 13	14775	15668	894	Hypothetical protein	*Azoarcus* sp.	60 (74)
*cel*D 14	15682	16680	999	Von Willebrand factor type A	*Variovorax paradoxus*	42 (62)
*cel*D 15	16853	17878	1026	Von Willebrand factor type A	*Frankia* sp.	47 (64)
*cel*D 16	17979	18416	438	Peroxiredoxin	*Citrobacter rodentium*	46 (62)
*cel*D 17	19753	18449	1305	2-oxoglutarate dehydrogenase, E2 subunit, dihydrolipoamide succinyltransferase	*Sphaerobacter thermophilus*	55 (70)
*cel*D 18	22043	19755	2289	2-oxoglutarate dehydrogenase, E1 subunit	*Roseiflexus castenholzii*	54 (71)
*cel*D 19	22970	22797	174	Hypothetical protein	No similarity found	
*cel*D 20	23122	24933	1812	Hypothetical protein	*Cellvibrio japonicus*	35 (54)
*cel*D 21	26918	25044	1875	Carbamoyltransferase	Desulfovibrio sp.	67 (79)
*cel*D 22	27076	26927	150	Hypothetical protein	Desulfovibrio magneticus	75 (89)
*cel*D 23	27491	27108	384	Hypothetical protein	Desulfovibrio magneticus	42 (71)
*cel*D 24	28747	27569	1179	Galactokinase	Roseiflexus castenholzii	48 (61)
*cel*D 25	29796	28744	1053	Galactose-1-phosphate uridylyltransferase	Rhodothermus marinus	63 (71)
*cel*D 26	31364	29955	1410	Endo-1,4-beta-xylanase	Streptomyces bingchenggensis	46 (59)
*cel*D 27	33877	31526	2352	Alpha-N-arabinofuranosidase	Paenibacillus sp.	34 (49)
*cel*D 28	34253	33978	276	Hypothetical protein	No similarity found	

Table 9. Predicted open reading frames in fosmid pXylP



Fosmid pXylP (31,200 bp)

Feature ID	Start	Stop	Length (bp)	Putative Function (Blast)	Organism	Identity (Similarity)
*xyl*P 1	456	28	429	Transcriptional regulator (MerR)	*Kineococcus radiotolerans*	59 (76)
*xyl*P 2	1755	523	1233	Putative carbamoyl-phosphate synthase	*Shewanella piezotolerans*	33 (50)
*xyl*P 3	3040	1760	1281	Major facilitator superfamily MFS_1	*Micromonospora* sp.	33 (47)
*xyl*P 4	3621	3037	585	Lysine decarboxylase	*Streptomyces pristinaespiralis*	59 (71)
*xyl*P 5	4277	3732	546	Hypothetical protein		
*xyl*P 6	4985	4284	702	Hypothetical protein	*Conexibacter woesei*	33 (47)
*xyl*P 7	5728	5081	648	Hypothetical protein	*Streptomyces roseosporus*	
*xyl*P 8	6882	5770	1113	Threonine synthase	*Achromobacter piechaudii*	38 (55)
*xyl*P 9	7679	6879	801	Precorrin-3B C17-methyltransferase	*Chlorobium tepidum*	33 (50)
*xyl*P 10	8866	7676	1191	Carbamoyl-phosphate synthase	*Shewanella piezotolerans*	27 (44)
*xyl*P 11	10057	8867	1191	Putative fusion protein	*Burkholderia pseudomallei*	28 (40)
*xyl*P 12	11105	10074	1032	Cysteine synthase	*Streptomyces roseosporus*	47 (66)

*xyl*P 13	12466	11747	720	Dimodular nonribosomal peptide synthetase	*Bacillus cereus*	54 (66)
*xyl*P 14	12942	14843	1902	Amidase	*Pseudomonas syringae*	50 (63)
*xyl*P 15	14840	15010	171	Hypothetical protein		
*xyl*P 16	15480	15106	375	Hypothetical protein	*Streptomyces coelicolor*	72 (81)
*xyl*P 17	16599	15616	984	Putative zinc-binding oxidoreductase	*Streptomyces coelicolor*	88 (92)
*xyl*P 18	16796	17803	1008	Transcriptional regulator, AraC family	*Streptomyces sp.*	87 (90)
*xyl*P 19	18341	19363	1023	Xylanase	*Streptomyces* sp.	84 (92)
*xyl*P 20	19531	20460	930	Acetylxylan esterase	*Streptomyces griseoflavus*	79 (89)
*xyl*P 21	22847	20409	2439	Magnesium or manganese-dependent protein phosphatase	*Streptomyces hygroscopicus*	50 (63)
*xyl*P 22	25131	22969	2163	Magnesium or manganese-dependent protein phosphatase	*Streptomyces avermitilis*	46 (61)
*xyl*P 23	26234	25386	849	Putative methyltransferase	*Streptomyces avermitilis*	93 (95)
*xyl*P 24	26474	26295	180	Hypothetical protein		
*xyl*P 25	27029	26643	387	Hypothetical protein	*Streptomyces bingchenggensis*	82 (89)
*xyl*P 26	27690	27325	366	Hypothetical protein	*Streptomyces pristinaespiralis*	43 (55)
xylP 27	27956	29104	1149	Hypothetical protein	*Streptomyces roseosporus*	69 (78)
*xyl*P 28	29973	29419	555	Dihydrofolate reductase	*Sanguibacter keddieii*	42 (58)
*xyl*P 29	30098	29982	117	Hypothetical protein		
*xyl*P 30	31148	30243	906	Hypothetical protein		

