`

CHARACTERIZATION OF LEVOGLUCOSAN UTILIZING BACTERIA AND THE ENZYME LEVOGLUCOSAN DEHYDROGENASE

by

AJAY S. ARYA

(Under the Direction of Mark A. Eiteman)

ABSTRACT

The gene for levoglucosan dehydrogenase, *lgdA*, was identified in *Pseudarthrobacter phenanthrenivorans* Sphe3 by searching the genome with known peptide sequences. The bacteria *Paenibacillus athensensis* MEC069^T, *Shinella sumterensis* MEC087^T, *Microbacterium marinilacus* MEC084, and *Klebsiella pneumoniae* MEC097 were isolated from soil and wastewater based on their ability to consume levoglucosan, and the genomes of these organisms were assembled and screened for *lgdA*. Identified *lgdA* were recombinantly expressed in *E. coli* BL21 (DE3)-RIPL cells, and the levoglucosan dehydrogenases (LGDH) were purified by FPLC and kinetically characterized based on their activity in the presence of levoglucosan and NAD^+. These enzymes were found to vary significantly. LGDH from *P. phenanthrenivorans* and *K. pneumoniae* showed a K_M for levoglucosan of 27 and 22 mM, respectively, and both enzymes showed no sign of substrate inhibition. The LGDH from *P. athensensis*, *S. sumterensis,* and *M. marinilacus* MEC084 showed K_M for levoglucosan of ~2 mM and were inhibited by levoglucosan at concentrations as low as 93 mM. All LGDH showed K_M for NAD^+ in the 0.3-1.0 mM range, except for the the enzyme from *S. sumterensis* with a K_M of 9.1 mM.

Purified *P. phenanthrenivorans* Sphe3 LGDH was crystallized, and some crystals were soaked with the substrates, levoglucosan or glucose, and NADH. The X-ray structures were solved by molecular replacement to resolve which residues make up the active site and how substrates bind to the active site. Based on the observed structures, the mechanism of catalysis in LGDH was proposed to be dependent on a His-Lys-Asp triad conserved in similar oxidoreducatases, in which His begins catalysis by taking the proton from hydroxyl-3 of levoglucosan and a hydride is shifted from levoglucosan C3 to the nicotinamide ring C4. Lys and Asp are likely to form a proton relay to remove the proton from His so that catalysis may continue.

INDEX WORDS: levoglucosan dehydrogenase, LGDH, x-ray structure, genomic sequencing, enzyme kinetics

CHARACTERIZATION OF LEVOGLUCOSAN UTILIZING BACTERIA AND THE ENZYME LEVOGLUCOSAN DEHYDROGENASE

by

AJAY S. ARYA

B. S., University of Florida, 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

CHARACTERIZATION OF LEVOGLUCOSAN UTILIZING BACTERIA AND THE ENZYME LEVOGLUCOSAN DEHYDROGENASE

by

AJAY S. ARYA

Major Professor:	Mark A. Eiteman
Committee:	Cory Momany William B. Whitman Joy Doran-Peterson

Electronic Version Approved:

Suzanne Barbour Ph.D.
Dean of the Graduate School
The University of Georgia
May 2019

DEDICATION

I dedicate this project to the scientist who opened my eyes to the microscopic life around us and inspired me to continue learning about microbes' place in the world, my grandfather, Mitan Arya.

ACKNOWLEDGEMENTS

I would like to thank Brandon Satinsky for his assistance in preparing genomic DNA for Illumina sequencing and his guidance on how to process the data; Dr. Mary-Ann Moran for giving me the opportunity to obtain genomic sequences; and Sarah Lee for her friendly assistance throughout my time in the Eiteman Lab.

I would also like to acknowledge my committee members Dr. Mark A. Eiteman, Dr. Cory Momany, Dr. William B. Whitman, and Dr. Joy Doran-Peterson. Thank you for welcoming me into your laboratories and being my mentors. In particular, I would like to acknowledge Dr. Eiteman for encouraging me to present my research whenever opportunities presented themselves so my skills as a public speaker would develop.

Finally, I would like to acknowledge my parents, Coleen Watkins and Prabhakar Arya, for their encouragement to pursue learning any skill that made me happy. Thank you for giving me so many opportunities as a child and your continued support as I've spent the past twelve years in college.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

An imminent need exists to reduce CO_2 emissions before climate change reaches an irreversible point. Since the industrial revolution, fossil fuels have been the primary source of anthropogenic CO_2, followed by human use of land and forestry. In 1750, the beginning of the industrial revolution, atmospheric CO_2 was approximately 277 ppm (Joos and Spahni, 2008), while in April 2019 the concentration was 409 ppm (Dlugokencky and Tans, 2019). Moreover, CO_2 emissions from fossil fuels have increased from an average of 3.1 ± 0.2 GtC yr^{-1} in the 1960s to 9.4 ± 0.5 GtC yr^{-1} during 2008-2017. Although the growth rate of CO_2 emissions has shown a recent decline, from 3.2% yr^{-1} in the 2000s to 1.5% yr^{-1} during 2008-2017 (Le Quere et al., 2018), the world is reaching a critical point in its history when every possible means to reduce CO_2 emissions should be considered.

Pyrolysis and Levoglucosan Formation

Fast pyrolysis is an attractive means of waste disposal as well as a method to convert biomass to fermentable sugars. Pyrolysis of cellulose begins at 300 °C and yields a char, volatile degradation products, and a tar fraction which upon hydrolysis contains levoglucosan as the dominant product (Shafizadeh and Fu, 1973) and cellobiosan as the second most abundant anhydrosugar (Lomax et al., 1991). As the process temperature is increased from 300 to 500 °C, the char fraction makes up less of the total weight while the levoglucosan-containing tar is increased (Shafizadeh et al., 1979).

The levoglucosan yield from cellulose is highly variable (5-80% of the total carbon) depending on the composition of the lignocellulosic biomass is being pyrolyzed, how the biomass was pretreated, and pyrolysis conditions (Maduskar et al., 2018). Lignocellulose is composed of cellulose (4-55%), hemicellulose (8-40%), lignin (7-26%), and the non-combustible minerals recovered from pyrolysis as ash (0.3-7%) (Sorek et al., 2014). Microwave pyrolysis of larch, used papers, and filter papers suggest that as the cellulose content a material increases, so does the amount of levoglucosan produced (Miura et al., 2001). During fast pyrolysis, the interactions in cellulose-lignin and cellulose-hemicellulose mixtures were found to have a negligible effect on the pyrolysis products. A native cellulose-hemicellulose mixture showed a similar result. A herbaceous native cellulose-lignin mixture showed a reduction in levoglucosan produced and an increase in the formation of low molecular weight compounds and furans. Woody native cellulose-lignin samples did not show the same reduction in levoglucosan, leading the authors to speculate that the higher degree of hydrogen bonding between lignin and cellulose in the herbaceous material inhibits the pyrolysis reactions that yield levoglucosan (Zhang et al., 2015). Pyrolysis of cellulose doped with varying inorganic salts showed a decrease in the formation of levoglucosan with following trend in relative effect: (a) cations: $K^+ > Na^+ > Ca^{2+} > Mg^{2+}$; (b) anions: $Cl^- > NO_3 \approx OH^- > CO_3^2 \approx PO_4^{3-}$. Furthermore, the addition of switchgrass ash to cellulose caused a shift to glyceraldehyde being the dominant pyrolysis product instead of levoglucosan (Patwardhan et al., 2010).

Pretreatment of lignocellulosic material is critical to maximize levoglucosan formation during fast pyrolysis. One of the main methods employed to improve levoglucosan yield is demineralization washing with water or mild acid pretreatment (Scott et al., 2001). The second prominently used method is to saturate lignocellulosic feedstocks with acid to form thermally

stable salts with alkaline and alkali earth metals naturally present, thereby reducing their impact on product formation. Impregnating woody lignocellulose with a 1% phosphoric acid solution leads to a levoglucosan content as high as 33.6% (Dobele et al., 2005) partitioned primarily into an aqueous phase (Bennett and Duff, 2005; Vitasari et al., 2011).

Levoglucosan Kinase

Levoglucosan kinase (LGK) has been the most extensively studied enzyme for conversion of levoglucosan in industrially useful organisms. LGK was first identified in a variety of yeasts and fungi including *Aspergillus terreus* K-26, *Sporobolomyces salmonicolor* IFO-0375, and *Cryptococcus albidus* IFO-0378 (Kitamura et al., 1990). LGK isolated from crude cell extracts phosphorylated levoglucosan to glucose 6-phosphate in the presence of magnesium ions (Fig. 1-1); LGK exhibited no phosphorylation of glucose, so it is a kinase specific to levoglucosan.

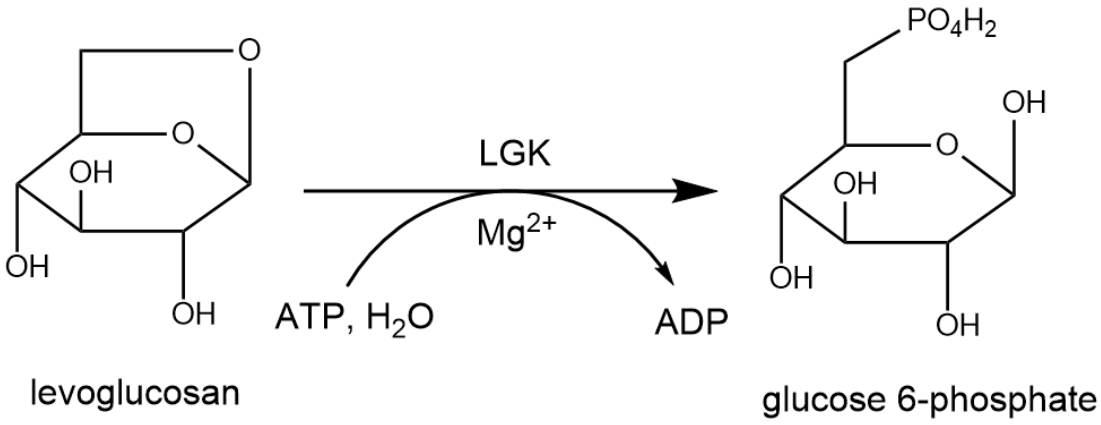


Figure 1-1. Conversion of levoglucosan to glucose 6-phosphate by levoglucosan kinase

Purified LGK from *S. salmonicolor* is specific for levoglucosan, exhibiting no activity on a variety of sugars (except for mannosan, 1,6-anhydro-β-D-mannopyranose, which was phosphorylated at 1% the relative rate as levoglucosan and has a similar intramolecular linkage

as levoglucosan). The K_Ms for levoglucosan and ATP are 85 mM and 0.19 mM, respectively. LGK is competitively inhibited by ADP with a K_i of 0.15 mM (Kitamura and Yasui, 1991). LGK from filamentous fungi have lower K_Ms for levoglucosan than LGK from yeasts: *A. niger* (60 mM) and *Penicillium herquei* (48 mM) compared with *Rhodotorula aurantiaca* (102 mM). The K_M for ATP is similar across different sources of LGK, and lies in the range of 0.21-0.35 mM (Xie et al., 2006). These enzymes exhibit optimal activity at pH 7-10, and a temperature of 30 °C.

Application of LGK

LGK was first cloned into *E. coli* DH5α from a cDNA library of *A. niger* CBX-209, yielding *Escherichia coli* capable of growth on defined medium with levoglucosan as the sole carbon source (Zhuang and Zhang, 2002). LGK isolated from a cDNA library of *Lipomyces starkeyi* YZ-215 expressed in *E. coli* BL21 DE3 cells enabled growth on levoglucosan as a sole carbon source (Dai et al., 2009). In both cases the heterologous enzyme exhibited a lower affinity for levoglucosan than enzyme isolated from its natural host (roughly double the K_M). The expression of codon optimized LGK in *E. coli* KO11 resulted in ethanol titers of 0.6 wt. % when grown on levoglucosan as the sole carbon source (Layton et al., 2011). Expression of codon optimized *lgk* in *E. coli* was shown feasible to produce styrene (Lian et al., 2016). By making single-site mutations in the codon-optimized LGK, expressing the mutated gene in *E. coli*, and monitoring the growth rate of these cells on levoglucosan, Klesmith and colleagues identified 86 mutations that improved growth rate more than 50% relative to cells expressing the codon-optimized gene, and 215 mutations that improved growth rate more than 20% (Klesmith et al., 2015). Ultimately, a combination of 15 mutations resulted in a 24-fold improvement in activity and a 15-fold improvement in growth rate. This enzyme was stable at a higher

temperature than the wild-type (41 °C compared to 34 °C) and exhibited double the relative catalytic efficiency (Klesmith et al., 2015). Codon-optimized *lgk* has been expressed in *Pseudomonas putida* KT2440 for the consumption of levoglucosan and cellobiosan (Linger et al., 2016), in *Corynebacterium glutamicum* to produce succinate from levoglucosan (Kim et al., 2015), and in *Rhodococcus jostii* RHA1to produce triacylglycerol (Xiong et al., 2016).

Conversion of levoglucosan using natural hosts of LGK has been extensively explored. *Aspergillus terreus* K26 produces itaconic acid from pure levoglucosan at rates and yields similar to glucose (Nakagawa et al., 1984). A γ-ray mutated strain of *Aspergillus niger* CBX-2 achieved an 87% conversion of levoglucosan to citric acid (Zhuang et al., 2001). *Schwanniomyces castelli* Labatt 1402, *Saccharomyces diastaticus* Labatt 1363, and *Candida utilis* CMI 23311 produced low yields of ethanol from partly purified levoglucosan (Prosen et al., 1993). *Rhodotorula glutinis* and *Rhodosporidium toruloides* were identified as oleaginous yeasts useful for lipid production directly from levoglucosan (Lian et al., 2013).

Previously known about levoglucosan dehydrogenase and levoglucosan consuming bacteria

Prior to the start of this work, little was known about the bacterial enzymes used to consume levoglucosan. *Arthrobacter* I-552 was isolated for its ability to grow on levoglucosan as a sole carbon source, and levoglucosan dehydrogenase (LGDH) was first isolated from the crude extracts of this organism by FPLC (Yasui et al., 1991; Nakahara et al., 1994). No LGK kinase activity was detected in *Arthrobacter* I-552, and levoglucosan dehydrogenase was isolated based on its activity in the presence of levoglucosan and NAD^+ (Nakahara et al., 1994). *Arthrobacter* I-552 was proposed t to convert levoglucosan following a three step reaction in which LGDH first uses NAD^+ to oxidize levoglucosan, producing 3-keto levoglucosan, NADH, and a proton. Then an unidentified enzyme facilitates the hydrolysis of 3-keto levoglucosan to 3-

keto glucose. Finally, LGDH uses the NADH and proton formed in the first step of the pathway to reduce 3-keto glucose to D-glucose (Figure 1-2, Nakahara et al., 1994). The partially purified LGDH from *Arthrobacter* sp. I-552 has a K_M of 14 mM for levoglucosan and a K_M of 0.47 mM for NAD (Nakahara et al., 1994).

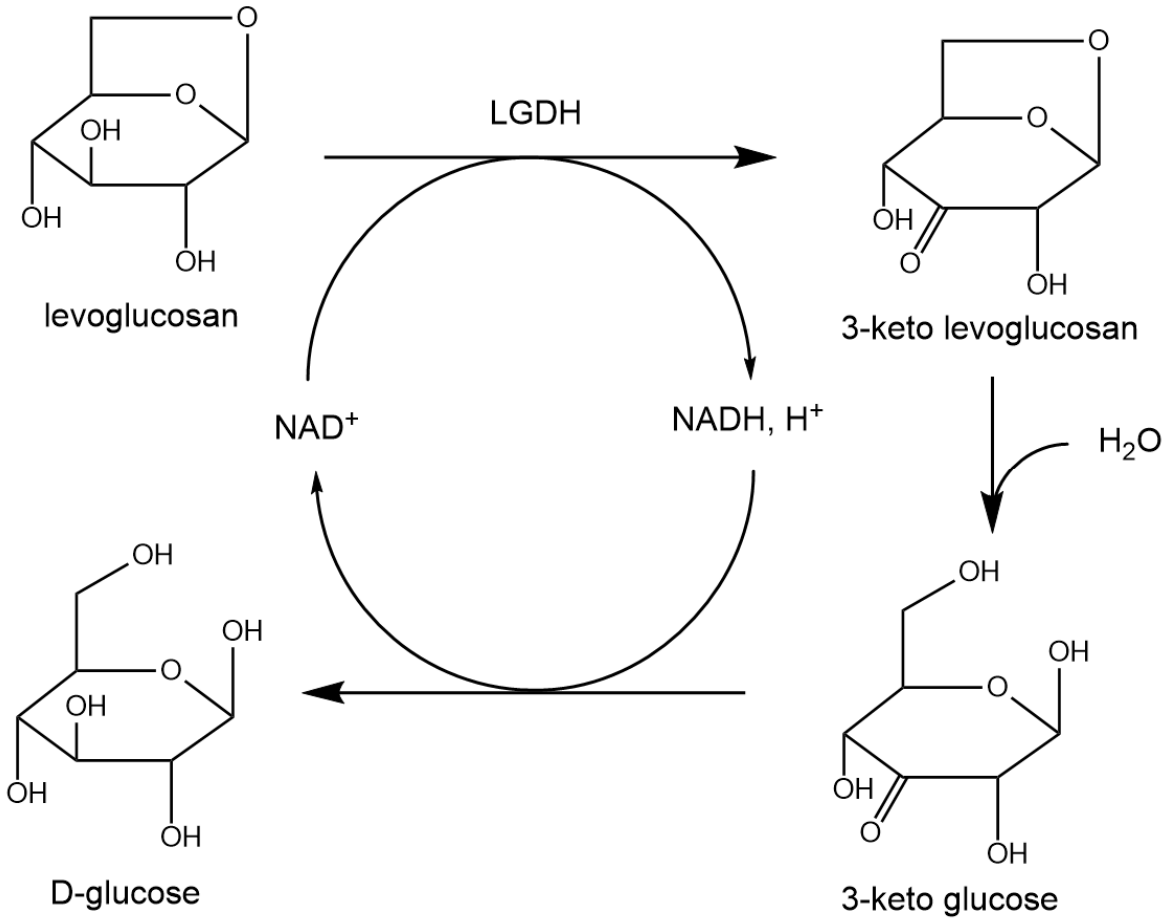


Figure 1-2. The proposed conversion of levoglucosan to glucose utilizing LGDH (Nakahara et al., 1994).

After cell lysis and the debris removal by centrifugation, LGDH activity was detected in the supernatant fraction, suggesting that LGDH resides in the cytosol and that levoglucosan enters the cell membrane by some unknown mechanism. Furthermore, since the expression of LGK (an intracellular enzyme) in *E. coli* enables this bacterium to use levoglucosan as its sole

carbon source (Zhuang and Zhang, 2002; Dai et al., 2009; Layton et al., 2011), levoglucosan likely enters cells through nonspecific channels or transporters. Since the description of LGDH in *Arthrobacter* sp. I-552, additional bacteria have been identified as being capable of growth on levoglucosan as their sole carbon source (Lin et al., 2016). Based upon their 16S rRNA gene sequences, these isolates were identified as *Sphingobacterium multivorum*, *Acinetobacter oleivorans* JC3-1, *Enterobacter* sp. SJZ-6, *Microbacterium* sp. FXJ8.207, and *Microbacterium* sp. FXJ8.203, as well as a commercially available strain *Enterobacter cloacae* DSM 16657.

For this research, a diverse set of techniques were used to probe levoglucosan consumption in bacteria. In chapter 2, environmental isolates were obtained by sampling soil and biochar where recent fires had occurred and sewage from water treatment facilities, then selecting for growth on levoglucosan as the sole carbon source. Five isolates were selected for further study and their genomic sequences were built by *de novo* assembly. The first described LGDH (Nakahara et al., 1994) was used as the basis to identify a *Pseudarthrobacter phenanthrenivorans* Sphe3 gene sequence as a likely LGDH. Recombinant expression and spectrophotometric enzyme assays confirmed the gene product to be LGDH. With the *P. phenanthrenivorans* levoglucosan dehydrogenase gene sequence (*lgdA*) known, *lgdA* was identified in the genomic sequences of our environmental isolates. Each LGDH was recombinantly expressed and purified to determine the kinetic parameters of their conversion of levoglucosan to 3-keto levoglucosan. The genomic regions surrounding each *lgdA* was also compared to identify other genes conserved, suggesting they also contribute to levoglucosan metabolism. In chapter 3, *P. phenanthrenivorans* LGDH was crystallized and successfully soaked with levoglucosan, glucose, and NADH, enabling us to observe what residues make up

the active site of the enzyme in X-ray diffraction models and propose a mechanism by which the enzyme converts its substrates.

