ALTERNATIVE FUNCTIONS FOR THE
BETA-GALACTOSIDASES OF MICROORGANISMS
FOUND IN ENVIRONMENTS WITHOUT LACTOSE

A Thesis in
Biochemistry, Microbiology, and Molecular Biology

by

Stephanie Ann Shipkowski

Copyright 2006 Stephanie Ann Shipkowski

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2006
The thesis of Stephanie Ann Shipkowski was reviewed and approved* by the following:

Jean E. Brenchley  
Professor of Microbiology/Biotechnology  
Thesis Advisor  
Chair of Committee  

Mary Ann Bruns  
Associate Professor of Soil Science/Microbial Ecology  

J. Gregory Ferry  
Professor of Biochemistry and Molecular Biology  

Allen T. Phillips  
Professor of Biochemistry  

Ming Tien  
Professor of Biochemistry  

Robert A. Schlegel  
Professor of Biochemistry and Molecular Biology  
Head of the Department of Biochemistry and Molecular Biology  

*Signatures are on file in the Graduate School.
ABSTRACT

Many microorganisms possess genes encoding glycoside hydrolase enzymes that allow them to catabolize carbohydrates present in their environments. Among the glycoside hydrolases, β-galactosidases are typically considered to function as lactases. However, many β-galactosidase producing microorganisms exclusively occupy habitats such as soil and water where the disaccharide lactose is not available. This suggests that these β-galactosidases hydrolyze substrates other than lactose and have new, unknown functions. My research objective was to examine possible natural functions of different glycosyl hydrolases with β-galactosidase activity in a group of phylogenetically related bacteria. I obtained hundreds of isolates by enriching for psychrophilic spore-forming organisms, which I then screened for β-galactosidase production. Genes encoding β-galactosidase activities were cloned, sequenced, and the encoded enzyme activities examined. My examination of these isolates, their genes encoding β-galactosidase activities, and the patterns of occurrence of these genes in fully sequenced genomes, led me to hypothesize that compounds from plants are the in vivo substrates.

One gene of special interest belonged to a glycoside hydrolase family (GHF) not known for having β-galactosidase activity (GHF 3). Because of this unique placement, the enzyme was purified and characterized. This enzyme also had β-glucosidase activity, a low thermal optimum, and was most active on aryl-glucosides. Although some bacterial enzymes with aryl-glucosidase activity are catabolic, others have been demonstrated to have roles in signaling or saprophytic interactions with plants through their actions on specific secondary metabolites from plants. Analysis of this β-glucosidase was published in Applied and Environmental Microbiology (Shipkowski, S. and J.E. Brenchley. 2005. Characterization of an unusual cold-active β-glucosidase belonging to Family 3 of the glycoside hydrolases from the psychrophilic isolate Paenibacillus sp. strain C7. 71(8): 4225-4232). During this work, examinations of more typical β-glucosidase (non-aryl) substrates such as cellobiose suggested analogous galactosidic compounds from plants could be substrates for β-galactosidases instead of lactose.

With these substrates in mind, I focused on the GHF 42 β-galactosidases because my interpretation of previous results relevant to this family did not provide evidence for lactose hydrolysis as an in vivo function. Therefore, I analyzed existing GHF 42 gene sequences and used conserved regions indicated by the alignments to design a PCR primer pair specific for this group of β-galactosidases, and demonstrated their utility using control templates. Next, I used the primers to screen the genomic DNA of bacterial isolates, and then screen the plasmid DNA of β-galactosidase-expressing genomic library transformants from these isolates. In addition to obtaining additional GHF 42 gene
sequences, my observations during the sequence alignments and analyses, led me to examine relevant genomic data. I gathered data for GHF 42 genes and adjacent sequence, organized the arrangements so that patterns were discernable, and observed a homologous gene arrangement shared by several organisms. The probable enzyme functions of genes located near genes of the GHF 42 group led to the hypothesis that the natural substrate for some of these β-galactosidases might be oligosaccharides produced by the degradation of the pectic plant polysaccharide, arabinogalactan type-I. I proposed a degradation pathway for this polysaccharide involving the functions of the additional proteins encoded by adjacent genes.

From the microorganisms with genomes containing the conserved gene arrangement, I selected *Bacillus subtilis* as a model for testing my hypothesis. *B. subtilis* has two GHF 42 genes. One, *lacA*, is in a putative polycistronic operon with other genes including a putative galactanase encoded by *galA*, whereas the second *yesZ*, is in a different gene arrangement. Because there is a genetic system for *B. subtilis*, it could be manipulated to test my prediction that the *lacA* gene could encode an enzyme that would hydrolyze the arabinogalactan type-I product yielded by *galA* galactanase activity. I first demonstrated that the addition of arabinogalactan type-I to *B. subtilis* cells increased the β-galactosidase activity more than the presence of many other sugars (including lactose) and plant polysaccharides. I also showed that *B. subtilis* grew on this polysaccharide as a sole carbon source, whereas *Escherichia coli* showed very little growth on galactan. To further clarify the role of the *lacA* β-galactosidase gene in the degradation of arabinogalactan type-I, it and the *galA* gene were cloned and expressed in (*lacZ*) *E. coli*; the combination of these genes allowed growth on galactan. Additionally, mutants were created in *B. subtilis* where independently the β-galactosidase genes (*lacA* and *yesZ*) or *galA* were interrupted by the insertion of a chloramphenicol resistance gene. The *galA::Cm^R* and *lacA::Cm^R* mutants no longer hydrolyzed X-Gal in the presence of arabinogalactan type-I and had a decreased ability to grow on this substrate. This supports the hypothesis that the oligomers of arabinogalactan type-I produced by the GalA GHF 53 enzyme are relevant natural substrates for this GHF 42 β-galactosidase enzyme. This is the first *in vivo* evidence for a reasonable function for the GHF 42 β-galactosidases, and allows us to start understanding the role they play in the environment. Similar functions for other GHF 42, as well as certain GHF 2, β-galactosidases are also suggested by genomic data. The evidence that some GHF 42 enzymes have not evolved for the function of lactose hydrolysis may limit their commercial use in the dairy industry, but opens opportunities involving modifications to the pectic substances found in the many plant materials we use.
TABLE OF CONTENTS