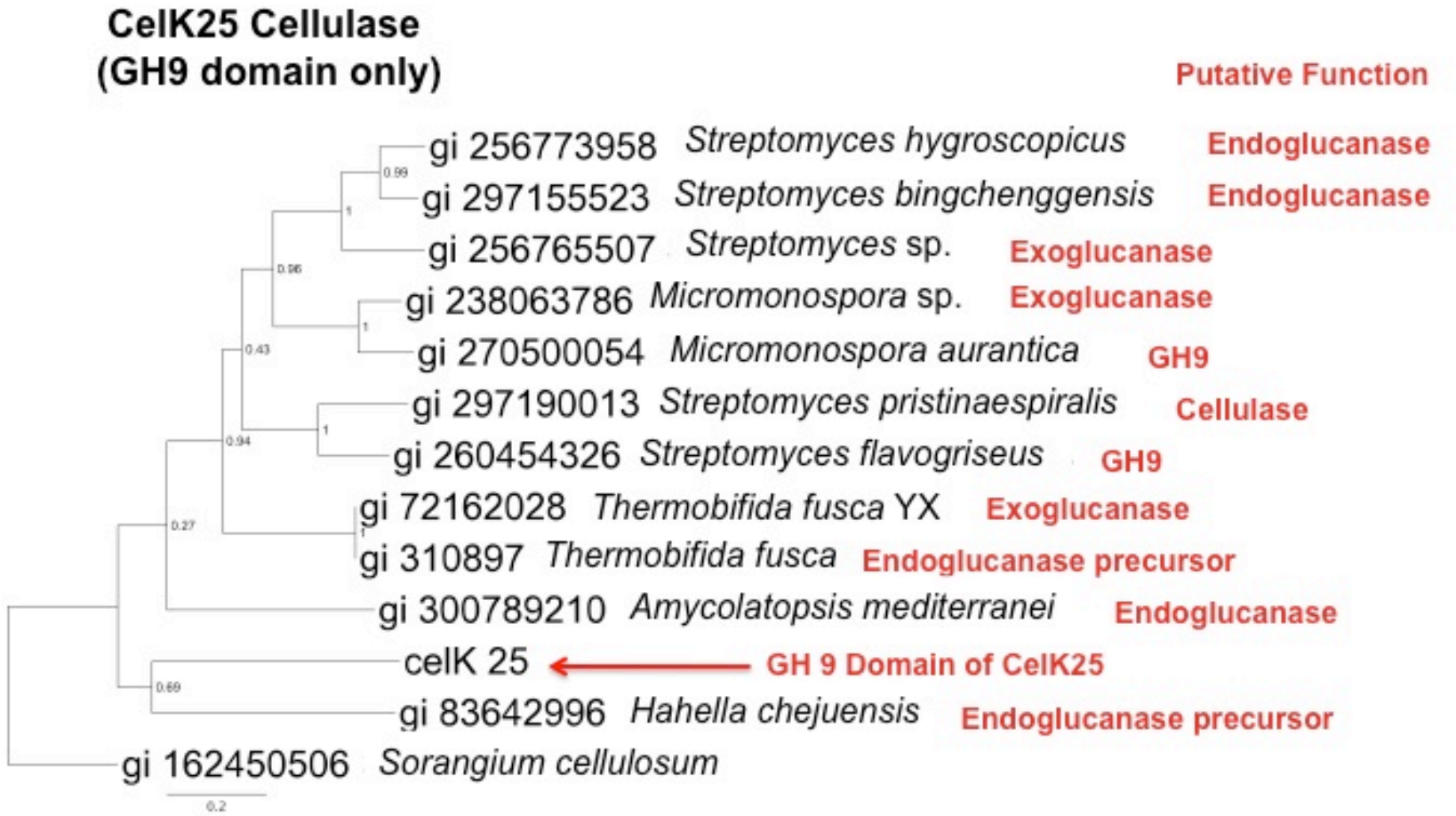


Figure 7. Phylogenetic tree of proteins related to CelK25 domain GH9

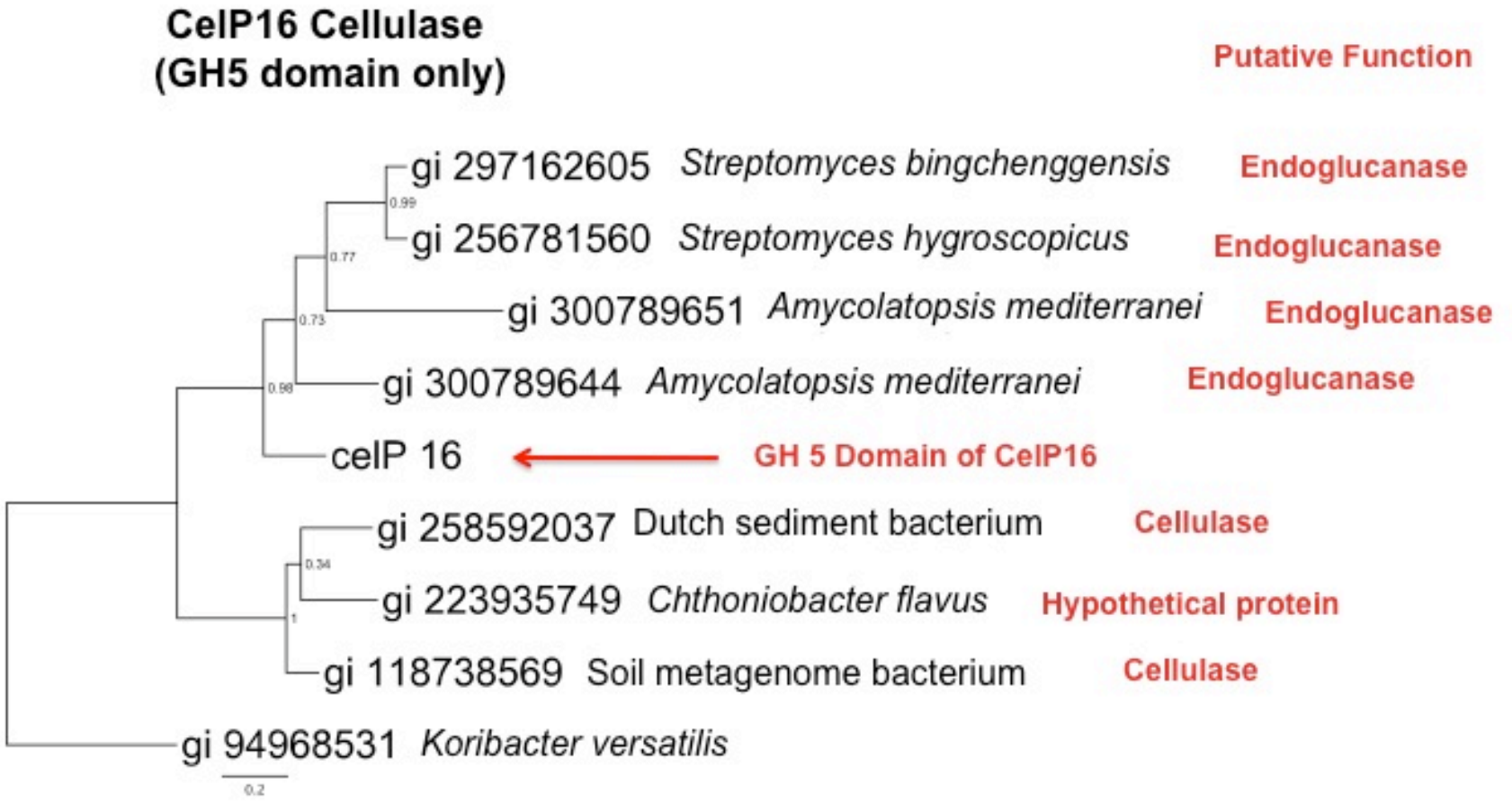


Figure 8. Phylogenetic tree of proteins related to CelP16 domain GH5

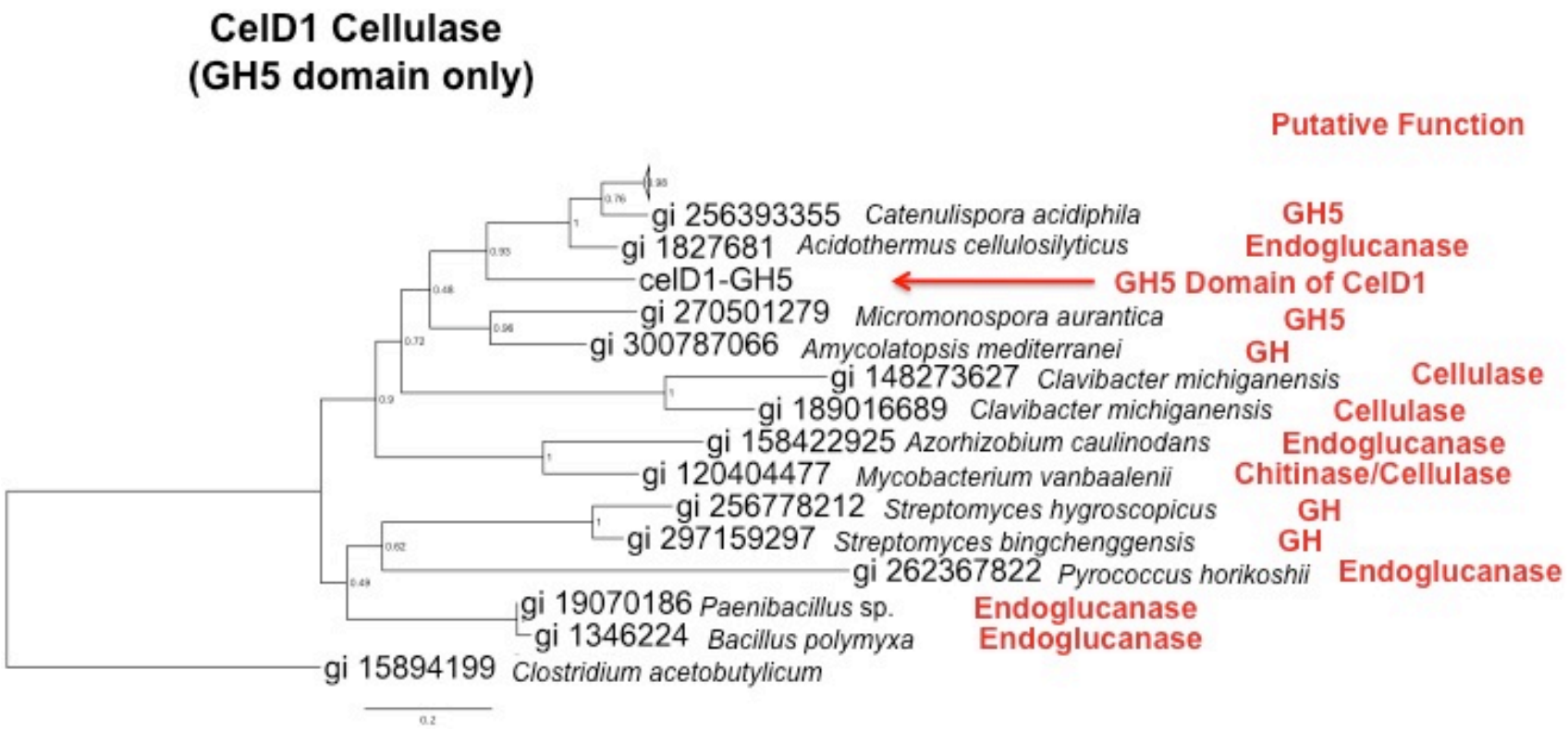


Figure 9. Phylogenetic tree of proteins related to CelD1 domain GH5

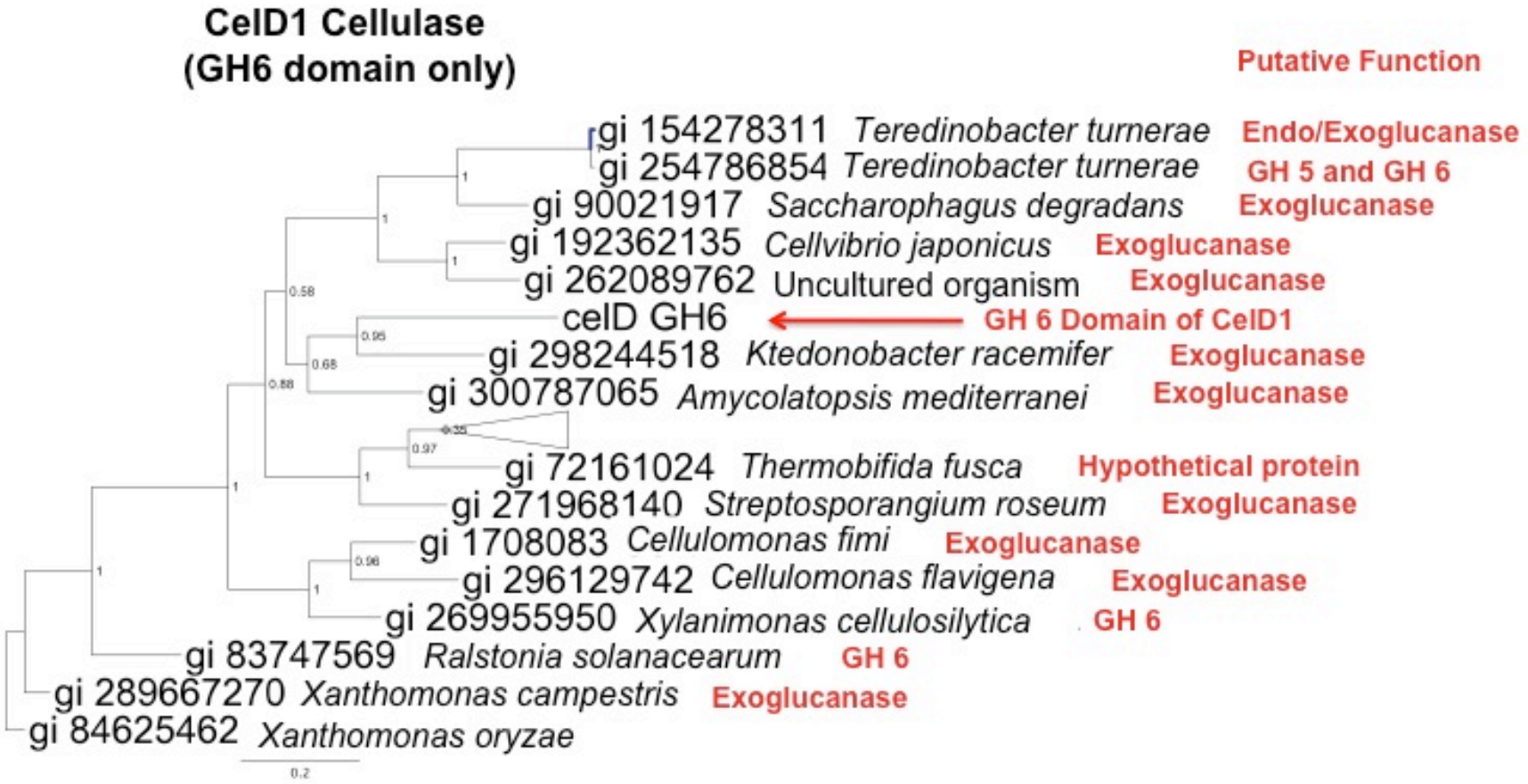


Figure 10. Phylogenetic tree of proteins related to CelD1 domain GH6

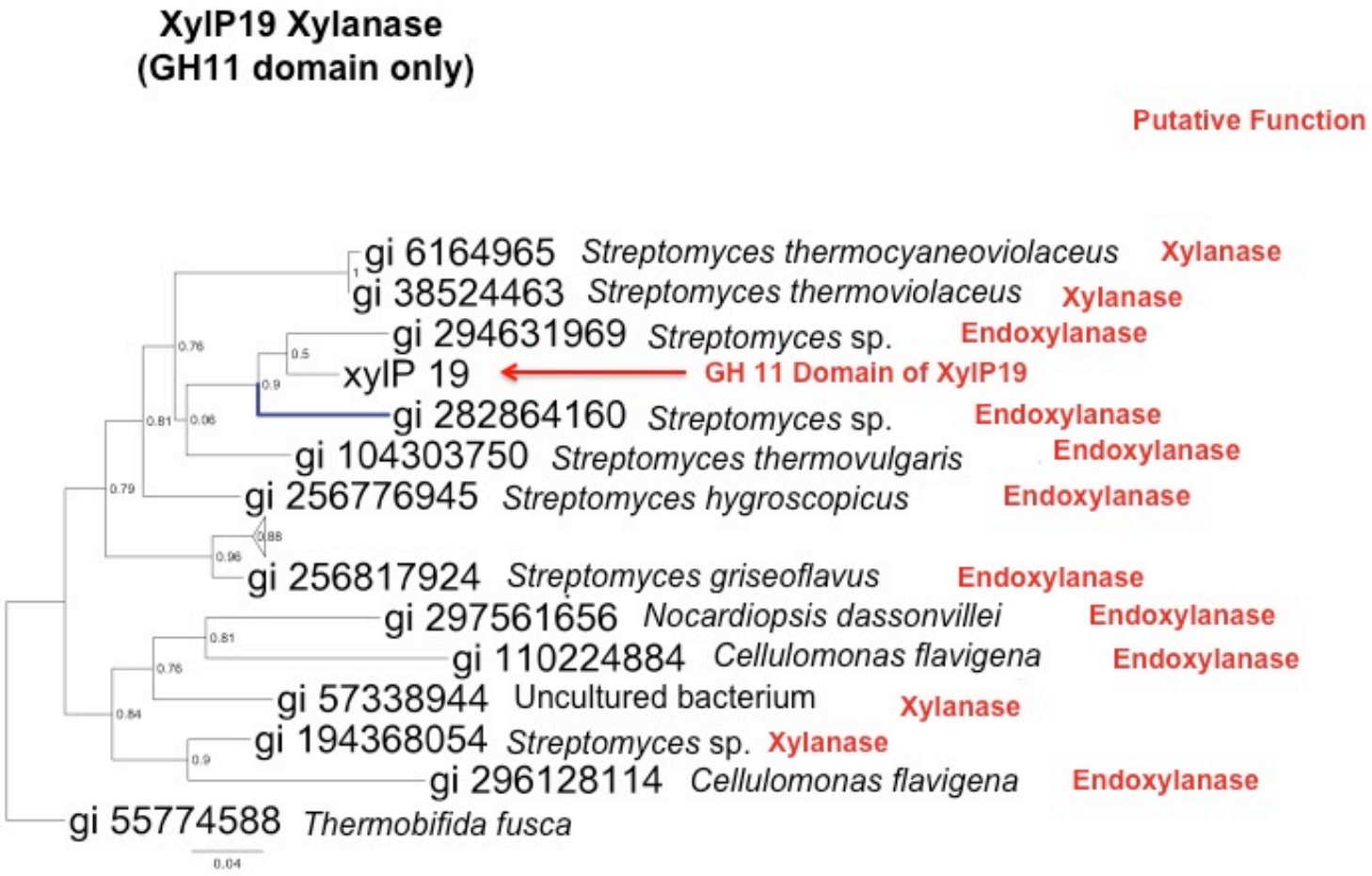


Figure 11. Phylogenetic tree of proteins related to XylP19 domain GH6

Chapter 3. Construction of an Autodisplay System in *Geobacter sulfurreducens* and *Shewanella oneidensis*

The need for more efficient and sustainable ways to produce energy, food and chemicals has driven scientists and engineers to redesign biological systems. Synthetic biology is an emerging field that creates biological building blocks to redesign existing biological processes. The goal of this project was to construct a synthetic system to display proteins on the surfaces of *Geobacter sulfurreducens* and *Shewanella oneidensis* using an autodisplay approach. A synthetic construct containing a secretion signal from *Vibrio cholerae* (CtxB), a *lacZ* gene from *E. coli* (passenger domain), and an AIDA-I autotransporter from *E. coli* (carrier domain) was designed in silico and synthesized. The resulting genetic construct was used to transform *E. coli*, *G. sulfurreducens* and *S. oneidensis*. The β-galactosidase activity in cytoplasmic/periplasmic and membrane fractions was determined for the three different organisms. The results obtained from the enzymatic activity assays indicated that the AIDA-I autodisplay system caused an increase in the accumulation of β-galactosidase in the membrane of *E. coli*, *G. sulfurreducens* and *S. oneidensis*. The β-galactosidase activity of *G. sulfurreducens* biofilms grown on a poised electrode demonstrated the potential application of the autodisplay system in electrochemical cell biocatalysis.

1. Introduction

Dissimilatory metal reducing bacteria (DMRB) are characterized by their ability to couple the oxidation of organic acids with the reduction of metal oxides such as iron (Fe^{+3}) and manganese (Mn^{+4}) (97, 150). The microbial metabolism of DMRB facilitates the cycling of metals and carbon in the environment (100). Furthermore, the microbial reduction of metal oxides and electrodes has practical applications in the area of bioremediation and electricity production (95, 96). *Shewanella oneidensis* and *Geobacter sulfurreducens* are the most extensively studied dissimilatory iron (Fe^{+3}) reducing bacteria. Besides using insoluble iron (Fe^{+3}), these organisms can support cellular growth by using poised electrodes as terminal electron acceptors. The ability of *Shewanella* and *Geobacter* to use insoluble terminal electron acceptors relies on electron transfer (ET) mechanisms localized in the cell membrane and beyond the cell surface (28, 151). The ability to place enzymes outside the cell, where they could interface with this unique electron transfer machinery, could enable new redox-based biocatalysis.

Recent reports have examined the transmembrane and extracellular electron transfer pathways utilized by *S. oneidensis* (28, 150, 151). The electron transfer mechanisms occur through a series of redox-active multi-heme *c*-type cytochromes that span the membrane. MtrF, a recently crystallized decaheme *c*-type cytochrome, is an example of a redox-active structure localized on the cell surface of *Shewanella* that facilitates electron transfer (24).

In contrast, the electron transfer mechanisms of *G. sulfurreducens* are not completely understood. It has been demonstrated that ET in *G. sulfurreducens* relies on cytochromes and a complex extracellular matrix that extends beyond the outer membrane entrapping redox-active *c*-type cytochromes to wire the electrons to the insoluble terminal acceptors (83, 136).

Considering the importance of the cell surface of *S. oneidensis* and *G. sulfurreducens,* it is critical to develop genetic tools for understanding the fundamentals of extracellular electron transfer mechanisms in DMRB. Furthermore, there is a potential application of the redox active space created throughout the microbial biofilm in redox biotechnology. This study is focused on the development of a genetic system to display proteins on the surface of *S. oneidensis* and *G. sulfurreducens*.

Surface display systems have been extensively used in *E. coli* and yeast for the production of vaccines, development of biomaterials, whole-cell biocatalysis, and other applications (41, 76, 77, 184). The typical surface display systems are based on outer membrane proteins (i.e. OmpA and LamB) and they are usually limited to display short peptides and small proteins (142, 170). In contrast, the use of autotransporters has shown to overcome this limitation by localizing proteins of up to 63-kDa on the surface of *E. coli* (85, 170). Autodisplay is a current surface display system based on autotransporter proteins to drive protein secretion throughout the membranes and into the cell surface

(59). In contrast to outer membrane proteins, autotransporter proteins can drive their own secretion throughout the membranes (49, 50).

The **a**dhesin-**i**nvolved in **d**iffused **a**dherence (AIDA-I) autotransporter protein has shown to facilitate the display of a wide variety of heterologous proteins on the surface of *E. coli* (58, 60, 81, 85). This report shows a strategy to access the surface of *Shewanella oneidensis* and *Geobacter sulfurreducens* using the AIDA-I autodisplay system. The ability to display functional proteins was evaluated by testing the enzymatic activity of cytoplasmic/periplasmic and membrane fractions. In addition, this report will demonstrate the application of the autodisplay system in *G. sulfurreducens* for performing whole-cell biocatalysis in electrochemical cells.

2. Materials and Methods

2.1 Bacterial Strains and Growth Conditions

The bacteria strains and plasmids used during this study are described in Table 1. *E. coli* strains and *S. oneidensis* MR-1 were grown aerobically in Luria-Bertani (LB) broth at 37°C and 30°C, respectively (121, 141). *G. sulfurreducens* PCA (ATTC 51573) was grown at 30°C in anaerobic mineral medium containing 20 mM acetate as the electron donor and 40 mM of fumarate as the electron acceptor (19, 108). *E. coli* and *S. oneidensis* carrying pMCS5-BB plasmids were grown aerobically in LB medium supplemented with 10 μg/ml and 20 μg/ml of gentamycin, respectively. *G. sulfurreducens* carrying pMCS5-BB was grown in anaerobic mineral medium supplemented with 20 μg/ml of gentamycin.

Table 1. Strains and plasmids used during this study

Strains, Plasmid or Primer	Characteristics or Sequence (5' - 3')	Source or restriction sites
Strains		
E. coli		
UQ950	*E. coli* DH5α λ(pir) host for cloning; F $^{-}$ Δ(*argF-lac)169* Φ80d*lacZ58*(Δ M15) *glnV44*(AS) *rfbD1 gyrA96*(NalR) *recA1 endA1 spoT1 thi-1 hsdR17 deoR* λ *pir*$^{+}$	(141)
WM3064	Donor strain for conjugation: *thrB1004 pro thi rpsL hsdS lacZ*ΔM15 RP4-1360 Δ(*araBAD*)*567* Δ*dapA1341*::[*erm pir*(wt)].	(141)
S. oneidensis MR-1	Wild type	(121)
G. sulfurreducens	Wild type (ATCC 51573)	(19)
Plasmids		
pBBR1MCS5	Mobilizable broad-host-range plasmid; Gm^{r}	(75)
pMCS5-BB	Mobilizable broad-host-range plasmid; Gm^{r}. pBBR1MCS-5 DNA fragment (NcoI-BstBI; 3,886 bp) ligated to a newly synthesized 396 bp biobrick fragment containing a pT-Lac promoter and a multiple cloning site (MCS).	This study
pMCS5BB-LacZ	pMCS5BB with a 3,090 bp DNA fragment in the XhoI –XmaI site containing the *lacZ* (NP_414878).	This study
pMCS5BB-Sec-LacZ	pMCS5BB with a 3,165 bp DNA fragment in the BamHI –XmaI site containing the secretion signal of the cholera toxin (Sec-ctxB) (ADP30957) fused to *lacZ* (NP_414878).	This study
pMCS5BB-Sec-LacZ-AidaI	pMCS5BB with a 4,515 bp DNA fragment in the BamHI–SpeI site containing the Sec-ctxB fused to *lacZ* and the AIDA-I autotransporter (Q03155).	This study
Primers		
F-LacZ	gcactcgagatgcaccatcaccatcaccatacc	Xho I
R-LacZ	tccccccgggtcacttctggcaccagaccagctgata	Xma I
F-Sec	gcaggatccatgattaagctcaagtttggcgtt	BamHI

2.2 Design and Construction of Multi Cloning Site (MCS) for pBBR1MCS5

A biobrick containing the pT-lac promoter and a multi-cloning site (MCS) was designed in Gene Designer Software and synthetically synthesized (DNA 2.0, California USA). The genetic features of the biobrick are described in Figure 1A. The biobrick was ligated in between the NcoI and BstBI sites of pBBR1MCS5 plasmid. The resulting engineered plasmid was called pMCS5-BB and it is described in Figure 1B.

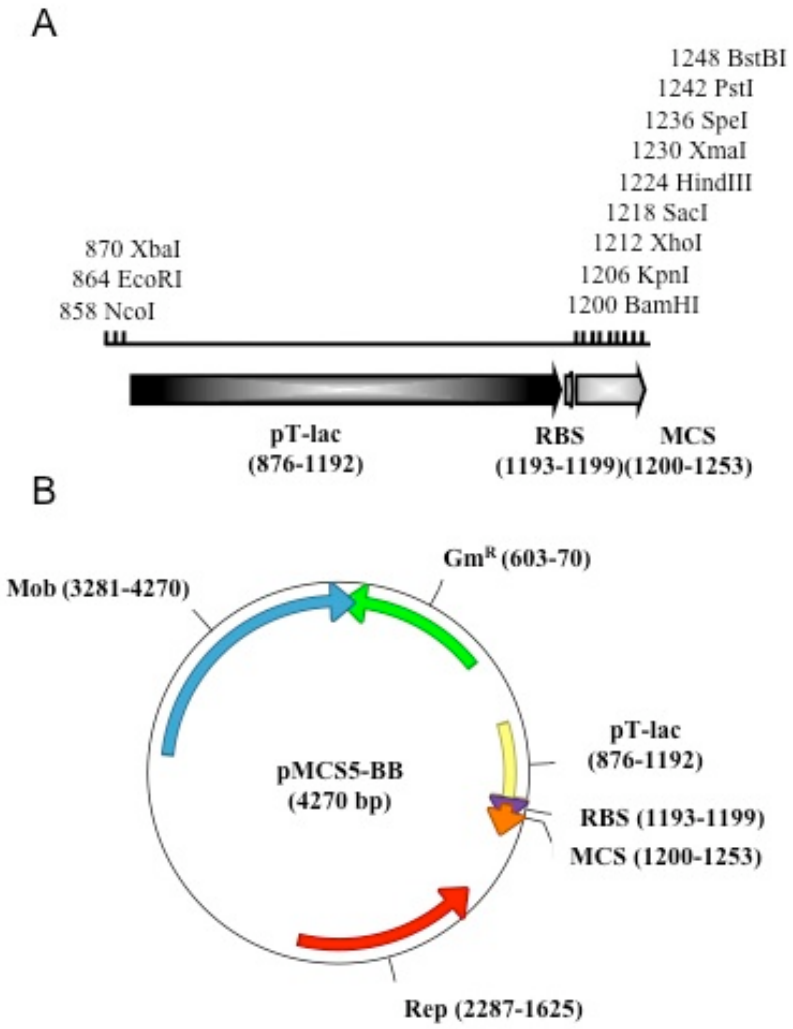


Figure 1. New genetic features found in pMCS5-BB plasmid. **A.** Biobrick element containing the p-TLac promoter, ribosomal binding site (RBS) and multi-cloning site (MCS). **B.** Map of pMCS5-BB plasmid harboring the new biobrik element, the mobilization gene (Mob), replication cassette (Rep), and the gentamycin cassette (Gm^R).

2.3 Designed and Synthetic Construction of Autodisplay System

The autodisplay system was designed as previously described by Jose and coworkers (2004) (60). The genetic features of the constructs used during this study are described in Figure 2. The secretion signal composed of amino acid 1 to amino acid 23 (1-23) of cholera toxin B subunit (CtxB) (accession number: ADP30957), the amino acid sequence (1-1024) of β-galactosidase (LacZ) (accession number: NP_414878) and the truncated amino acid sequence (840-1286) of the AIDA-I autotransporter (accession number: Q03155) were obtained from GeneBank. The autodisplay genetic system was designed using Gene Designer Software 2.0.138 and its codon usage was optimized for *G. sulfurreducens*. The autodisplay construct (pMCS5BB-Sec-LacZ-AidaI) was synthesized (DNA 2.0, CA, USA). The resulting construct was first transformed into *E. coli* WM3064 donor strain and it was used for mating *G. sulfurreducens* and *S. oneidensis* (136). The plasmid carrying the *lacZ* gene (pMCS5BB-LacZ) was constructed by PCR amplifying from pMCS5BB-Sec-LacZ-AidaI with forward primer (F-LacZ) and reverse primer (R-LacZ) using the following thermocycler parameters: 95°C for 4 minutes followed by 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 3 minutes and a final extension at 72°C for 10 minutes reaction. The plasmid pMCS5BB-Sec-LacZ was constructed by PCR amplifying from pMCS5BB-Sec-LacZ-AidaI with forward primer (F-Sec) and reverse primer (R-LacZ) using the same thermocycler conditions as previously described. The resulting ≈3-Kb DNA fragments were ligated in between the XhoI and XmaI or BamHI and XmaI sites of pMCS5-BB plasmid.