References

J. Dai, Z. Yu, Y. He, L. Zhang, Z. Bai, Z. Dong, Y. Du, H. Zhang. 2009. Cloning of a noel levoglucosan kinase gene from *Lipomyces starkeyi* and its expression in *Escherichia coli*. *World J Microbiol Biotechnol*. 25:1589-1595.

E. Dlugokencky and P. Tans. Trends in atmospheric carbon dioxide, National Oceanic & Atmospheric Administration, Earth System Research Laboratory (NOAA/ESRL), available at: http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html, last access: 14 April 2019.

E. Kim, Y. Um, M. Bott, H. M. Woo. 2015. Engineering of *Corynebacterium glutamicum* for growth and succinate production from levoglucosan, a pyrolytic sugar substrate. *FEMS Microbiology Letters.* 362(19).

J. R. Klesmith, J. Bacik, R. Michalcyzk, T. A. Whitehead. 2015. Comprehensive sequence-flux mapping of a levoglucosan utilization pathway in *E. coli*. *ACS Synth Biol.* 4:1235-1243.

Y. Kitamura, Y. Abe, T. Yasui. 1991. Metabolism of levoglucosan (1,6-anhydro-β-D-glucopyranose) in microorganisms. *Agric Biol Chem*. 55(2):515-521.

J. Lian, M. Garcia-Perez, S. Chen. 2013. Fermentation of levoglucosan with oleaginous yeasts for lipid production. *Bioresource Technology*. 133:183-189.

J. Lian, R. McKenna, M. R. Rover, D. R. Nielson, Z. Wen, L. R. Jarboe. 2016. Production of biorenewable styrene: utilization of biomass-derived sugars and insights into toxicity. *J Ind Microbiol Biotechnol*. 43:595-604.

J. G. Linger, S. E. Hobdey, M. A. Franden, E. M. Fulk, G. T. Beckham. 2016. Conversion of levoglucosan and cellobiosan by *Pseudomonas putida* KT2440. *Metabolic Engineering Communications.* 3:24-29.

J. A. Lomax, J. M. Commandeur, P.W. Arisz, J. A. Boon. 1991. Characterization of oligomers and sugar ring-cleavage products in the pyrolysate of cellulose. *J Anal App Pyrol.* 19: 65-79.

S. Maduskar, V. Maliekkal, M. Neurock, P. J. Dauenhauer. 2018. On the yield of levoglcusaon from cellulose pyrolysis. *ACS Sustain Chem Eng*. 6(5):7017-25.

M. Miura, H. Kaga, T. Yoshida, K. Ando. 2000. Microwave pyrolysis of cellulosic materials for the production of anhydrosugars. *J Wood Sci*. 47:502-506.

M. Nakagawa, Y. Sakai, T. Yasui. 1984. Itaconic acid fermentation of levoglucosan. *J Ferment Technol*. 62(3).

P. R. Patwardhan, J. A. Satrio, R. C. Brown, B. H. Shanks. 2010. Influence of inorganic salts on the primary pyrolysis products of cellulose. *Bioresour Technol*. 101:4646-4655.

D. S. Scott, L. Patterson, J. Piskorz, D. Radlein. 2001. Pretreatment of poplar wood for fast pyrolysis: rate of cation removal. *J Anal Appl Pyrolysis*. 57(2):169-76.

F. Shafizadeh and Y. L. Yu. 1973. Pyrolysis of cellulose. *Carbohydrate Res.* 29:113-122.

F. Shafizadeh, R. H. Furneaux, T. G. Cochran, J. P. Scholland, Y. Sakai. 1979. Production of levoglucosan and glucose from pyrolysis of cellulosic materials. *J Appl Polym Sci.* 23:3525.

N. Sorek, T. H. Yeats, H. Szemenyei, H. Youngs, C. R. Somerville. 2014. The implications of lignocellulosic biomass chemical composition for the production of advanced biofuels. *Bioscience*. 64(3):192-201.

X. Xiong, J. Lian, X. Yu, M. Garcia-Perez, S. Chen. 2016. Engineering levoglucosan metabolic pathway in *Rhodococcus jostii* RHA1 for lipid production. *J Ind Microbiol Biotechnol*. 43:1551-1560.

J. Zhang, Y. S. Choi, C. G. Yoo, T. H. Kim, R. C. Brown, B. H. Shanks. 2014. Cellulose-hemicellulose and cellulose-lignin interactions during fast pyrolysis. *ACS Sustainable Chem Eng*. 3:292-301.

X. L. Zhuang, H. X. Zhang, J. Z. Yang, H. Y. Qi. 2001. Preparation of levoglucosan by pyrolysis of cellulose and its citric acid fermentation. *Bioresour Technol.* 79:63-66.

CHAPTER 2

Isolation and Characterization of Bacteria which use Levoglucosan as the Sole Carbon Source[1]

A. S. Arya, M. T. H. Hang, M. A. Eiteman, W. B. Whitman. To be submitted to *Appl Environ Microbiol*

Abstract

Bacteria were isolated from wastewater and soil samples containing charred wood remnants based on their ability to use levoglucosan as a sole carbon source and on their activity of levoglucosan dehydrogenase (LGDH). A comparative 16S rRNA analysis showed these bacteria to be members of the genera *Microbacterium, Paenibacillus*, *Shinella*, and *Klebsiella*. Genomic sequencing of these isolates and determining the Average Nucleotide Identity (ANI) in comparison to other bacteria of the same genus verified that two isolates were of novel species, *Paenibacillus athensensis* MEC069^T and *Shinella sumterensis* MEC087^T, while the remaining isolates were *Microbacterium marinilacus* MEC084 and *Klebsiella pneumoniae* MEC097. The genetic sequence of LGDH, *lgdA*, was found in the genomes of these four isolates as well as *Pseudarthrobacter phenanthrenivorans* Sphe3, with the activity of the *P. phenanthrenivorans* LGDH experimentally verified through recombinant expression in *E. coli*. Comparison of the putative genes surrounding *lgdA* in the isolate genomes suggests several other genes may facilitate the bacterial catabolism of levoglucosan, including a putative sugar isomerase and several transport proteins.

Introduction

During high temperature, fast pyrolysis of lignocellulosic biomass, complex organic molecules are dehydrated, and substantial water is generated as a by-product. Specifically, at temperatures greater than 300°C cellulose is readily depolymerized and dehydrated to a mixture of anhydrosugars and their polysaccharides, the most abundant being levoglucosan (1,6-anhydro-β-D-glycopyranose, Shafizadeh, 1973) and the disaccharide of glucose and levoglucosan, cellobiosan (Lomax et al., 1991). Although lignocellulose generates less levoglucosan than pure cellulose, impregnating woody lignocellulose with a 1% w/v phosphoric acid solution leads to a levoglucosan content as high as 33.6% w/w (Dobele et al., 2005) partitioned primarily into an aqueous phase (Bennett and Duff, 2005; Vitasari et al., 2011).

Levoglucosan is a potentially fermentable sugar. However, it cannot naturally be metabolized by typical production microorganisms (*e.g., Saccharomyces cerevisiae*, *Escherichia coli*). Chemically, levoglucosan can be hydrolyzed using sulfuric acid at 110°C to readily fermentable glucose (Bennett and Duff, 2005, Bennett et al., 2009). Those microorganisms that can metabolize levoglucosan use two different strategies. As a rule, fungi able to metabolize levoglucosan use levoglucosan kinase (LGK) to hydrate and phosphorylate the anhydro-sugar simultaneously to glucose-6-phosphate (Kitamura and Yasui, 1991; Prosen et al., 1993; Xie et al., 2006). LGK is specific for levoglucosan, and common hexoses such as glucose, mannose and galactose are neither phosphorylated nor do they inhibit the enzyme (Kitamura and Yasui, 1991). LGKs generally have poor affinity for levoglucosan with a K_M of about 50-110 mM, while the K_M for ATP is 0.2 mM (Kitamura and Yasui, 1991; Xie et al., 2006). ADP, but not glucose-6-phosphate, is strongly inhibitory, having a K_I of 0.15 mM (Kitamura and Yasui, 1991). Genes encoding LGK from *Aspergillus niger* (Zhuang and Zhang, 2002) and *Lipomyces starkeyi* (Dai et

al., 2009) have been expressed in *E. coli*. Recently 15 mutations in the LGK of *L. starkeyi* have resulted in a 15-fold increase in growth rate on levoglucosan and an over 24-fold improvement in enzyme activity in an *E. coli* host (Klesmith et al., 2015).

In contrast to fungal metabolism, very little is known about prokaryotic utilization of levoglucosan. One bacterium, *Arthrobacter* sp. I-552, has been described that metabolizes levoglucosan (Yasui et al., 1991; Nakahara et al., 1994). This bacterium does not express an LGK to phosphorylate levoglucosan, but instead is proposed to follow a three-step conversion to glucose. Levoglucosan is first oxidized via a NAD-dependent dehydrogenase (LGDH) to 3-keto levoglucosan. An enzymatic hydration is proposed to form 3-keto glucose, which is then reduced to glucose by LGDH (Nakahara et al., 1994). The partially purified LGDH from *Arthobacter* sp. I-552 has a K_M of 14 mM for levoglucosan and a K_M of 0.47 mM for NAD (Nakahara et al., 1994). After cell lysis and the debris removal by centrifugation, LGDH activity was detected in the supernatant fraction, suggesting that LGDH resides in the cytosol and that levoglucosan enters the cell membrane by some unknown mechanism. Furthermore, the expression of LGK (an intracellular enzyme) in *E. coli*, enables *E. coli* to use levoglucosan as its sole carbon source (Zhuang and Zhang, 2002; Dai et al., 2009; Layton et al., 2011), suggesting that levoglucosan enters cells through nonspecific channels or transporters. Since the description of LGDH in *Arthrobacter* sp. I-552, additional bacteria have been identified as being capable of growth on levoglucosan as their sole carbon source (Lin et al., 2016). Based upon their 16S rRNA gene sequences, these isolates were identified as *Sphingobacterium multivorum, Acinetobacter oleivorans* JC3-1, *Enterobacter* sp. SJZ-6, *Microbacterium* sp. FXJ8.207, and *Microbacterium* sp. FXJ8.203, as well as strain *Enterobacter cloacae* DSM 16657.

The goal of this study was to isolate and characterize additional bacteria that use levoglucosan as the sole carbon source. Furthermore, using the genomic sequence of each isolate, this study aims to identify shared genes which likely play a role in levoglucosan catabolism. Levoglucosan has been previously reported to be a major component in the fine particulate emissions from the combustion of wood in residential fireplaces (Schauer et al., 2001) and prescribed forest fires (Lee et al., 2005). We therefore reasoned that levoglucosan-degrading bacteria might be found at sites where lignocellulosic materials were burned and sought bacteria from soil samples near recent campfires and forest fires.

Materials and Methods

Liquid Medium

Defined Levoglucosan Isolation Medium (LIM) consisted of (per L): 2.00 g levoglucosan, 1.50 g $KH_2PO_4 \cdot 7H_2O$, 3.10 g $Na_2HPO_4 \cdot 7H_2O$, 0.50 g NH_4Cl, 0.50 g $(NH_4)_2SO_4$, 47 mg $CaCl_2 \cdot 2H_2O$, 20 mg $FeSO_4 \cdot 7H_2O$, 90 mg $MgCl_2 \cdot 6H_2O$, 7.5 mg nitriloacetic acid, 15 mg $MgSO_4 \cdot 7H_2O$, 2.5 mg $MnSO_4 \cdot H_2O$, 5.0 mg NaCl, 0.41 mg $CoCl_2 \cdot 2H_2O$, 0.50 mg $ZnSO_4 \cdot H_2O$, 0.46 mg $NiCl_2 \cdot 6H_2O$, 50 µg $CuSO_4 \cdot 5H_2O$, 50 µg H_3BO_3, 50 µg $Na_2MoO_4 \cdot 2H_2O$, 50 µg $Na_2WO_4 \cdot 2H_2O$, 40 µg $Al_2(SO_4)_3$, 5 µg Na_2SeO_3, 120 µg pyridoxine·HCl, 60 µg thioctic acid, 60 µg p-aminobenzoic acid, 24 µg folic acid, 60 µg calcium pantothenate, 60 µg nicotinic acid, 60 µg thiamine·HCl, 24 µg biotin, 6 µg riboflavin, and 1.2 µg cyanocobalamin. Agar plates contained LIM medium with 18 g/L agar.

Isolation

Samples were taken from several sources: silty topsoil and charred wood from campfire pit (Nantahala National Forest, NC, USA; GPS location: 35.01050, -83.25258), silty topsoil heavy in organic matter from campfire (Russell Farmstead, Mountain Rest, SC, USA; GPS location:

34.91107, -83.17371), raw wastewater from a chemical company (Athens, GA, USA; GPS location: 33.98023, -83.32673), and raw wastewater (Middle Oconee Water Reclamation Center, Athens, GA, USA; GPS location: 33.90955, -83.39059). For each sample in duplicate, 0.3 g or 0.3 mL were placed in 3 mL LIM in sterile screw-top tubes which were continuously shaken at 37°C. Twice after consecutive 24 h periods, 0.10 mL from each tube was transferred into 3 mL fresh LIM and similarly shaken. After 24 h growth in the third tube, samples were streaked onto LIM-agar plates, and individual colonies were isolated.

Analyses

The optical density (OD) measured at 600 nm (DU-650 spectrophotometer, Beckman Instruments, San Jose, CA) was used to monitor cell growth. High performance liquid chromatography using refractive index detection and a Coregel 64-H ion-exclusion column (Transgenomic Ltd., Glasgow, United Kingdom) with a mobile phase of 4 mN (2 mM) H_2SO_4 was used for analysis of organic chemicals (Eiteman and Chastain, 1997).

Growth rates were measured in 250 mL baffled shake flasks shaken at 250 rpm (19 mm pitch) containing 50 mL of LIM. About 6-8 samples during exponential growth were used to calculate the growth rate from a plot of the logarithm of OD versus time.

Levoglucosan kinase activity was measured by a coupled assay with pyruvate kinase and lactate dehydrogenase (Kitamura et al., 1991). The assay mixture initially contained 0.1 M levoglucosan, 50 mM triethanolamine·HCl, 0.1 M KCl, 1.12 U/mL pyruvate kinase (Sigma-Aldrich Co., St. Louis, MO), 1.6 U/mL lactate dehydrogenase (Sigma-Aldrich Co., St. Louis, MO), 2 mM ATP, 2 mM PEP, 10 mM $MgCl_2$, and 160 µM NADH. A codon optimized levoglucosan kinase gene from *L. starkeyi* YZ-215 was expressed in *E. coli* XL1-Blue cells and used as a positive control (Layton et al., 2011). Levoglucosan dehydrogenase activity was measured by the direct

formation of NADH (Nakahara et al., 1994), and the assay mixture contained 0.1 M Tris, 2 mM sodium NAD, and 0.1 M levoglucosan, pH 9.0. For either enzyme, one unit (IU) of activity was defined as the amount of enzyme required to produce 1 μmol of NADH per min. Activity and specific enzyme activity (IU per dry cell weight) were measured by growing a 50 mL culture in LIM in a 250 mL baffled shake flask to an OD of 1.0. A portion of the culture was used to prepare a crude cell extract for activity measurement, and a portion was centrifuged, the pellet washed and resuspended twice before drying in an oven at 60°C for 24 h for fry cell weight.

Illumina Sequencing and Genome Assembly

Overnight cultures of each isolate grown in LIM were used to isolate genomic DNA as detailed in the instructions of the Promega Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI). Genomic DNA in EB buffer was then used to create NGS DNA libraries (prepared by the Georgia Genomics Facility, Athens, GA) which were then sequenced using paired-end 2X150 base pair reads on a HiSeq 2500 sequencer (Genomic Services Lab at HudsonAlpha Institute for Biotechnology, Huntsville, AL).

Low quality reads were removed from the Illumina sequencing data using Bowtie 2 (Langmead et al., 2012). Genomes were then assembled using the A5-miseq pipeline (Coil et al., 2015). The mol % G + C of each organism was determined from the final genome assemblies. 16S rRNA gene sequences were identified in the genome assemblies and used for a phylogenetic analysis of each isolate. Average nucleotide identity between isolate genomes and other species was determined with the ANI Calculator (www.ezbiocloud.net, Yoon et al., 2017). All draft genomes were submitted to GenBank and annotated with the NCBI Prokaryotic Genome Annotation Pipeline (Tatusova et al., 2013).

Phenotypic and Biochemical Comparison

Based on the results of the genomic analysis, reference strains were selected for phenotypic comparison to our isolates. Type strains of related genera were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ) for comparisons. These include *Paenibacillus polymyxa* (DSM 36, ATCC 842), *Microbacterium lacticum* (DSM 20427, ATCC 8180), *Klebsiella pneumoniae* subsp. *pneumoniae* (DSM 30104, ATCC 13883), and *Shinella zoogloeoides* (DSM 287, ATCC 19623).

Isolates and acquired reference strains were introduced to a variety of carbohydrates to determine which substances could be used as sole energy sources. The ability of each organism to metabolize a single substrate as the carbon source was tested with the following compounds: D-arabinose, cellobiosan, D-fructose, D-galactose, D-glucose, myo-inositol, lactose, levoglucosan, D-mannitol, mannose, D-sorbitol, sucrose, and D-xylose. All substrate test solutions (3 mL in sterilized test tubes) were identical to LIM except that levoglucosan was replaced by the target substrate at 20 mM, and 4 mg/L bromocresol purple was added to detect acid production. Each organism was grown at 37°C for 48 h. The following conventional growth media were also used during this study: SIM stab for the determination of H_2S production, indole production, and motility; Lysogeny Broth (LB) agar in GasPAK EZ Pouch System (Becton, Dickinson and Company, New Jersey, United States) to test for anaerobic growth. The catalase test was performed by growing each isolate overnight on LIM agar at 37°C, then streaking one colony onto a clean microscope slide and adding several drops of 3% H_2O_2 and watching for vigorous bubbling; the oxidase test was performed in a similar manner except cells were collected on a sterile cotton swab then a few drops of freshly prepared Kovac's Oxidase solution (1% tetra-methyl-*p*-phenylenediamine dihydrochloride in deionized water) were added directly to the swab, and cells were observed for a change in color to purple-black.

To examine growth on cellobiosan as a sole carbon source, *Klebsiella pneumoniae* subsp. *pneumoniae* (DSM 30104, ATCC 13883) and one of the isolates (MEC097, DSM 27296) were each grown from a single colony overnight in 5.0 mL LIM containing 10 mM glucose instead of levoglucosan. These overnight cultures were used to inoculate 50 mL LIM containing 10 mM cellobiosan in 250 mL shake flasks in triplicate. Samples for HPLC and OD analyses were examined each hour until no increase in OD was observed.

Identification of lgdA in Genome Assemblies and Comparison of Neighboring Genes

Previously published partial peptide sequences of LGDH (Nakahara et al., 1994) were used in a tBLASTn search of GenBank. The closest result, gene locus Asphe10730 of assembly GCA_000189535.1 of *Pseudarthrobacter phenanthrenivorans* Sphe3, was cloned into pET-28b and recombinantly expressed in *E. coli* BL21 (DE3)-RIPL cells (Agilent Technologies, Santa Clara, CA). Crude cell extracts were assayed for LGDH activity as previously described. Genomes from levoglucosan isolates were then searched by BLAST using the experimentally identified *lgdA* gene. Genes common in all genomes near *lgdA* were aligned to compare their relative identities.

Recombinant Expression of LGDH, Purification, and Spectrophotometric Characterization of Enzymes

A modified version of the pET28b vector (pET28b(+)-CHSapKO-CH.BspQI) was used for the expression of *lgdA* under the control of a T7-promoter (Galloway et al., 2013). Briefly, the modifications within pET28b-CHSap include a SapI site at position 3108 knocked out, double overlapping SapI sites positioned between the start codon and the double stop codon, and a polyhistidine tag at the 3' location relative to the reading frame for the production of C-terminally histagged protein. Identified *lgdA* were PCR amplified from genomic DNA using Phusion

polymerase (New England Biolabs Inc., Ipswich, MA) and the primers listed in Table 2-1. PCR fragments were purified on DNA Clean and Concentrate columns (Zymo Research Co., Irvine, CA) and eluted using nuclease-free water. 50 ng of pET28b(+)-CHSapKO-CH.BspQI vector (Galloway et al., 2013) and 27 ng of PCR product (1:2 molar ratio) were digested together with 1 unit of BspQI (New England Biolabs Inc., Ipswich, MA) and subsequently ligated with Fast-Link^{TM} T_4 DNA ligase (Epicentre, Madison, WI) as previously described (Galloway et al., 2013). After ligation, 1 µL of the ligation reaction was transformed into 40 µL *E. coli* strain XL1-Blue (Stratagene, San Diego, CA) by electroporation, and cells were grown on LB agar containing kanamycin. Colonies were screened for the presence of pET28b-CHSap containing *lgdA* by restriction analysis and PCR verification. DNA sequencing was used to verify the sequence of *lgdA* in each construct (pET28b-Ch*lgdA*).