List of Figures viii
List of Tables x
List of Frequently Used Abbreviations xi
Acknowledgements xii

Chapter 1. Introduction 1
  1.1 Overview 2
  1.2 Reasons to investigate β-galactosidases 2
  1.3 Classification of glycoside hydrolases 4
  1.4 Known functions of β-galactosidases 7
  1.5 Dissertation organization 9
  1.6 References 15

Chapter 2. Characterization and isolation of psychrophilic bacteria belonging to Bacillales and examination of their β-galactosidases 16
  2.1 Summary 17
  2.2 Introduction 18
  2.3 Results 22
    2.3.1 Characterization of isolates 22
    2.3.2 Thermal dependency of β-galactosidase activities 24
    2.3.3 Analysis of the cloned β-galactosidase genes and adjacent sequence 25
  2.4 Discussion 28
    2.4.1 Characterization of isolates 28
    2.4.2 Thermal optima of β-galactosidases 29
    2.4.3 Analysis of the cloned β-galactosidase genes and adjacent sequence 30
  2.5 Materials and Methods 33
  2.6 References 37
Chapter 3. Characterization of an unusual cold-active β-glucosidase belonging to family 3 of the glycoside hydrolases from the psychrophilic isolate Paenibacillus sp. C7

3.1 Summary 42
3.2 Introduction 44
3.3 Results 46
  3.3.1 Characterization of the C7 isolate 46
  3.3.2 Cloning of a gene encoding β-galactosidase activity 48
  3.3.3 Analysis of the bglY gene 48
  3.3.4 Analyses of bglY and neighboring sequence regions 49
  3.3.5 Enzyme purification 51
  3.3.6 Effects of temperature and pH on activity 52
  3.3.7 Effects of metal ions on activity 55
  3.3.8 Substrate preference studies 56
  3.3.9 Kinetic studies 58
3.4 Discussion 59
  3.4.1 Characterization of C7 and BglY and comparison to other GHF 3 enzymes 59
  3.4.2 Possible functions of BglY 61
3.5 Materials and Methods 64
3.6 References 71

Chapter 4. β-galactosidases from Glycoside Hydrolase Family 42: Biochemical and physiological perspectives, potential substrates, and detection via specifically designed primers

4.1 Summary 76
4.2 Introduction 78
4.3 Results 80
  4.3.1 Literature Review: Biochemical characteristics of GHF 42 80
  4.3.2 Literature Review: Data regarding lactose as a substrate for GHF 42 enzymes 84
  4.3.3 Primer design 88
  4.3.4 Primer testing within vector background 90
  4.3.5 Primer testing within genomic DNA background 91
  4.3.6 Use of primers to screen genomic DNA from isolates 92
  4.3.7 Phylogeny of isolates 93
  4.3.8 Analysis of heterologously expressed GHF 42 β-galactosidases 96
4.4 Discussion 98
  4.4.1 Possible substrates for GHF 42 enzymes 98
  4.4.2 Design and testing of GHF 42-specific primers 98
  4.4.3 Phylogeny of isolates identified as possessing GHF 42 genes via PCR screen 100
  4.4.4 Analysis of heterologously expressed GHF 42 β-galactosidases. 101
4.5 Materials and Methods 103
4.6 References 108
Chapter 5.  β-galactosidases from GHF 42: Ecological perspectives and potential substrates

5.1 Summary 114
5.2 Introduction 115
5.3 Results and Discussion 116
  5.3.1 GHF 42 enzymes: distribution and phylogenetic relationships 116
  5.3.2 Examination of GHF 42 gene arrangements 120
  5.3.3 GHF 53 associations with GHF 42 and GHF 2 129
  5.3.4 GHF 53 and β-galactosidase synergy 131
  5.3.5 Galactan-galactosidase relationships in literature 134
  5.3.6 Other possible functions 136
5.4 Conclusion 138
5.5 Materials and methods 140
5.6 References 142

Chapter 6. A natural function for a Glycoside Hydrolase Family 42 β-galactosidase of Bacillus subtilis

6.1 Summary 148
6.2 Introduction 150
6.3 Results 153
  6.3.1 Production of β-galactosidase activity in B. subtilis 153
  6.3.2 Construction of vectors and knockouts 154
  6.3.3 Physiological effects on E. coli 155
  6.3.4 Creation of B. subtilis mutants 156
  6.3.5 Comparison of wild type and mutants 157
  6.3.6 LacA β-galactosidase purification 159
  6.3.7 LacA β-galactosidase characterization 160
6.4 Discussion 163
6.5 Materials and methods 168
6.6 References 175

Chapter 7. Summary 177
**List of Figures**

1-1. Gene arrangement for lactose utilization in *E. coli*. 8

1-2. Classification of $\beta$-glucosidase substrates as a model for potential $\beta$-galactosidase substrates. 12

2-1. Expected genera of isolates resulting from the enrichment process. 21

2-2. Motile colony pattern formation by isolate *Sporosarcina* sp. CRE9. 22

2-3. Homology of $\beta$-galactosidases and adjacent ORFs. 27

3-1. Phylogenetic relationships of 16S rDNA sequences of isolate C7 and related *Paenibacillus* spp. 47