The construct pMCS5BB-Sec-LacZ-AidaI was designed to target β-galactosidase to the cell surface and contained the *lacZ* gene, flanked by a secretion signal from cholera toxin (CtxB) and the AIDA-I autotransporter. The predicted molecular weight of this hybrid protein was 168-kDa. The construct pMCS5BB-Sec-LacZ was designed to target β-galactosidase to the periplasm, and contained the secretion signal (CtxB) fused to the N-terminal of *lacZ*. The construct called pMCS5BB-LacZ was designed to express the β-galactosidase in the cytoplasm (37). The constitutive pT-Lac promoter drove the expression of all hybrid proteins. These three constructs and the empty plasmid pMCS5-BB were used to independently transform *E. coli*, *S. oneidensis* and *G. sulfurreducens* cells as previously described.

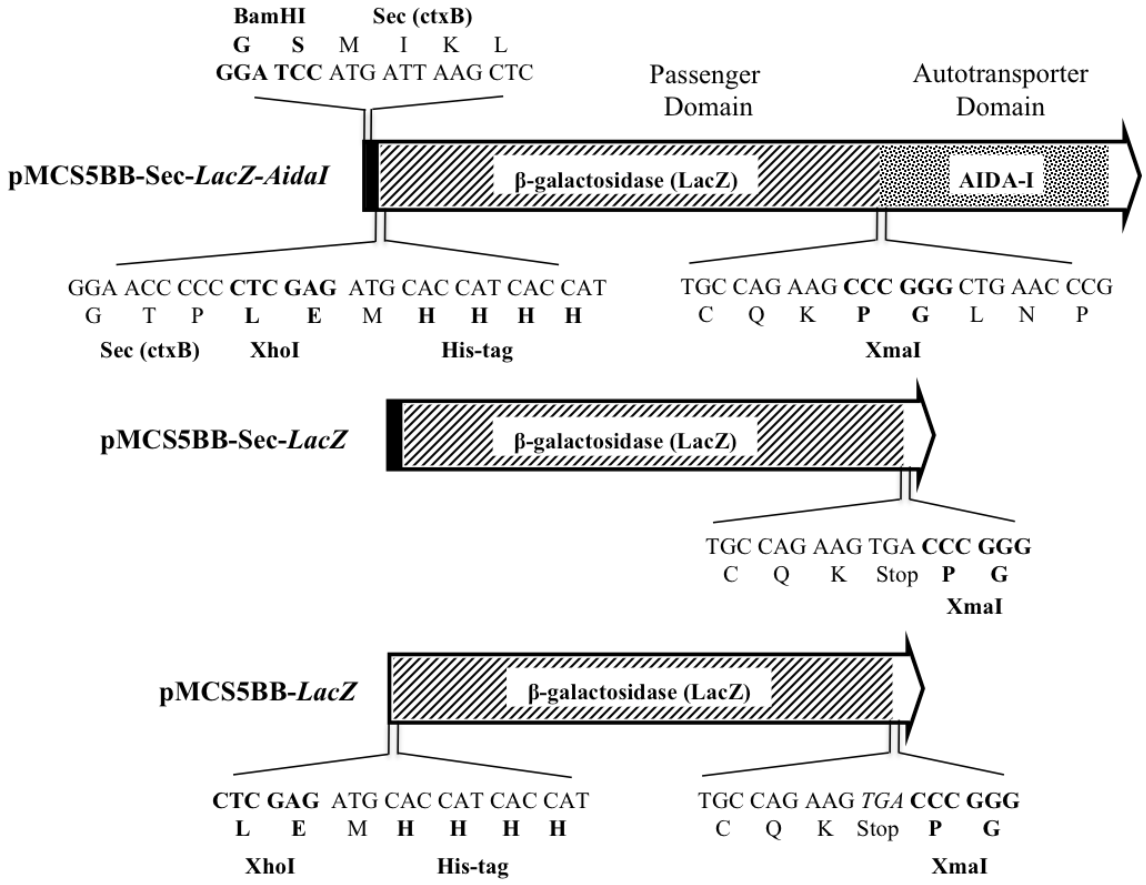


Figure 2. Genetic features of gene fusions and protein hybrids examined during this study.

2.4 Preparation of Cytoplasmic and Membrane Fractions

E. coli UQ950 and *S. oneidensis* MR-1 cells were grown in 100 ml of medium overnight as previously described (Section 3.1). The cells of *G. sulfurreducens* were grown overnight in 1 L of mineral medium as previously described (Section 3.1). The resulting cells were harvested by centrifugation at 6000 x *g*. The cell pellets were washed twice with 30 ml of potassium phosphate buffer (100 mM, pH 7.2) and resuspended in fresh potassium phosphate buffer (100 mM, pH 7.2) supplemented with 0.5 mM of PMSF and DNAse I crystals. The resulting cell suspensions were sonicated. To remove intact cells, the cell lysates were centrifuged twice once at 13,000 x *g* and once 20,000 x *g* for 15 minutes each. Samples of clarified lysates were taken for protein and enzymatic assays. The clarified cell lysates were centrifuged at 100,000 x *g* for 1 hr to separate the cytoplasm and periplasm from the membrane. The resulting supernatant was considered the cytoplasmic/periplasmic fraction. The membrane pellets were washed with 25 ml of potassium phosphate buffer (100 mM, pH 7.2) and they were centrifugated at 100,000 x *g* for an additional hour. The membrane pellets were resuspended in 1 ml of potassium phosphate buffer (100 mM, pH 7.2) using 18-gauge needle. The lysates, cytoplasmic/periplasmic and membrane fractions were subjected to protein determination and enzymatic β-galactosidase assays.

2.5 Protein Determination and Analysis

Protein samples and standards were diluted in 0.2 N NaOH and heated at 65°C for 30 minutes. The protein concentration in the resulting samples was determined using the bicinchoninic (BCA) assay kit (Thermo Scientific, Rockford, IL, USA).

2.6 Determination of β-galactosidase Activity

The enzymatic β-galactosidase (LacZ) activity was determined using a modified version of the ONPG assay developed by Miller (117). Briefly, a reaction containing known amounts of protein diluted in 225 µl of potassium phosphate buffer (100 mM, pH 7.2) and 75 µl of 10 mM *o*-nitrophenyl-β-D-galactoside (ONPG) was combined in a 96-well microplate. The β-galactosidase activity was determined by measuring the absorbance of the reactions at 410 nm (Abs_{410}) over time. The specific β-galactosidase activity was expressed as Units (µMol min^{-1} mg^{-1} protein) or mUnits (0.001 Units). The extinction coefficient used to determine the concentration of released nitrophenol was 1.46 mM^{-1} cm^{-1} and it was experimentally calculated under the same assaying conditions. The activity recovery percentages were calculated based on the sum of the total activity in soluble (cytoplasmic/periplasmic) and membrane fractions.

2.7 Biocatalysis in Electrochemical Cells

The electrochemical bioreactors were assembled as previously described (108). The three-electrode bioreactors were incubated at 30°C in a water bath and continuously purged with N_2/CO_2. Anaerobic bioreactors containing 10 ml of mineral medium

(supplemented with 40 mM acetate and without electron acceptor) were connected to a 16-channel potentiostat with EZ-Lab Software (VMP; Bio-Logic, Knoxville, TN) to run chronoamperometry routines. The anaerobic reactors were inoculated with 50% (vol/vol) *G. sulfurreducens* cells (OD_{600} 0.5-0.55) and mineral medium (supplemented with 40 mM acetate and without electron acceptor) at a potential of +0.24V versus the standard hydrogen electrode (SHE). The cells were incubated until reaching plateau and then mineral medium was removed and replaced with 7.5 ml of new mineral medium containing 20 mM acetate. Once the cells had recovered from the medium swap (≈1 hr), 2.5 ml of 10 mM ONPG solution prepared in mineral medium (supplemented with 20 mM electron donor and without electron acceptor) was added to the bioreactors. The β-galactosidase activity was determined by taking 300 µl samples of supernatant every hour and by calculating the changes in absorbance at 410 nm (Abs_{410}) over time in a 96-well plate spectrophotometer. After the last sample was taken, the electrodes were removed from the bioreactors and placed into tubes containing 1 ml of 0.2 N NaOH for protein analysis. Resulting tubes were vortex and heated at 65°C for 30 minutes in a water bath. The protein concentration of resulting samples was determined using the bicinchoninic (BCA) assay kit as previously described. The specific β-galactosidase activity was expressed as Units (µMol min^{-1} mg^{-1} protein) or mUnits (0.001 Units).

3. Results

3.1 Design and Construction of Autodisplay System in *E. coli, S. oneidensis,* and *G. sulfurreducens*.

The genetic components required for displaying β-galactosidase on the surface of the *E. coli*, *S. oneidensis* and *G. sulfurreducens* were examined using three independent plasmids (Figure 2). Each of these plasmids was designed to target the expression of β-galactosidase to three different compartments of the cell including the cell surface, the periplasm and the cytoplasm.

3.2 Overexpression of β-galactosidase in *E. coli, S. oneidensis,* and *G. sulfurreducens*

The *E. coli*, *S. oneidensis* and *G. sulfurreducens* harboring each of the four plasmids were grown and processed as previously described (Section 3.1 and Section 3.4). The resulting cell lysates were fractionated into soluble (cytoplasmic/periplasmic) and membrane fractions. Each fraction was subjected to protein determination and was assayed for β-galactosidase activity. Detailed information about the protein concentration and specific activity of each fraction can be found in the supplementary information Table 2-4.

The specific β-galactosidase activities of the cell lysates are described in Figure 3. The highest specific β-galactosidase activity was obtained in *E. coli* lysates followed by *S. oneidensis* and *G. sulfurreducens*. There was no significant difference in the specific

activity of *E. coli* cells expressing cytoplasmic (869 ± 110 mUnits mg $^{-1}$ protein) and autodisplayed β-galactosidase (1140 ± 199 mUnits mg $^{-1}$ protein). Meanwhile, the addition of the secretion signal caused a 50% decrease in the specific β-galactosidase activity of *E. coli* lysates.

The activity of lysates produced from *S. oneidensis* cells expressing cytoplasmic β-galactosidase was 233 ± 57 mUnits mg $^{-1}$ protein. The presence of the secretion signal caused a 93% decrease in the specific β-galactosidase activity of *S. oneidensis* lysates. The activity of *S. oneidensis* lysates produced from cells expressing autodisplayed β-galactosidase was 50 ± 5 mUnits mg $^{-1}$ protein.

The activity of *G. sulfurreducens* lysates produced from cells expressing β-galactosidase in the cytoplasm was 5 ± 0.5 mUnits mg $^{-1}$ protein. The addition of the secretion signal resulted in a 60% increase in the specific β-galactosidase activity. The highest specific activity was obtained in lysates produced from *G. sulfurreducens* cells expressing the autodisplayed β-galactosidase (16 ± 4 mUnits mg $^{-1}$ protein).

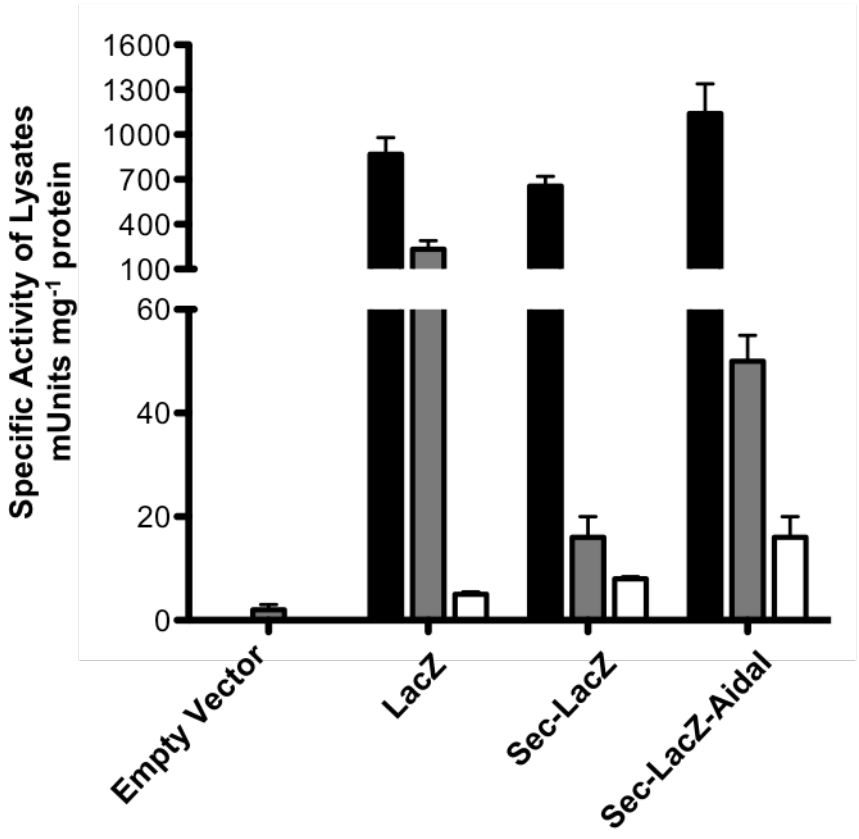


Figure 3. Specific β-galactosidase activity of lysates produced from ■ *E. coli*, ▩ *S. oneidensis*, and □ *G. sulfurreducens* cells harboring empty vector (pMCS5BB), LacZ (pMCS5BB-LacZ), Sec-LacZ (pMCS5BB-Sec-LacZ) and Sec-LacZ-AidaI (pMCS5BB-Sec-LacZ-AidaI) plasmids.

3.3 Specific β-galactosidase Activity of Membrane Fractions Recovered from *E. coli, S. oneidensis,* and *G. sulfurreducens*

The β-galactosidase activities of membrane fractions recovered from *E. coli* cells are described in Figure 4A. When the secretion signal was fused to the *lacZ* gene (pMCS5BB-Sec-LacZ), the specific activity increased from 0.15 to 1.21 mUnits mg $^{-1}$ protein. Fusion of the autotransporter protein at the C-terminal of the LacZ protein increased the specific activity of the membrane fraction 11.4-fold. As expected, the fractions obtained from cells harboring the empty vector did not show significant β-galactosidase activity.

The percentages of β-galactosidase activity recovered in the membrane fractions of *E. coli* are described in Figure 4D. *E. coli* cells expressing cytoplasmic β-galactosidase (pMCS5BB-LacZ) accumulated 0.7 % of their total activity in their membrane fraction. When the secretion signal was present, the ratio of activity increased 25-fold. The addition of the autotransporter protein only increased the accumulation of β-galactosidase in the membrane fraction by 3-fold compared to the cytoplasmic construct.

The specific activities obtained from membrane fractions of *S. oneidensis* are described in Figure 4B. The specific activity of membrane fractions recovered from cells expressing cytoplasmic β-galactosidase was 95 ± 33 mUnits mg $^{-1}$ protein. When the secretion signal was fused to the *lacZ* gene, the specific activity of membrane fractions decreased 2-fold. In contrast, the addition of the autotransporter resulted in a 1.5-fold

increase in the specific β-galactosidase activity of the membrane fraction. There was no activity observed in the membrane fraction of cells harboring empty plasmid.

The percentages of β-galactosidase activity recovered in the membrane fractions of *S. oneidensis* are described in Figure 4E. The cells expressing cytoplasmic β-galactosidase accumulated 9% of their total activity in their membrane. When the secretion signal was present, the amount of β-galactosidase activity accumulated in the membrane increased 7-fold. The addition of the autotransporter protein resulted in the recovery of 70% of the total β-galactosidase activity in the membrane fraction.

The activity results obtained from *G. sulfurreducens* cells are described in Figure 4C. There was not significant specific activity in membrane fractions recovered from cells expressing cytoplasmic β-galactosidase. When the secretion signal was present, the specific activity of the membrane fractions was 19 ± 3 mUnits mg $^{-1}$ protein. The addition of the autotransporter protein at the C-terminal of the LacZ protein increased 31-fold the specific activity of the membrane fraction.

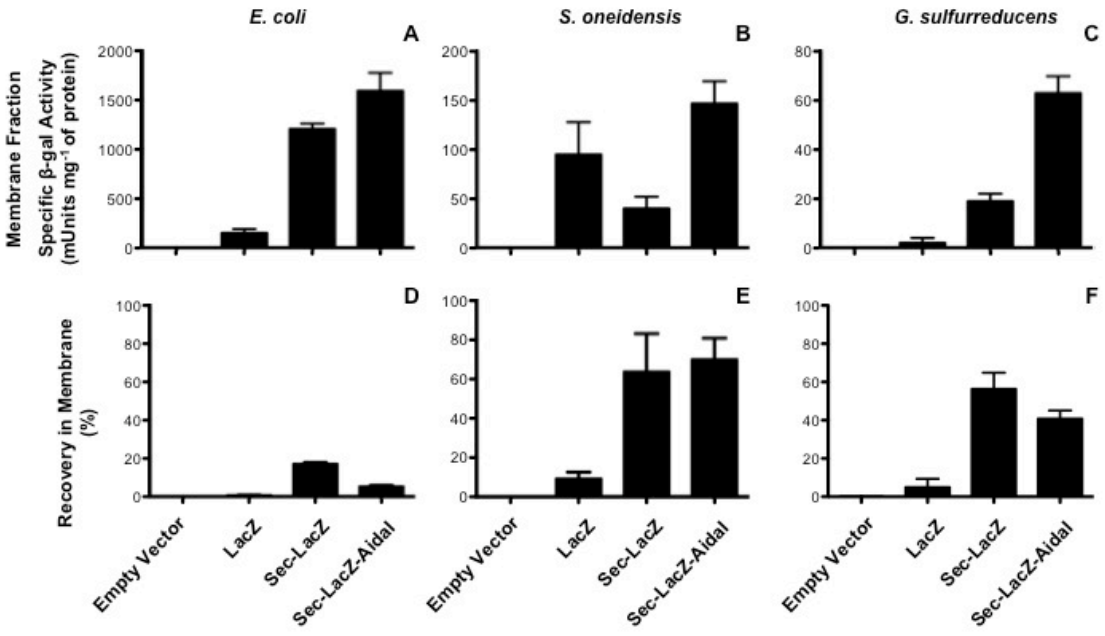


Figure 4. Specific β-galactosidase activity and percentage activity recovered in membrane fractions of *E. coli*, *S. oneidensis* and *G. sulfurreducens* cells harboring empty vector (pMCS5BB), LacZ (pMCS5BB-LacZ), Sec-LacZ (pMCS5BB-Sec-LacZ) and Sec-LacZ-AidaI (pMCS5BB-Sec-LacZ-AidaI) plasmids.