Table 2-1. Primers used for the recombinant expression of levoglucosan dehydrogenase (*lgdA*) from isolates in pET28b-SapKO vector with a C-terminal histag.

Organism	Primer name	Sequence	T_M
P. phenanthrenivorans Sphe3	MEC189-C-term-F	GG*GCTCTTC*AATGATGCAGAACCTCAACGTCG	67
	MEC189-C-term-R	GG*GCTCTTC*AGTGGGCGGAGATCTGCGGAAC	
P. athensensis MEC069	MEC069-C-term-F	GG*GCTCTTC*GATGGGAAAGGCGCATTC	62
	MEC069-C-term-R	GG*GCTCTTC*AGTGGATCGTTTTCACCCATTGC	
P. athensensis MEC069	MEC069-SOE-F-GALG	GGAAGGCGCCCTCGGCGCTATTCTTAGCTTCC	60
	MEC069-SOE-R-GALG	AATAGCGCCGAGGGCGCCTTCCTCAATATACTTTTTCGCC	
M. marinilacus MEC084	MEC084-C-term-F	GG*GCTCTTC*CATGCAGGACCTCAACATCG	63
	MEC084-C-term-R	GC*GCTCTTC*AGTGGCCATCGGAGGAGCTCTC	
S. sumterensis MEC087	MEC087-C-term-F	GG*GCTCTTC*GATGACCAAAGTGATGAACGTC	60
	MEC087-C-term-R	GG*GCTCTTC*CGTGTCCCACCTTCTCCCACAG	
K. pneumoniae MEC097	MEC097-C-term-F	GG*GCTCTTC*GATGAAAACACTGAATGTAGGTATG	58
	MEC097-C-term-R	CC*GCTCTTC*CGTGGATTTCGTCAACCGAAACC	

BspQI recognition sequences are italicized. Note that in the middle of the *P. athensensis lgdA* the DNA sequence coding for the residues 185-188 (GALG) make a strong hairpin structure which appears to inhibit protein production in BL21 cells. MEC069 *lgdA* was first cloned using primer pair MEC069-C-term-F and MEC069-C-term-R, then digested and ligated to the pET28b(+)CH-SapKO.BspQI vector. Primers MEC069-SOE-F-GALG and MEC069-SOE-R-GALG are used to introduce silent mutations (underlined nucleotides) that should make hairpin formation less favorable in *E. coli* by amplifying the pET-lgdA vector from the

point of mutations creating one double stranded piece of DNA, which is then recircularized using NEBuilder HiFi Master Mix (New England Biolabs Inc., Ipswich, MA).

To generate recombinant LGDH, 40 µL aliquots of *E. coli* strain BL21 Codon Plus (DE3)-RIPL (Stratagene, San Diego, CA) were transformed with each pET28b-Ch*lgdA* by electroporation and subsequently plated on LB + Kan agar. After growing overnight, 5-10 colonies were inoculated into 2.5 mL Superbroth (35 g/L tryptone, 20 g/L yeast extract, 5 g/L NaCl, a 5 mL 1 N NaOH/L)with 100 µg/mL kanamycin in a test tube, and then the culture incubated at 37°C, 300 rpm, for 4-6 hours. A volume of 5 mL was used to inoculate 2 liter shake flasks containing 500 mL PA-5052 autoinduction medium (Studier, 2005) with 100 µg/mL kanamycin, which were incubated at room temperature (22-24°C) and 300 rpm for 18-20 h. Cells were harvested by centrifugation (11,000 × g for 20 min at 4°C) and resuspended in ~30 mL of 50 mM Tris, 25 mM imidazole, 500 mM NaCl, pH 8.0. Resuspended cells were lysed with a French Press, and the cell-free extract centrifuged (60,000 × g for 20 min). The supernatant was purified with a Pharmacia ÄKTA Purifier HPLC system at room temperature. The cell extract was first loaded (2 mL/min) on a 5 mL Histrap HP column (GE Life Sciences, Marlborough, MA) equilibrated with 50 mM Tris, 25 mM imidazole, 500 mM NaCl, pH 8.0. After washing unbound protein from the column, the histagged LGDH was eluted with a gradient of imidazole from 25 mM to 500 mM over 15 column volumes. The protein eluted as one large peak, which was collected and dialyzed first into 1 L of 50 mM Tris buffer (pH 8.0), 250 mM NaCl, 5 mM β-mercaptoethanol (BME), then into 2 L 50 mM Tris buffer (pH 8.0), 5 mM BME. After dialyzing for 4 hours in each solution, LGDH was further purified by anion-exchange chromatography using a 5 mL HiTrap Q column (GE Life Sciences, Marlborough, MA) equilibrated with 50 mM Tris, pH 8.0. Protein was eluted by using a linear gradient of 0-1.0 M NaCl over 15 column volumes. LGDH eluted as one large, sharp peak which was dialyzed first in 1 L 50 mM Tris (pH 8.0), 250 mM NaCl, 5 mM BME, then in 2 L 50

mM Tris (pH 8.0), 5 mM BME. Fractions from both purifications were stored at -20°C until SDS-PAGE analysis of protein purity.

Purified LGDH was assayed for activity using a spectrophotometric assay (Nakahara et al., 1994) with 0.1 M Tris (pH 9.0 at room temperature), 0.1-10.0 mM NAD^+, and 1-300 mM levoglucosan. Each reaction mixture was incubated at 30°C in a Peltier heater for 2 min, then the reaction initiated by the addition of 100 µL of 0.05-0.10 mg LGDH/mL which had separately been incubated 30°C. Absorbance at 340 nm was recorded every 4 seconds for 2 min for the calculation of the reaction rate. Kinetic parameters were determined with an R script utilizing the minpack.lm plug-in for the Levenberg-Marquardt nonlinear least-squares algorithm (Elzhov et al., 2016).

Results

Isolation

Upon enrichment at 37°C, samples from several soils associated with wood fires as well as water treatment facilities yielded five bacterial strains able to grow on levoglucosan as the sole carbon source (Table 2-2). Isolate MEC084 came from a mixture that was primarily charred wood with silty topsoil in a campfire pit (Nantahala National Forest, NC, USA). Isolates MEC087 and MEC089 were also from campfire sediment, however this sample had browner silty topsoil and was heavy in organic matter (Russell Farmstead, Mountain Rest, SC, USA). The remaining two isolates came from wastewater: MEC069 was from the wastewater leaving a chemical plant (Athens, GA, USA) while MEC097 was isolated from the raw wastewater entering a wastewater treatment facility (Middle Oconee Wastewater Treatment Facility, Athens, GA, USA).

Table 2-2. Bacterial isolates which are able to metabolize levoglucosan as the sole carbon source at 37°C.

Isolate Designation	Gram Stain	Cell Morphology	Colonial Morphology	Source	Closest 16S rRNA BLAST match[a]
MEC069	+	bacilli, alone or in chains of 2-3	circular, raised, entire, smooth, dull, opaque, orange, small	North Oconee Water Reclamation Facility, Athens, GA, USA	*Paenibacillus*
MEC084	+	cocco-bacilli, mostly in chains or clumps	circular, flat, entire, smooth, opaque, white-cream, small	Nantahala National Forest, NC, USA	*Microbacterium lacusdiani*
MEC087	-	small cocci in chains and clumps	circular, raised, entire, smooth, glistening, opaque, white, pinpoint	Russell Farmstead, SC, USA	*Shinella zoogloeoides*
MEC089	-	small cocci in chains	circular, raised, entire, smooth, glistening, opaque, white, pinpoint	Russell Farmstead, SC, USA	*Shinella zoogloeoides*
MEC097	-	bacilli, often in chains of 2-3	circular, umbonate, entire, smooth, mucoid glistening, opaque, medium, white	Raw Wastewater, Water Reclamation Center, Athens, GA, USA	*Klebsiella pneumoniae*

[a] If 99% identity was obtained, then only the closest species matched to the 16S rRNA sequence is shown. Otherwise, the genus with the greatest identity is shown

Initial characterization

The maximum specific growth rates on levoglucosan and levoglucosan-utilizing enzyme activities of the five isolates were determined (Table 3). The growth rates on levoglucosan can be categorized into two groups: MEC097 had a high growth rate on levoglucosan (~ 0.8 h^{-1}, doubling time 52 minutes) while the remaining isolates achieved maximum specific growth rates of about 0.4 h^{-1} (doubling time ~ 104 minutes). All isolates showed levoglucosan dehydrogenase (LGDH) activity, while no levoglucosan kinase activity was detected. The greatest specific LGDH activity of 0.32 IU/g dry cell weight was measured in the isolate showing the greatest specific growth rate (MEC097), while the remaining activities ranged between 0.024 and 0.12 IU/g dry cell weight.

Table 2-3. Maximum specific growth rates on levoglucosan and levoglucosan dehydrogenase activity for five bacterial isolates.

Isolate	Maximum Specific Growth Rate [a] (h^{-1})	Specific Levoglucosan Dehydrogenase Activity [a] (IU/g dry cell wt)
MEC069	0.47 (0.11)	0.057 (0.028)
MEC084	0.43 (0.02)	0.026 (0.006)
MEC087	0.40 (0.04)	0.090 (0.042)
MEC089	0.38 (0.08)	0.12 (0.05)
MEC097	0.80 (0.10)	0.32 (0.16)

[a] Mean of 3-4 independent samples. Standard deviation of measurements is shown in parentheses.

Taxonomic identification of isolates

In order to understand how widespread levoglucosan catabolism is in the bacterial domain, the genomic DNA from each isolate was Illumina sequenced (assembly statistics, GenBank accession numbers presented in Table 2-4). 16S rRNA sequences were used in a BLAST search of the GenBank database. MEC069 was closely related to *Paenibacillus marchantiophytorum* R55^T (KP056549.1) and *Paenibacillus yunnanensis* YN2^T (KJ914577.1) with 95.72% and 94.11% sequence similarity, respectively. MEC084 was closely related to *Microbacterium lacusdiani* (NR_149217.2) and *Microbacterium marinilacus*^{T} (AB286020.1) with 99.86% and 98.90% sequence similarity, respectively. MEC087 and MEC089 share 100% sequence identity and are closely related to *Shinella zoogloeoides* strain PD7-5 (AB506123.1) and *Shinella* sp. HZN7 (CP015736.1) with 99.86% and 98.45% sequence similarity, respectively. MEC097 had 99.36% sequence similarity with *Klebsiella pneumoniae* ATCC 13883 (X87276.1). Phylogenetic trees based on the 16S rRNA gene sequences and constructed by the maximum-likelihood method show the relationship between MEC069 and the *Paenibacillus* (Fig. 2-1), MEC084 and the *Microbacterium* (Fig. 2-2), and MEC087/MEC089 and the *Shinella* (Fig. 2-3).

Table 2-4. Genome assembly statistics

Isolate	GenBank Accession Number	Number of Contigs	Coverage	N50	Total Sequence Length	PGAP Predicted Coding Sequences
Paenibacillus MEC069	MYFO00000000	125	19	120,355	5,618,969	4,841
Microbacterium MEC084	MYFP00000000	121	42	51,425	3,045,012	2,798
Shinella MEC087	MYFQ00000000	140	55	170,660	5,365,482	5,160
Shinella MEC089	MYFN00000000	194	41	165,155	5,355,183	5,166
Klebsiella pneumoniae MEC097	MWQX00000000	183	28	122,944	5,513,061	5,298

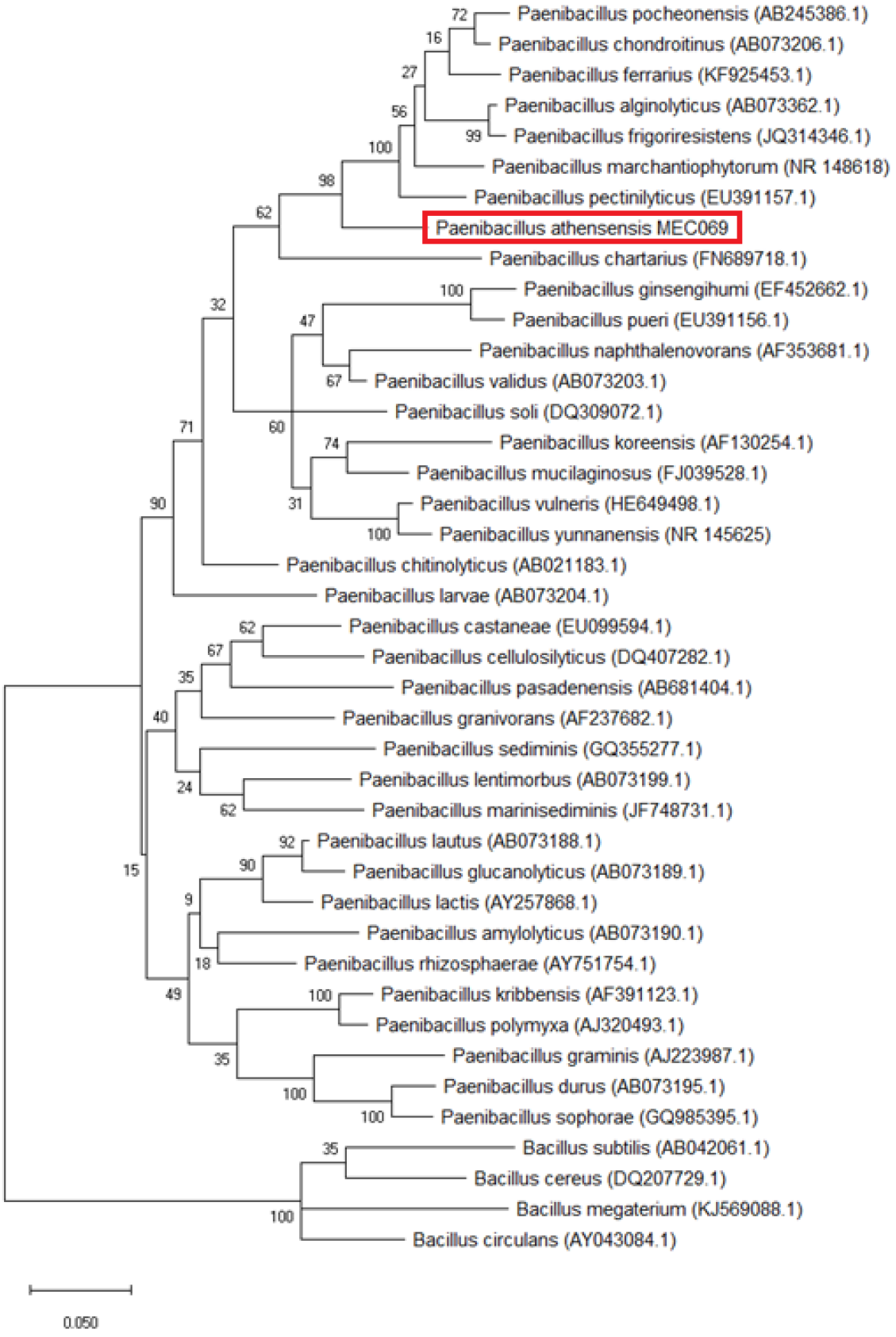

Paenibacillus pocheonensis (AB245386.1)
Paenibacillus chondroitinus (AB073206.1)
Paenibacillus ferrarius (KF925453.1)
Paenibacillus alginolyticus (AB073362.1)
Paenibacillus frigoriresistens (JQ314346.1)
Paenibacillus marchantiophytorum (NR 148618)
Paenibacillus pectinilyticus (EU391157.1)
Paenibacillus athensensis MEC069
Paenibacillus chartarius (FN689718.1)
Paenibacillus ginsengihumi (EF452662.1)
Paenibacillus pueri (EU391156.1)
Paenibacillus naphthalenovorans (AF353681.1)
Paenibacillus validus (AB073203.1)
Paenibacillus soli (DQ309072.1)
Paenibacillus koreensis (AF130254.1)
Paenibacillus mucilaginosus (FJ039528.1)
Paenibacillus vulneris (HE649498.1)
Paenibacillus yunnanensis (NR 145625)
Paenibacillus chitinolyticus (AB021183.1)
Paenibacillus larvae (AB073204.1)
Paenibacillus castaneae (EU099594.1)
Paenibacillus cellulosilyticus (DQ407282.1)
Paenibacillus pasadenensis (AB681404.1)
Paenibacillus granivorans (AF237682.1)
Paenibacillus sediminis (GQ355277.1)
Paenibacillus lentimorbus (AB073199.1)
Paenibacillus marinisediminis (JF748731.1)
Paenibacillus lautus (AB073188.1)
Paenibacillus glucanolyticus (AB073189.1)
Paenibacillus lactis (AY257868.1)
Paenibacillus amylolyticus (AB073190.1)
Paenibacillus rhizosphaerae (AY751754.1)
Paenibacillus kribbensis (AF391123.1)
Paenibacillus polymyxa (AJ320493.1)
Paenibacillus graminis (AJ223987.1)
Paenibacillus durus (AB073195.1)
Paenibacillus sophorae (GQ985395.1)
Bacillus subtilis (AB042061.1)
Bacillus cereus (DQ207729.1)
Bacillus megaterium (KJ569088.1)
Bacillus circulans (AY043084.1)
0.050

Figure 2-1. Phylogenetic tree of *Paenibacillus* 16S rRNA gene sequences constructed using the maximum-likelihood method. Members of the *Bacillus* were used as an outgroup. *Paenibacillus* presented were chosen from a tree of all *Paenibacillus* 16S rRNA to represent major clades within the genus as well as show the relationship between *P. athensensis* MEC069 (red box) and organisms for which the genomic sequence is available. The tree with the highest log likelihood (-11446.96) is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 41 nucleotide sequences. There were a total of 1645 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

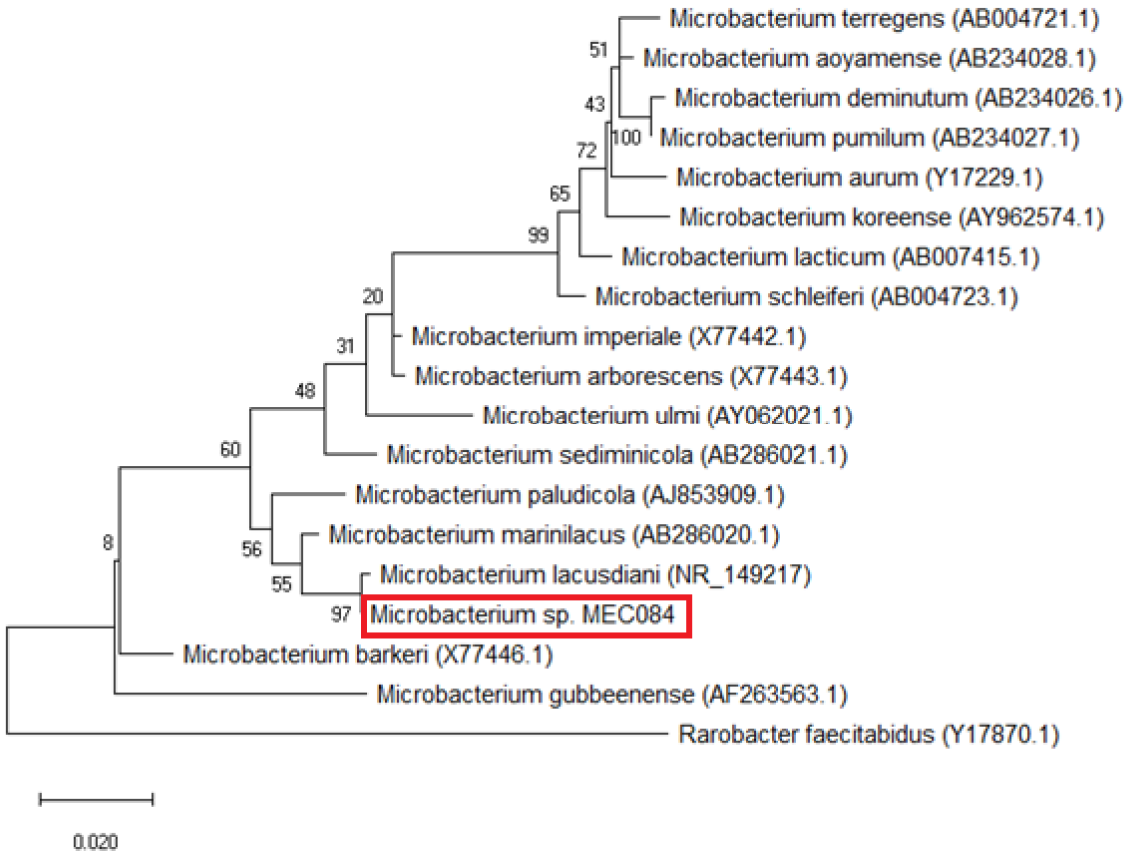


Figure 2-2. Phylogenetic tree of *Microbacterium* 16S rRNA gene sequences constructed using the maximum-likelihood method. *Rarobacter faecitabidus* DSM 4813^{T} was used as an outgroup. *Microbacterium* presented were chosen from a tree of all *Microbacterium* 16S rRNA to represent major clades within the genus as well as show the relationship between *M. lacusdiani* MEC084 (red box) and organisms for which the genomic sequence is available. The tree with the highest log likelihood (-4371.19) is drawn to scale, with branch lengths measured in the number of substitutions per site and the percentage of trees in which the associated taxa clustered together is shown next to the branches. The analysis involved 19 nucleotide sequences. There were a total of 1546 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