3-2. Features of the insert cloned from isolate C7. 50

3-3. Thermal dependency of activity of the purified BglY enzyme. 53

3-4. Thermostability of purified BglY. 53

3-5. Effects of pH on ONPGlu hydrolysis by purified BglY. 54

3-6. Structures for some of the substrates used in assays with purified BglY. 57

4-1. Sources for theoretical $\beta$-galactosidase substrates. 88

4-2. Ribbon model structure showing locations of primer encoded regions. 89

4-3. Primer guided process for cloning GHF 42 genes. 93

4-4. Composite 16S rDNA phylogenetic tree showing relationship of isolates with known species. 95

5-1. Habitat of microorganisms with sequenced genomes with GHF 42 genes. 118

5-2. Phylogenetic relationships between GHF 42 enzymes. 122

5-3. Gene arrangements of GHF 42 enzymes overlaid on phylogenetic tree. 125

5-4. Gene arrangements containing GHF 42 and GHF 53. 128
5-5. Chemical structures of proposed and known β-galactosidase substrates. 128
5-6. Phylogenetic relationships between GHF 53 enzymes. 130
5-7. Hypothesized functionality suggested by GHF 42/53 gene arrangements. 132
6-1. Plasmid inserts from *B. subtilis* genome. 154
6-2. Effects of pH on ONPG hydrolysis by purified LacA. 160
6-3. Thermal dependence of activity of purified LacA. 161
6-4. Thermostability of purified LacA. 161
List of Tables

1-1. Activities and distributions of several relevant GHFs. 6
2-1. Characteristics of some X-Gal hydrolyzing isolates. 23
2-2. Greenland ice-core isolates and their β-galactosidase thermal optima. 23
2-3. Attributes of heterologously expressed β-galactosidases. 24
3-1. Purification scheme for (His-tagged) BglY from E. coli. 51
3-2. Relative activity of BglY on various chromogenic substrates. 52
3-3. Effects of ions on the activity of EDTA-treated BglY enzyme. 55
3-4. Relative activity of BglY on various aryl substrates. 57
3-5. Kinetic values determined for purified BglY with different substrates. 58
4-1. Biochemical properties of characterized GHF 42 enzymes. 81
4-2. GHF 42 enzyme substrate and kinetic data. 83
4-3. PCR Experiments and controls. 91
4-4. Selected GHF 42 positive isolates and related organisms. 94
4-5. Transformants from GHF 42 positive organisms. 97
5-1. Inconsistent distribution of GHF 42 within genera. 119
6-1. Expected effects of mutations on the proposed pathway. 152
6-2. Some polysaccharides tested for upregulation of β-galactosidase activity. 153
6-3. Physiological effects of LacZ, LacA, GalA and YesZ expression in E. coli. 156
6-4. B. subtilis induced hydrolysis of X-Gal by various substrates. 158
6-5. Growth of wild type and mutants on minimal media plates. 159
6-6. Purification scheme for (His-tagged) LacA from E. coli. 160
6-7. Effects of ions on the activity of LacA enzyme. 162
6-8. Relative activity of LacA on various chromogenic substrates. 162
6-9. Primers used. 170
## List of Frequently Used Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH</td>
<td>Glycoside Hydrolase</td>
</tr>
<tr>
<td>GHF</td>
<td>Glycoside Hydrolase Family</td>
</tr>
<tr>
<td>ONPG or ONPGal</td>
<td>(o)-nitrophenyl galactopyranoside</td>
</tr>
<tr>
<td>PNPG or PNPGal</td>
<td>(p)-nitrophenyl galactopyranoside</td>
</tr>
<tr>
<td>ONPGlu</td>
<td>(o)-nitrophenyl glucopyranoside</td>
</tr>
<tr>
<td>PNPGlu</td>
<td>(p)-nitrophenyl glucopyranoside</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-(\beta)-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>X-Gal</td>
<td>5-bromo-4-chloro-3-indolyl-(\beta)-D-galactoside</td>
</tr>
<tr>
<td>X-Glc</td>
<td>5-bromo-4-chloro-3-indolyl-(\beta)-D-glucoside</td>
</tr>
<tr>
<td>EC</td>
<td>Enzyme Commission</td>
</tr>
<tr>
<td>CAZy</td>
<td>Carbohydrate-Active enZYmes</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani (medium)</td>
</tr>
<tr>
<td>TSA</td>
<td>trypticase soy agar without dextrose</td>
</tr>
<tr>
<td>R2A</td>
<td>not an abbreviation – name for a low nutrient medium</td>
</tr>
<tr>
<td>B.</td>
<td>the genus <em>Bacillus</em></td>
</tr>
<tr>
<td>P.</td>
<td>the genus <em>Paenibacillus</em></td>
</tr>
<tr>
<td>S.</td>
<td>the genus <em>Sporosarcina</em></td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>ABC</td>
<td>ATP-binding cassette (transporter)</td>
</tr>
<tr>
<td>COG</td>
<td>Clusters of Orthologous Groups</td>
</tr>
<tr>
<td>MEGA</td>
<td>Molecular Evolutionary Genetics Analysis</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyltransferase</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank my advisor Jean Brenchley for challenging me, patience, encouragement in helping me to develop a balance between independence and dialogue, and guidance in achieving clarity in my work and writing.

I also thank Allen T. Phillips, Mary Ann Bruns, James G. Ferry, and Ming Tien for agreeing to serve on my committee and for thought provoking ideas. I thank Dr. Phillips especially for borrowed chemicals, and for help during manuscript preparation regarding both broad scientific concepts as well as technical details. I thank Dr. Bruns for helping me develop my hypotheses, and for enthusiasm towards examining a cross-disciplinary problem.