The percentages of β-galactosidase activity recovered in the membrane fractions of *G. sulfurreducens* are described in Figure 4F. The cells expressing cytoplasmic β-galactosidase accumulated 5% of their total activity in their membrane. When the secretion signal was present, the total amount of β-galactosidase recovered in the membrane increased by 14-fold. The addition of the autotransporter protein increased the amount of total β-galactosidase in the membrane to 40%.

3.4 Functional Expression of β-galactosidase in Electrochemical Cells

Geobacter sulfurreducens biofilms were grown on poised electrodes to determine whether the surface display system could potentially be used to catalyze reactions in an electrochemical cell. The specific activities of *G. sulfurreducens* cells attached to electrodes are described in Figure 5. The cells harboring the cytoplasmic β-galactosidase did not produce significant specific activity. When a secretion signal was fused to the *lacZ* gene, the specific activity increased by 9-fold. The addition of the autotransporter protein domain increased the activity by 31-fold, resulting in 62.7 ± 31.2 mUnits of β-galactosidase per mg $^{-1}$ protein.

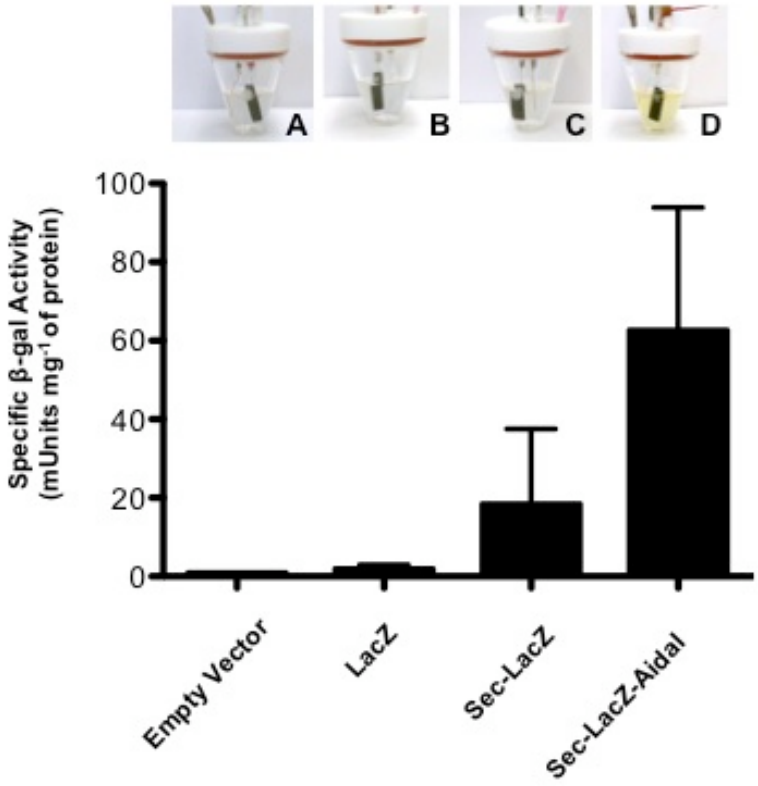


Figure 5. Specific β-galactosidase activity produced in electrochemical cells by biofilms of *G. sulfurreducens* harboring **A**. Empty vector (pMCS5BB) **B**. LacZ (pMCS5BB-LacZ plasmid) **C**. Sec-LacZ (pMCS5BB-Sec-LacZ plasmid) and **D.** Sec-LacZ-AidaI (pMCS5BB-Sec-LacZ-AidaI plasmid). The pictures of the bioreactors were taken after collecting the last aliquot for the enzymatic analysis.

4. Discussion

Dissimilatory metal reducing bacteria have a wide number of biotechnological applications including bioremediation, electricity production and electrosynthesis (10, 95, 98, 139). The development of genetic tools to access the outer membrane and the extracellular matrix is crucial for understanding the electron transfer mechanisms utilized by *G. sulfurreducens* and *S. oneidensis.* Furthermore, engineering the cell surface of these organisms represents an opportunity to improve and to expand their biocatalytic performance in electrochemical cells. This is the first report that demonstrates the recombinant expression of a functional cytoplasmic protein in the membrane of *S. oneidensis* and *G. sulfurreducens*. Furthermore, β-galactosidase is the largest protein that has been displayed by an AIDA-I autotransporter protein (170).

The localization and functionality of the β-galactosidase were demonstrated with activity assays of soluble (cytoplasmic/periplasmic) and membrane fractions. The membrane fractions recovered from cells harboring the autodisplay system (pMCS5BB-Sec-LacZ-AidaI) showed the highest specific activity (Fig4A-C). *S. oneidensis* cells harboring the autodisplay system localized the highest percentage (70%) of their total β-galactosidase in their membrane fraction followed by *G. sulfurreducens* (40%) and *E. coli* (5%) (Fig4A-C).

The percentage of total β-galactosidase activity localized on the cell surface of *E. coli* was similar to results obtained from a previous study in where the AIDA-I system was

used to display sorbitol dehydrogenase on the surface of *E. coli* (59, 60). Despite the fact that lower expression levels were found in *S. oneidensins* and *G. sulfurreducens*, these organisms seemed to accumulate higher percentages of their total β-galactosidase in their membrane than *E. coli*.

In contrast to *E. coli*, the secretion systems in *S. oneidensis* and *G. sulfurreducens* are not well understood. However, it is well documented that Fe^{+3} reducing microorganism such as *S. oneidensis* and *G. sulfurreducens*, rely on outer membrane and extracellular proteins for microbial respiration (115, 150). For example, the secretion of a putative multicopper protein (OmpB) to the cell surface of *G. sulfurreducens* has been shown to rely on the expression of OxpG, a protein predicted to be part of an atypical type II secretion system (114). The typical type II secretion system has been found involved in the secretion of *c*-type cytochromes, MtrC and OmcA, to the surface of *S. oneidensis* (149). Therefore, the presence of complex secretions system may allow these microorganisms to localize higher percentage of the total protein in their membrane.

The use of surface display systems in microbial fuel cells has been recently proposed as an alternative for enzyme-based biofuel cells (36, 163). While the enzyme-based biofuel cells used purified enzymes to catalyze redox reactions at the surface of electrodes, microbial fuel cells exploit the microbial metabolism to oxidize a wide variety of substrates including ethanol and cellulose (26, 31). Recent studies performed in yeast have used surface display of redox enzymes (glucose oxidase, bilirubin oxidase, and

laccase) for catalyzing reactions in microbial fuel cells (36, 163). The display of glucose oxidase on the surface of yeast has shown to be more stable over time than the purified enzyme preparation assayed under the same conditions (36). Furthermore, the surface display systems in yeast have allowed the development of self-regenerated hybrid microbial fuel cell (163). Since yeast does not attach nor transfer electrons directly to the electrode surface, redox mediators have to present for supporting continuous flow of electrons between the cells and the electrode surface.

G. sulfurreducens and *S. oneidensis* are platforms for microbial fuel cell biocatalysis as these organisms naturally attach and transfer electrons to electrode surfaces (6, 10). Furthermore, the flow of electrons coming out of the cells during microbial respiration could potentially be used to catalyze redox reactions. The activity results obtained from *G. sulfurreducens* cells grown in electrochemical cells demonstrated the potential application of the AIDA-I based autodisplay system for whole cell biocatalysis. Future studies should focus on demonstrating the versatility of the display system.

Surface display systems have a wide range of applications that now can be explored in Fe^{+3} reducing bacteria. The expression of enzymes to degrade insoluble substrates such as cellulose may be key for coupling the oxidation of complex carbohydrates with electricity production using a biofilms of a single organism on the electrode surface. Furthermore, the expression of redox proteins at the surface of *G. sulfurreducens* and *S. oneidensis* may help improve the electrochemical properties of these organisms.

5. Supplementary Information

Table 2. Protein concentration and specific β-galactosidase activity recovered from *E. coli* cells.

Strain	Construct	Fraction	Protein concentration (mg/ml)	Volume (mL)	Total Protein (mg)	Protein Recovery[1] (%)	Specific Activity (mUnits/mg protein)	Total mUnits	Specific Activity Recovery[2] (%)
E. coli	**MCS5BB**	Lysate	2.96 ± 0.21	7	20.7	100	0 ± 0	0	
		Soluble	2.46 ± 0.06	7	17.2	83	0 ± 0	0	0
		Membrane	1.15 ± 0.04	1	1.1	6	0 ± 0	0	0
	MCS5BB-LacZ	Lysate	4.01 ± 0.20	7	28.0	100	869 ± 110	24 373	
		Soluble	3.53 ± 0.13	7	24.7	88	984 ± 5	24 327	99.3
		Membrane	1.12 ± 0.04	1	1.1	4	145 ± 36	162	0.7
	MCS5BB-Sec-LacZ	Lysate	3.64 ± 0.32	7	25.5	100	656 ± 64	16 734	
		Soluble	2.98 ± 0.28	7	20.9	82	480 ± 95	10 010	82.9
		Membrane	1.71 ± 0.19	1	1.7	7	1 206 ± 56	2 062	17.1
	MCS5BB-Sec-LacZ-AIDA-I	Lysate	3.05 ± 0.18	7	21.38	100	1 140 ± 199	24 384	
		Soluble	2.76 ± 0.21	7	19.32	90	1 157 ± 110	22 355	94.7
		Membrane	0.78 ± 0.13	1	0.78	4	1 595 ± 185	1 247	5.3

[1]The protein recovery percentages were calculated based on the experimental values obtained from the lysate, soluble (cytoplasmic/periplasmic), and membrane fractions.
[2]The activity recovery percentages were calculated based on the sum of the total activity in soluble (cytoplasmic/periplasmic) and membrane fractions.

Table 3. Protein concentration and specific β-galactosidase activity recovered from *S. oneidensis* cells.

Strain	Construct	Fraction	Protein concentration (mg/ml)	Volume (mL)	Total Protein (mg)	Protein Recovery[1] (%)	Specific Activity (mUnits/mg protein)	Total mUnits	Specific Activity Recovery[2] (%)
S. oneidensis	**MCS5BB**	Lysate	6.11 ±0.89	7	42.8	100	2 ± 1	71	
		Soluble	4.09 ±0.50	7	28.7	67	2 ± 0	44	93.5
		Membrane	7.31 ±0.72	1	7.3	17	0 ± 1	3	6.5
	MCS5BB-LacZ	Lysate	4.28 ±0.19	7	30.0	100	233 ± 57	6 989	
		Soluble	2.91 ±0.10	7	20.3	68	251 ± 59	5 112	90.8
		Membrane	5.46 ±0.58	1	5.5	18	95 ± 33	518	9.2
	MCS5BB-Sec-LacZ	Lysate	4.38 ±0.79	7	30.7	100	16 ± 4	500	
		Soluble	2.93 ±0.52	7	20.5	67	7 ± 1	144	36.2
		Membrane	6.34 ±0.87	1	6.3	21	40 ± 12	255	63.8
	MCS5BB-Sec-LacZ-AIDA-I	Lysate	5.39 ± 0.41	7	37.7	100	50 ± 5	1 895	
		Soluble	3.85 ± 0.41	7	27.0	72	18 ± 4	476	29.9
		Membrane	7.60 ± 1.11	1	7.6	20	147 ± 23	1 118	70.1

[1]The protein recovery percentages were calculated based on the experimental values obtained from the lysate, soluble (cytoplasmic/periplasmic), and membrane fractions.
[2]The activity recovery percentages were calculated based on the sum of the total activity in soluble (cytoplasmic/periplasmic) and membrane fractions.

Table 4. Protein concentration and specific β-galactosidase activity recovered from *G. sulfurreducens* cells.

Strain	Construct	Fraction	Protein concentration (mg/ml)	Volume (mL)	Total Protein (mg)	Protein Recovery[1] (%)	Specific Activity (mUnits/mg protein)	Total mUnits	Specific Activity Recovery[2] (%)
G. sulfurreducens	**MCS5BB**	Lysate	2.07 ± 0.80	22	45.5	100	0 ± 0	0	
		Soluble	1.60 ± 0.58	22	35.2	77	0 ± 0	0	0
		Membrane	5.45 ± 1.66	1	5.4	12	0 ± 0	0	0
	MCS5BB-LacZ	Lysate	2.58 ± 0.11	22	56.9	100	5 ± 0	301	
		Soluble	2.03 ± 0.14	22	44.6	78	6 ± 1	273	95
		Membrane	6.12 ± 1.16	1	6.1	11	2 ± 2	14	5
	MCS5BB-Sec-LacZ	Lysate	4.84 ± 0.91	7	33.9	100	8 ± 0	256	
		Soluble	3.58 ± 0.42	7	25.0	74	4 ± 0	93	43.6
		Membrane	6.38 ± 0.60	1	6.4	19	19 ± 3	120	56.4
	MCS5BB-Sec-LacZ-AIDA-I	Lysate	2.07 ± 0.12	22	45.5	100	16 ± 4	732	
		Soluble	1.73 ± 0.19	22	38.2	84	10 ± 1	369	59.2
		Membrane	4.03 ± 0.26	1	4.0	9	63 ± 7	255	40.8

[1]The protein recovery percentages were calculated based on the experimental values obtained from the lysate, soluble (cytoplasmic/periplasmic), and membrane fractions.
[2]The activity recovery percentages were calculated based on the sum of the total activity in soluble (cytoplasmic/periplasmic) and membrane fractions.

Chapter 4. Construction of a Unique Expression Strategy for Engineering *G. sulfurreducens* Biofilms

The cell envelope of *G. sulfurreducens* is crucial for cellular respiration. The construction of genetic systems that provide access to this cellular compartment could allow exploiting the flux of electrons in redox biocatalysis. This reports describes a novel strategy for targeting proteins to the cell surface of *G. sulfurreducens*. This approach involves using the OmcZ *c*-type cytochrome of *G. sulfurreducens* for targeting proteins to the extracellular matrix of *Geobacter*. The green fluorescent protein (eGFP) was fused in-frame to the C-terminal of OmcZ, and monitored using confocal microscopy. Confocal microscope imaging and immunolabeling demonstrated the functional expression of eGFP on the surface of *G. sulfurreducens.* This report demonstrates for the first time the use of a bacterial *c*-type cytochrome to engineer biofilms of *G. sulfurreducens*.

1. Introduction

Dissimilatory metal reducing bacteria (DMRB) are characterized by their ability to gain energy from coupling the oxidation of organic acids with the reduction of insoluble metal oxides such as iron (Fe^{+3}) and manganese (Mn^{+4}) (97, 150). The microbial reduction of metals is an important biochemical process that contributes to the cycling of metals and carbon in the environment (100). Furthermore, the use of dissimilatory metal reducing bacteria (DMRB) has practical applications in the area of bioremediation and production of alternative energy. Members of the *Geobacteraceae* family have been found to be predominant and responsible for iron (Fe^{+3}) reduction occurring at anaerobic aquifer and sediments (25, 156). *Geobacter sulfurreducens,* a delta proteobacteria, has been a model organism to study dissimilatory iron (Fe^{+3}) reduction.

The electron transfer mechanisms used by *G. sulfurreducens* are not fully understood. Previous studies have demonstrated that the electron transfer mechanisms rely on redox active *c*-type cytochromes for the reduction of insoluble electron acceptors (115, 150). The genome of *G. sulfurreducens* contains more than 100 putative *c*-type cytochromes (33). However, the role of only some of these proteins has been studied during insoluble iron (Fe^{+3}) reduction. Some of the outer membrane proteins involved in the reduction of insoluble electron acceptors in *G. sulfurreducens* include OmcZ, OmcS, OmcB and OmpJ (2, 57, 82, 115).

The extracellular electron transfer mechanisms have been studied in biofilms of *G. sulfurreducens* cells grown on poised electrode surfaces. These studies have led to the identification of OmcZ, an octaheme *c*-type cytochrome, as essential for the development of biofilms on electrode surfaces. A transcriptomics study performed on electrode biofilms identified that the transcription of the *omcZ* gene increased 5.7-fold when *G. sulfurreducens* cells were grown using an anode as terminal electron acceptor (125).

A recent study has identified a cluster of genes (*xap*) in *G. sulfurreducens* responsible for the synthesis of an extracellular polysaccharide matrix (136). The presence of this polysaccharide material in *G. sulfurreducens* has shown to support biofilm formation and anchoring of cytochromes beyond the cell membrane. Mutant *G. sulfurreducens* cells impaired to produce a wild type polysaccharide matrix were unable to use Fe(III) oxyhydroxide as electron acceptor, and their extracellular matrix was devoid of *c*-type cytochromes including OmcZ. The *c*-type cytochrome OmcZ is a predicted octa-heme *c*-type cytochrome comprised by 473 amino acids. Immunogold labeling studies performed on anodic biofilms of *G. sulfurreducens* cells have shown that OmcZ is localized in the extracellular matrix formed throughout the biofilm (56, 136).

Biochemical studies have identified the presence of two protein populations of OmcZ on the extracellular matrix of *G. sulfurreducens* cells harvested from current-producing anodic biofilms (56, 125). These two OmcZ populations have predicted molecular weights of 30-kDa and 50-kDa. N-terminal and LC-MS sequences have indicated that 30-

kDa protein is a truncated version of the full-length 50-kDa OmcZ (57). The 30-kDa OmcZ form is composed by 282 amino acids including a signal peptide and eight predicted heme *c*-binding motifs. Since antibodies have only been raised for the 30-kDa form of OmcZ, the presence of the remaining 191 amino acids has only been identified in the 50-kDa OmcZ form. The processing of OmcZ occurs between residues 328 and 329 (57).