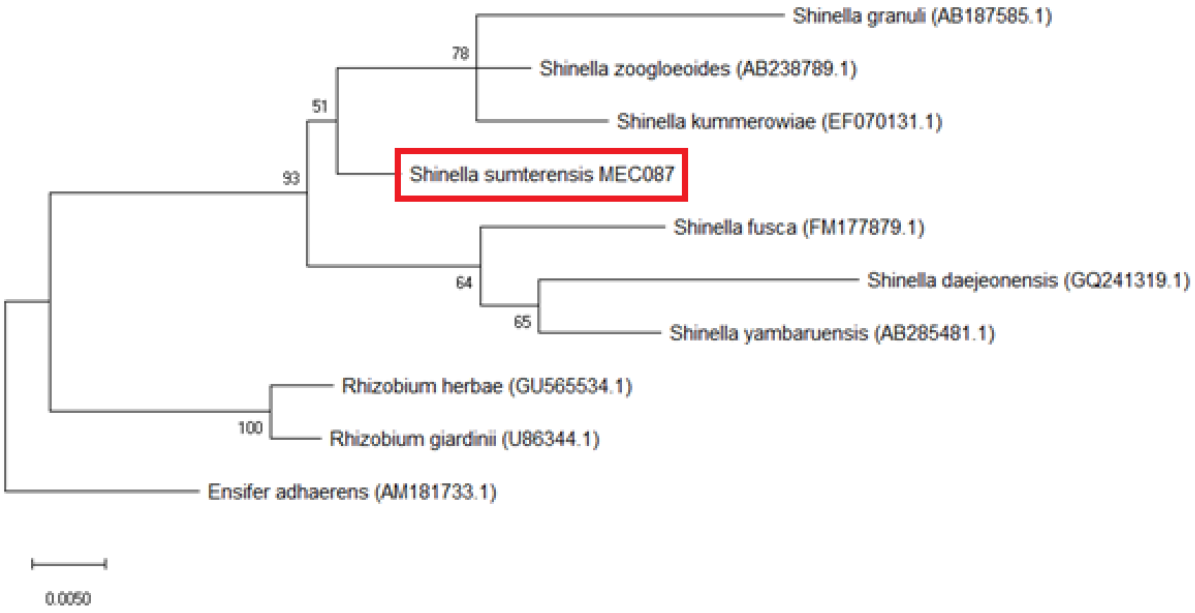


Figure 2-3. Phylogenetic tree of *Shinella* 16S rRNA gene sequences constructed using the maximum-likelihood method. *Ensifer adhaerens* LMG 20216^{T}, *Rhizobium giardinii, Rhizobium herbae* were used as an outgroup. All 16S rRNA genes available from officially named *Shinella* were used to show relatedness to *S. sumterensis* MEC087 (red box). The tree with the highest log likelihood (-3279.84) is drawn to scale, with branch lengths measured in the number of substitutions per site and the percentage of trees in which the associated taxa clustered together is shown next to the branches. The analysis involved 10 nucleotide sequences. There were a total of 1530 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

The ANI between levoglucosan-utilizing isolates and members of the same genus were calculated when genomes were available. For MEC069 notable ANI were observed with *Paenibacillus ginsenghumi*, *Paenibacillus pasadenensis*, *P. pectinolyticus*, and *P. alginolyticus* with values of 71.74%, 71.53%, 71.35%, and 70.18%, respectively (Table 2-5). MEC084 had notable ANI with *Microbacterium barkeri*, *Microbacterium indicum*, *Microbacterium paludicola* with identities of 77.14%, 77.09%, and 75.83%, respectively (Table 2-6). MEC087 and MEC089 were compared with *S. zoogloeoides* and *Shinella* sp. HZN7, having ANI of 83.53% and 84.24%, respectively (Table 2-7). When MEC097 was compared with *Klebsiella pneumoniae* ATCC 13883^{T}, the ANI was 99.05%.

Table 2-5. ANI of *Paenibacillus athensensis* MEC069 and Related Bacteria

Bacterium	1	2	3	4	5
1: *P. athensensis MEC069*	1				
2: *P. alginolyticus* DSM 5050 NZ_AUGY01000001.1	0.7018	1			
3: *P. pectinolyticus* KCTC 13222 GCA_001700435.1	0.7135	0.7729	1		
4: *P. larvae* subsp. *larvae* DSM 25430 NC_023134.1	0.6899	0.6687	0.6686	1	
5: *P. mucilaginosus* 3016 NC_016935.1	0.7106	0.6704	0.6713	0.6871	1
P. jilunlii^{T} CGMCC 1.10239 GCF_900102965.1	0.6874	0.6646	0.6765	0.6799	0.6940
P. pasadenensis^{T} DSM 19293 GCF_000422485.1	0.7153	0.6928	0.6815	0.6690	0.7137
P. ginsengihumi^{T} DSM 21568 GCA_000380965.1	0.7174	0.6709	0.6841	0.6840	0.7281

Table 2-6. ANI of *Microbacterium lacusdiani* MEC084 and Related Bacteria

Bacterium	1	2	3	4
1: *M. MEC084*	1			
2: *M. barkeri* 2011-R4 NZ_AKVP01000020.1	0.7714	1		
3: *M. paraoxydans* DSM 15019 LT629770.1	0.7628	0.8374	1	
4: *M. maritypicum* MF109 NZ_ATAO01000206.1	0.7561	0.8393	0.8232	1
M. paludicola CC3 GCA_001887285.1	0.7583			
M. mangrovi MUSC 115T GCA_000802305.1	0.7589			
M. gubbeenense DSM 15944T GCA_000422745.1	0.7633			
M. indicum DSM 19969T GCA_000422385.1	0.7709			

Note: *M. lacticum* is the type species of the genus *Microbacterium*. However its genome sequence is not currently available.

Table 2-7. ANI of *Shinella sumterensis* MEC087, MEC089, and Related Bacteria

Bacterium	1	2	3	4	5	6
1: *S. sumterensis MEC087*	1					
2: *S. sumterensis MEC089*	0.9997	1				
3: *Shinella zoogloeoides* DD12 GCF_000496935.2	0.8353	0.8356	1			
4: *Shinella* sp. HZN7 GCA_001652565.1	0.8424	0.8423	0.9061	1		
5: *R. leguminosarum* bv. Trifolii WSM1689 NZ_CP007045.1	0.7632	0.7640	0.7609	0.7679	1	
6: *R. giardinii* bv. Giardinii H152 NZ_KB902578.1	0.7639	0.7655	0.7615	0.7742	0.7507	1

Phenotypic and Biochemical Comparison

Carbohydrate utilization of each isolate as well as their corresponding type strains is presented in Table 2-8.

Table 4. Phenotypic characterization of bacterial isolates which metabolize levoglucosan as the sole carbon source (indicated with an asterisk *) and reference strains used in this study.

Strain	1*	2	3*	4	5*	6*	7	8*	9
DNA G+C (% mol) content	55.7	44.9	71.7	74.9	63.4	63.4	63.4	57.3	57.1
Utilization of:									
L-arabinose	-	-	-	+	+	+	+	+	+
cellobiosan	+	+	+	+	+	+	+	+	+
fructose	+, a	+, a	+, a	+	+	+	+	+, a	-
D-galactose	+, a	+, a	+	+	+	+	+	+, a	+, a
D-glucose	+, a	+, a	+	+	+	+	+	+, a	+, a
myo-inositol	+	-	+	-	+	+	+	+, a	+, a
lactose	+, a	-	+	+, a	+	+	-	+	+
levoglucosan	+	-	+	-	+, a	+, a	-	+	-
mannitol	+, a	+, a	+	+, a	+	+	+	+, a	+, a
D-mannose	+, a	+, a	+	+	+	+	+, a	+	
sorbitol	-	+, a	-	+	+	+	+	+, a	+, a
sucrose	+, a	-	+, a	+, a	+, a	+, a	+	+, a	+, a
D-xylose	+	+, a	+	+	+	+	+	+, a	-

Strains:

1, MEC069; 2, *Paenibacillus polymyxa* ATCC 842 (DSM 36); 3, MEC084; 4, *Microbacterium lacticum* ATCC 8180 (DSM 20427);

5, MEC087; 6, MEC089; 7, *Shinella zoogloeoides* ATCC 19623 (DSM 287);

8, MEC097; 9, *Klebsiella pneumoniae* subsp. *pneumoniae* ATCC 13883 (DSM 30104).

Symbols: + grows on substrate or exhibits trait, - does not grow on substrate or does not exhibit trait, *a* produced acid

Paenibacillus MEC069 and *Paenibacillus polymyxa* exhibited many similar properties on a variety of carbohydrates. Both bacteria grew and produced gas on fructose, galactose, glucose, mannitol, and mannose, grew in the presence of cellobiosan, and did not utilize L-arabinose. While MEC069 grew in the presence of lactose and sucrose, and produced acid in the presence of myo-inositol and levoglucosan, *P. polymyxa* did not grow in the presence of any of these four carbohydrates. While *P. polymyxa* grew and produced acid in the presence of xylose, MEC069 grew without producing acid.

Microbacterium MEC084 and *M. lacticum* were similar on most of the carbohydrates examined. Both bacteria grew in the presence of cellobiosan, galactose, glucose, mannose, and xylose, and both bacteria produced acid in the presence of sucrose. MEC084 did not utilize L-arabinose or sorbitol, while *M. lacticum* did; *M.* MEC084 utilized myo-inositol and levoglucosan, while *M. lacticum* did not. Although both bacteria grew in the presence of lactose and mannitol, only *M. lacticum* produced acid.

For the grouping of *Shinella* strains, MEC087 and MEC089 were compared to *S. zoogloeoides* (ATCC 19623). MEC087 and MEC089 were identical in every way observed, being capable of growth on all of the tested carbohydrates. *S. zoogloeoides* exhibited growth on all carbohydrates examined except lactose and levoglucosan. While only MEC087 and MEC089 produced acid from levoglucosan and sucrose, *S. zoogloeoides* alone produced acid in the presence of mannose.

Klebsiella pneumoniae MEC097 were Gram-negative bacilli arranged in chains, facultatively anaerobic, nonmotile, catalase- and oxidase- positive. Colonies were circular, mucoid, glistening, white, opaque, umbonate, with entire edges after growth for 24 hours on LIM at 37°C. Growth was observed on arabinose, cellobiosan, lactose, and levoglucosan. Growth with

acid production occurred on fructose, galactose, myo-inositol, mannitol, sorbitol, sucrose, and xylose. With the exception of growth on levoglucosan as a sole carbon source, MEC097 exhibited identical phenotypes to *K. pneumoniae* ATCC 13883^T for all traits tested.

Comparison of the Growth of Klebsiella strains on Cellobiosan

The isolates and all corresponding type strains showed growth on cellobiosan, but only the isolates grew when levoglucosan was the sole carbon source (Table 2-8), demonstrating that growth on cellobiosan is not a predictor for growth on levoglucosan. To examine this difference in greater detail, the two *Klebsiella* strains (isolate MEC097 and *Klebsiella pneumoniae* subsp. pneumoniae ATCC 13883^T) were grown on 10 mM cellobiosan, and this substrate, products, and OD were measured (Fig. 2-4). Both strains reached their maximum OD and consumed most of the cellobiosan within 14 h. The type strain reached a maximum OD of 2.3 ± 0.3 and accumulated 9.8 ± 1.1 mM levoglucosan, essentially equimolar to the amount of cellobiosan consumed. The isolate *K. pneumoniae* MEC097 attained a maximum OD of 4.0 ± 0.4 and in contrast with the type strain, no levoglucosan was detected during the course of cellobiosan degradation.

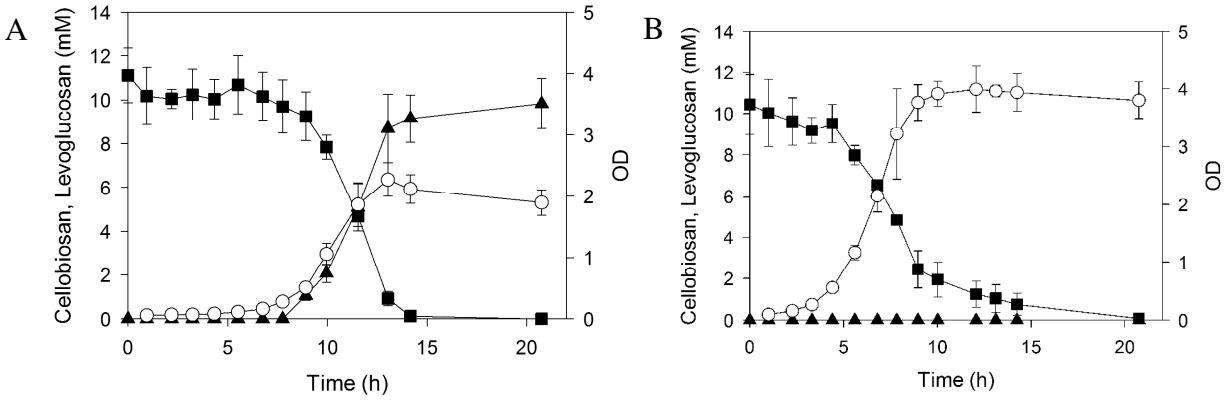


Figure 2-4. Growth of (a) *Klebsiella pneumoniae* subs. *Pneumoniae*^{T} and (b) *K. pneumoniae* MEC097 on 10 mM cellobiosan as the sole carbon source: cellobiosan (■), levoglucosan (▲), OD (○).

Identification of lgdA in Genome Assemblies and Comparison of Neighboring Genes

The two published N-terminal peptide sequences (Nakahara et al., 1991) have 100% sequence identity with bases 7-60 and bases 697-753 of CDS ADX2256.1 in *P. phenanthrenivorans* Sphe3 (CP002379.1). This gene is 1,173 bp in length and was annotated as a putative dehydrogenase containing a NAD-binding Rossman fold. Expression of this gene in *E. coli* BL21 (DE3)-RIPL cells produced an enzyme that converted NAD^+ to NADH in the presence of levoglucosan. The genomes of levoglucosan-utilizing bacteria were BLAST with the ADX2256.1 sequence, and each genome had at least one gene with 66-73 % sequence identity (Fig. 2-5, Table 2-9). Expression of *lgdA* from each isolate in *E. coli* MG1655 did not confer the ability to consume levoglucosan. The contigs containing each putative *lgdA* were aligned with Mauve and revealed several genes conserved among all bacteria considered (Fig. 2-6). These four genes are predicted by GenBank to be a monosaccharide ABC transporter substrate-binding protein of the CUT2 family (Table 2-10, sequence identity 53-70 %), a monosaccharide ABC transporter membrane protein of the CUT2 family (Table 2-11, sequence identity 58-72 %), a sugar phosphate isomerase/epimerase (Table 2-12, sequence identity 44-56 %), and a putative dehydrogenase known to be levoglucosan dehydrogenase (Table 2-9). The contig containing the *lgdA* of *K. pneumoniae* MEC097 also has an IS3 transpose (gene locus B4U61_26630) 3899 nucleotides downstream of gene locus B4U61_26650 (Fig. 2-6).

Table 2-9. Sequence identity among various translated *lgdA*

Isolate	1	2	3	4	5
1: *P. phenanthrenivorans* Sphe3	1				
2: *P. athensensis* MEC069^T	0.7132	1			
3: *M. marinilacus* MEC084	0.7282	0.7158	1		
4: *S. sumterensis* MEC087/MEC089	0.6694	0.6771	0.6771	1	
5: *K. pneumoniae* MEC097	0.7018	0.7519	0.6607	0.6875	1

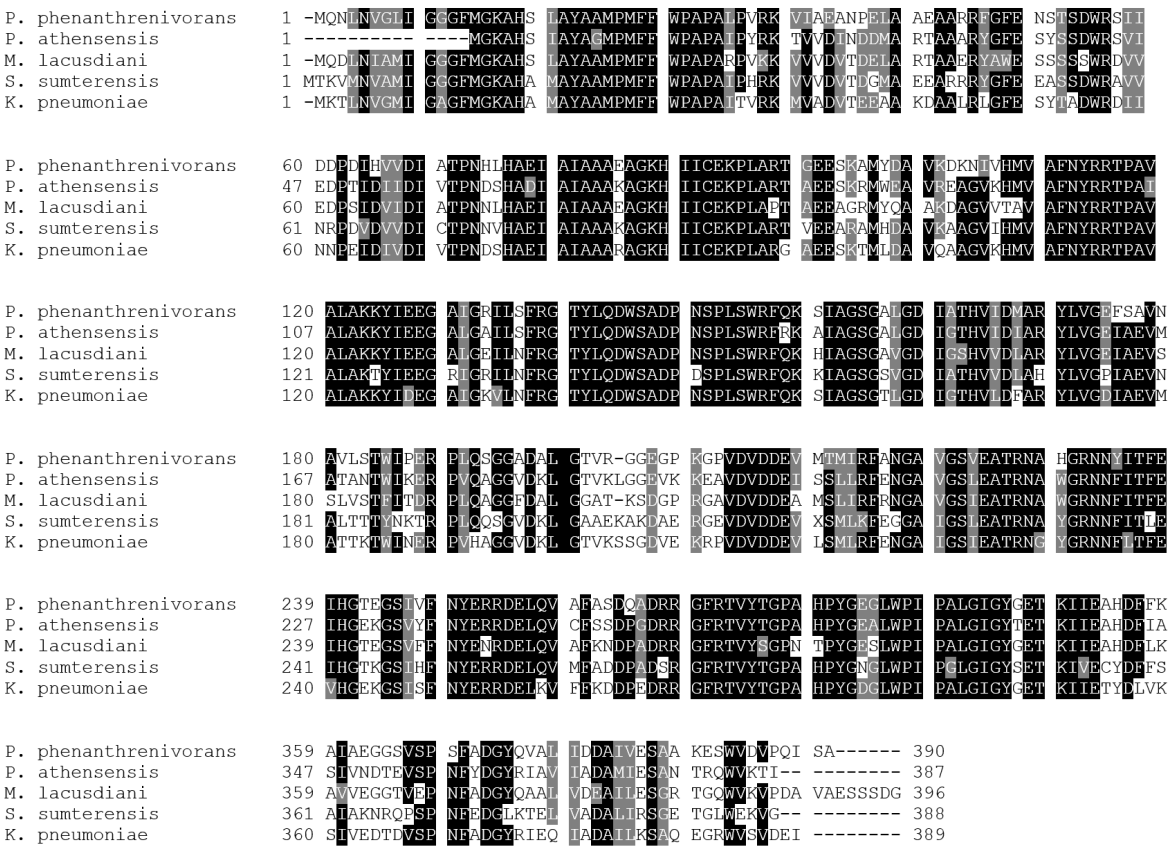


Figure 2-5. Alignment of translated *lgdA* genes from bacterial isolates and *Pseudarthrobacter phenanthrenivorans* Sphe3. Black highlighting indicates a residue conserved in 80% of the sequences, grey highlighting indicates similar residues.