I am also grateful for the hearty greetings and chemical substrates I received from the late John Pazur. Without understanding the enzymatic hydrolysis of the simple polysaccharide starch, where would we find a foundation for the analysis of more complex polysaccharides? In a similar sentiment, I would also like to thank all the researchers who were brave enough to include in their publications the data that didn’t fit, but provided me with a trail of clues that ultimately led to answers to the questions I was asking.

I also thank past and present Brenchley lab members for helpful discussions, ideas, lab meetings, amusement, and help in preparing manuscripts. I would specifically like to thank Kevin Gutshall for mentoring me through my rotation, Peter S. Sheridan for phylogenetic instruction and curmudgeon Fridays, Jennifer Loveland-Curtze for helping me to hear myself think, Vanya Miteva for use of her GIC isolates and help with techniques, Nick Panasik, Jr. for guiding me into the madness of immense data analysis
projects, James Coker for assurance, help with enzyme work, and workarounds, and
Jennifer Biddle for caffeine and method suggestions.

I also must thank my personal computer support team, Arthur Shipkowski, Don
Kenepp, and Bill Szewc, for helping to prevent, solve, and recover from various
hardware, software, and wetware computer issues.

I also appreciate the many people who inspired, challenged, and educated me in
science including instructors: Dr Porter, Dr. Brenchley, Dr. Kahn, Dr. Schaich, Dr.
Chase, Dr. Eveleigh, Mr. Crane, Mr. Griffith; my parents, Margaret and Vincent Rybicki;
my husband, Arthur Shipkowski; as well as Bea Plummer, Jan Bachofer, Autumn Finlan
and many others.

Funding for my dissertation came from the Department of Energy grant DE-
FG02-93ER20117 and NSF/IGERT DGE-9972759 for the Biogeochemical Research
Initiative for Education. I thank Mary Ann Bruns for her advisory role to me within this
program and I thank the BRIE program for the opportunity to have greater freedom in my
scientific explorations by providing the finances and the training that allowed me to have
confidence and resources for exploring a project outside the traditional boundaries of the
Brenchley laboratory.
Chapter 1

Introduction
1.1 Overview

The β-galactosidase, LacZ, of the lac operon from Escherichia coli is well-known and frequently used in blue/white transformant screening systems. This lactose-hydrolyzing enzyme is considered by many to be the quintessential β-galactosidase. However, this familiarity with the lac operon may lead to assumptions about other β-galactosidases that are incorrect. This introduction presents the reasons why β-galactosidases are studied, the classification systems applied to the numerous β-galactosidases, and our current knowledge regarding in vivo functions for these enzymes. It also describes the overall organization of the dissertation and explains how the developments in each chapter led to the next. I first explain my initial experimental goals and show how these led to insights regarding the in vivo substrates of β-galactosidases. In the chapters that follow, I provide the first solid in vivo evidence supporting an environmentally plausible function for a group of β-galactosidases whose purpose was previously unknown. Finally, I present a summary of the impact these results may have on our studies of (and with) these enzymes, as well as interpretations of the roles of enzymes with related functions.

1.2 Reasons to investigate β-galactosidases

The LacZ β-galactosidase of E. coli, is familiar historically as the model used in understanding the regulation of the lactose utilization operon. Currently a few studies continue to explore the in vivo role of other β-galactosidases, but most explore them as tools for molecular biology experiments. The routine use of LacZ as a molecular marker has resulted in the development of numerous methodologies based on β-galactosidase activity, and the availability of many chromogenic substrates such as X-Gal (5-bromo-4-
chloro-3-indolyl-β-D-galactoside) and ONPG (o-nitrophenyl-β-D-galactopyranoside).

These and other substrates make it convenient to screen for new isolates with β-galactosidase activity as well as E. coli (lacZ) transformants containing new microbial β-galactosidase genes. One of the reasons for investigating additional β-galactosidases (and hosts) is to make possible the use of β-galactosidase-based techniques under conditions unsuitable for LacZ (e.g., high temperature or high salt applications). Of special interest to the Brenchley research program are cold-active β-galactosidases, for which there are several applications. One potential application is in cloning systems specifically designed for thermolabile enzymes. These systems would consist of a host capable of growing efficiently (i.e. “much better than E. coli”) at low temperatures (i.e. 15°C and below) and a compatible expression vector encoding a β-galactosidase with high activity (i.e. “higher than LacZ”) at these temperatures. A second use for cold-active β-galactosidases is to help understand how cold-active enzymes function at low temperatures that restrict the activity of mesophilic enzymes. A third role for cold-active β-galactosidases is in the food industry, for instance, using cold-active β-galactosidases to hydrolyze lactose in milk during refrigeration for consumption by lactose intolerant individuals.

Although we often view β-galactosidases as tools for humans, they are ultimately borrowed tools from microorganisms that use them to degrade carbohydrates as carbon sources. This function is not only critical for the microorganism’s survival, but also contributes to the global carbon cycle. Initial detection of β-galactosidases, either in their native microorganism or heterologously expressed in transformants, usually involves screening for hydrolysis of synthetic chromogenic substrates. Although this is beneficial, the subsequent investigation of the enzymes often neglects characterization using
naturally occurring substrates. This type of enzyme characterization leaves many unanswered questions regarding finer details concerning substrate specificity, physiological roles, and effects on the carbon cycle in the environment. Many ecological studies focus on the enzymes that degrade earth’s most abundant glucose polymer, cellulose. However, many glycoside hydrolases other than cellulases still play significant roles even though they degrade less abundant environmental carbohydrates. For instance, if not for β-N-acetyl-glucosaminidases the world would be buried in insect and crab shells. The study of β-galactosidases and other carbohydrate-degrading enzymes is important for understanding the environments in which they can be found by answering many other questions: how do these enzymes fit into the intricate carbon cycle? To what degree do various microorganisms producing these enzymes contribute to this process? Do all β-galactosidases function in lactose utilization like LacZ? Which glycosylated compounds present in the environment do these enzymes interact with? And foremost, what are the natural substrates of these enzymes?