The integrity and composition of the outer membrane and extracellular matrix of DMRB is essential for dissimilatory iron (Fe^{+3}) reduction. The development of genetic tools to access this redox active space is crucial for understanding the fundamentals of cellular respiration. Furthermore, there is an opportunity to exploit the redox active space created throughout the biofilm of *G. sulfurreducens* in redox biocatalysis (3, 36, 163). This work describes the development of a novel genetic system that involves using an in-frame fusion with *c*-type cytochrome OmcZ to target the expression of a recombinant protein to extracellular matrix form throughout *G. sulfurreducens* biofilms.

2. Material and Methods

2.1 Bacterial Strains and Growth Conditions

The bacteria strains and plasmids used during this study are described in Table 1. *G. sulfurreducens* PCA (ATTC 51573) was grown at 30°C in anaerobic mineral medium containing 20 mM acetate as the electron donor and 40 mM of fumarate as the electron acceptor (19, 108). *G. sulfurreducens* cells harboring the kanamycin resistance cassette (Kan^R) were grown anaerobically in mineral medium supplemented with 200 μg/ml of kanamycin (Table 1).

Table 1. Strains and plasmids used during this study

Strains, Plasmid or Primer	Characteristics or Sequence (5' - 3')	Source
Strains		
G. sulfurreducens	Wild type (ATCC 51573)	(19)
OmcZ-eGFP-Kan^R	Genomic C-terminal in-frame fusion of eGFP in *omcZ* (GSU2076) gene followed by the kanamycin (Kan^R) resistance cassette.	This study
OmcZ-Kan^R	Genomic insertion of the kanamycin (Kan^R) resistance cassette downstream of the *omcZ* (GSU2076) gene.	This study
Plasmids		
pBBR1MCS-2	Mobilizable broad-host-range plasmid; Km^r.	(75)
pUCBB-eGFP	Plasmid derived from pUCMmod; Amp^R	(174)
pBBRBB-eGFP	Mobilizable broad-host range plasmid; Km^r	(174)
Primers		
F-OmcZ	gaagccgtcctgaccatcgtcggctccagc	
R-OmcZ	ccgtttgactttcttcggagccgccaggattttgc	
F-EGFP	ctccgaagaaagtcaaacggatggtgagcaagggcgaggagc	
R-EGFP	agctggcaattccggttcgcttttacttgtacagctcgtccatg	
F-Kan	aagcgaaccggaattgccagctggggcgcc	
R-Kan	ccccgttggtcggaaaggtatcagaagaactcgtcaagaaggc	
F-NC	tacctttccgaccaacggggcaggggattc	
R-NC	cgcaacctggtcggcaccgatgtgctgcgc	
F-C- OmcZ	gaagccgtcctgaccatcgtcggctccagc	
R-C-OmcZ	ttaccgtttgactttcttcggagccgccaggattttgc	
F-C- Kan	ctccgaagaaagtcaaacggtaaaagcgaaccggaattgccagct	
R-C-Kan	ccccgttggtcggaaaggtatcagaagaactcgtcaagaaggc	
F-C2- OmcZ	tacctttccgaccaacggggcaggggattc	
R-C2-OmcZ	cgcaacctggtcggcaccgatgtgctgcgc	

2.2 Construction of a Genetic System in *G. sulfurreducens* Using an In-frame *c*-type Cytochrome Fusion

The last ≈500 bp of the *omcZ* gene (GSU2076) were amplified from wild type *G. sulfurreducens* using forward (F-OmcZ) and reverse (R-OmcZ) primers. The downstream region of the *omcZ* gene was amplified from wild type *G. sulfurreducens* with forward (F-NC) and reverse (R-NC) primers. The eGFP gene (720 bp) was amplified from pUCBB-EGFP plasmid using forward (F-EGFP) and reverse (R-EGFP) primers. The kanamycin cassette (940 bp) was amplified from pBBRMCS-2 plasmid with forward (F-Kan) and reverse (R-Kan) primers. The four resulting DNA fragments were gel purified and they were assembled into a single ≈3-Kb DNA fragment using primerless PCR reaction and the following thermocycler parameters; 95°C for 4 minutes followed by 15 cycles of 96°C for 40 seconds, 47°C for 1 minute and 72°C for 5 minutes and a final extension at 72°C for 10 minutes reaction. The linear ≈3-Kb fragment was then amplified using forward (F-OmcZ) and reverse (R-NC) primers and the following thermocycler parameters; 95°C for 4 minutes followed by 40 cycles of 96°C for 40 seconds, 47°C for 1 minute and 72°C for 5 minutes and a final extension at 72°C for 10 minutes reaction. The resulting ≈3-Kb linear DNA fragment was electroporated into wild type *G. sulfurreducens* and resistant colonies were selected on mineral medium agar plates supplemented with 200 μg/ml of kanamycin. The resistant colonies were grown in 10 ml of mineral medium and the insertion in the genome of *G. sulfurreducens* was confirmed through PCR amplification and sequencing.

2.3 Electrochemical Analysis

The electrochemical bioreactors were assembled as previously described (108). The three-electrode bioreactors were incubated at 30°C in a water bath and continuously purged with N_2/CO_2. Anaerobic bioreactors containing 10 ml of mineral medium (supplemented with 40 mM acetate and without electron acceptor) were connected to a 16-channel potentiostat with EZ-Lab Software (VMP; Bio-Logic, Knoxville, TN, USA) to run chronoamperometry routines. The anaerobic reactors were inoculated with 50% (vol/vol) of wild type and mutant *G. sulfurreducens* cells (OD_{600} 0.5-0.55) and mineral medium (supplemented with 40 mM acetate and without electron acceptor) at a potential of +0.24V versus the standard hydrogen electrode (SHE).

2.4 Confocal Microscopy

The biofilms grown on poised electrodes were imaged using a Nikon C1 spectral imaging confocal microscope (Nikon, Japan). The electrodes were harvested and the biofilms were stained with SYTO 62 (Invitrogen, Carlsbad, CA, USA) in 100 mM phosphate buffer supplemented with 0.04 g/L of $CaCl_2 \bullet H_2O$ and 0.2 g/L of $MgSO_4 \bullet 7H_2O$. The electrodes were incubated with the SYTO 62 dye for 30 minutes under oxygen conditions to allow staining and proper folding of eGFP.

2.5 Immunofluorescence Labeling

The immunofluorescence eGFP labeling was performed by the Imaging Center at the University of Minnesota. Immediately after harvesting, the electrodes were washed with

100 mM phosphate buffer and fixed with a mixture of formaldehyde and glutaraldehyde for 60 min at room temperature. Then, the electrodes were washed three times (5 minutes each) with 100 mM phosphate buffer supplemented with 4% (w/v) of sucrose. Additionally, the electrodes were washed three additional times (15 minutes each) with 100 mM phosphate buffer supplemented with 4% (w/v) of sucrose and 50 mM NH_4Cl buffer. After incubating electrodes in a blocking buffer for 30 minutes, the electrodes were labeled using a polyclonal anti-GFP primary antibody (abcam, Cambridge, MA, USA) at 4°C overnight. The resulting electrodes were washed five times with 1x PBS (5 minutes each) and incubated at 4°C for 4 hrs with secondary antibody. For the immunofluorescence labeling, the concentration of secondary antibody was 1:10,000, and the goat anti-rabbit antibody was conjugated with Alexa568 red fluorescent dye (Invitrogen, Carlsbad, CA, USA).

3. Results

3.1 Detection of Functional eGFP in *G. sulfurreducens* Biofilms

The wild type *G. sulfurreducens* and mutants harboring the in-frame OmcZ-eGFP fusion (OmcZ-eGFP-Kan^R) were grown on poised electrodes. The expression of eGFP was detected using confocal microscopy (Figure 1-2). When eGFP was fused to the *omcZ* gene, *G. sulfurreducens* cells exhibited green fluorescence throughout the biofilm (Figure 1). There was no green fluorescence detected in biofilm of wild type *G. sulfurreducens* cells grown under the same conditions (Figure 2).

The current production of wild type *G. sulfurreducens* cells and the mutant cells harboring the in-frame fusion of OmcZ-eGFP-Kan^R and OmcZ-Kan^R was described in Figure 3. After 40 hours of incubation, wild type *G. sulfurreducens* cells produced 2000 μA on 4 cm^2 of electrode surface. When the kanamycin cassette was inserted downstream of the *omcZ* gene (OmcZ-Kan^R), the current production decreased 80% compared to wild type *G. sulfurreducens* cells. The cells harboring the in-frame fusion of eGFP at the C-terminal of OmcZ followed by the Kan^R cassette (OmcZ-eGFP-Kan^R) produced the same amount of current as the cells harboring only the Kan^R cassette (OmcZ- Kan^R).

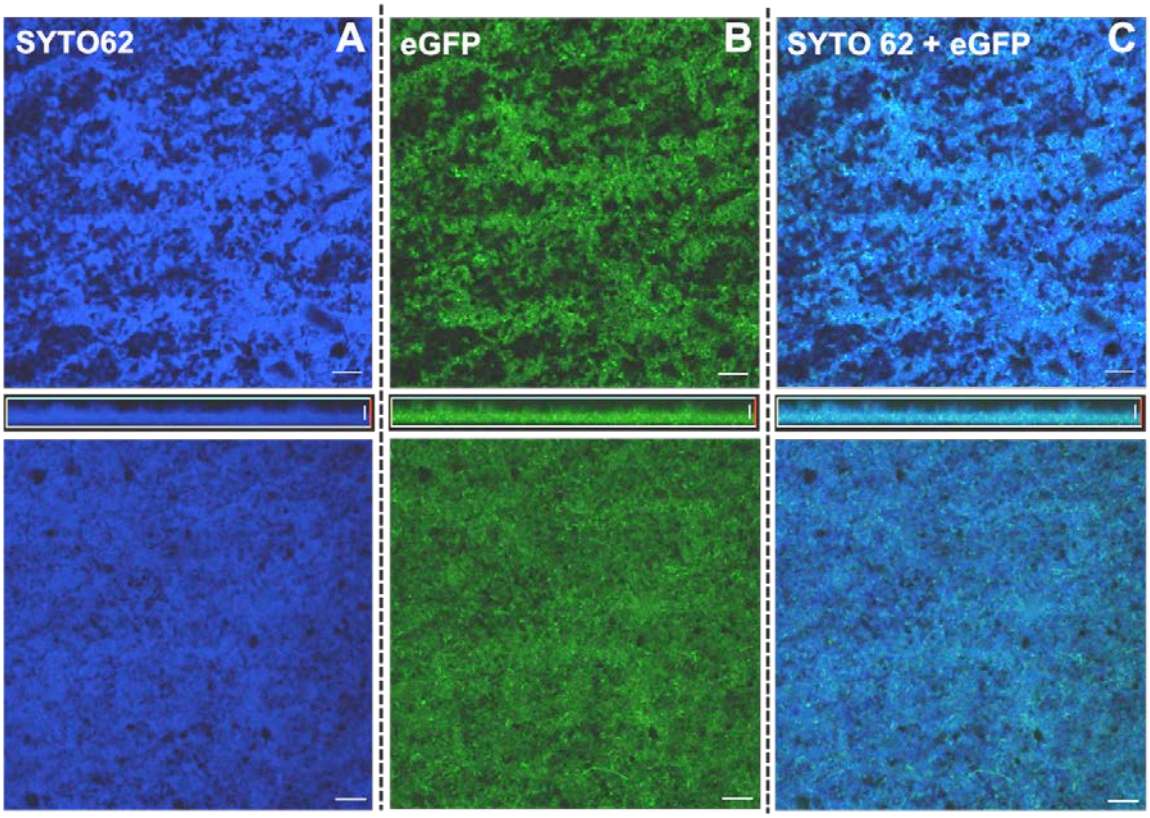


Figure 1. Confocal microscope images of a biofilm from *G. sulfurreducens* expressing an in-frame eGFP fusion with OmcZ (GSU 2076). Top panels are slide projections (bar= 10 μm), middle panels are side projections (bar= 5 μm), and bottom panels are maximum projections (bar= 10 μm). **A.** The red fluorescence emitted by SYTO 62 appears in color blue. **B**. The green fluorescence emitted by eGFP appears in color green. **C**. The blue and green fluorescence are overlapped.

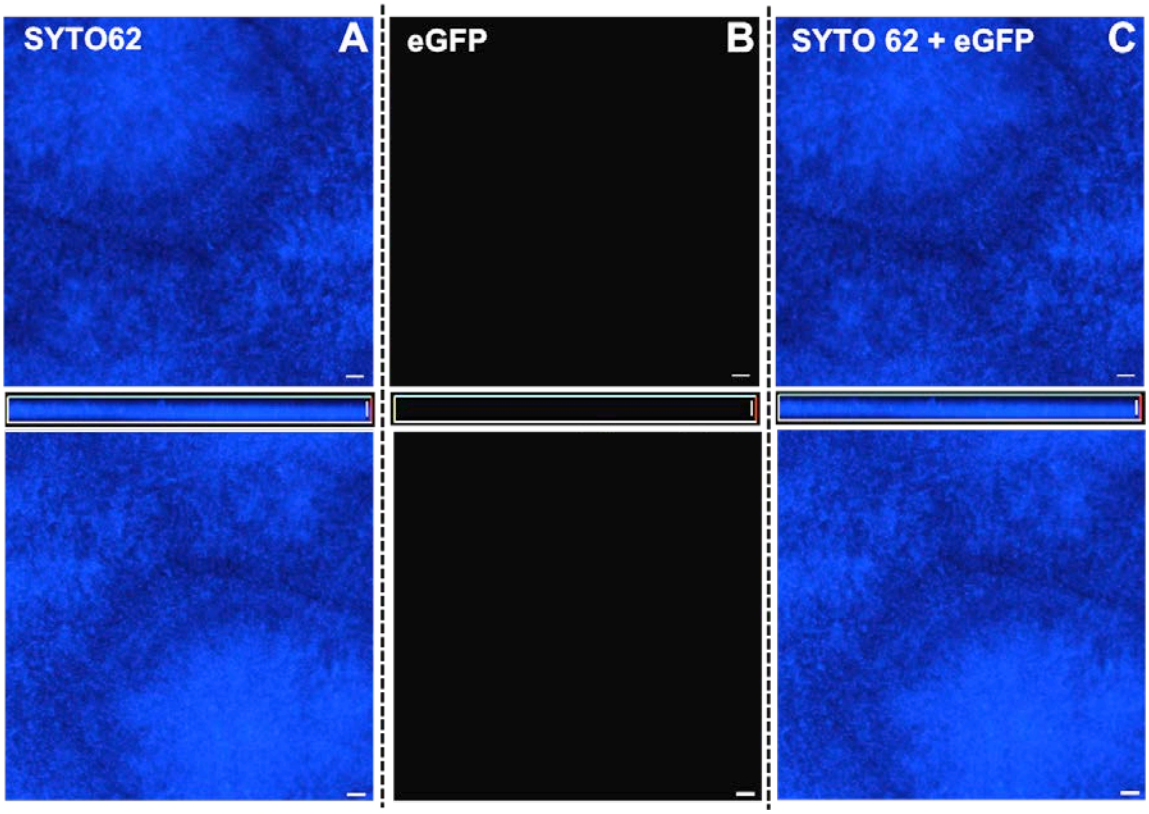


Figure 2. Confocal microscope images of wild type *G. sulfurreducens* biofilms (bar = 10μm). Top panels are slide projections, middle panels are side projections and bottom panels are maximum projections. **A**. The red fluorescence emitted by SYTO 62 appears in color blue. **B**. The green fluorescence emitted by eGFP appears in color green. **C**. The red and green fluorescence are overlapped.

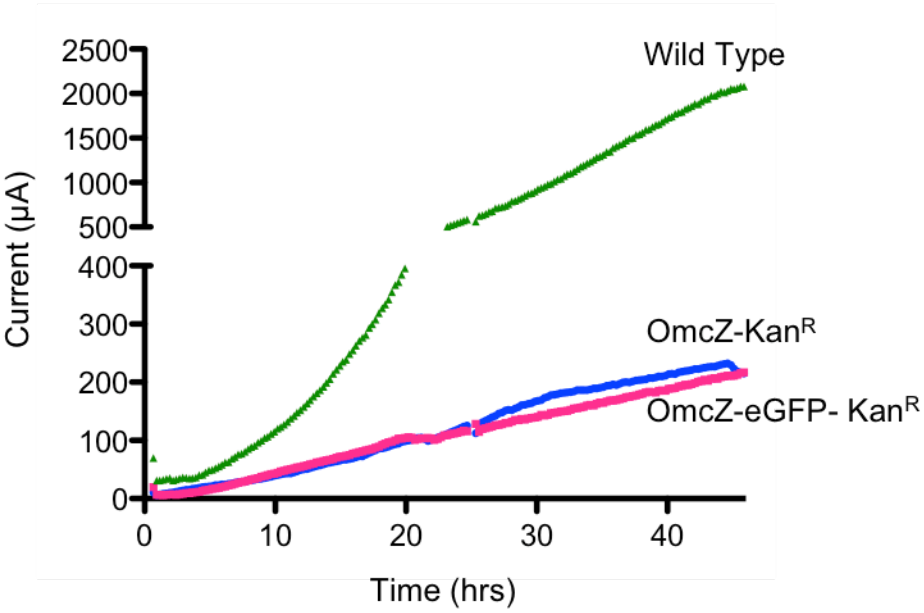


Figure 3. Current production of wild type, OmcZ-Kan^R and OmcZ-eGFP-Kan^R *G. sulfurreducens* cells grown on poised electrodes (+0.24 V) versus SHE.

3.2 Surface Localization of eGFP

The localization of eGFP on the surface of *G. sulfurreducens* cells was demonstrated using immunofluorescence labeling (Figure 4). Fixed cells incubated with primary antibody were scraped from electrodes and treated with secondary antibody. The resulting cells were subjected to confocal imaging. The imaging results revealed that eGFP remained fluorescent in fixed cells (Figure 4A). Furthermore, the immunolabeling of eGFP revealed the presence of the red fluorescent antibody in similar locations as the eGFP (Figure 4B-C). The red fluorescence observed in cells scraped from electrodes suggested that eGFP localized on the surface and matrix of the *G. sulfurreducens* cells (Figure 4A-C). There was no green or red fluorescence observed in images obtained from *G. sulfurreducens* expressing cytoplasmic eGFP. The green fluorescence has not been detected when the expression of the eGFP gene is driven by constitutive promoters such as t-lac or lac. Therefore, an alternative negative control could be *G. sulfurreducens* cells expressing eGFP under the OmcZ promoter or using a constitutive promoter but in a different expression host (i.e. *Shewanella* or *E. coli*).