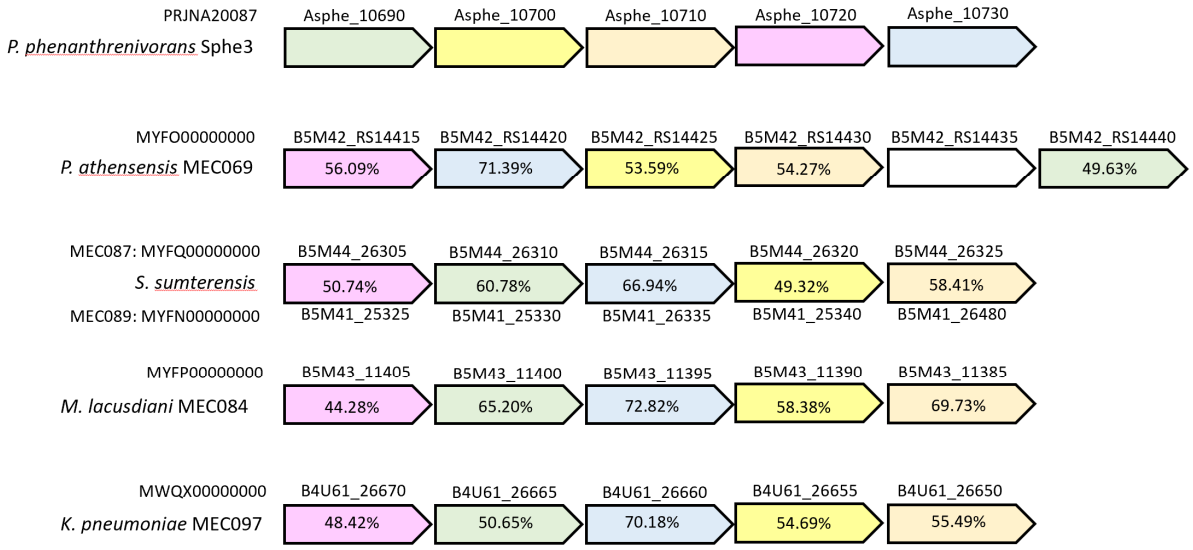


Figure 2-6. Genes neighboring *lgdA* in the genomes of levoglucosan utilizing isolates. *P. phenanthrenivorans* Sphe3 is shown at the top with *lgdA* in blue. Genes homologous across genomes are colored the same, and the sequence identity between the translated protein and *P. phenanthrenivorans* homolog are presented in each arrow. Pink genes are *lgdB*, the putative isomerase responsible for the conversion of 3-keto levoglucosan to 3-keto glucose. The orange, yellow, and green genes encode respectively for a putative monosaccharide ABC transporter substrate-binding protein of the CUT2 family, a monosaccharide ABC transporter membrane protein of the CUT2 family, and an ABC transporter ATP-binding protein. Gene loci are presented above each gene with the genome accession number above the organism's name (in the case of *S. sumterensis* the loci and accession number of strain MEC089 are presented below the genes).

Table 2-10. Sequence identity among putative monosaccharide ABC transporter substrate binding proteins of the CUT2 family homologous to Asphe3_10690.

Isolate	1	2	3	4	5
1: *P. phenanthrenivorans* Sphe3	1				
2: *P. athensensis* MEC069^T	0.5593	1			
3: *M. lacusdiani* MEC084	0.6963	0.5707	1		
4: *S. sumterensis* MEC087/MEC089	0.6267	0.5957	0.6196	1	
5: *K. pneumoniae* MEC097	0.5627	0.5301	0.5606	0.5918	1

Table 2-11. Sequence identity among putative monosaccharide ABC transporter membrane protein homologous to Asphe3_10710.

Isolate	1	2	3	4	5
1: *P. phenanthrenivorans* Sphe3	1				
2: *P. athensensis* MEC069^T	0.5854	1			
3: *M. lacusdiani* MEC084	0.7180	0.5895	1		
4: *S. sumterensis* MEC087/MEC089	0.6699	0.6045	0.6560	1	
5: *K. pneumoniae* MEC097	0.6132	0.6911	0.5926	0.6204	1

Table 2-12. Sequence identity among putative sugar phosphate isomerase/epimerases homologus to Asphe3_10720.

Isolate	1	2	3	4	5
1: *P. phenanthrenivorans* Sphe3	1				
2: *P. athensensis* MEC069^T	0.5609	1			
3: *M. lacusdiani* MEC084	0.4428	0.4579	1		
4: *S. sumterensis* MEC087/MEC089	0.5074	0.4610	0.4207	1	
5: *K. pneumoniae* MEC097	0.4842	0.5566	0.4484	0.4521	1

Determination of LGDH Kinetic Parameters from Various Hosts through Recombinant Expression and Enzyme Assays

With a probable *lgdA* identified in each isolate's genome, we sought to compare these enzymes' kinetic parameters. Purified recombinant LGDH (using a histag at the C-terminus of each protein) was used to determine the reaction rate under 27 different concentrations of levoglucosan and NAD^+. These rates were then used to compute the K_M and k_{cat} and inhibition parameter K_I if appropriate (Table 2-13).

Table 2-13. Kinetic parameters of purified recombinant LGDH from various hosts.

Strain	k_{CAT}	K_M (levoglucosan) (mM)	K_M (NAD) (mM)	K_I (levoglucosan) (mM)
P. phenanthrenivorans	52 (3)	27 (4)	1.0 (0.2)	
P. athensensis MEC069	84 (4)	1.7 (0.4)	0.31 (0.05)	320 (50)
M. lacusdiani MEC084	63 (7)	1.8 (0.8)	0.63 (0.18)	470 (150)
S. sumterensis MEC087	27 (5)	2.1 (0.8)	9.1 (2.2)	93 (21)
K. pneumoniae MEC097	58 (3)	22 (3)	0.79 (0.12)	

Values presented are the calculated value (standard error having $p<0.05$). Only when a substrate inhibition model showed significance is the K_I value also shown.

The kinetic parameters of the five isolated LGDH vary significantly. The *S. sumterensis* LGDH is slowest in terms of turnover, with k_{CAT} of 27 s^{-1}, $K_{M(NAD)}$ 9.1 mM, $K_{M(LG)}$ 2.1 mM, and $K_{I(LG)}$ 93 mM. *P. phenanthrenivorans* and *K. pneumoniae* LGDH are intermediate, with k_{CAT} of 52 and 58 s^{-1} respectively, with neither showing signs of substrate inhibition. LGDH of *P. athensensis* and *M. lacusdiani* are the fastest acting enzymes, having k_{CAT} of 84 and 63 s^{-1}. The LGDH of *P. phenanthrenivorans* and *K. pneumoniae* did not show any substrate inhibition, while the other three did show substrate inhibition.

Discussion

The five bacterial isolates enriched in this study on the basis of their ability to metabolize levoglucosan as the sole carbon source, represent four distinct species, two of which are novel species. The isolates are proposed to be *Paenibacillus athensensis* MEC069^T, *Microbacterium lacusdiani* MEC084, *Shinella sumterensis* MEC087^T, and *Klebsiella pneumoniae* MEC097.

P. athensensis was determined to be a novel species based on its 16S rRNA sequence and the ANI of its genome compared to closely related *Paenibacillus* (Supplementary Table 2). Among officially named *Paenibacillus*, *P. pectinilyticus* and *P. alginolyticus* were the closest relatives with 95.77% and 94.95% 16S rRNA gene similarity, respectively, suggesting that *P. athensensis* MEC069^T is a unique species (Fig. 2-1). The ANI between these bacteria and MEC069 were 71.35%, and 70.18%, confirming that they are distinct species of *Paenibacillus*.

Considering the identical colony appearance and isolation from the same source (Table 2-2), growth rate and specific enzyme activity on levoglucosan (Table 2-3), substrate utilization (Table 2-8), ANI compared to bacteria of the same genus (Table 2-) as well as the similarities in genome size and composition, it is likely MEC087 and MEC089 represent clones of *Shinella sumterensis*. While *Shinella sumterensis* exhibited many similarities with *S. zoogloeoides*

including 99.86% 16S rRNA sequence similarity, the ANI between these two bacteria was 83.53%, indicating that the two organisms are likely of the same genus, but of distinct species.

While MEC084 may be a novel species among the *Microbacterium*, this strain may be an isolate of *M. lacusdiani*. From the 16S rRNA sequences, MEC084 has 99.86% similarity to *M. lacusdiani* and a 98.90% sequence similarity with *M. marinilacus*. Notable differences between *M. lacusdiani* and MEC084 are that *M. lacusdiani* grows on D-sorbitol, does not grow on D-mannose, myo-inositol as sole carbon sources, was isolated from a freshwater environment, and has a G + C content of 70.4 mol % (Zhang et al., 2017). With no genetic sequence from *M. lacusdiani* available for comparison other than the 16S rRNA gene, one cannot definitively conclude these are unique bacteria.

Isolate MEC097 is the only known strain of *K. pneumoniae* capable of utilizing levoglucosan as its sole carbon source. Comparing 16S rRNA sequences, there is a 99.25% sequence similarity between MEC097 and *K. pneumoniae* DSM 30104^{T} (X87276.1); the ANI between the two strains (using assembly GCA_000281755.1) is 0.9905 making it clear they are same species. On plate media, both organisms have the same mucoid, white colonies. Aside from MEC097 utilizing fructose, levoglucosan, and D-xylose as sole carbon sources when the type strain does not, both strains can utilize L-arabinose, cellobiosan, D-galactose, D-glucose, myo-inositol, lactose, mannitol, sorbitol, and sucrose. MEC097 appears to be a unique strain of *K. pneumoniae*.

Although isolates exhibited similar biochemical activities to their closest match by phylogenetic analysis, surprisingly none of the reference strains were capable of growth on levoglucosan. This difference in substrate utilization between close phylogenetic neighbors

implies that the isolates have a novel metabolic pathway which may have been shared through horizontal gene transfer or convergent evolution.

Among the five isolates described in this study, *K. pneumoniae* subsp. MEC097 exhibited both the fastest growth rate on levoglucosan (0.80 ± 0.10 h^{-1}) and specific LGDH activity ($0.32 \pm$ 0.16 IU/g dry cell weight). In contrast, *P. athensensis* MEC069, *M. lacusdiani* MEC084, *S. sumterensis* MEC087, and *S. sumterensis* MEC089 all had growth rates roughly in the range of 0.30–0.50 h^{-1} and roughly 15–30% the specific LGDH activity of *K. pneumoniae* MEC097.

Interestingly, there is no clear correlation between the maximum specific growth rate of isolates on levoglucosan and the kinetic parameters of each LGDH (Table 2-13), suggesting that enzyme expression level, levoglucosan transport, or another factor may be controlling the rate of cell growth on levoglucosan. For example, the measured LGDH activity during growth was the greatest for *K. pneumoniae* MEC097, although the k_{CAT} is an intermediate value compared to other LGDHs. In contrast, *P. athensensis* MEC069 and *M. lacusdiani* MEC084 contain the most active LGDH enzymes, but these strains exhibited the lowest specific LGDH activity per gram dry cell weight.

We hypothesized that if LGDH is oxidizing the C3 hydroxyl group of levoglucosan, then organisms capable of growth on levoglucosan might also be able to metabolize myo-inositol which has a similar arrangement of hydroxyls. Each isolate was indeed able to use myo-inositol as the sole carbon source. Among the reference strains, however, all strains except *P. polymyxa* and *M. lacticum* were also able to grow on myo-inositol. Since most reference strains grew on myo-inositol but none were able to metabolize the similarly structured levoglucosan, the reference strains likely use a distinct enzymatic pathway for the metabolism of myo-inositol. These

observations do not conclusively determine whether the isolates use the LGDH to metabolize myo-inositol.

Cellobiosan is a disaccharide consisting of one β-D-glucose subunit and a levoglucosan subunit bound by an 1,4-linkage. All bacteria examined were able to metabolize cellobiosan, even those unable to metabolize levoglucosan. The observation that reference strains differed in cellobiosan and levoglucosan consmption suggests two possible pathways to degrade cellobiosan: (1) bacteria convert cellobiosan to cellobiose which is then directly metabolized, or (2) bacteria cleave cellobiosan into levoglucosan and glucose monomers, and then grow on one or both of the resulting monosaccharides. Our study comparing the growth of *K. pneumoniae* MEC097 and *K. pneumoniae* subsp. *pneumoniae* on cellobiosan (Fig. 4) supports the conclusion that these bacteria cleave cellobiosan into glucose and levoglucosan. In particular, the reference strain accumulated an equimolar amount of levoglucosan during cellobiosan degradation and only attained one-half the cell density as MEC097 on cellobiosan. Cellobiose was not detected during the growth of either strain. Neither glucose nor levoglucosan accumulated during growth of MEC097 on cellobiosan, indicating that if cellobiose is indeed cleaved on this isolate, then these two monosaccharides are metabolized simultaneously. While the ability to catabolize levoglucosan may be fairly uncommon among bacteria, the β-galactosidase activity necessary to cleave cellobiosan into its component monomers appears to be much more common. This observation is worth noting as researchers continue the development of bacterial strains for the utilization of pyrolysis products.

Comparison of *lgdA* in the isolate genomes as well as comparison of the neighboring genes yields several insights into bacterial metabolism of levoglucosan. While the arrangement and total number of genes in each *lgdA* operon were variable between isolates, five genes were conserved

among all isolates (Fig. 2-6). PGAP annotated these genes as a putative dehydrogenase (*lgdA*, blue, shown in this work to be levoglucosan dehydrogenase), a sugar phosphate isomerase/epimerase (*lgdB*, pink), a monosaccharide ABC transporter membrane protein of the CUT2 family (yellow), a monosaccharide ABC transporter substrate-binding protein of the CUT2 family (orange), and an ABC transporter ATP-binding protein (green). While previous studies which demonstrate that *E. coli* will grow on levoglucosan merely by overexpression of levoglucosan kinase (Zhuang and Zhang, 2002; Dai et al., 2009; Layton et al., 2011) indicate that no specialized levoglucosan transporters are necessary for levoglucosan to cross the cell membrane of *E. coli*, the substrate-binding periplasmic protein, transmembrane protein, and ATP-binding protein components of ABC transporters consistently identified in the five *lgd* operons suggest that the genes clustered with *lgdA* and *lgdB* may be components of ABC transporters specific to levoglucosan. Indeed, these genes are orthologous to the components of the ribose ABC-transporter of *Caldanaerobacter subterraneus* subsp. *tengcongensis* and the myo-inositol ABC transport of *Caulobacter vibrioides* (Herrou and Crosson, 2013). Furthermore, the presence of a conserved putative isomerase adjacent to *lgdA* suggests that during the catabolism of levoglucosan this may be the enzyme responsible for hydrolysis of 3-keto levoglucosan. The product of *lgdB* is orthologous to tagatose-3-epimerase (Shin et al., 2017) and 5-keto-L-gluconate epimerase *iolO* (Rodionova et al., 2013) from *Thermotoga maritima*, D-tagatose epimerase from *Pseudomonas cichorii* (Bosshart et al., 2015), D-psicose 3-epimerase from *Clostridium cellulolyticum* (Chan et al., 2012), inosose isomerase *iolI* from *B. subtilis* (Zhang et al., 2002), among other isomerases. Many of these enzymes require a divalent cation such as Mn^{2+}, Mg^{2+}, or Zn^{2+} in their active site to isomerize a keto-sugar. Several also do not require a phosphorylated substrate, an observation which is consistent with the proposed mechanism of catalyzing the conversion of 3-keto

levoglucosan to 3-keto glucose (Nakahara et al., 1994). Determining which gene products are necessary for levoglucosan metabolism and what reactions they mediate would be a logical next step from this research.

In the *K. pnuemoniae* MEC097 genome, the close proximity between an IS3 transpose and the five genes described above suggests that the genes entered the genome by horizontal gene transfer. No available *Klebsiella* genomes contain *lgdA*, so it can be reasoned that *lgdA* and other genes that contribute to levoglucosan consumption were integrated into the MEC097 genome by horizontal gene transfer and not a group of genes that likely evolved with *K. pneumoniae*. In the genome assemblies of *P. athensensis* MEC069, *M. lacusdiani* MEC084, *S. sumterensis* MEC087, and *S. sumterensis* MEC089 no transposases are found in the contigs containing the five conserved genes.

Description of *Paenibacillus athensensis* sp. nov.

Paenibacillus athensensis (a.thens.en'sis. N.L. masc. adj. athensensis, in recognition of Athens, GA, USA, the location from which the type strain was isolated.)

Cells Gram-negative staining cocco-bacilli, facultatively anaerobic, motile, catalase- and oxidase- negative. Colonies are small, cream-beige colored, opaque, dull with entire margins after 24 h of growth on LIM at 37°C. Growth was observed on cellobiosan, myo-inositol, levoglucosan, and xylose. Growth and acid production occurs on fructose, galactose, glucose, lactose, mannitol, mannose, and sucrose. Growth was not observed with arabinose or sorbitol. The DNA G+C content is 55.7 mol %. The type strain, MEC069^T (= DSM 28474^T = ATCC TSD-105) was isolated from the wastewater of a chemical manufacturing company (Athens, GA, USA; GPS location: 33.98023, -83.32673).

Description of *Shinella sumterensis* sp. nov.

Shinella sumterensis (sum'ter.en'sis N.L. fem. adj. sumterensis, in recognition of Sumter National Forest, SC, USA, the location from which the type strain was isolated.)

Cells are Gram-negative cocci arranged in long chains, facultatively anaerobic, motile, catalase- and oxidase- positive. Colonies are small, pinkish-white, glistening, smooth, with entire margins after growth for 24 h on LIM at 37°C. Growth was observed on arabinose, cellobiosan, fructose, galactose, glucose, myo-inositol, lactose, levoglucosan, mannitol, mannose, sorbitol, sucrose, and xylose. The DNA G+C content is 63.4 mol %. The type strain, MEC087^T (= DSM 27294^T = ATCC TSD-106) was isolated from silty topsoil heavy in organic matter from a campfire pit (Russell Homestead, SC, USA; GPS location: 34.91107, -83.17371).

Acknowledgments

The Sarah Lee for technical assistance, and Mary-Ann Moran and Brandon Satinsky for assistance in obtaining Illumina sequencing of isolate genomes.

References

N. M. Bennett, S. J. B. Duff. 2005. Fortification of spent sulphite pulping liquor with levoglucosan derived from pyrolysis oil. *ASAE Paper* PNW05-1015.

N. M. Bennett, S. S. Helle, S. J. B. Duff. 2009. Extraction and hydrolysis of levoglucosan from pyrolysis oil. *Biores. Technol.* 100:6059-6063.

N. Bose, P. Greenspan, C. Momany. 2002. Expression of recombinant human betaine: homocysteine S-methyltransferase for X-ray crystallographic studies and further characterization of interaction with S-adenosylmethionine. *Protein Expr. Purif.* 25(1):73-80.

D. Coil, J. Guillaume, A. E. Darling. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*. 31(4):587-9.

J. Dai, Z. Yu, Y. He., L. Zhang, Z. Bai, Z. Dong, Y. Du, H. Zhang. 2009. Cloning of a novel levoglucosan kinase gene from *Lipomyces starkeyi* and its expression in *Escherichia coli*. *World J. Microbiol. Biotechnol.* 25:1589-1595.

A. C. E. Darling, B. Mau, F. R. Blattner, N. T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394-1403.

G. Dobele, G. Rossinskaja, T. Diazbite, G. Telysheva, D. Meier, O. Faix. 2005. Application of catalysis for obtaining 1,6-anhydrosaccharides from cellulose and wood by fast pyrolysis. *J. Anal. Appl. Pyrolysis* 74:401-405.

T. V. Elzhov, K. M. Mullen, A-N. Spiess, B. Bolker. 2016. R-Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds. https://CRAN.R-project.org/package=minpack.lm

N. R. Galloway, H. Toutkoushian, M. Nune, N. Bose, C. Momany. 2013. Rapid cloning for protein crystallography using type IIS restriction enzymes. *Cryst Growth Des.* 13(7):2833-2839.

J. Herrou and S. Crosson. 2013. Myo-inositol and D-ribose ligand discrimination in an ABC periplasmic binding protein. *J Bacteriol.* 195:2379-2388.

J. Howard, C. Longley, A. Morrison, D. Fung. 1993. Process for isolating levoglucosan from pyrolysis liquids. Canadian Patent 2,084,906.

A. Kageyama, Y. Takahashi, Y. Matsuo, H. Kasai, Y. Shizuri, S. Omura. 2007. *Microbacterium sediminicola* sp. nov. and *Microbacterium marinilacus* sp. nov., isolated from marine environments. *Int. J. Syst. Evol. Microbiol.* 57:2355-2359.

J. R. Klesmith, J. P. Bacik, R. Michalczyk, T. A. Whitehead. 2015. Comprehensive sequence-flux mapping of a levoglucosan utilization pathway in *E. coli*. *ACS Synth. Biol.*4: 1235-1243.

Y. Kitamura, Y. Abe, T. Yasui. 1991. Metabolism of levoglucosan (2,6-anhydro-β-D-glucopyranose) in microorganisms. *Agric. Biol. Chem.* 55(2):515-521.

Y. Kitamura, T. Yasui. 1991. Purification and some properties of levoglucosan (1,6-anhydro-β-D-glucopyranose) kinase from the yeast *Sporobolomyces salmonicolor*. *Agric. Biol. Chem.* 55(2):523-529.

B. Langmead, S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9:357-9.

D. S. Layton, A. Ajjarapu, D. W. Choi, L. R. Jarboe. 2011. Engineering ethanologenic *Escherichia coli* for levoglucosan utilization. *Bioresour. Technol.* 102(17):8318-8322.

S. Lee, K. Baumann, J. Schauer, R. Sheeshley, L. Naeher, S. Meinardi, D. Blake, E. Edgerton, A. Russell, M. Clements. 2005. Gaseous and particulate emissions from prescribed burning in Georgia. *Environ. Sci. Tech.* 39(23):9049-56.