### 1.3 Classification of glycoside hydrolases

Enzymes with β-galactosidase activity, including LacZ, belong to the larger group known as glycoside hydrolases, which hydrolyze the bond between a carbohydrate and another moiety. The IUB (International Union of Biochemistry) Enzyme Commission (EC) nomenclature (9), assigned by activity, for β-galactosidases is 3.2.1.23. Lactose hydrolysis (EC 3.2.1.108) is a distinct and more specific function; not all β-galactosidases are lactases. In 1991, Henrissat published an alternative nomenclature that took into account the structural features of enzymes, based on the concept of sequence similarities being congruent with folding similarities, so that divergent and convergent evolution of
function could be considered (2). Using sequence similarities, hydrophobic cluster
analysis, and structures when available, Henrissat’s original 35 families have grown to
almost 100 glycoside hydrolase families (GHFs). This classification reveals that enzymes
capable of β-galactosidase activity have evolved at least four times, as represented by
their occurrence in GHFs 1, 2, 35, and 42 (3). Even though LacZ (GHF 2) is capable of
lactose hydrolysis, the co-evolution of lactase activity in the other three families is often
assumed rather than experimentally confirmed.

An online database containing representatives of the GHFs with links to their
NCBI gene and protein entries, and other information, is found at CAZy
(http://afmb.cnrs-mrs.fr/CAZY/index.html) (1), which stands for Carbohydrate-Active
Enzymes. Although all β-galactosidases belong to GHF 1, 2, 35 or 42, not all GHF 1 and
2 enzymes are β-galactosidases (Table 1-1). Also, although examples of GHF 1, 2, and 35
are found in all domains of life, the distribution of β-galactosidase activity within these
families differs for each domain. β-galactosidase activity in the domain Archaea is found
mostly in GHF 1, whereas GHF 2 β-galactosidases (like LacZ) are frequently found in
Bacteria, and most GHF 35 β-galactosidases belong to Eukaryota (Table 1-1). GHF 42
possesses only β-galactosidase activity and consists mostly of sequences of bacterial
origin. A single archaeal representative is known, but there are currently no known
eukaryotic GHF 42 enzymes. Examples of in vivo lactase activity have been amply
demonstrated for GHF 2 (LacZ) and GHF 1 (human lactase phlorizin hydrolase), but not
for GHF 35 or GHF 42.
Table 1-1 Activities and distributions of several relevant GHFs

<table>
<thead>
<tr>
<th>GHF</th>
<th>Activities*, summarized as per CAZy</th>
<th>Representatives found in:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>bacteria</td>
</tr>
<tr>
<td>1</td>
<td>β-glucosidase (21)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>β-galactosidase (23)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>6-phospho-β-galactosidase (85)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>6-phospho-β-glucosidase (86)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>β-mannosidase (25)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>β-glucuronidase (31)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>and others…</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>β-galactosidase (23)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>β-mannosidase (25)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>β-glucuronidase (31)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>β-glucosidase (21)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>β-N-acetylhexosaminidase (52)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>xylan-β-1,4-xylosidase (37)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>α-L-arabinofuranosidase (55)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>unknown</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>and others…</td>
<td>—</td>
</tr>
<tr>
<td>35</td>
<td>β-galactosidase (23)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Unknown / other</td>
<td>X</td>
</tr>
<tr>
<td>42</td>
<td>β-galactosidase (23)</td>
<td>X</td>
</tr>
</tbody>
</table>

X - known examples
| - - activity not yet demonstrated

* Numbers in parentheses are the EC 3.2.1.#

Since family assignment is made by sequence homology, not by activity, new open reading frames (ORFs) with no experimentally-confirmed activity can be quickly categorized into families. This classification method has orphaned biochemical data for some enzymes where their family designation is unknown because no sequence data are available. Analogously, the rapid increase in genome sequencing has given rise to many classified ORFs lacking biochemical characterization of the encoded enzymes. Although the general physiologies of the host microorganisms of these putative enzymes are known, it is difficult to find specific physiological data (e.g. is lactose utilized?) for the exact strains whose genomes were sequenced. Simultaneously, there are many cases where much more could be learned about previously classified enzymes if adjacent sequence information, which is easily available from full genome sequences, was known.
1.4 Known functions of β-galactosidases