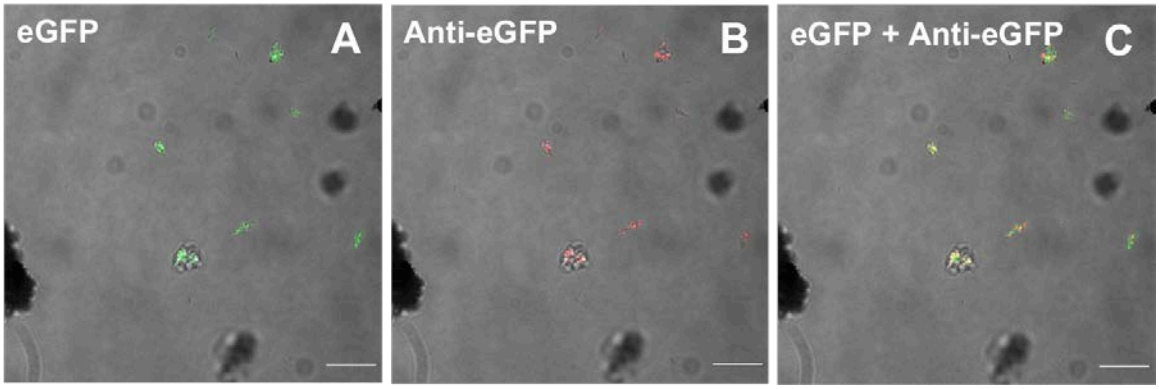


Figure 4. Immunofluorescence labeling of eGFP on the surface of *G. sulfurreducens* cells (bars = 10μm). A) *G. sulfurreducens* expressing eGFP fused to *c*-type cytochrome OmcZ with visible and green fluorescence confocal microscope channel overlapped. B) *G. sulfurreducens* expressing eGFP fused to *c*-type cytochrome OmcZ labeled with red fluorescent secondary antibody ($Alexa_{568}$). C) Surface localization of eGFP and red fluorescent antibody ($Alexa_{568}$) on the *G. sulfurreducens* cells.

4. Discussion

The typical surface display systems have relied on using membrane-associated proteins as carriers to facilitate the translocation of heterologous proteins across the cell membrane (142, 170). Even though an extensive number of outer membrane proteins have been used to display protein on the surface of bacteria, the use of bacterial outer membrane *c*-type cytochromes has never been reported. This work describes a unique strategy to target heterologous proteins to the surface and extracellular matrix of *G. sulfurreducens*. This system exploits an in-frame protein fusion with *c*-type cytochrome OmcZ to translocate eGFP onto the surface of *G. sulfurreducens*. The surface localization of eGFP was confirmed with confocal microscopy (Figure 1-2) and immunofluorescence labeling (Figure 4).

Previous studies identified the presence of two protein populations of OmcZ on the outer membrane and extracellular matrix of *G. sulfurreducens* biofilms (56, 125). These two OmcZ populations had molecular weighs of 30-kDa and 50-kDa. Later biochemical studies demonstrated that the 30-kD form was a truncated version of the full length 50-kDa OmcZ (57). Considering this together, the green fluorescence observed in the OmcZ-eGFP-Kan^{R} *G. sulfurreducens* biofilms suggested the presence of the 50-kDa form of OmcZ and/or the entrapment of the cleaved 20-kDa proteins in the extracellular matrix.

There is an strong interest in using genetically engineered organisms on microbial fuel cells (3). Microbial fuel cells exploit the metabolism of bacteria for the production of

electricity. The drawbacks of using microbial fuel cells are related with the low current densities as a result of the inefficient microbial electron transfer mechanisms to solid surfaces. Previous reports have shown that surface display systems in yeast have potential applications catalyzing reactions in microbial fuel cells. However, these systems continue to rely on redox mediators to transfer electrons to the electrodes (163).

In contrast to other organisms, *G. sulfurreducens* naturally attaches to electrode surfaces and it uses direct electron transfer mechanisms for supporting cellular respiration. These microbial capabilities eliminate the need of immobilizing cells on electrode surfaces and adding electron transfer mediators to microbial fuel cell. The display of proteins on the surface of *G. sulfurreducens* could facilitate coupling the oxidation of alternative substrates with the production of electricity. Furthermore, the expression of redox enzymes may allow modification of the existent electron transfer mechanisms used by the cell.

The results obtained from the chronoamperometry experiments suggested that the genomic insertion of the kanamycin antibiotic cassette produced slower growing cells and thinner anodic biofilms than those produce by wild type *G. sulfurreducens* cells (Figure 3). Nevin and coworkers reported similar lost of current when they replaced the *omcZ* gene with the Kan resistance cassette (125). In that case the lost of current was attributed to the lack of OmcZ expression since complementation of OmcZ restored the current production to wild type levels. It needs to be determined if the current production

of the genomic OmcZ-Kan^R insertion of *G. sulfurreducens* biofilms can be restored by complementing the mutation with one of the genes located downstream of OmcZ. Those results could help determine if the in-frame fusion caused a polar mutation.

This reports shows for the first time the functional expression of eGFP in young *G. sulfurreducens* biofilms. The functional expression of eGFP was demonstrated using confocal imaging. Furthermore, the immunofluorescent labeling suggested that the OmcZ-eGFP protein fusion was localized outside the cell.

Chapter 5. Conclusions and Future Directions

1. Conclusions and Future Perspectives

1.1 Metagenomics

The first project of this thesis was focused on the discovery of unique biocatalysts involved in the hydrolysis of complex cellulosic and hemicellulosic polysaccharides using a functional metagenomics approach. The most attractive advantage of using functional metagenomics is the opportunity to mine the genetic material of uncultured organisms for novel enzymes. However, this opportunity has come with a series of challenges that include developing efficient methods for extracting DNA fragments from environmental samples, finding suitable vectors for cloning DNA, finding compatible expressions hosts and designing sensitive screening techniques to detect the presence of the enzymes of interest. For the past ten years, those challenges have been overcome at a certain extent and the use of metagenomics has allowed identifying functional enzymes in a wide range of complex environments (15, 35, 52, 79, 131).

The first chapter of this thesis described a project that tested the odds of finding novel cellulases and hemicellulases from soil using a functional metagenomic approach. During this work, metagenomic libraries were made from soil found in places continuously exposed to biomass. High-throughput screening assays were used to identify the expression of functional enzymes throughout the libraries. The screening assays employed traditional soluble substrates such as carboxymethyl cellulose (CMC) and non-traditional substrates such as insoluble cellulosic and hemicellulosic substrates cross-

linked with azo dyes. The screening assays identified a group of clones with putative glycosyl hydrolase activity. The most unique clones were further studied to identify the genes responsible for the observed phenotypes. The genes with putative glycosyl hydrolase activity were independently cloned into expression vectors and the glycosyl hydrolase activity of resulting enzymes was demonstrated. These enzymes showed hydrolytic activity towards cellulosic or hemicellulosic substrates. While some enzymes shared similarity to proteins identified in other metagenomic projects, other proteins showed to be unique. During this project, an endoglucanase that contains unique domain architecture was fully characterized. This work demonstrated the use of functional metagenomics coupled to high-throughput screening to identify unique and functional cellulases in the soil metagenome of Minnesota.

For the past five years we have learned that the possibility of finding novel genes using metagenomics has come with a high price. These difficulties are reflected in the low hitting rates of finding novel genes reported in our results, and those of others. This problem in part could be attributed to the lack or low gene expression attained by the host of the metagenomic libraries. Furthermore, low gene expression levels could be the result of a large number of factors including, differences in codon usage, promoter recognition, start/stop codon recognition sites, protein folding machinery, and protein secretion machinery. These factors stress the great importance of the transcription and translation machinery of the selected expression host. Since most metagenomic libraries today continue using *E. coli* as expression host, the possibilities of finding novel enzymes

continue to be low and limited to enzymes that can be expressed in that particular organism. Recent studies have shown the development of alternative shuttle vectors for constructing environmental libraries (109, 112). The development of these broader vector/host systems has shown some promise for increasing the possibilities of finding novel genes.

The slim possibilities of finding novel cellulases, as it is for other enzymes, in metagenomic libraries could also be due to the low sensitivity of existent screening assays. This difficulty is directly related to low gene expression levels. The assays used for our study were based on the detection of a particular gene product, which guaranteed that the positive clones harbored a functional gene. However, once again it limited our possibilities to find only genes that could be expressed in *E. coli*.

The recalcitrant and insoluble nature of cellulose imposes a great challenge for designing sensitive assays. However, there is still the possibility that developing more sensitive assays could have given a greater opportunity to identify novel genes. Some alternatives for developing more sensitive high-throughput screening techniques could be to focus on coupling two independent reactions in one single assay. For instance, detection of cellulose hydrolysis could possibly be identified by adding an enzyme that will detect the presence of cellobiose such as cellobiose dehydrogenase in the reaction. The same principle could be applied for detecting xylan hydrolysis and xylose detection using xylose dehydrogenase (88). Meanwhile, screening under different pH and temperature

conditions could also have increased the outcomes of finding novel genes. The pH optimal of a great number of cellulases is acidic; therefore there is a possibility of increasing the number of positive outcomes of finding a novel gene by screening under different growing conditions (pH and temperature).

The future of functional metagenomic will have to come with improvements in the construction and screening of metagenomic libraries. Those improvements will have to offer the possibility of learning more about the nature of the DNA inserts, using broader expression vectors, multiple expression hosts, and designing novel strategies to tune the gene expression.

Metagenomics continue to be a powerful tool used in industry for the discovery of novel biocatalysis. Companies such as Verenium/Diversa have employed metagenomic related approaches to discover and to produce industrially relevant enzymes. This company holds thousands of patents describing processes for making genomic libraries, screening technologies, improving biocatalysts and producing them in a large scale.

This is an exciting time for microbiologist, ecologists and for all those individuals interested in studying the microbial diversity in the world. Emerging technologies including metagenomics and high-throughput sequencing have given us the opportunity to look beyond our eyes and to describe the entire microbial population living on the planet Earth. A recent initiative called the Earth Microbiome Project has the ambitious goal of describing the microbial world living in our planet Earth. Strong efforts are being

put on analyzing the large amounts of data generated from the sequencing technologies to not only describe the complexity of the microbial population but also to make better assessments about the relevance of these organisms in a particular environment.

1.2 Surface Display Systems in Iron-Reducing Bacteria

One of the central topics studied in the Bond Lab is dissimilatory iron reduction and electrochemistry. In the past years we have gained more knowledge about the indirect electron transfer mechanism used by *Shewanella oneidensis* for iron reduction. This mechanism supports the theory that reduction of insoluble metals relies on the presence of redox active *c*-type cytochromes. The presence of flavin compounds that function as a shuttle has been found to be critical for long-range electron transfer. Meanwhile, there is still little known about the possible direct electron mechanism used by *S. oneidensis.* Therefore, the possible direct electron transfer mechanisms that occur at the interface between the cells of *S. oneidensis* and the insoluble metals remain unknown.

The electron transfer mechanisms used by *Geobacter* differ from the electron transfer mechanism described in *Shewanella*. Even though *Geobacter* genome contains more than 100 putative *c*-type cytochromes, there is still very little known about the roles of most of these proteins during the reduction of insoluble metals. Electrochemical techniques have demonstrated that *G. sulfurreducens* can transfer electron directly to insoluble metals without the need of soluble or shuttling molecules. Furthermore, it has been demonstrated that the direct electron transfer mechanism is supported by other strategies including the

formation of a polysaccharide matrix that anchors cytochromes. This polysaccharide matrix is suspected to support the long-range reduction of insoluble metals in *G. sulfurreducens* biofilms. This latest study stress the importance of the interface between the outer membrane and terminal electron acceptor for understanding the extracellular electron transfer mechanism used by *Geobacter*.

The results presented in Chapter 3 and Chapter 4 described two independent approaches for engineering the cell surface and biofilms of metal reducing bacteria. The first approach involved using an autodisplay system to engineer the surface of *Geobacter* and *Shewanella*. During this approach, an autodisplay system that has shown to work in *E. coli* was synthesized and optimized for expression in *G. sulfurreducens*. The autodisplay system was evaluated in *E. coli*, *G. sulfurreducens* and *S. oneidensis*. The results obtained from the activity assays indicated that the autodisplay system caused the accumulation of functional enzyme in the cell membrane of the three organisms. The results of these experiments suggested that the autodisplay system could be used to engineer the cell envelope of *G. sulfurreducens* and *S. oneidensis.* In addition, *G. sulfurreducens* cells expressing the autodisplay system were grown on an electrode surface, and were able to catalyze the hydrolysis of a soluble compound. These results suggested the potential application of the autodisplay system for whole-cell biocatalysis in electrochemical cells. The autodisplay system was tested with an enzyme that hydrolyze a soluble substrate, future work will have to focus on demonstrating the hydrolysis of insoluble substrates.

The implementation of the autodisplay system in *Geobacter* and *Shewanella* demonstrated to be a valuable tool for accessing the membrane of these organisms. The results of these experiments showed that we are a step closer in the expansion of the biocatalytic capabilities and redox properties of these organisms.

The second approach involved developing a unique expression strategy for engineering *G. sulfurreducens* biofilms. The purpose of this system was to engineer the interfacial matrix formed between the cell envelope and the insoluble electron acceptor. This approach exploited the used of a *c*-type cytochrome OmcZ, as a carrier for targeting proteins to the extracellular space.

The use of *c*-type cytochromes for the expression of recombinant proteins beyond the cells membrane is a unique approach for engineering biofilms of *G. sulfurreducens*. The confocal microscope imaging and immunofluorescence labeling results obtained during our experiments suggested the functional expression of a fluorescent protein on the surface of *G. sulfurreducens*. Furthermore, the results of this project demonstrated for the first time the potential use of a *c*-type cytochrome to access the extracellular matrix of *G. sulfurreducens*. In summary, the present work described two independent strategies for engineering the membrane and biofilms of metal reducing bacteria.

The next steps should focus on demonstrating that the surface display tools can be used to answer fundamental questions regarding the electron transfer mechanisms and, to exploit

the redox active space created throughout the biofilm of *G. sulfurreducens* in redox biocatalysis. For example, these systems could be used to express redox enzymes such as cytochromes on the membrane or the extracellular matrix. The expression of cytochromes could be detected by using a heme stain procedure or voltammetric techniques.

The use of genetically engineered microorganisms for whole-cell biocatalysis in microbial fuel cells is an emerging field. Microbial fuel cells exploit microbial metabolism for the conversion of biochemical energy into electricity. Recent studies in yeast have shown the potential of engineered microorganisms to catalyze redox reactions in a microbial fuel cell setting. In contrast to those systems, *Geobacter* and *Shewanella* are organisms that naturally attach and transfer electrons to electrode surfaces, which removes the need for adding expensive mediators. Furthermore, the flow of electrons coming in or out of the cells during microbial respiration could potentially be used to catalyze redox reactions.

Next steps should also focus on using the surface display systems to modify the nature of the interface between cells and the electrode. The expression of proteins that catalyze the synthesis of polymers such as silica may have the potential of creating more conductive biofilms. The modification of electrode surfaces with silica has shown to improve the redox capabilities of different analytes. Therefore, it would be interesting to determine if the presence of silica in *Geobacter* biofilms grown on electrode surfaces improve their redox properties.

2. References

1. **Aden, A., M. Ruth, K. Ibsen, J. Jechura, K. Neeves, J. Sheehan, B. Wallace, L. Montague, and A. Slayton.** 2002. Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover.
2. **Afkar, E., G. Reguera, M. Schiffer, and D. R. Lovley.** 2005. 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 Microbiol **5:**41.
3. **Alfonta, L.** 2010. Genetically engineered microbial fuel cells. Electroanalysis **22:**822-831.
4. **Armstrong, F. A.** 1999. Electron transfer and coupled processes in protein film voltammetry. Biochem Soc Trans **27:**206-210.
5. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**.
6. **Baron, D., E. LaBelle, D. Coursolle, J. A. Gralnick, and D. R. Bond.** 2009. Electrochemical measurement of electron transfer kinetics by *Shewanella oneidensis* MR-1. J Biol Chem **284:**28865-28873.
7. **Bayer, E., Y. Shoham, and R. Lamed.** 2006. Cellulose-decomposing bacteria and their enzyme system, 3rd ed, vol. 2. Springer, New York.
8. **Bayer, E. A., J. P. Belaich, Y. Shoham, and R. Lamed.** 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol **58:**521-554.
9. **Benz, I., and M. A. Schmidt.** 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. Infect Immun **57:**1506-1511.
10. **Bond, D. R., and D. R. Lovley.** 2003. Electricity production by *Geobacter sulfurreducens* attached to electrodes. Appl Environ Microbiol **69:**1548-1555.
11. **Bond, D. R., and D. R. Lovley.** 2005. Evidence for involvement of an electron shuttle in electricity generation by *Geothrix fermentans*. Appl Environ Microbiol **71:**2186-2189.
12. **Boraston, A. B., D. N. Bolam, H. J. Gilbert, and G. J. Davies.** 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J **382:**769-781.
13. **Boraston, A. B., A. L. Creagh, M. Alam, J. M. Kormos, P. Tomme, C. A. Haynes, A. J. Warren, and D. G. Kilburn.** 2001. Binding specificity and

thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. Biochemistry **40:**6240-6247.