J. Lian, J. Choi, Y. S. Tan, A. Howe, Z. Wen, L. R. Jarboe. 2016. Identification of soil microbes capable of utilizing cellobiosan. *PloS ONE*. 11(2).

F. A. Loewus and S. Kelly. 1962. Conversion of glucose to inositol in parsley leaves. *Biochem. Biophys. Res. Commun.* 7:204-208.

M. W. Loewus and F. A. Loewus. 1980. The C-5 hydrogen isotope-effect in myo-inositol synthase as evidence for the myo-inositol oxidation-pathway. *Carbohydr. Res.* 82:333-342.

J. A. Lomax, J. M. Commandeur, P.W. Arisz, J. A. Boon. 1991. Characterization of oligomers and sugar ring-cleavage products in the pyrolysate of cellulose. *J. Anal. App. Pyrol.* 19: 65-79.

A. L. Majumder, M. D. Johnson, S. A. Henry. 1997. 1L-myo-inositol-1-phosphate synthase. *Biochim. Biophys. Acta.* 1348:245-256.

M. E. Migaud and J. W. Frost. 1995. Inhibition of myo-inositol-1-phosphate synthase by a reaction coordinate intermediate. *J. Amer. Chem. Soc.* 117(18):5154-5155.

L. Moens. 1994. Isolation of levoglucosan from pyrolysis oil derived from cellulose. World Patent 9,405,704.

K. Nakahara, Y. Kitamura, Y. Yamagishi, H. Shoun, T. Yasui. 1994. Levoglucosan dehydrogenase involved in the assimilation of levoglucosan in *Arthrobacter* sp. I-552. *Biosci. Biotech. Biochem.* 58(12):2193-2196.

E. M. Prosen, D. Radlein, J. Piskorz, D. S. Scott, R. L. Legge. 1993. Microbial utilization of levoglucosan in wood pyrolysate as a carbon and energy source. *Biotechnol. Bioeng.* 42:538-541.

N. Saitou and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.

J.J. Schauer, M.J. Kleeman, G.R. Cass, B.R. Simoneit. 2001. Measurement of emissions from air pollution sources. 3. C1–C29 organic compounds from fireplace combustion of wood, *Environ. Sci. Technol.* 35(9): 1716–1728.

S. D. Scott, J. Piskorz, D. Radlein, P. Majerski. 1996. Process for the production of anhydrosugars from lignin and cellulose containing biomass by pyrolysis. U. S. Patent 5,395,455.

F. Shafizadeh and Y. L. Fu. 1973. Pyrolysis of cellulose. *Carbohydr. Res.* 29:113-122.

F. W. Studier. 2005. Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expr. Purif.* 41: 207-234.

K. Tamura, M. Nei, and S. Kumar. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl Acad. Sci. USA*. 101:11030-11035.

K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: Molecular Evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.

C. R. Vitasari, G. W. Meindersma, A. B. De Haan. 2011. Water extraction of pyrolysis oil: The first step for the recovery of renewable chemicals. *Biores. Technol.* 102:7204-7210.

H. Xie, X. Zhuang, Z. Bai, H. Qi, H. Zhang. 2006. Isolation of levoglucosan-assimilating microorganisms from soil and an investigation of their levoglucosan kinases. *World J. Microbiol. Biotechnol.* 22:887-892.

T. Yasui, Y. Kitamura, K. Nakahara, Y. Abe. 1991. Metabolism of levoglucosan (1,6-anhydro-β-D-glucopyranose) in bacteria. *Agric. Biol. Chem.* 55(7):1927-1929.

S. H. Yoon, S. M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo, J. Chun. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67(5):1613-1617.

B. H. Zhang, N. Salam, J. Cheng, H. Q. Li, J. Y. Yang, D. M. Zha, Q. G. Guo, W. J. Li. 2016. *Microbacterium lacusdiani* sp. nov., a phosphate-solubilizing novel actinobacterium isolated from mucilaginous sheath of *Microcystis*. *J Antibiot (Tokyo).* 70(2):147-151.

X. Zhuang, H. Zhang. 2002. Identification, characterization of levoglucosan kinase, and cloning and expression of levoglucosan kinase cDNA from *Aspergillus niger* CBX-209 in *Escherichia coli*. *Prot. Express Purif.* 26:71-81.

CHAPTER 3

Determining the Structure of Levoglucosan Dehydrogenase with X-Ray Crystallography

A. S. Arya, C. Momany, M. A. Eiteman. To be submitted to *J Appl Biotechnol.*

Abstract

LGDH crystals of the P 43 21 2 space group were soaked in artificial mother liquors containing levoglucosan or glucose, with or without NADH, to observe how these substrates interact with the active site residues. Three structures were obtained of LGDH dimers complexed with levoglucosan, with levoglucosan and NADH, and with glucose and NADH. In the last, only one subunit of the dimer was occupied by NADH. These structures support the hypothesis that LGDH performs a hydride-shift on hydroxyl-3 of levoglucosan or glucose to form 3-keto levoglucosan or 3-keto glucose, respectively.

Introduction

In bacteria the anhydrosugar levoglucosan is proposed to be oxidized to 3-keto levoglucosan by levoglucosan dehydrogenase with NAD+ as the electron acceptor (LGDH, Nakahara et al., 1994). A BLAST search of the partial peptide sequences (Nakahara et al., 1994) TGNLNVGLIGGGFMFKAHSL and TRGGEGPKGPVDVDDEVMTM on NCBI or PDB revealed that the protein is homologous to members of the Gfo/Idh/MocA family, an oxidoreductase family known for utilizing a Rossman fold to bind NADP(H) or NAD(H). Compared to levoglucosan, reduced activity of LDGH has been observed on glucose (5% relative activity), galactose (1%), fructose (1%), sorbose (30%), xylose (18%), arabinose (1%), and ribose (8%). With the goal of establishing the mechanism by which LGDH oxidizes these sugars and what residues in the active site may make LGDH specific for levoglucosan, histagged LGDH was purified, crystallized, and then frozen in artificial mother liquor containing 30% w/v of each carbohydrate. Crystals soaked in levoglucosan, levoglucosan and NADH, and glucose and NADH diffracted well enough to resolve carbohydrate binding to the active site.

Results and Discussion

Overall Structure

All crystals of LGDH in this study are in space group P 4_3 2_1 2, with LGDH forming a dimer. Apo-enzyme crystals were soaked in a mother liquor made from 1.8 M triammonium citrate and 30% w/v of each carbohydrate of interest as a cryo-protectant. Structures of the active site containing levoglucosan, glucose, and levoglucosan with NADH were solved to better understand the catalytic mechanism of LGDH. A monomer of LGDH is composed of nine α-helices and fifteen β-strands (Fig. 3-1), while the asymmetric unit of all crystals in this study contains a LGDH dimer (Fig. 3-2). Refinement statistics are presented in Table 3-1.

Table 3-1. Data collection and refinement statistics

	LGDH-levo	LGDH-levo-NADH	LGDH-gluc-NADH
Wavelength	1	1	1
Resolution range	37.1 - 2.15 (2.23 - 2.15)	41.2 - 2.20 (2.27 - 2.20)	33.8 - 1.80 (1.86 - 1.80)
Space group	P 43 21 2	P 43 21 2	P 43 21 2
Unit cell	92.4 92.4 167.3 90 90 90	95.1 95.1 167.6 90 90 90	95.5 95.5 171.8 90 90 90
Unique reflections	40107 (3903)	39333 (3420)	68141 (6885)
Completeness (%)	99.83 (99.59)	98.07 (86.80)	91.79 (94.28)
Wilson B-factor	30.48	35.28	17.64
Reflections used in refinement	40075 (3902)	39296 (3418)	68140 (6885)
Reflections used for R-free	2011 (202)	1974 (171)	3414 (343)
R-work (%)	18.85 (23.53)	16.56 (21.01)	22.19 (24.93)
R-free (%)	22.19 (28.22)	20.00 (26.61)	23.63 (26.79)
Number of non-hydrogen atoms	6523	6469	6731
macromolecules	5906	5916	5966
ligands	377	354	68
solvent	240	199	697
Protein residues	774	775	782
RMS(bonds)	0.005	0.005	0.008
RMS(angles)	1.01	1.06	1.20
Ramachandran favored (%)	95.6	95.3	96.4
Ramachandran allowed (%)	3.8	3.3	3.0
Ramachandran outliers (%)	0.6	1.4	0.6
Rotamer outliers (%)	1.7	1.9	1.5
Clashscore	4.1	5.6	6.4
Average B-factor	46.5	48.1	21.8
macromolecules	44.3	46.6	20.6
ligands	87.3	75.6	22.5
solvent	38.1	42.7	31.9

Statistics for highest resolution shell are shown in parentheses.

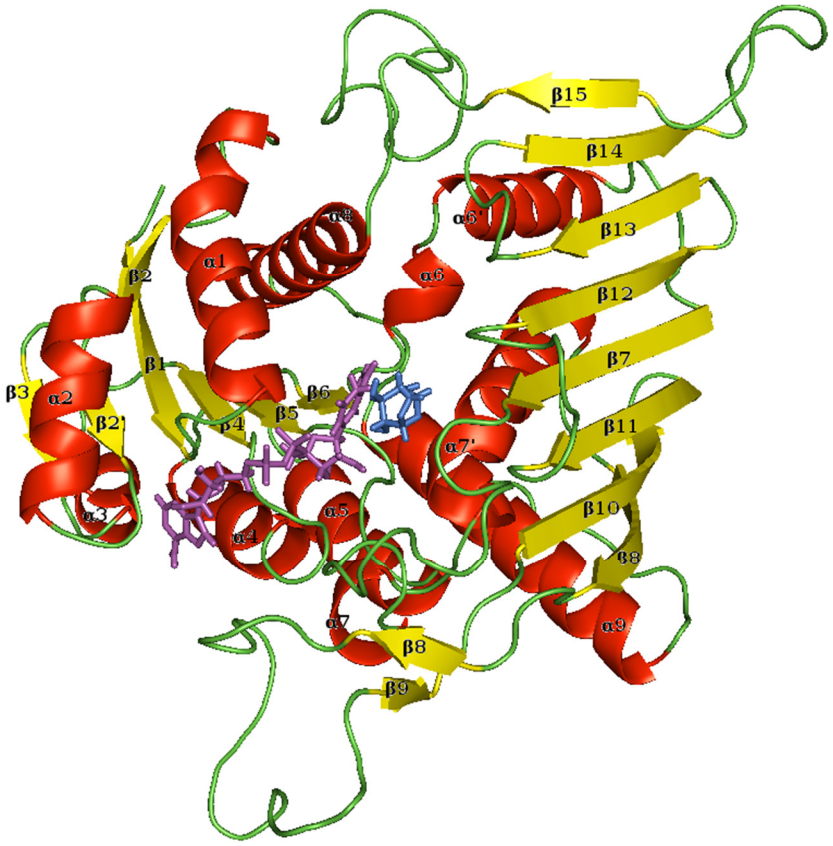


Figure 3-1. Cartoon representation of the LGDH monomer. The secondary structure is highlighted showing alpha helices (red), beta strands (yellow), loops (green), and the substrates levoglucosan (cyan) and NADH (magenta).

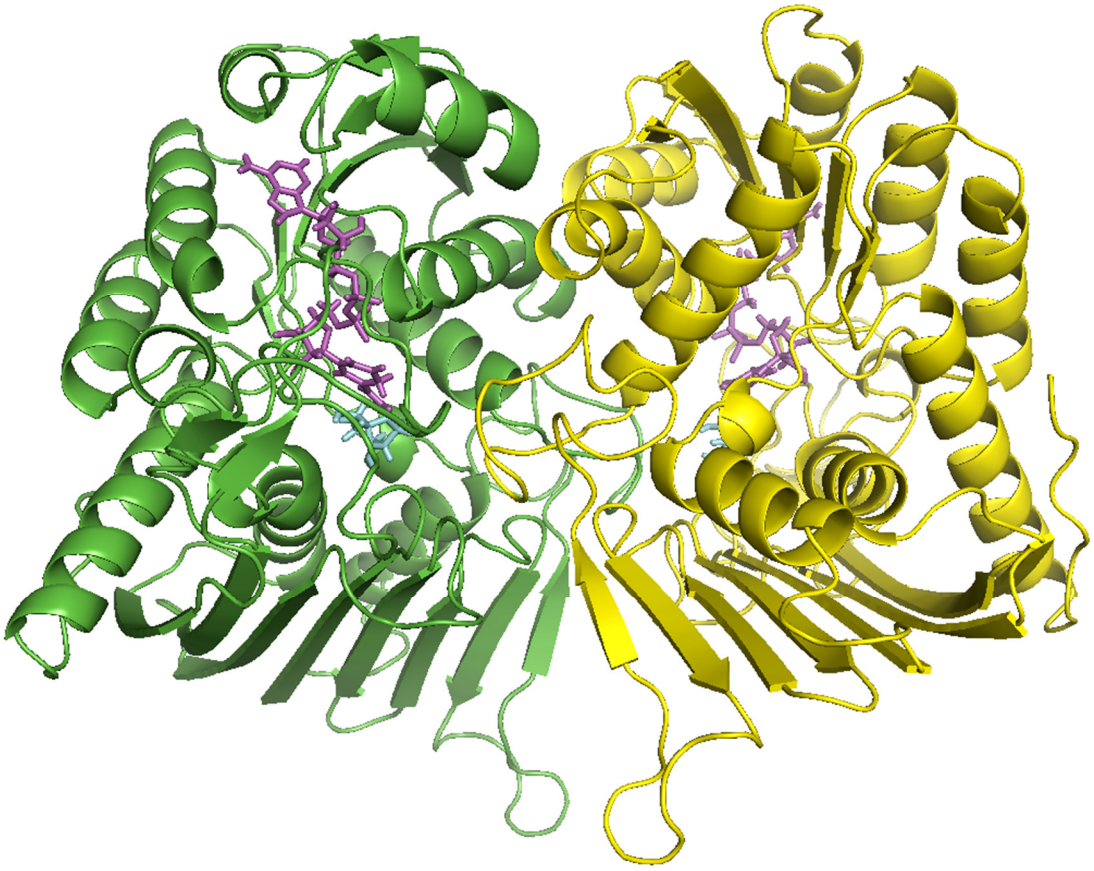


Figure 3-2. LGDH dimer as seen in the asymmetric unit of all crystal structures in this study. Pictured are Chain A (yellow), chain B (green), NADH (magenta), and levoglucosan (cyan).

Analysis with PISA (Krissinel and Henrick, 2007) demonstrated that the observed asymmetric unit, with the chain A and chain B related by a x,y,z symmetry operation (Fig. 3-2), is likely how the protein binds in physiological conditions. The greatest Complexation Significance Score was 0.511 with an interface surface area of 1751.7 $Å^2$, and 15 potential hydrogen bonds and 12 potential salt bridges across the interface. LGDH likely exists as a tetramer, with copies of chain A and B related by a crystallographic -y+1, -x+1, -z+1/2 transformation (Fig. 3-3). The interface area between each chain and its copy was estimated as 1507.9-1527.2 $Å^2$, with 5-7 potential hydrogen bonds, 12 potential salt bridges, and complexation significance scores of 0.445 in both cases.

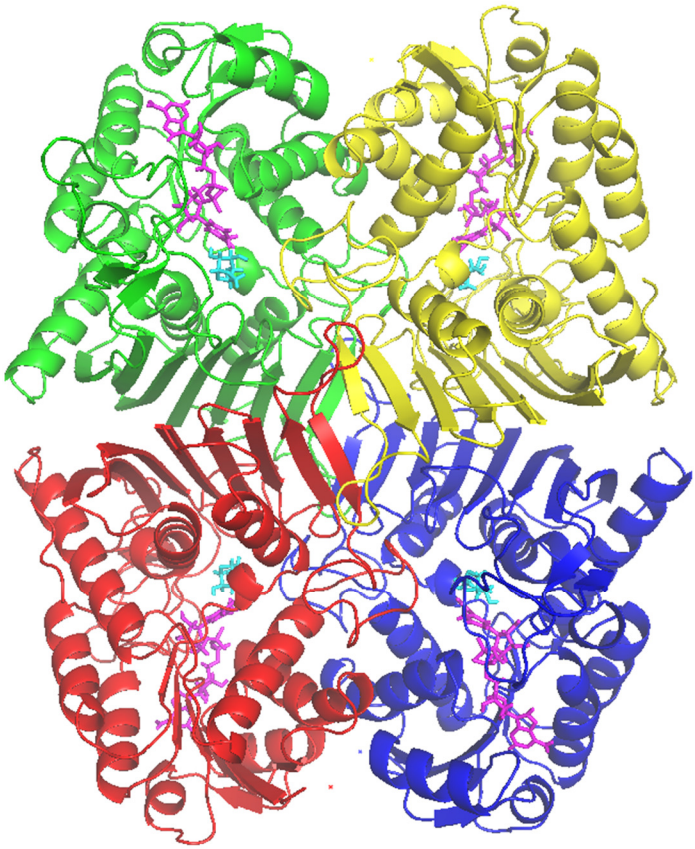


Figure 3-3. LGDH quaternary structure predicted by PISA. Each LGDH monomer is a different color.

Active Site Structure and Conservation

All residues within 5 Å of levoglucosan and NADH in the binding pocket were considered the active site (Fig. 3-4A and 3-4B). These residues are also highlighted in the alignment of known LGDHs (Fig. 3-5).

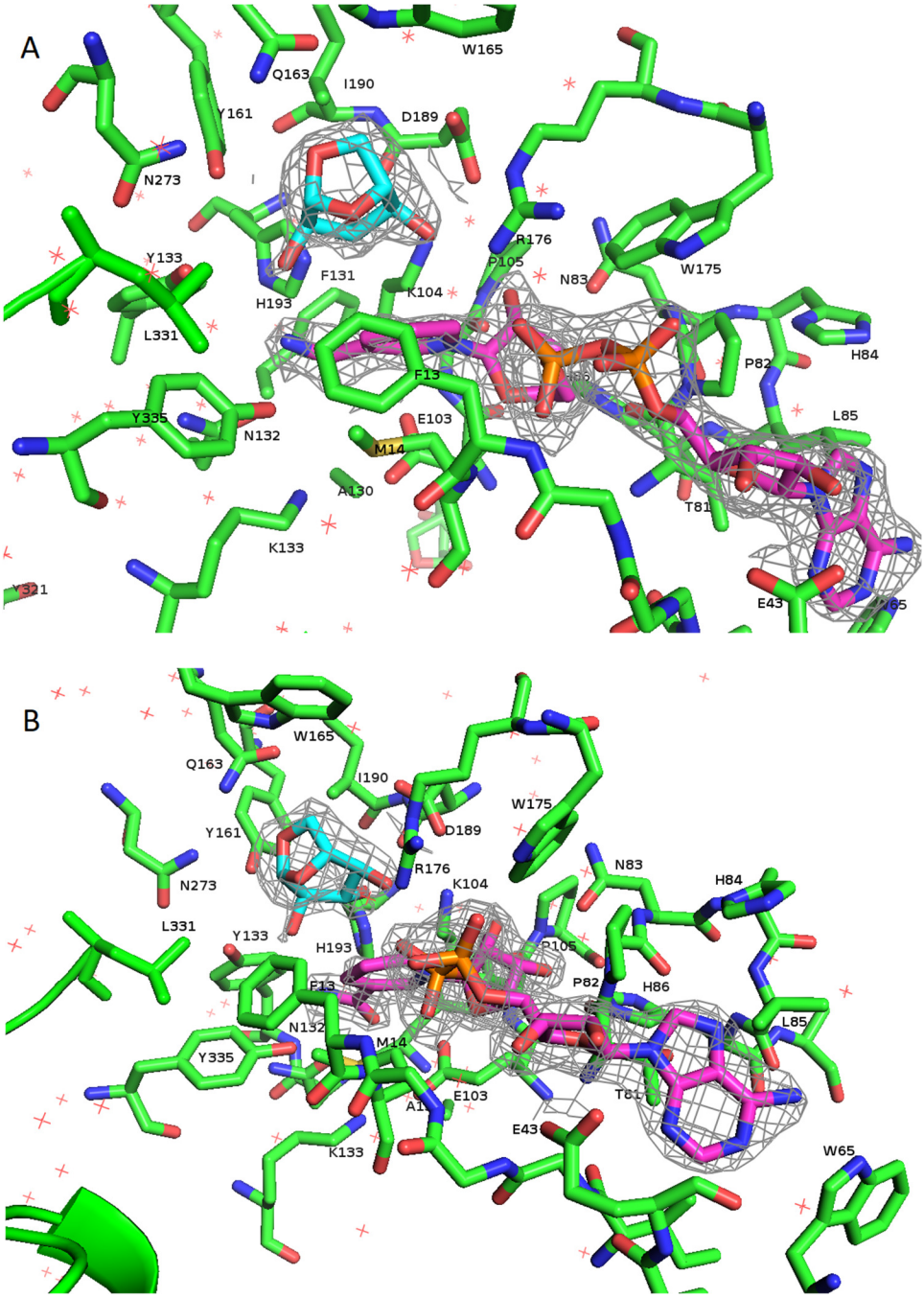


Figure 3-4. Levoglucosan and NADH as seen in the active site of LGDH. All residues with side chains potentially interacting with substrates are labelled. Electron density maps ($2F_o$-F_c) of the substrates are shown at 1 σ r.m.s.d contour levels. A and B different views of the same model with A, the side of the active site where levoglucosan binds, and B, the side of the active site near the adenine group of NADH.