Lack of data regarding the activity of β-galactosidases on non-synthetic substrates makes it more difficult to suggest explicit physiological substrates and metabolic roles for β-galactosidases. The best-studied example of a β-galactosidase with a known in vivo role is LacZ, which functions in E. coli in the utilization of lactose as a carbon source. Lactose has been found only in the milk of mammals, and since E. coli and many bacteria with enzymes related to LacZ are found in the intestines of mammals that consume milk, this is a reasonable environmentally-relevant function, with an abundance of supporting experimental data and a well-understood operon. In this operon (Fig 1-1), the lactose permease gene, lacY, and the poorly understood acetyltransferase encoded by the lacA gene follow the β-galactosidase gene lacZ. Upstream but not transcribed as part of the operon, lacI encodes the repressor that binds to the promoter DNA, repressing expression of the lacZ, Y and A genes. When an inducer (technically a derepressor) binds to LacI, the protein is released from the DNA allowing expression of the lac operon. The in vivo inducer of the lac operon is allolactose, a side product of lactose hydrolysis. In the laboratory, we often substitute a similar chemical, IPTG (isopropyl-β-D-thiogalactopyranoside), which acts as a gratuitous inducer that is not hydrolyzed. Once expressed, the lactose permease imports lactose into the cell where LacZ hydrolyzes it intracellularly.
Lactose as a free sugar is generally not found in soil, yet many soil microorganisms possess β-galactosidase enzymes. Thus, it is unlikely that lactose is the natural substrate for many of these β-galactosidases, nor is lactose the only known substrate for enzymes with β-galactosidase activity. For instance, several activities have been found for (extracellular) GHF 35 enzymes: (exo-) hydrolysis of β-1,3 linked galactose (as found on arabinogalactan proteins) (EC 3.2.1.145) by bacterial and plant enzymes (4, 5, 8, 10), exo-β-1,4-galactanase activity on pectic galactan by a plant enzyme (no E.C. number) (6), and exo-β-D-glucosaminidase activity contributing to chitin degradation by an archaeal enzyme (no E.C. number) (7). However, this limited list of potential substrates is probably not exhaustive, and we still do not know what explicit substrate(s) most individual β-galactosidases use in vivo. Without knowing the functions of these enzymes, we do not understand their environmental significance, nor can we assess whether they might have commercial applications. Of all the GHFs containing β-galactosidases, we are especially limited in our understanding of GHF 42 because a physiological role has not been demonstrated for any enzyme in this group.
1.5 Dissertation Organization

My experimental goal was an evaluation of closely-related β-galactosidase enzymes that I expected to possess a uniform function, which could be explored by looking for consistencies in growth and up-regulation of β-galactosidase activity on a variety of substrates. I intended to do this by enriching for psychrophilic (cold-loving) bacteria from a narrow phylogenetic group and cloning from them a number of hypothetically orthologous β-galactosidase genes. Because the discerning the unknown function of the GHF 42 β-galactosidases was of the most interest, I planned to selectively work with the subset of enzymes belonging to this group, as determined by sequencing.

The phylogenetic group I chose was the order Bacillales, which contains endospore-forming organisms. I enriched for psychrophilic members of this group by using a heat treatment to kill vegetative cells (but not endospores), screened the resulting isolates for β-galactosidase activity, and cloned β-galactosidase genes from them. My initial survey isolated bacterial strains used for some of the analyses in later chapters and involved brief characterization of several β-galactosidases (Chapter 2). One enzyme was investigated in detail because it had significant β-galactosidase activity (at least on chromogenic substrates), but did not belong to any of the GHF families described above (1, 2, 35 or 42), but instead to GHF 3 (Chapter 3). The activity of this enzyme on chromogenic galactosidases seemed to be due to the preference of the enzyme for aryl-moieties (such as o-nitrophenyl), a common feature of the many glucosides that are secondary plant metabolites. A version of Chapter 3 was published by the journal Applied and Environmental Microbiology (Shipkowski, S. and J.E. Brenchley. 2005. Characterization of an unusual cold-active β-glucosidase belonging to Family 3 of the
glycoside hydrolases from the psychrophilic isolate *Paenibacillus* sp. strain C7. AEM 71(8): 4225-4232).

The results of the initial survey and my attempts to discern the function of the GHF 3 enzyme made me very curious about the possible function of GHF 42 enzymes. I first reviewed the data pertaining to lactose as a substrate for characterized GHF 42 enzymes to clarify that there are no demonstrations of a GHF 42 gene being sufficient and necessary for the growth of the bacterium (originally possessing the gene) on lactose (Chapter 4). This indicated that finding a reasonable environmental (non-lactose) substrate for these β-galactosidases would be a valuable contribution.

I believed additional GHF 42 genes and enzymes would help me to attain this goal. I therefore constructed primers to better identify which X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) hydrolyzing isolates and *E. coli* transformants possessed genes belonging to GHF 42. I then tested and used these primers to guide the cloning process (Chapter 4). Using the classification system of GHF 3 substrates as a model (Fig 1-2, panel A), I planned to discern a potential function for GHF 42 using induction studies with specific isolates combined with biochemical characterization of their enzymes. At the broadest point of comparison, both GHF 3, which contains β-glucosidases, and some of the GHFs containing the β-galactosidases (GHFs 1 & 2) also contain enzymes with other activities (Table 1-1). But sub-classified within β-glucosidase activity, GHF 3 enzymes can act on three types of substrates, with examples of each originating from plants: 1. aryl-substrates like salicin 2. disaccharides like cellobiose, and 3. oligosaccharides like cellotetraose (Fig 1-2, panel A). This is a far greater variety of known substrates than that typically considered for the β-galactosidases (lactose, ONPG).
Figure 1-2. Classification of β-glucosidase substrates as a model for potential β-galactosidase substrates. A. β-glucosidases of GHF 3 can act on (1.) aryl-substrates, (2.) disaccharides, and (3.) oligosaccharides. The blue circles represent glucose molecules as per the CFG (Consortium for Functional Glycomics) standard. Some of the di- and oligoglucosides can be obtained by enzymatic degradation of larger polysaccharides, generally known as glucans. We have trivial names for many of these substrates, shown in parentheses. B. By analogy, β-galactosidases might have activity on (1.) aryl-galactosides, (2.) β-galactobioses, and (3.) oligogalactosides, which might in turn originate from the enzymatic degradation of galactans (equivalent trivial names like cellulose do not exist). The yellow circles represent galactose molecules.
Figure 1-2 Classification of \( \beta \)-glucosidase substrates as a model for potential \( \beta \)-galactosidase substrates