14. **Braun, V., and K. Rehn.** 1969. Chemical characterization, spatial distribution and function of a lipoprotein (murein- lipoprotein) of the *E. coli* cell wall. Eur J Biochem **10:**426-438.
15. **Brennan, Y., W. N. Callen, L. Christoffersen, P. Dupree, F. Goubet, S. Healey, M. Hernandez, M. Keller, K. Li, N. Palackal, A. Sittenfeld, G. Tamayo, S. Wells, G. P. Hazlewood, E. J. Mathur, J. M. Short, D. E. Robertson, and B. A. Steer.** 2004. Unusual microbial xylanases from insect guts. Appl Environ Microbiol **70:**3609-3617.
16. **Brown, S.** 1992. Engineered iron oxide-adhesion mutants of the *Escherichia coli* phage lambda receptor. Proc Natl Acad Sci U S A **89:**8651-8655.
17. **Brutinel, E. D., and J. A. Gralnick.** 2012. Shuttling happens: soluble flavin mediators of extracellular electron transfer in *Shewanella*. Appl Microbiol Biotechnol **93:**41-48.
18. **Butler, J. E., F. Kaufmann, M. V. Coppi, C. Nunez, and D. R. Lovley.** 2004. MacA a diheme *c*-type cytochrorne involved in Fe(III) reduction by *Geobacter sulfurreducens*. J Bacteriol **186:**4042-4045.
19. **Caccavo Jr, F., D. J. Lonergan, D. R. Lovley, M. Davis, J. F. Stolz, and M. J. McInerney.** 1994. *Geobacter sulfurreducens* sp. nov., a hydrogen- and acetate-oxidizing dissimilatory metal-reducing microorganism Appl Environ Microbiol **60:**3752-3759.
20. **Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat.** 2009. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res **37:**233-238.
21. **Charbit, A., J. Boulain, A. Ryter, and M. Hofnung.** 1986. Probing the topology of a bacterial membrane protein by genetic insertion of a foreign epitope; expression at the cell surface. EMBO J **5:**3029-3037.
22. **Chiriac, A. I., E. M. Cadena, T. V. Vidal, A. L. Torres, P. Diaz, and F. I. J. Pastor.** 2010. Engineering a family 9 processive endoglucanase from *Paenibacillus barcinonensis* displaying a novel architecture. Appl Microbiol Biotechnol **86:**1125-1134.
23. **Chundawat, S. P. S., G. T. Beckham, M. E. Himmel, and B. E. Dale.** 2011. Deconstruction of lignocellulosic biomass to fuels and chemicals. Annual Review of Chemical and Biomolecular Engineering **2:**121-145.
24. **Clarke, T., M. Edwards, A. Gates, A. Halla, G. F. Whitea, J. Bradleya, C. L. Reardonb, L. Shib, A. S. Beliaevb, M. J. Marshallb, Z. Wangb, N. J. Watmougha, J. K. Fredricksonb, J. M. Zacharab, J. N. Butta, and D. J. Richardson.** 2011. Structure of a bacterial cell surface decaheme electron conduit. Proc Natl Acad Sci U S A **108:**9384-9389.
25. **Coates, J. D., E. J. P. Phillips, D. J. Lonergan, H. Jenter, and D. R. Lovley.** 1996. Isolation of *Geobacter* species from diverse sedimentary environments. Appl Environ Microbiol **62:**1531-1536.

26. **Cooney, M., V. Svoboda, C. Lau, G. Martin, and S. D. Minteer.** 2008. Enzyme catalysed biofuel cells. Energy Environ Sci **1:**320-337.
27. **Coppi, M. V., C. Leang, S. J. Sandler, and D. R. Lovley.** 2001. Development of a genetic system for *Geobacter sulfurreducens*. Appl Environ Microbiol **67:**3180-3187.
28. **Coursolle, D., and J. A. Gralnick.** 2010. Modularity of the Mtr respiratory pathway of *Shewanella oneidensis* strain MR- 1. Mol Microbiol **77:**995-1008.
29. **Daniel, R.** 2005. The metagenomics of soil. Nat Rev Microbiol **3:**470-478.
30. **Daniel, R.** 2004. The soil metagenome - a rich resource for the discovery of novel natural products. Curr Opin Biotechnol **15:**199-204.
31. **Davis, F., and S. P. Higson.** 2007. Biofuel cells - Recent advances and applications. Biosens Bioelectron **22:**1224-1235.
32. **Deshpande, M. V., K. Eriksson, and L. G. Pettersson.** 1984. An assay for selective determination of exo-1, 4,-[beta]-glucanases in a mixture of cellulolytic enzymes. Anal Biochem **138:**481-487.
33. **Ding, Y. R., K. K. Hixson, M. A. Aklujkar, M. S. Lipton, R. D. Smith, D. R. Lovley, and T. Mester.** 2008. Proteome of *Geobacter sulfurreducens* grown with Fe(III) oxide or Fe(III) citrate as the electron acceptor. Biochim Biophys Acta **1784:**1935-1941.
34. **Duan, C., and J. Feng.** 2010. Mining metagenomes for novel cellulase genes. Biotechnol Lett **32:**1765-1775.
35. **Feng, Y., C. Duan, H. Pang, X. Mo, C. Wu, Y. Yu, Y. Hu, J. Wei, J. Tang, and J. Feng.** 2007. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. Appl Microbiol Biotechnol **75:**319-328.
36. **Fishilevich, S., L. Amir, Y. Fridman, A. Aharoni, and L. Alfonta.** 2009. Surface display of redox enzymes in microbial fuel cells. J Am Chem Soc **131:**12052–12053.
37. **Fowler, A. V., and I. Zadin.** 1970. The amino acid sequence of beta galactosidase. J Biol Chem **245:**5032-5041.
38. **Francisco, J. A., C. Earhart, and G. Georgiou.** 1992. Transport and anchoring of beta-lactamase to the external surface of *Escherichia coli*. Proc Natl Acad Sci U S A **89:**2713-2717.
39. **Francisco, J. A., C. Stathopoulos, R. A. J. Warren, D. G. Kilburn, and G. Georgiou.** 1993. Specific adhesion and hydrolysis of cellulose by intact *Escherichia coli* expressing surface anchored cellulase or cellulose binding domains. Nat Biotechnol **11:**491-495.
40. **Freudl, R., S. MacIntyre, M. Degen, and U. Henning.** 1986. Cell surface exposure of the outer membrane protein OmpA of *Escherichia coli* K-12. J Mol Biol **188:**491-494.
41. **Georgiou, G., H. L. Poetschke, C. Stathopoulos, and J. A. Francisco.** 1993. Practical applications of engineering Gram-negative bacterial cell surfaces. Trends Biotechnol **11:**6-10.

42. **Ghose, T. K.** 1987. Measurement of cellulase activities. Pure Appl Chem **59:**257-268.
43. **Gilad, R., L. Rabinovich, S. Yaron, E. A. Bayer, R. L. Lamed, H. J. Gilbert, and Y. Shoham.** 2003. CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. J Bacteriol **185:**391-398.
44. **Gnansounou, E., and A. Dauriat.** 2010. Techno-economic analysis of lignocellulosic ethanol: a review. Bioresour Technol **101:**4980-4991.
45. **Halter, R., J. Pohlner, and T. F. Meyer.** 1984. IgA protease of Neisseria gonorrhoeae: isolation and characterization of the gene and its extracellular product. EMBO J **3:**1595-1601.
46. **Handelsman, J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev **68:**669-685.
47. **Hartmans, S., H. T. de Vries, P. Beijer, R. L. Brady, M. Hofbauer, and A. J. Haandrikman.** 2004. Production of oxidized guar galactomannan and its applications in the paper industry, vol. 864. American Chemical Society.
48. **Healy, F., R. Ray, H. Aldrich, A. Wilkie, L. O. Ingram, and K. T. Shanmugam.** 1995. Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. Appl Microbiol Biotechnol **43:**667-674.
49. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. Fernandez, and D. Ala'Aldeen.** 2004. Type V protein secretion pathway: the autotransporter story. Microbiol Mol Biol Rev **68:**692-744.
50. **Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro.** 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol **6:**370-8.
51. **Hendriks, A. T. W. M., and G. Zeeman.** 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresour Technol **100:**10-18.
52. **Hess, M., A. Sczyrba, R. Egan, T.-W. Kim, H. Chokhawala, G. Schroth, S. Luo, D. S. Clark, F. Chen, T. Zhang, R. I. Mackie, L. A. Pennacchio, S. G. Tringe, A. Visel, T. Woyke, Z. Wang, and E. M. Rubin.** 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science **331:**463-467.
53. **Holmes, D. E., K. T. Finneran, R. A. O'Neil, and D. R. Lovley.** 2002. Enrichment of members of the family *Geobacteraceae* associated with stimulation of dissimilatory metal reduction in uranium-contaminated aquifer sediments. Appl Environ Microbiol **68:**2300-2306.
54. **Humbird, D., and A. Aden.** 2009. Biochemical production of ethanol from corn stover: 2008 state of technology model. National Renewable Energy Laboratory.
55. **Huson, D. H., D. C. Richter, S. Mitra, A. F. Auch, and S. C. Schuster.** 2009. Methods for comparative metagenomics. BMC Bioinformatics **10**.
56. **Inoue, K., C. Leang, A. Franks, T. L. Woodard, K. P. Nevin, and D. R. Lovley.** 2010. Specific localization of the *c*-type cytochrome OmcZ at the anode surface in current-producing biofilms of *Geobacter sulfurreducens*. Environ Microbiol Rep **3:**211-217.

57. **Inoue, K., X. Qian, L. Morgado, B. C. Kim, T. Mester, M. Izallalen, C. A. Salgueiro, and D. R. Lovley.** 2010. Purification and characterization of OmcZ, an outer-surface, octaheme *c*-type cytochrome essential for optimal current production by *Geobacter sulfurreducens*. Appl Environ Microbiol **76:**3999-4007.
58. **Jose, J., R. Bernhardt, and F. Hannemann.** 2001. Functional display of active bovine adrenodoxin on the surface of *E. coli* by chemical incorporation of the [2Fe-2S] cluster. Chembiochem **2:**695-701.
59. **Jose, J., and T. F. Meyer.** 2007. The autodisplay story, from discovery to biotechnical and biomedical applications. Microbiol Mol Biol Rev **71:**600-619.
60. **Jose, J., and S. von Schwichow.** 2004. Autodisplay of active sorbitol dehydrogenase (SDH) yields a whole cell biocatalyst for the synthesis of rare sugars. Chembiochem **5:**491-499.
61. **Jung, H., J. Lebeault, and J. Pan.** 1998. Surface display of *Zymomonas mobilis* levansucrase by using the ice-nucleation protein of *Pseudomonas syringae*. Nat Biotechnol **16:**576-580.
62. **Jung, H., J. Park, S. Park, J. Lebeault, and J. Pan.** 1998. Expression of carboxymethylcellulase on the surface of *Escherichia coli* using *Pseudomonas syringae* ice nucleation protein. Enzyme Microb Technol **22:**348-354.
63. **Karp, A., and I. Shield.** 2008. Bioenergy from plants and the sustainable yield challenge. New Phytol **179:**15-32.
64. **Kataeva, I., V. N. Uversky, and L. G. Ljungdahl.** 2003. Calcium and domain interactions contribute to the thermostability of domains of the multimodular cellobiohydrolase, CbhA, a subunit of the *Clostridium thermocellum* cellulosome. Biochem J **372:**151-161.
65. **Kawahara, H.** 2002. The structures and functions of ice crystal-controlling proteins from bacteria. J Biosci Bioeng **94:**492-496.
66. **Kennedy, J., N. D. O'Leary, G. S. Kiran, J. P. Morrissey, F. O'Gara, J. Selvin, and A. D. W. Dobson.** 2011. Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. J Appl Microbiol **111:**787-788.
67. **Kim, B., X. Qian, L. Ching, M. Coppi, and D. Lovley.** 2006. Two putative *c*-type multiheme cytochromes required for the expression of OmcB, an outer membrane protein essential for optimal Fe(III) reduction in *Geobacter sulfurreducens*. J Bacteriol **188:**3138-3142.
68. **Kim, B.-C., and D. R. Lovley.** 2008. Investigation of direct vs. indirect involvement of the *c*-type cytochrome MacA in Fe(III) reduction by *Geobacter sulfurreducens*. FEMS Microbiol Lett **286:**39-44.
69. **Kim, C., H. Choi, and D. S. Lee.** 1993. Purification and biochemical properties of an alkaline pullulanase from alkalophilic *Bacillus* sp. S-1. Biosci Biotechnol Biochem **57:**1632-1637.
70. **Kim, S., C. Lee, B. Han, M. Kim, Y. Yeo, S. Yoon, B. Koo, and H. Jun.** 2008. Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiol Lett **282:**44-51.

71. **Kim, S.-J., C.-M. Lee, B.-R. Han, M.-Y. Kim, Y.-S. Yeo, S.-H. Yoon, B.-S. Koo, and H.-K. Jun.** 2008. Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiology Letters **282:**44-51.
72. **Klauser, T., J. Krämer, K. Otzelberger, J. Pohlner, and T. F. Meyer.** 1993. Characterization of the *Neisseria* Igaβ-core: the essential unit for outer membrane targeting and extracellular protein secretion. J Mol Biol **234:**579-593.
73. **Klemm, P., and M. A. Schembri.** 2000. Fimbriae assisted bacterial surface display of heterologous peptides. Int J Med Microbiol **290:**215-221.
74. **Kotrba, P., L. Rulíšek, and T. Ruml.** 2011. Bacterial Surface Display of Metal-Binding Sites. Springer.
75. **Kovach, M. E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop, and K. M. Peterson.** 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene **166:**175-176.
76. **Kuroda, K., and M. Ueda.** 2011. Cell surface engineering of yeast for applications in white biotechnology. Biotechnol Lett **33:**1-9.
77. **Kuroda, K., and M. Ueda.** 2011. Molecular design of the microbial cell surface toward the recovery of metal ions. Curr Opin Biotechnol **22:**427-433.
78. **Kurokawa, J., E. Hemjinda, K. Arai, T. Kimura, K. Sakka, and K. Ohmiya.** 2002. *Clostridium thermocellum* cellulase CelT, a family 9 endoglucanase without an Ig-like domain or family 3c carbohydrate-binding module. Appl Microbiol Biotechnol **59:**455-461.
79. **Kwon, E. J., Y. S. Jeong, Y. H. Kim, S. K. Kim, H. B. Na, J. Kim, H. D. Yun, and H. Kim.** 2010. Construction of a metagenomic library from compost and screening of cellulase and xylanase positive clones. Journal of the Korean Society for Applied Biological Chemistry **53:**702-708.
80. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.
81. **Lattemann, C. T., J. Maurer, E. Gerland, and T. F. Meyer.** 2000. Autodisplay: functional display of active beta-lactamase on the surface of *Escherichia coli* by the AIDA-I autotransporter. J Bacteriol **182:**3726-3733.
82. **Leang, C., M. V. Coppi, and D. R. Lovley.** 2003. OmcB, a *c*-type polyheme cytochrome, involved in Fe(III) reduction in *Geobacter sulfurreducens*. J Bacteriol **185:**2096-2103.
83. **Leang, C., X. Qian, T. Mester, and D. R. Lovley.** 2010. Alignment of the *c*-type cytochrome OmcS along pili of *Geobacter sulfurreducens*. Appl Environ Microbiol **76:**4080-4084.
84. **Lee, S., J. H. Choi, and Z. Xu.** 2003. Microbial cell-surface display. Trends Biotechnol **21:**45-52.
85. **Li, C., Y. Zhu, I. Benz, M. Schmidt, W. Chen, A. Mulchandani, and C. Qiao.** 2008. Presentation of functional organophosphorus hydrolase fusions on the surface of *Escherichia coli* by the AIDA- I autotransporter pathway. Biotechnol Bioeng **99:**485-490.

86. **Li, L., D. G. Kang, and H. Cha.** 2004. Functional display of foreign protein on surface of *Escherichia coli* using N- terminal domain of ice nucleation protein. Biotechnol Bioeng **85:**214-221.
87. **Li, L., S. Taghavi, S. M. McCorkle, Y. Zhang, M. G. Blewitt, R. Brunecky, W. S. Adney, M. E. Himmel, P. Brumm, C. Drinkwater, D. A. Mead, S. G. Tringe, and D. van der Lelie.** 2011. Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases. Biotechnol Biofuels **4**.
88. **Liang, B., L. Li, M. Mascin, and A. Liu.** 2011. Construction of xylose dehydrogenase displayed on the surface of bacteria using ice nucleation protein for sensitive d-xylose detection. Anal Chem **84:**275-282.
89. **Lindow, S. E., D. C. Arny, and C. D. Upper.** 1982. Bacterial ice nucleation: a factor in frost injury to plants. Plant Physiol **70:**1084-1089.
90. **Lindow, S. E., E. Lahue, A. G. Govindarajan, N. J. Panopoulos, and D. Gies.** 1989. Localization of ice nucleation activity and the iceC gene product in *Pseudomonas syringae* and *Escherichia coli*. Mol Plant Microbe Interact **2:**262-272.
91. **Lloyd, J. R., E. L. Blunt-Harris, and D. R. Lovley.** 1999. The periplasmic 9.6-kilodalton *c*-type cytochrome of *Geobacter sulfurreducens* is not an electron shuttle to Fe(III). J Bacteriol **181:**7647-9.
92. **Lloyd, J. R., C. Leang, A. L. H. Myerson, M. V. Coppi, S. Cuifo, B. Methe, S. J. Sandler, and D. R. Lovley.** 2003. Biochemical and genetic characterization of PpcA, a periplasmic *c*-type cytochrome in *Geobacter sulfurreducens*. Biochem J **369:**153-161.
93. **Lommel, M., and S. Strahl.** 2009. Protein O-mannosylation: conserved from bacteria to humans. Glycobiology **19:**816-828.
94. **Lorenz, P., and J. Eck.** 2005. Metagenomics and industrial applications. Nat Rev Microbiol **6:**510-516.
95. **Lovley, D. R.** 1995. Bioremediation of organic and metal contaminants with dissimilatory metal reduction. J Ind Microbiol **14:**85-93.
96. **Lovley, D. R.** 2006. Bug juice: harvesting electricity with microorganisms. Nat Rev Microbiol **4:**497-508.
97. **Lovley, D. R.** 1991. Dissimilatory Fe (III) and Mn (IV) reduction. Microbiol Rev **55:**259-287.
98. **Lovley, D. R.** 2006. Dissimilatory Fe (III)-and Mn (IV)-reducing prokaryotes.
99. **Lovley, D. R., S. J. Giovannoni, D. C. White, J. E. Champine, E. J. P. Phillips, Y. A. Gorby, and S. Goodwin.** 1993. *Geobacter metallireducens* gen. nov. sp. nov., a microorganism capable of coupling the complete oxidation of organic compounds to the reduction of iron and other metals. Arch Microbiol **159:**336-344.
100. **Lovley, D. R., D. E. Holmes, and K. P. Nevin.** 2004. Dissimilatory Fe (III) and Mn (IV) reduction. Adv Microb Physiol **49:**219-286.
101. **Lovley, D. R., and D. J. Lonergan.** 1990. Anaerobic oxidation of toluene, phenol, and p-cresol by the dissimilatory iron-reducing organism GS-15. Appl Environ Microbiol **56:**1858-1864.