Figure 3-5. Alignment of known LGDHs with active site residues highlighted.

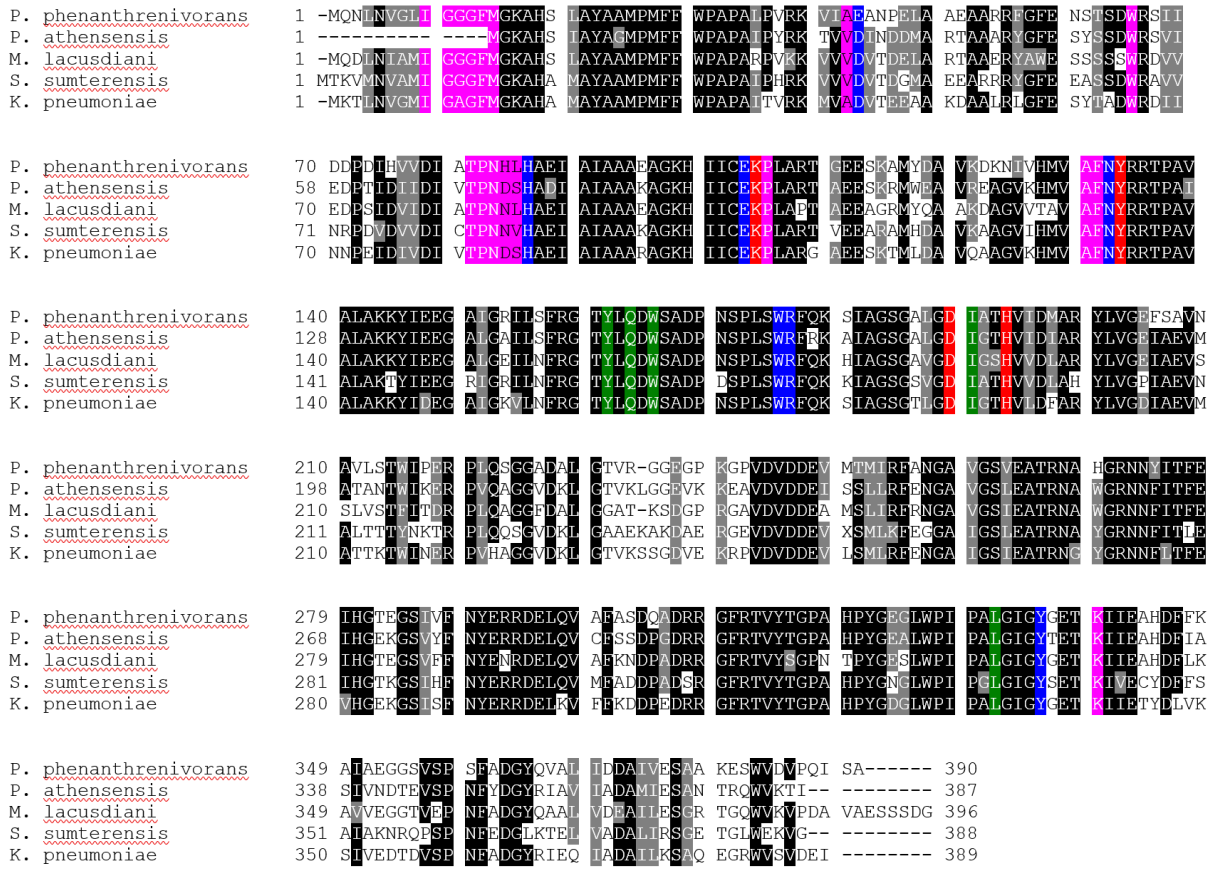


Fig. 2-5, with active site residues highlighted. Black highlighting indicates a residue conserved in 80% of the sequences, grey highlighting indicates similar residues. Active site residues were defined as those within 5 Å of either substrate in the *P. phenanthrenivorans* Sphe3 LGDH. Residues highlighted in red hydrogen bond with carbohydrates in the active site or make a van der Waal contact, green residues are near the carbohydrate, blue residues hydrogen bond NAD(H), and pink residues are near NAD(H).

Carbohydrate Interactions with the Active Site

The ketosugars investigated bind to the LGDH active site by hydrogen bonding to Tyr133, His193, Lys104, Asp189, and Arg176. Differences in the observed electron density for the substrates occurred as a result of the presence of NADH in the active site. Levoglucosan is sharply defined in the active site of crystals soaked without NADH (Fig. 3-6A). Hydrogen bonds are observed between hydroxyl-2 of levoglucosan and Tyr133, hydroxyl-3 and His193 and Lys104, and hydroxyl-4 with Asp189 and Arg176. In crystals soaked with NADH (Fig. 3-6B), the electron density of levoglucosan is not as well defined, being more spherical on the side of the diether ring of the sugar. The presence of NADH similarly seemed to alter the interactions between glucose and the active site. One crystal soaked in 30% glucose and ~1.9 nM NADH produced a model with one LGDH monomer containing no NADH (Fig 3-6C), and the other LGDH monomer containing NADH (Fig 3-6D). The well-defined electron density of glucose in the active site without NADH shows glucose in a conformation setting hydroxyl-3 within 3.0 Å of His193, the residue likely to facilitate the hydride transfer from hydroxyl-3 to NAD^+. In contrast, the electron density is not clearly defined in the active site containing NADH, but it appears hydroxyl-4 hydrogen bonds with His193. Binding of ketosugars in multiple conformations to the active site would explain the relatively high K_M for levoglucosan of 14.0 mM (Nakahara et al., 1994) to 27.0 mM (Chapter 2), and 5% relative activity on glucose (Nakahara et al., 1994). The electron density of levoglucosan in the active site with NADH suggests levoglucosan may be interacting with the active site in two conformations, the conformation observed without NADH present in which hydroxyl-2 and Tyr133 hydrogen bond (Fig. 3-7A), or the alternative with levoglucosan flipped 180° along hydroxyl-3 causing

hydrogen bonds to form between hydroxyl-4 and Tyr133 as well as hydroxyl-2 and Asp189 and Arg176 (Fig. 3-7B).

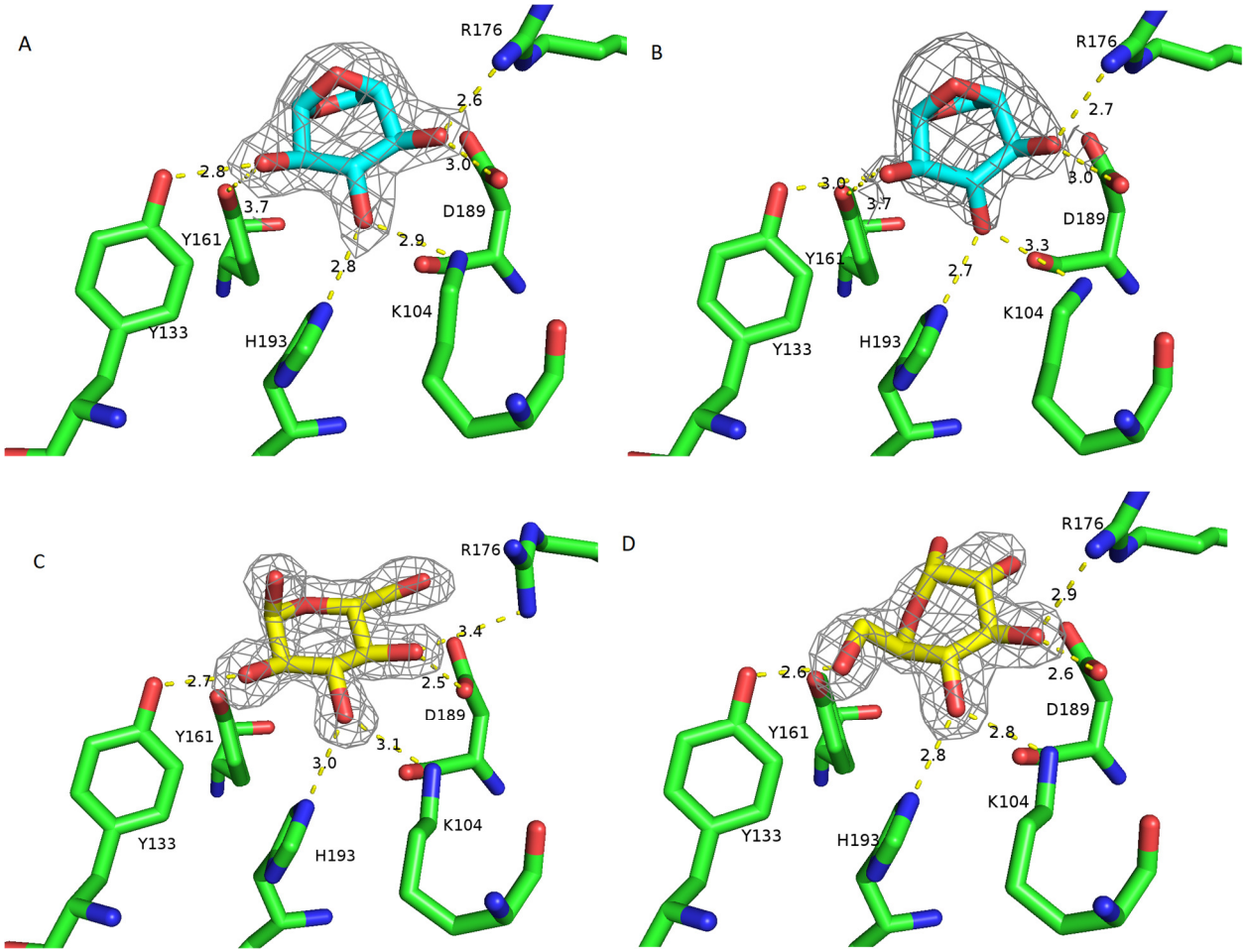


Figure 3-6.

A, the expected conformation for conversion of levoglucosan to 3-keto levoglucosan in which hydroxyl-3 is directed down to His193, leaving the proton on C3 on the side of the carbohydrate near NADH (not pictured). B, an alternative conformation of levoglucosan in the active site in which hydroxyls 2, 3, and 4 are still able to hydrogen bond with Tyr161, His193, Lys104, Asp189, and Arg176, however the proton on C3 is now facing away from NADH (not pictured).

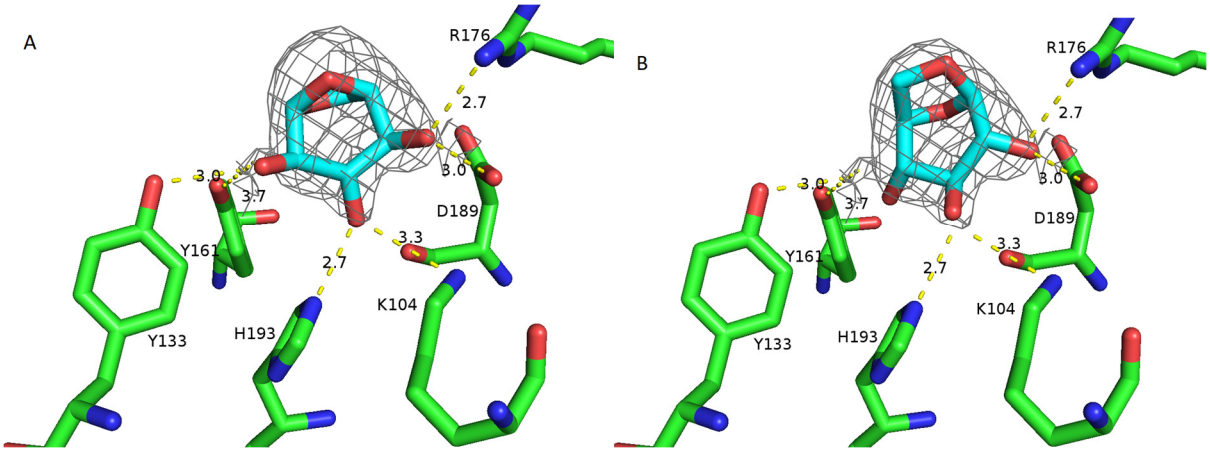


Figure 3-7. Two conformations of levoglucosan in the active site of LGDH in the presence of NADH. A, the expected conformation for conversion of levoglucosan to 3-keto levoglucosan in which hydroxyl-3 is directed down to His193, leaving the proton on C3 on the side of the carbohydrate near NADH (not pictured). B, an alternative conformation of levoglucosan in the active site in which hydroxyls 2, 3, and 4 are still able to hydrogen bond with Tyr161, His193, Lys104, Asp189, and Arg176, however the proton on C3 is now facing away from NADH (not pictured).

NAD(H) Interactions with the Active Site

NADH electron density was well defined in all structures containing the cofactor. Hydrogen bonding likely occurs between NAD(H) and Glu43 and the two hydroxyls of the ribose adjacent to adenine, His86 and a hydroxyl group of the adenine adjacent to the nicotinamide ring, Glu103 and the nicotinamide amide, Asn132 and the nicotinamide amide, Trp175 and a phosphate group, Arg176 and the hydroxyl group of ribose closest to the nicotinamide group, and Tyr335 and the nicotinamide amide group (Fig. 3-4A and B).

Residues 222 through 234 are a Poorly Defined Loop Region

The loop between β-strand 8 and β-strand 9, Glu222 through Gly234, consistently had poorly defined electron density. These residues were excluded from the final refinement of the models, and any discernable electron density was filled with waters. However, one can speculate on how the residues connect β-strand 8 and β-strand 9 in each model (Fig. 3-8). Glu222 through Gly234 likely comprise a mobile loop that closes down on NAD(H) and possibly levoglucosan, closing the active site and blocking access to the active site from water molecules. Interestingly, in an alignment of all the known LGDHs, this loop is one region of the protein sequence that shows relatively low levels of identity compared to the rest of the sequence. Perhaps the movement of this loop, opening and closing the active site, explains the difference in reaction rates observed between the LGDH described in Chapter 2.

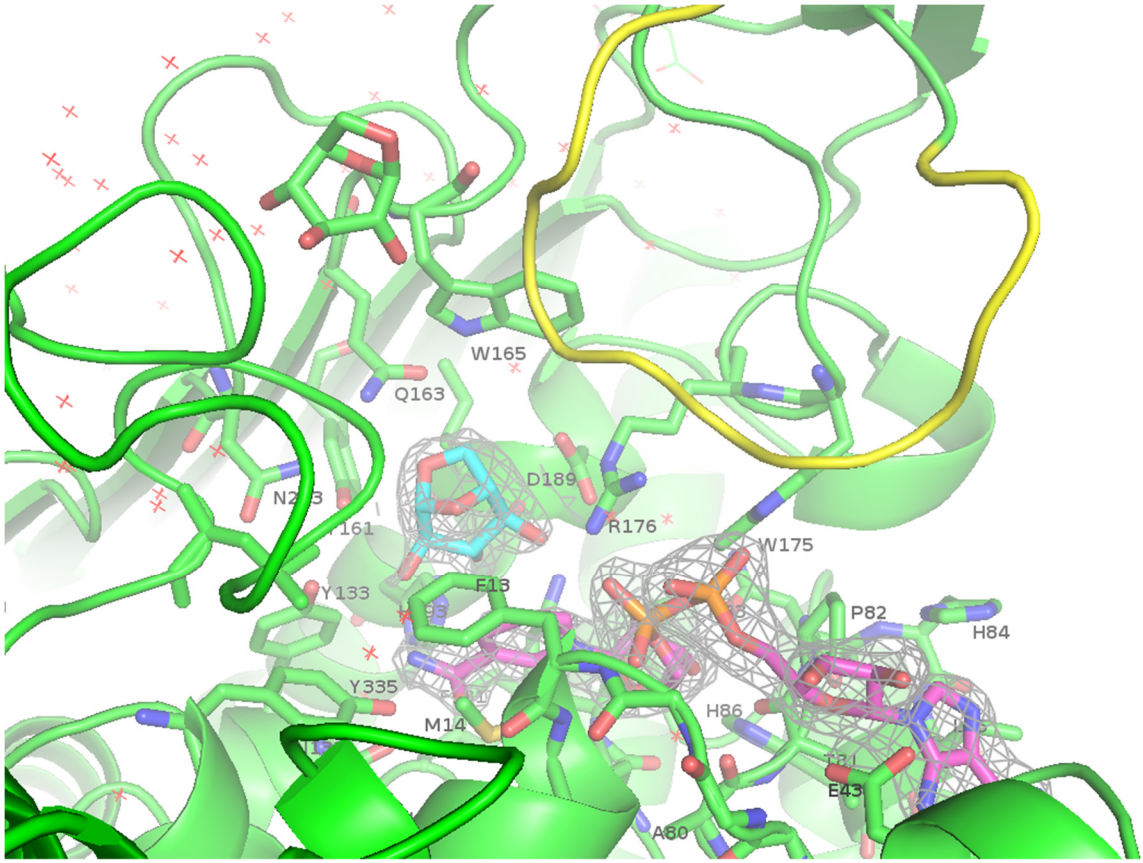


Figure 3-8. One possible conformation of loop Glu222 through Gly234 along the active site, closing on NAD(H) and levoglucosan. The disordered loop (yellow) is pictured with no side chains shown because they were not predicted to form hydrogen bonds with substrates in the models.

Mechanistic Implications

Active site residues His193, Lys104, and Asp189 are likely to function as a catalytic triad as seen in myo-inositol dehydrogenase (van Straaten et al., 2010), 1,5-anhydrofructose reductase (Dambe et al., 2006; Kuhn et al., 2006), and glucose-fructose oxidoreductase, which uses Tyr217 instead of His (Kingston et al., 1996; Nurizzo et al., 2001). With the myo-inositol dehydrogenase of *Bacillus subtilis* (IDH), the mutations H176A and D172N showed 250 and 30-fold reduction in k_{cat}, respectively (Daniellou et al., 2007). Spatially, the arrangement of the three residues is nearly identical (Fig. 3-9), with the distance from His to Lys being 4.0 Å in LGDH (4.1 Å in IDH), close enough to transfer a proton, which may then be transferred from Lys to the Asp 3.4 Å away (4.3 Å in IDH). The mechanism of these residues is pictured in Figure 3-10.