### A. \( \beta \)-glucosidase substrates

1. **Aryl-glucosides**
   - ONPGlu
   - Salicin
   - Coniferin
   - Pelargonidin-3-glucoside (a glucosylated anthocyanin)

2. **Disaccharides**
   - \( \beta_4 \)
   - \( \beta_2 \)
   - \( \beta_3 \)
   - \( \beta_6 \)
   - \( \beta_1,4 \)-glucobiose (cellulobiose)
   - \( \beta_1,2 \)-glucobiose (sophorose)
   - \( \beta_1,3 \)-glucobiose (laminaribiose)
   - \( \beta_1,6 \)-glucobiose (gentiobiose)

3. **Oligosaccharides**
   - \( \beta_4 \)
   - \( \beta_4 \)
   - \( \beta_4 \)
   - \( \beta_1,4 \)-glucotetraose (cellulotetraose)
   - \( \beta_1,3 \)-glucotriose (laminaritriose)

---

**Polysaccharidic sources for \( \beta \)-glucosidase substrates**

- \( \beta_4 \)
  - \( \beta_1,4 \)-glucan (cellulose) (lichenan)
- \( \beta_2 \)
  - \( \beta_1,2 \)-glucan (produced by a few proteobacteria)
- \( \beta_3 \)
  - \( \beta_1,3 \)-glucan (laminarin - brown algae) (curdlan - bacteria) (callobose - plants)
- \( \beta_6 \)
  - \( \beta_1,6 \)-glucan (pustulan - fungi)

### B. Theoretical \( \beta \)-galactosidase substrates

1. **Aryl-galactosides**
   - ONPG
   - Pelargonidin-3-galactoside (a galactosylated anthocyanin)

2. **Disaccharides**
   - \( \beta_4 \)
   - \( \beta_2 \)
   - \( \beta_3 \)
   - \( \beta_6 \)
   - \( \beta_1,4 \)-galactobiase
   - \( \beta_1,2 \)-galactobiase
   - \( \beta_1,3 \)-galactobiase
   - \( \beta_1,6 \)-galactobiase

3. **Oligosaccharides**
   - \( \beta_4 \)
   - \( \beta_4 \)
   - \( \beta_4 \)
   - \( \beta_1,4 \)-galactotetraose
   - \( \beta_1,3 \)-galactotriose

---

**Potential polysaccharidic sources for \( \beta \)-galactosidase substrates**

- \( \beta_4 \)
  - \( \beta_1,4 \)-galactan
- \( \beta_2 \)
  - \( \beta_1,2 \)-galactan
- \( \beta_3 \)
  - \( \beta_1,3 \)-galactan
- \( \beta_6 \)
  - \( \beta_1,6 \)-galactan
By analogy, $\beta$-galactosidases might have activity on similar compounds from plants that have galactose instead of glucose (Fig 1-2, Panel B). Several polysaccharides that release GHF 3 $\beta$-glucosidase substrates through the action of other enzymes are known (cellulose yields cellobiose, laminarin yields laminaribose, etc) and generally referred to as $\beta$-glucans (Fig 1-2, Panel A). Equivalently, GHF 42 $\beta$-galactosidases might be acting on the oligosaccharides that compose $\beta$-galactans (Fig 1-2, Panel B) or other more complex polysaccharides (plant gums and pectic substances) that could have shorter branches containing $\beta$-linked galactose.

During the sequence analyses needed to design the GHF 42 primers, I noticed some gene patterns that might represent a conserved operon structure for GHF 42 $\beta$-galactosidases. I proposed that clues relating to one of these substrate types might be found by taking advantage of the tendency of bacteria to organize genes related to the same function in an adjacent fashion - as operons. This led me to further explore the gene arrangements surrounding GHF 42 genes. Since the onset of my project, many more sequences containing GHF 42 genes have become available as a result of genome sequencing projects. Analyses of GHF 42 distribution within all sequenced genomes and the phylogeny of the GHF 42 enzymes provide some information about potential functions of these enzymes. Comparison of gene arrangements containing GHF 42, both from genomes and other sequencing work, confirms the existence of a conserved relationship between several types of genes highly suggestive of a particular function. These studies (Chapter 5), combined with the perspective gained from studying the enzyme described in Chapter 3, led me to hypothesize that oligosaccharides released from the polysaccharide arabinogalactan (type-I) are substrates for some GHF 42
enzymes. This is distinct from the less complex activities from GHF 35 enzymes, which can act directly on galactans in an exo- fashion, without the prior activity of an endo-acting enzyme. This hypothesis is tested in Chapter 6, which demonstrates a function for a \(\beta\)-galactosidase, LacA, other than lactose hydrolysis in a classically studied microorganism, \textit{B. subtilis}. I also attempted to ascribe functions to previously hypothetical proteins in \textit{B. subtilis} that likely participate in the degradation and utilization of arabinogalactan (type-I). The summary (Chapter 7) describes the impact these results may have on our future studies of and with these enzymes.
1.6 References


Chapter 2

Characterization and isolation of psychrophilic bacteria belonging to *Bacillales* and examination of their β-galactosidases
2.1 Summary