102. **Lovley, D. R., and E. J. Phillips.** 1986. Organic matter mineralization with reduction of ferric iron in anaerobic sediments. Appl Environ Microbiol **51:**683-689.
103. **Lovley, D. R., and E. J. P. Phillips.** 1988. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. Appl Environ Microbiol **54:**1472-1480.
104. **Majander, K., T. K. Korhonen, and B. Westerlund-Wikstrom.** 2005. Simultaneous display of multiple foreign peptides in the FliD capping and FliC filament proteins of the *Escherichia coli* flagellum. Appl Environ Microbiol **71:**4263-4268.
105. **Malvankar, N. S., M. Vargas, K. P. Nevin, A. E. Franks, C. Leang, B.-C. Kim, K. Inoue, T. Mester, S. F. Covalla, J. P. Johnson, V. M. Rotello, M. T. Tuominen, and D. R. Lovley.** 2011. Tunable metallic-like conductivity in microbial nanowire networks. Nat Nanotechnol **6:**573-579.
106. **Margeot, A., B. Hahn-Hagerdal, M. Edlund, R. Slade, and F. Monot.** 2009. New improvements for lignocellulosic ethanol. Curr Opin Biotechnol **20:**372-380.
107. **Marsili, E., D. B. Baron, I. D. Shikhare, D. Coursolle, J. A. Gralnick, and D. R. Bond.** 2008. *Shewanella* Secretes flavins that mediate extracellular electron transfer. Proc Natl Acad Sci U S A **105:**3968-3973.
108. **Marsili, E., J. Rollefson, D. B. Baron, R. M. Hozalski, and D. R. Bond.** 2008. Microbial biofilm voltammetry: direct electrochemical characterization of catalytic electrode-attached biofilms. Appl Environ Microbiol **74:**7329-7337.
109. **Martinez, A., S. J. Kolvek, C. L. T. Yip, J. Hopke, K. A. Brown, I. A. MacNeil, and M. S. Osburne.** 2004. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. Appl Environ Microbiol **70:**2452-2463.
110. **Maurer, J., J. Jose, and T. F. Meyer.** 1997. Autodisplay: one-component system for efficient surface display and release of soluble recombinant proteins from *Escherichia coli*. J Bacteriol **179:**794-804.
111. **Maurer, J., J. Jose, and T. F. Meyer.** 1999. Characterization of the essential transport function of the AIDA-I autotransporter and evidence supporting structural predictions. J Bacteriol **181:**7014-7020.
112. **McMahon, M. D., C. Guan, J. Handelsman, and M. G. Thomas.** 2012. Metagenomic analysis of *Streptomyces lividans* reveals host-dependent functional expression. Appl Environ Microbiol **78:**3622-3629.
113. **McMillan, J. D.** 1994. Enzymatic conversion of biomass for fuels production, vol. 566. ACS Symposium Series, Golden, CO.
114. **Mehta, T., S. E. Childers, R. Glaven, D. R. Lovley, and T. Mester.** 2006. A putative multicopper protein secreted by an atypical type II secretion system involved in the reduction of insoluble electron acceptors in *Geobacter sulfurreducens*. Microbiology **152:**2257-2264.

115. **Mehta, T., M. V. Coppi, S. E. Childers, and D. R. Lovley.** 2005. Outer membrane *c*-type cytochromes required for Fe(III) and Mn(IV) oxide reduction in *Geobacter sulfurreducens*. Appl Environ Microbiol **71:**8634-8641.
116. **Miller, G. L.** 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem **31:**426-428.
117. **Miller, J. H.** 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
118. **Mingardon, F., J. D. Bagert, C. Maisonnier, D. L. Trudeau, and F. H. Arnold.** 2011. Comparison of family 9 cellulases from mesophilic and thermophilic bacteria. Appl Environ Microbiol **77:**1436-1442.
119. **Mosier, N., C. Wyman, B. Dale, R. Elander, Y. Y. Lee, M. Holtzapple, and M. Ladisch.** 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol **96:**673-686.
120. **Murai, T., M. Ueda, M. Yamamura, H. Atomi, Y. Shibasaki, N. Kamasawa, M. Osumi, T. Amachi, and A. Tanaka.** 1997. Construction of a starch-utilizing yeast by cell surface engineering. Appl Environ Microbiol **63:**1362-1366.
121. **Myers, C. R., and K. H. Nealson.** 1988. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor Science **240:**1319-1321.
122. **Nam, K. H., S. Kim, and K. Y. Hwang.** 2009. Crystal structure of CelM2, a bifunctional glucanase-xylanase protein from a metagenome library. Biochem Biophys Res Commun **383:**183-186.
123. **Narita, J., K. Okano, T. Tateno, T. Tanino, T. Sewaki, M. Sung, H. Fukuda, and A. Kondo.** 2006. Display of active enzymes on the cell surface of *Escherichia coli* using PgsA anchor protein and their application to bioconversion. Appl Microbiol Biotechnol **70:**564-572.
124. **Nealson, K. H., and J. Scott.** 2006. Ecophysiology of the genus *Shewanella*, Third Edition ed, vol. 6. Springer, New York.
125. **Nevin, K. P., B.-C. Kim, R. H. Glaven, J. P. Johnson, T. L. Woodard, B. A. Methe, R. J. DiDonato, S. F. Covalla, A. E. Franks, A. Liu, and D. R. Lovley.** 2009. Anode biofilm transcriptomics reveals outer surface components essential for high density current production in *Geobacter sulfurreducens* fuel cells. PLoS One **4:**e5628.
126. **Nevin, K. P., and D. R. Lovley.** 2000. Lack of production of electron-shuttling compounds or solubilization of Fe(III) during reduction of insoluble Fe(III) oxide by *Geobacter metallireducens*. Appl Environ Microbiol **66:**2248-2251.
127. **Nevin, K. P., and D. R. Lovley.** 2002. Mechanisms for accessing insoluble Fe(III) oxide during dissimilatory Fe(III) reduction by *Geothrix fermentans*. Appl Environ Microbiol **68:**2294-2299.
128. **Nogawa, M., M. Goto, H. Okada, and Y. Morikawa.** 2001. L-Sorbose induces cellulase gene transcription in the cellulolytic fungus *Trichoderma reesei*. Curr Genet **38:**329-334.
129. **Notenboom, V., A. B. Boraston, D. G. Kilburn, and D. R. Rose.** 2001. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga*

maritima xylanase 10A in native and ligand-bound forms. Biochemistry **40:**6248-6256.

130. **Ockerman, P.** 1968. Identity of β-glucosidase,β-xylosidase and one of the β-galactosidase activities in human liver when assayed with 4-methylumbelliferyl-β-D-glycosides studies in case of Gaucher's Disease. Biochim Biophys Acta **165:**59-62.
131. **Pang, H., P. Zhang, C. J. Duan, X. C. Mo, J. L. Tang, and J. X. Feng.** 2009. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. Curr Microbiol **58:**404-408.
132. **Pettersen, R. C.** 1984. The chemical composition of wood, vol. 207. American Chemical Society.
133. **Qian, X., G. Reguera, T. Mester, and D. R. Lovley.** 2007. Evidence that OmcB and OmpB of *Geobacter sulfurreducens* are outer membrane surface proteins. FEMS Microbiol Lett **277:**21-27.
134. **Richter, H., K. P. Nevin, H. Jia, D. A. Lowy, D. R. Lovley, and L. M. Tender.** 2009. 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. Energ Environ Sci **2:**506-516.
135. **Riesenfeld, C. S., P. D. Schloss, and J. Handelsman.** 2004. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet **38:**525-552
136. **Rollefson, J., C. Stephen, M. Tien, and D. R. Bond.** 2011. Identification of an extracellular polysaccharide network essential for cytochrome anchoring and biofilm formation in *Geobacter sulfurreducens*. J Bacteriol **193:**1023-1033.
137. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, C. J, J. Handelsman, and R. M. Goodman.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol **66:**2541.
138. **Rooney, W., J. Blumenthal, B. Bean, and J. E. Mullet.** 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels Bioproducts and Biorefining **1:**147-157.
139. **Ross, D. E., J. M. Flynn, D. B. Baron, J. A. Gralnick, and D. R. Bond.** 2011. Towards electrosynthesis in *Shewanella*: energetics of reversing the Mtr pathway for reductive metabolism. Plos One **6:**e16649.
140. **Sakon, J., W. S. Adney, M. E. Himmel, S. R. Thomas, and P. A. Karplus.** 1996. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. Biochemistry **35:**10648-10660.
141. **Saltikov, C., and D. Newman.** 2003. Genetic identification of a respiratory arsenate reductase. Proc Natl Acad Sci U S A **100:**10983-10988.

142. **Samuelson, P., E. Gunneriusson, P. Nygren, and S. Stahl.** 2002. Display of proteins on bacteria. J Biotechnol **96:**129-154.
143. **Schmidt, T. M., E. DeLong, and N. R. Pace.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol **173:**4371-4378.
144. **Schubot, F. D., I. A. Kataeva, J. Chang, A. K. Shah, L. G. Ljungdahl, J. P. Rose, and B. Wang.** 2004. Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*. Biochemistry **43:**1163-1170.
145. **Schwarz, W.** 2001. The cellulosome and cellulose degradation by anaerobic bacteria. Applied microbiology and biotechnology **56:**634-649.
146. **Serrano, J., L. Higgins, B. A. Witthuhn, L. B. Anderson, T. Markowski, G. W. Holcombe, P. A. Kosian, J. J. Korte, J. E. Tietge, and S. J. Degitz.** 2010. In vivo assessment and potential diagnosis of xenobiotics that perturb the thyroid pathway: Proteomic analysis of *Xenopus laevis* brain tissue following exposure to model T4 inhibitors. Comp Biochem Physiol Part D Genomics Proteomics **5:**138-150.
147. **Shallom, D., and Y. Shoham.** 2003. Microbial hemicellulases. Curr Opin Microbiol **6:**219-228.
148. **Shi, H., and W. W. Su.** 2001. Display of green fluorescent protein on *Escherichia coli* cell surface. Enzyme Microb Technol **28:**25-34.
149. **Shi, L., S. Deng, M. J. Marshall, Z. Wang, D. W. Kennedy, A. C. Dohnalkova, H. M. Mottaz, E. A. Hill, Y. A. Gorby, A. S. Beliaev, D. J. Richardson, J. M. Zachara, and J. K. Fredrickson.** 2008. Direct involvement of type II secretion system in extracellular translocation of *Shewanella oneidensis* outer membrane cytochromes MtrC and OmcA. J Bacteriol **190:**5512-5516.
150. **Shi, L., D. J. Richardson, Z. Wang, S. N. Kerisit, K. M. Rosso, J. M. Zachara, and J. K. Fredrickson.** 2009. The roles of outer membrane cytochromes of *Shewanella* and *Geobacter* in extracellular electron transfer. Environ Microbiol Rep **1:**220-227.
151. **Shi, L., T. C. Squier, J. M. Zachara, and J. K. Fredrickson.** 2007. Respiration of metal (hydr)oxides by *Shewanella* and *Geobacter*: a key role for multihaem *c*-type cytochromes. Mol Microbiol **65:**12-20.
152. **Shimazu, M., A. Mulchandani, and W. Chen.** 2001. Cell surface display of organophosphorus hydrolase using ice nucleation protein. Biotechnol Prog **17:**76-80.
153. **Singhania, R. R., R. K. Sukumaran, A. K. Patel, C. Larroche, and A. Pandey.** 2010. Advancement and comparative profiles in the production technologies using solid-state and submerged fermentation for microbial cellulases. Enzyme Microb Technol **46:**541-549.
154. **Smith, G. P.** 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science **228:**1315-1317.
155. **Smith, G. P., and V. A. Petrenko.** 1997. Phage display. Chem Rev **97:**391-410.

156. **Snoeyenbos-West, O. L., K. P. Nevin, R. T. Anderson, and D. R. Lovley.** 2000. Enrichment of *Geobacter* species in response to stimulation of Fe(III) reduction in sandy aquifer sediments. Microb Ecol **39:**153-167.
157. **Stathopoulos, C., G. Georgiou, and C. F. Earhart.** 1996. Characterization of *Escherichia coli* expressing an Lpp'OmpA (46-159)-PhoA fusion protein localized in the outer membrane. Appl Microbiol Biotechnol **45:**112-119.
158. **Steele, H. L., K. Jaeger, R. Daniel, and W. R. Streit.** 2009. Advances in recovery of novel biocatalysts from metagenomes. J Mol Microbiol Biotechnol **16:**25-37.
159. **Steele, H. L., and W. R. Streit.** 2005. Metagenomics: advances in ecology and biotechnology. FEMS Microbiol Lett **247:**105-111.
160. **Stein, L. Y., M. T. La Duc, T. J. Grundl, and K. H. Nealson.** 2001. Bacterial and archaeal populations associated with freshwater ferromanganous micronodules and sediments. Environ Microbiol **3:**10-18.
161. **Strycharz-Glaven, S. M., R. M. Snider, A. Guiseppi-Elie, and L. M. Tender.** 2011. On the electrical conductivity of microbial nanowires and biofilms. Energ Environ Sci **4:**4366-4379.
162. **Su, G., H. M. Brahmbhatt, J. Wehland, M. Rohde, and K. N. Timmis.** 1992. Construction of stable LamB-Shiga toxin B subunit hybrids: analysis of expression in *Salmonella typhimurium* aroA strains and stimulation of B subunit-specific mucosal and serum antibody responses. Infect Immun **60:**3345-3359.
163. **Szczupak, A., D. Kol-Kalmanz, and L. Alfonta.** 2012. A hybrid biocathode: surface display of O2-reducing enzymes for microbial fuel cell applications. Chem Commun (Camb) **48:**49-51.
164. **Tanskanen, J., T. K. Korhonen, and B. Westerlund-Wikstrom.** 2000. Construction of a multihybrid display system: flagellar filaments carrying two foreign adhesive peptides. Appl Environ Microbiol **66:**4152-4156.
165. **Teather, R. M., and P. J. Wood.** 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol **43:**777.
166. **Torsvik, V., J. Goksoyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. Appl Environ Microbiol **56:**782.
167. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin.** 2005. Comparative metagenomics of microbial communities. Science **308:**554-557.
168. **Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield.** 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:**37-43.
169. **Valls, M., S. Atrian, V. De Lorenzo, and L. A. Fernandez.** 2000. Engineering a mouse metallothionein on the cell surface of *Ralstonia eutropha* CH34 for immobilization of heavy metals in soil. Nat Biotechnol **18:**661-665.

170. **van Bloois, E., R. Winter, H. Kolmar, and M. W. Fraaije.** 2011. Decorating microbes: surface display of proteins on *Escherichia coli*. Trends Biotechnol **29:**79-86.
171. **van Tilbeurgh, H., M. Claeyssens, and C. K. Bruyne.** 1982. The use of 4-methylumbelliferyl and other chromophoric glycosides in the study of cellulolytic enzymes. FEBS Lett **149:**152-156.
172. **Veiga, E., V. De Lorenzo, and L. A. Fernandez.** 1999. Probing secretion and translocation of a β- autotransporter using a reporter single- chain Fv as a cognate passenger domain. Mol Microbiol **33:**1232-1243.
173. **Venter, J. C., K. Remington, J. H. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66-74.
174. **Vick, J. E., E. T. Johnson, S. Choudhary, S. E. Bloch, F. Lopez-Gallego, P. Srivastava, I. B. Tikh, G. T. Wawrzyn, and C. Schmidt-Dannert.** 2011. Optimized compatible set of BioBrick (TM) vectors for metabolic pathway engineering. Appl Microbiol Biotechnol **92:**1275-1286.
175. **Voget, S., H. Steele, and W. Streit.** 2006. Characterization of a metagenome-derived halotolerant cellulase. Journal of biotechnology **126:**26-36.
176. **Voget, S., H. L. Steele, and W. R. Streit.** 2006. Characterization of a metagenome-derived halotolerant cellulase. J Biotechnol **126:**26-36.
177. **von Canstein, H., J. Ogawa, S. Shimizu, and J. R. Lloyd.** 2008. Secretion of flavins by *Shewanella* species and their role in extracellular electron transfer. Appl Microbiol Biotechnol **74:**615-623.
178. **Wang, J., and Y. Chao.** 2006. Immobilization of cells with surface-displayed chitin-binding domain. Appl Environ Microbiol **72:**927-931.
179. **Warnecke, F., P. Luginbuhl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernandez, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature **450:**560-565.
180. **Wehmeier, S., A. Varghese, S. Gurcha, B. Tissot, M. Panico, P. Hitchen, H. R. Morris, G. S. Besra, A. Dell, and M. C. M. Smith.** 2009. Glycosylation of the phosphate binding protein, PstS, in *Streptomyces coelicolor* by a pathway that resembles protein O- mannosylation in eukaryotes. Mol Microbiol **71:**421-433.
181. **Westerlund-Wikström, B.** 2000. Peptide display on bacterial flagella: principles and applications. Int J Med Microbiol **290:**223-230.
182. **Woese, C. R.** 1987. Bacterial evolution. Microbiol Mol Biol Rev **51:**221-271.

183. **Wolber, P., C. A. Deininger, M. W. Southworth, J. Vandekerckhove, M. V. Montagu, and G. J. Warren.** 1986. Identification and purification of a bacterial ice-nucleation protein. Proc Natl Acad Sci U S A **83:**7256-7260.
184. **Wu, C. H., A. Mulchandani, and W. Chen.** 2008. Versatile microbial surface-display for environmental remediation and biofuels production. Trends Microbiol **16:**181-188.
185. **Yang, B., Z. Dai, S. Ding, and C. E. Wyman.** 2011. Enzymatic hydrolysis of cellulosic biomass. Biofuels **2:**421-450.
186. **Yim, S., H. Jung, J. Pan, H. Kang, T. Ahn, and C. Yun.** 2006. Functional expression of mammalian NADPH-cytochrome P450 oxidoreductase on the cell surface of *Escherichia coli*. Protein Expr Purif **49:**292-298.
187. **Zhang, F., J. Chen, W. Ren, G. Nie, H. Ming, S. Tang, and W. Li.** 2011. Cloning, expression and characterization of an alkaline thermostable GH9 endoglucanase from *Thermobifida halotolerans* YIM 90462T. Bioresour Technol **102:**10143-10146.
188. **Zhang, Y. H. P., M. E. Himmel, and J. R. Mielenz.** 2006. Outlook for cellulase improvement: screening and selection strategies. Biotechnol Adv **24:**452-481.