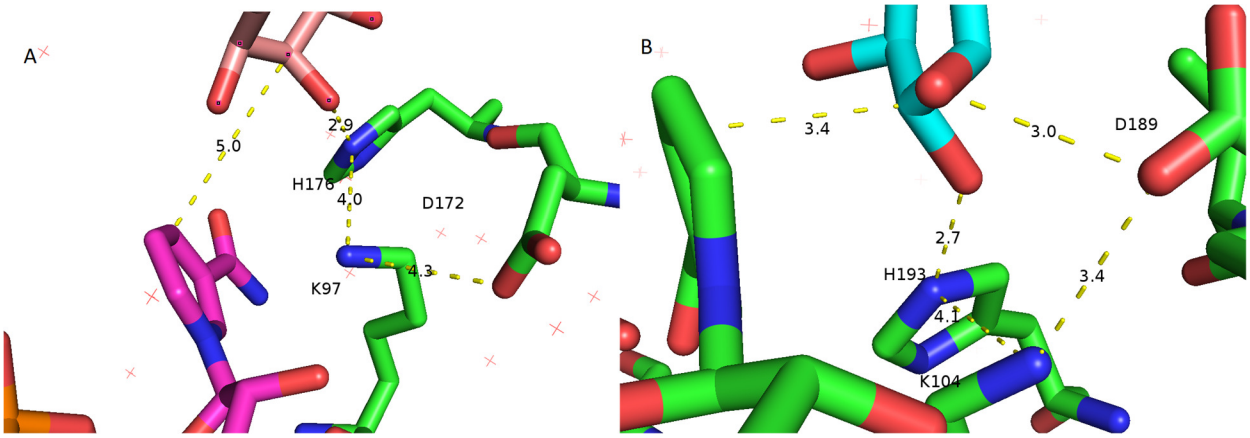
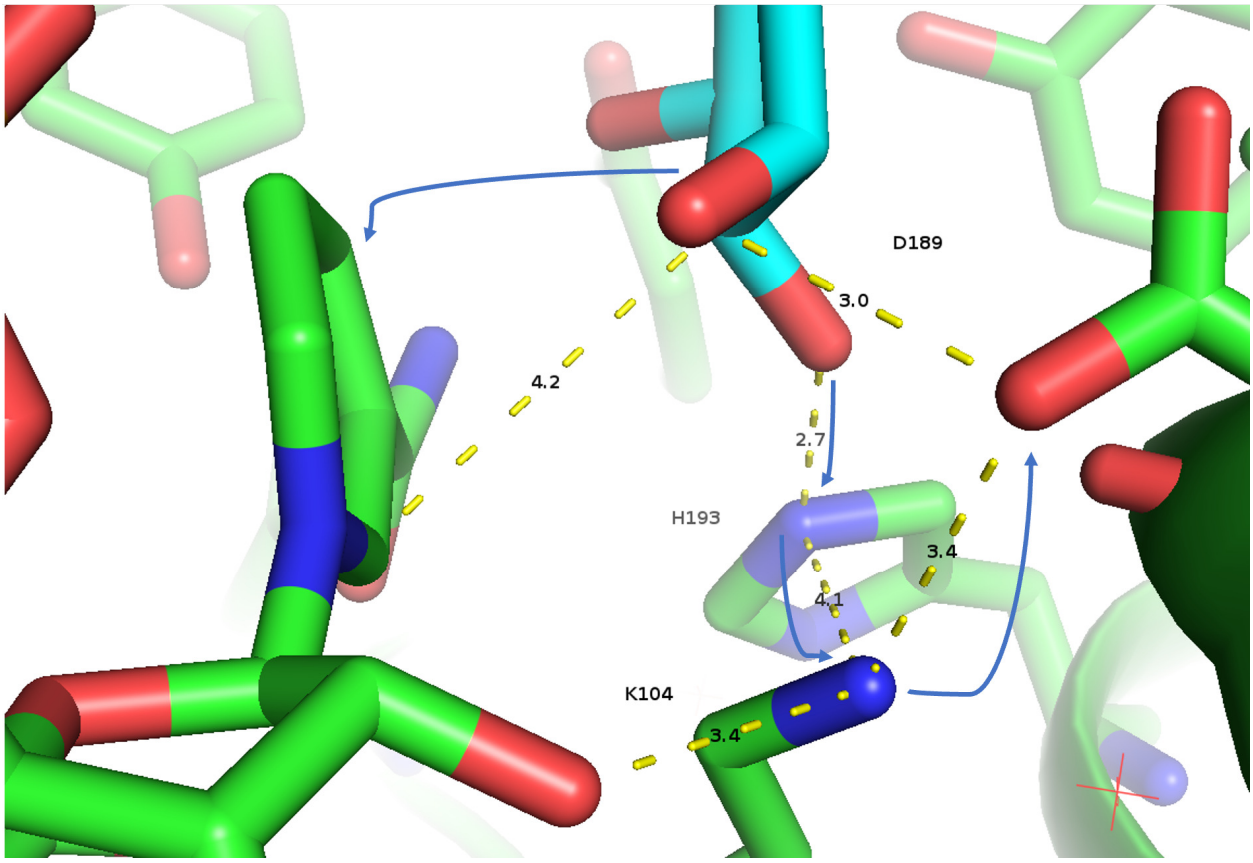


Figure 3-9. The His-Lys-Asp catalytic triad shown with NADH and a sugar substrate. A, IDH active site with myo-inositol at the top, NADH on the left. B, LGDH active site with levoglucosan at the top, NADH on the left.

Figure 3-10. The hypothesized catalytic mechanism of LGDH.



Catalysis is initiated by proton transfer from hydroxyl-3 of levoglucosan to His193, along with a hydride shift from C3 of levoglucosan to C4 of the nicotinamide ring in NAD^+. Lys104 and Asp189 likely function as a proton relay to shuttle the proton bound to His193.

Materials and Methods

Protein expression and purification

The sequence of the *P. phenanthrenivorans* Sphe3 gene was cloned into a pET28b vector with a C-terminal polyhistidine tag as previously described in Chapter 1 (Bose et al., 2002). The plasmid was then transformed into *E. coli* BL21(DE3) RIPL cells (Stratagene, San Diego, CA) that were plated on LB agar plates containing 40 µg/mL kanamycin. After overnight growth, ~5-10 colonies were picked and used to inoculate 5 mL of terrific broth (12 g/L tryptone, 24 g/L yeast extract, 4 mL glycerol/L, 0.017 M KH_2PO_4, 0.072 M K_2HPO_4) containing 40 µg/mL kanamycin and allowed to grow for 2-3 h before being used to inoculate 500 mL PA-5052 autoinduction medium (Studier, 2005). The autoinduction culture was grown for 20-24 h at room temperature before cells were harvested by centrifugation (8000 × g for 20 min at 4°C), then resuspended in a buffer containing 50 mM Tris, 500 mM NaCl, 25 mM imidazole, 5 mM β-mercaptoethanol, pH 8.4. Cells were lysed by French pressure Cell (vendor, location) three times at 16000 psi, then centrifugation (60,000 × g for 20 min at 4°C) to remove the cell debris. The supernatant was loaded into the superloop of an ÄKTA Purifier and loaded onto a 5 mL HisTrap HP nickel-chelate column (GE Life Sciences, Marlborough, MA) equilibrated with the binding buffer (50 mM Tris, 500 mM NaCl, 25 mM imidazole, 5 mM β-mercaptoethanol, pH 8.4). The column was washed with 10 column volumes of binding buffer before bound LGDH was eluted with a linear gradient from 25 to 500 mM imidazole. LGDH eluted at about 75-150 mM imidazole. The purified protein was then dialyzed in a buffer containing 50 mM Tris, 250 mM NaCl, 5 mM β-mercaptoethanol, pH 8.4 for at least 4 h, then transferred to a final dialysis buffer containing no NaCl. After dialysis was complete, the protein was loaded onto a HiTrap Q HP anion exchange column (GE Life Sciences, Marlborough, MA) equilibrated with 50 mM

Tris, 5 mM β-mercaptoethanol, pH 8.4. The protein was eluted with a linear gradient from 0 to 500 mM NaCl and LGDH eluted at about 200-275 mM NaCl. Purified LGDH was dialyzed against 50 mM Tris, 5 mM β-mercaptoethanol, pH 8.4 and concentrated to 15.7 mg/mL with a Vivaspin Turbo 15 centrifugal concentrator (Sartorius Stedim Biotech GmbH, Goettingen, Germany).

Crystallization and data collection

P. phenanthrenivorans LGDH was crystallized using the microbatch technique. 96 well plates were filled with Al's Oil (Hampton Research, Aliso Viejo, CA) then 3 µL of the concentrated LGDH solution were pipetted into each well. Crystallization screens were performed using the following commercially available kits from Hampton Research: Crystal Screen I, Crystal Screen II, Index, Salt Rx1, and Salt Rx2. To obtain diffraction quality crystals, to each 3 µL drop of protein solution were added 2 µL of crystallization buffer (1.8 M tri-ammonium citrate pH 7.0 from Hampton Research Index Crystallization Kit) were mixed well by pipetting. The microbatch plates were stored at 23°C and checked every few days for crystal formation. The optimal bipyramidal crystals used in this study formed after 2 weeks of incubation, sometimes from wells that precipitated upon addition of the crystallization buffer.

Artificial mother liquors were formulated so concentrated crystallization buffer could be added to protein buffer in a 4:1 ratio, resulting in a liquor containing the crystallization buffer at a slightly greater concentration than initial crystallization conditions upon mixing the micro-batch wells. When it was desired to see the interactions between LGDH and a small carbohydrate, the sugar would be added to the concentrated crystallization buffer at 37.5% w/v, so its final concentration in the artificial mother liquor would be 30%, and the carbohydrate would serve as the cryoprotectant. To soak crystals with NADH, the artificial mother liquor

required an increased pH of 8.6 for NADH stability. A concentrated crystallization buffer contained levoglucosan with 1.9 nM NADH. Crystals were soaked overnight at 23 °C before freezing. All data sets were collected on the 22BM beam line of Argonne National Laboratory (Lemont, IL). Diffraction data were processed and scaled using HKL2000 (Otwinowski and Minor, 1997). The 3 structures were collected at 1.8 to 2.2 Å resolution.

Structure determination and refinement

The initial phase of *P. phenanthrenivorans* LGDH was determined by the molecular replacement method with PHASER (McCoy et al., 2007) in the CCP4 program suite (Winn et al., 2011). The structure of PDB ID:4H3V, an uncharacterized oxidoreductase from *Kribbella flavida* (Michalska et al., 2012) was the most similar protein in PDB at the time (37% sequence identity, E-value: 1.84E-64) and was used as a molecular replacement search molecule. Manual model building was carried out using Coot (Emsley et al., 2010) and refinements were performed with Phenix (Adams et al., 2010). Initially, the model was subjected to real-space refinement, in addition to refinement of individual B-factors and occupancies. Once the structure secondary structure of both LGDH chains in the unit cell fit the $2(F_0-F_C)$ map, real space refinement was longer used, further refinement was carried out using group B-factors and TLS parameters with TLS groups defined using the TLSMD server (Painter and Merritt, 2006). Images were prepared using PyMOL (Schrodinger, LLC, New York).

References

P. V. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* D66:213-221.

N. Bose, P. Greenspan, C. Momany. 2002. Expression of recombinant human betaine: homocysteine S-methyltransferase for X-ray crystallographic studies and further characterization of interaction with S-adenosylmethionine. *Protein Expr. Purif.* 25(1):73-80.

R. Daniellou, H. Y. Zheng, D. M. Langill, D. A. R. Sanders, D. R. J. Palmer. 2007. Probing the promiscuous active site of myo-inositol dehydrogenase using synthetic substrates, homology modeling, and active site modification. *Biochemistry*. 46:7369-7477.

T. R. Dambe, A. M. Kuhn, T. Brossette, F. Giffhorn, A. J. Scheidig. 2006. Crystal structure of NADP(H)-dependent 1,5-anhydro-D-fructose reductase from *Sinorhizobium morelense* at 2.2 Å resolution: construction of a NADH-accepting mutant and its application in rare sugar synthesis. *Biochemistry*. 45:10030-10042.

P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 66:486–501.

R. L. Kingston, R. K. Scopes, E. N. Baker. 1996. The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure*. 4:1413-1428.

E. Krissinel and K. Henrick. 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372: 774-797.

A. Kuhn, S. K. Yu, F. Giffhorn. 2006. Catabolism of 1,5-anhydro-D-fructose in *Sinorhizobium morelense* S-30.7.5.: discovery, characterization, and overexpression of a new 1,5-anhydro-D-fructose reductase and its application in rare sugar analysis and rare sugar synthesis. *Appl. Environ. Microbiol.* 72:1248-1257.

K. Nakahara, Y. Kitamura, Y. Yamagishi, H. Shoun, T. Yasui. 1994. Levoglucosan dehydrogenase involved in the assimilation of levoglucosan in *Arthrobacter* sp. I-552. *Biosci. Biotech. Biochem.* 58(12):2193-2196.

D. Nurizzo, D. Halbig, G. A. Sprenger, E. N. Baker. 2001. Crystal structures of the precursor form of glucose-fructose oxidoreductase from fructose oxidoreductase from *Zymomonas mobilis* and its complexes with bound ligands. *Biochemistry*. 40:13857-13867.

A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read. 2007. Phaser crystallographic software. *J Appl Crystallogr.* 40:658-674.

K. Michalska, J. C. Mack, S. M. McKnight, M. Endres, A. Joachimiak, (Midwest Center for Structure Genomics). 2012. PDB ID: 4H3V Crystal structure of oxidoreductase domain protein from Kribella flavida.

Z. Otwinowski and W. Minor. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307-326.

J. Painter and E. A. Merritt. 2006. *TLSMD* web server for the generation of multi-group TLS models. *J Appl Cryst.* 39:109-111.

K. E. van Straaten, H. Zheng, D. R. J. Palmer, D. A. R. Sanders. 2010. Structural investigation of myo-inositol dehydrogenase from *Bacillus subtilis*:implications for catalytic mechanism and inositol dehydrogenase classification. *Biochem. J.* 432:237-247.

M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson. 2011. Overview of the *CCP4* suite and current developments. *Acta Crystallogr D Biol Crystallogr*. D67:235-242.

CHAPTER 4

Conclusions and Future Directions

Conclusions

This work demonstrates that *lgdA* and other genes likely to contribute to levoglucosan consumption are found throughout the bacterial domain, even though levoglucosan consumption is not a common growth characteristic (Table 2-8). *lgdA* has been experimentally identified in five organisms (Table 2-13): *P. phenanthrenivorans* Sphe3, *P. athensensis* MEC069^T, *M. lacusdiani* MEC084^T, *S. sumterensis* MEC087^T, and *K. pneumoniae* MEC097. These organisms are members of the Firmicutes, Actinobacteria, and Proteobacteria. The facts that these are distantly related organisms and that levoglucosan consumption has not been identified in many organisms more closely related to these bacteria, suggest that the genes for levoglucosan consumption have been shared by horizontal gene transfer. The contig containing *lgdA* in the genomic assembly of *K. pneumoniae* MEC097 contains an IS3 transposase ~4 kb downstream of the putative periplasmic sugar-binding protein of the ABC transporter found with *lgdA*. Based on these observations, the genes critical for levoglucosan consumption likely entered the *K. pneumoniae* genomes by horizontal gene transfer.

In addition to levoglucosan, fast pyrolysis of lignocellulosic materials produces a large fraction of cellobiosan (Lomax et al., 1991; Patwardhan et al., 2011). Cellobiosan is produced in yields as high as 30% of the total levoglucosan yield (Piskorz et al., 2000), so ideallya microbe growing on pyrolysis-derived sugars could consume both anhydrosugars. Bacteria have been isolated specifically for the ability to use cellobiosan as a sole carbon source (Lian et at., 2016). *Pseudomonas putida* KT2440 heterologously expressing *lgk* from *A. niger* is able to consume cellobiosan with any one of four β-glucosidases from *Thermotoga maritima, Agrobacterium* sp., *Phanaerochate chryosporium*, or *A. niger* added exogenously to growth medium (Linger et al., 2016). The previous work with *P. putida* KT2440 and the experiments on *K. pneumoniae* ATCC

13883^T and *K. pneumoniae* MEC097 grown on cellobiosan show that the presence of β-glucosidase activity enables microbes to consume the glucose subunit of cellobiosan, while the presence of LGDH and its downstream enzymes (or LGK) enables complete consumption of cellobiosan (Fig. 2-4). The isolation of bacteria solely on their ability to use cellobiosan as a sole carbon source and finding that they can also grow on levoglucosan as a sole carbon source suggests that the ability to grow on cellobiosan is a unique growth characteristic that is accompanied by the ability to consume levoglucosan (Lian et al., 2016). In contrast, all strains tested showed growth on cellobiosan even if they could not grow on levoglucosan (Table 2-8), suggesting that the β-glucosidase activity is able to cleave the β-1,4-glycosidic linkage in cellobiosan are in a variety of microbes. The cleavage of cellobiosan by exogenous β-glucosidases from *T. maritima, Agrobacterium* sp., *P. chryosporium*, or *A. niger* also imply that these β-glucosidases are promiscuous enzymes.

P. phenanthrenivorans LGDH has a tertiary structure typical of members of the GFO/IDH/MocA family (Fig. 3-1), known for using NAD(P)H to oxidize or reduce substrates. The N-terminus of these enzymes contain a Rossman-fold used to bind NAD(P)H, while the carbohydrate side of the substrate binding pocket is in the C-terminus. Similar to the catalytic mechanism of myo-inositol dehydrogenase from *Bacillus subtilis* (van Straaten et al., 2010), LGDH likely relies on a His-Lys-Asp catalytic triad to oxidize/reduce levoglucosan, with His being the catalytic base that initiates the conversion of sugar substrate, and the nearby Lys and Asp forming a proton relay.

Even though ~89% of active site residues are conserved across the identified LGDHs and the peptide sequence identity between any two is 66-75%, activity varies significantly, with K_M

for levoglucosan in the 1.7-27 mM range, K_M for NAD^+ of 0.3-9.1 mM, and kcat of 27-84 s^{-1}. LGDH with K_M for levoglucosan on the order of 2.0 mM also exhibit inhibition at [LG]≥ 93 mM.

Future Directions

The first logical step from this work is to characterize other genes involved in levoglucosan consumption in bacteria. Because *lgdB* and three components to an ABC sugar transporter were consistently observed to be with *lgdA* in the genomic assemblies of levoglucosan-utilizing strains, one hypothesis is that *lgdB* is the hydrolase that converts 3-keto levoglucosan to 3-keto glucose. This hypothesis could be supported *in vitro* by purifying recombinant *lgdA* and *lgdB* separately from *E. coli* BL21 (DE3)-RIPL, then combining the two enzymes in a reaction containing levoglucosan and an excess of NAD^+, followed by analysis of carbohydrates present by HPLC, GC-MS, or NMR spectroscopy. Such experiments should also test the efficacy of this enzyme mixture with different divalent cations added as all of the *lgdB* homologs identified require specific divalent cations for activity. Two *in vivo* methods are also plausible for testing this hypothesis: (A) plasmid-based expression of *lgdA* and other necessary genes in a strain that does not naturally consume levoglucosan (e.g., *E. coli* or *K. pneumoniae* ATCC 13883^T), or (B) making deletions in a levoglucosan consuming strain and identifying which knockouts prevent growth on levoglucosan. Cloning the *lgdA* and surrounding genes from *K. pneuominae* MEC097 to a plasmid, then transforming *K. pneumoniae* ATCC 13883^T with that *lgdA*-containing vector may be the simplest method as the two microbes are the same species and will likely translate the DNA in the same way. Alternatively, genomic deletions can be made in *K. pneumoniae* MEC097 using the λ Red recombinase system (Datsenko and Wanner, 2000; Jung et al., 2012).

Another step is verification of the function of the putative ABC transporter proteins shown to be conserved with *lgdA* and *lgdB*. If any of those gene products are an ATP-binding cassette

transporter specific for levoglucosan, they may be useful in other strains being engineered for growth on levoglucosan, whether they are using LGK or LGDH.

To complete the description of LGDH activity, the enzyme mechanism should be probed with mutations of the active site residues, particularly His193, Lys104, and Asp189. When Sugiura published their crystallographic structures of LGDH (Sugiura et al., 2018) I was trying to purify the same enzyme with the mutations H193F, K104M, and D189N and found that H193F and K104M were insoluble. Simple substitutions with alanine may be more soluble and show the contribution of each of these residues towards catalysis.

One final direction would be to engineer *K. pneumoniae* MEC097 for the conversion of pyrolysis derived anhydrosugars to 2,3-butanediol. *Klebsiella* are known for producing industrially significant amounts of 2,3-butanediol (Ji et al., 2011). The primary concern when working *K. pneumoniae* is that it is an opportunistic pathogen, known to cause bacterial pneumonia (Carpenter, 1990), and nosocomial urinary tract infections. A protocol to make avirulent *K. pneumoniae* by knocking out *wabG*, a gene involved lipopolysaccharide production, details how to use the λ Red recombinase system (Datsenko and Wanner, 2000) to delete the gene (Jung et al., 2012). Furthermore, Jung describes the use of *K. pneumoniae* 2242Δ*wabG* to produce 2,3-butanediol from glucose. *K. pneumoniae* MEC097 was the fastest growing isolate on levoglucosan (Table 2-3). If *K. pneumoniae* MEC097 was made avirulent, it would be an ideal strain of conversion of pyrolysis-derived anhydrosugars to 2,3-butanediol.

References

J. L. Carpenter. 1990. Klebsiella pulmonary infections: occurence at one medical center and review. *Rev Infect Dis* 12:672–682.

K. A. Datsenko and B. L. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci*. 97(12):6640-5.

X. J. Ji, H. Huang, P. K. Ouyang. (2011) Microbial 2,3-butanediol production: a state-of-the-art review. *Biotechnol Adv.* 29:351–364.

S-G. Jung, J-H. Jang, A-Y. Kim, M-C. Lim, B. Kim, J. Lee, Y-R Kim. 2012. Removal of pathogenic factors from 2,3-butanediol-producing *Klebsiella* species by inactivating virulence-related *wabG* gene. *Appl Microbiol Biotechnol.* 97:1997-2007.

P. R. Patwardhan, D. L. Dalluge, B. H. Shanks, R. C. Brown. 2011. Distinguishing primary and secondary reactions of cellulose pyrolysis. *Bioresourc. Technol.* 102(8):5265-5269.

J. Piskorz, P. Majerski, D. Radlein, A. Vladars-Usas, D. S. Scott. 2000. Flash pyrolysis of cellulose for production of anhydro-oligomers. *J. Anal. Appl. Pyrol.* 56:145-166.