I obtained β-galactosidases from closely-related psychrophilic bacteria belonging to three families within the order *Bacillales*, which contains phylogenetically related endospore-forming bacteria. The bacteria I worked with were either characterized as belonging to this group by other laboratory members or were ones that I isolated through a specific enrichment process. I designed this enrichment to select for members in four *Bacillales* genera (*Bacillus*, *Paenibacillus*, *Sporosarcina*, and *Brevibacillus*). Vegetative cells in samples were killed by a heat treatment and the spores allowed to germinate aerobically at low temperatures (2 - 10°C). Subsequently, I screened the enrichment isolates and other laboratory isolates belonging to the same or closely related genera for β-galactosidase activity. I cloned β-galactosidase encoding genes from some of these isolates and characterized their enzymes. Sequencing was performed on the cloned inserts to determine which of four possible glycoside hydrolase families (GHFs) the cloned β-galactosidase genes belonged to and to determine whether any of the adjacent genes gave indications of *in vivo* function. Of the six genes cloned, two belonged to GHF 42, the β-galactosidase group of the highest interest for this project. One enzyme did not belong to any of the expected families, but was of interest, and its characterization will be described later (Chapter 3).
2.2 Introduction

Psychrophiles are typically isolated from environments that are considered by humans to be cold. These environments include the permanently cold (< 5°C) Arctic and Antarctic regions, permafrost, glaciers, and much of the Earth’s oceans, as well as seasonally cold soils and waters. Microorganisms can function in the cold because of physiological adaptations, including enzymes that have differences in their amino acid sequence that ultimately allow higher catalytic activities at lower temperatures (as compared with those found in mesophilic organisms). Cold active enzymes are of interest for determining which structural differences (and which underlying amino acid differences) allow them to function at low temperatures and because some may have industrial applications.

Early explorations into the thermal dependency of enzymes compared enzyme pairs from phylogenetically distant organisms (even from different domains of life (eg Bacteria and Archaea (32, 41). Due to this approach, the significance of the results obtained have been questioned (26). Initially, my project was designed to examine the basis of cold-activity under conditions that were expected to reduce the influence of random changes and alternate selective pressures on the comparisons by restricting the comparisons to a single GHF (glycoside hydrolase family) of enzymes in a single bacterial order. The work in this chapter was directed towards achieving this initial goal. However, as the project progressed, my focus switched to addressing the unknown function(s) of the chosen enzyme group, the GHF 42 β-galactosidases. This chapter describes isolates, cloned β-galactosidase genes, and data relevant to the work performed in later chapters, but which has been reanalyzed. It includes the results of comparisons to
both amino acid and DNA sequences that were not available at the time this experimentation was performed.

GHF 42 was chosen for comparisons because of the ease of screening for transformants using the chromogen X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside), because of their small size compared to other β-galactosidases, and because a majority of GHF 42 genes belonged (then and now) to Gram-positive organisms suggesting that they would be a good source of these genes. The phylogenetic group of focus consisted of three families in the order Bacillales (Planococcaceae, Paenibacillaceae, and Bacillaceae), and more specifically the psychrophilic bacteria belonging to these groups. Some of these bacteria were from other laboratory members’ projects and were identified by 16S rDNA analysis; others I isolated through selective enrichments.

In these enrichments a heat treatment kills the vegetative cells, even those belonging to order Bacillales, but allows the heat-resistant spores to survive. Following heat treatment, the samples are plated onto media where spores can germinate and cells grow aerobically at low temperature. Members from several other subgroups of spore-forming Bacillales will not grow because the required thermophilic, halophilic, microaerophilic, anaerobic, or other specific conditions are not met, and so isolates from only the four genera Paenibacillus, Sporosarcina, Bacillus, and Brevibacillus were expected (Fig 2-1). Isolates belonging to these last two genera were anticipated to occur less frequently since these consist mainly of meso- and thermophilic species and are represented infrequently in diversity studies of psychrophilic environments. Although endospores from both thermophilic and psychrophilic microorganisms may be present in
any environment, the samples were collected from northern climates, in winter, with the expectation that spores from psychrophiles would be present in greater numbers.

In this chapter, I describe the characterization of several new isolates and the creation of genetic libraries from these bacteria. In total, six $\beta$-galactosidase genes were cloned: four originated from the spore-specific enrichment isolates, one from an isolate identified by other phenotypic characteristics, and one from a bacterium isolated by Miteva et al. (23). The thermal dependencies of the $\beta$-galactosidase activities were examined in the six enzymes heterologously expressed in *E. coli*. Sequencing revealed that two of the cloned $\beta$-galactosidase genes belonged to the group of interest (GHF 42) and that one unexpectedly belonged to another group, GHF 3.
Figure 2-1. Expected genera of isolates resulting from the enrichment process.

Key:
- Black – groups containing spore-formers
- **Bold** – expected genera
- Grey – group is non-spore forming, will be killed by heat treatment
- Brown – required enrichment conditions not met for group
  (e.g., thermophilic, strictly anaerobic, halophilic… )

**Firmicutes**
- **Mollicutes**
- **Clostridia**
- **Bacilli**
  - Lactobacillales
  - **Bacillales**
    - **Listeriaceae**, **Caryophanaceae**, **Turicibacteraceae**, **Staphylococcaceae**
    - **Alicyclobacillaceae**, **Thermoactinomycetaceae**, **Sporolactobacillaceae**
    - **Planococcaceae**
      - **Planococcus**, **Planomicrobium**, **Kurthia**, **Filibacter**
    - **Sporosarcina**
    - **Paenibacillaceae**
      - **Paenibacillus**, **Brevibacillus**
    - **Bacillaceae**
      - **Exiguobacterium**
      - **Geobacillus**, **Ureibacillus**, **Anoxybacillus**, **Amphibacillus**
      - **Halobacillus**, **Virgibacillus**, **Gracilibacillus**, **Marinibacillus**
      - **Oceanobacillus**, **Salinibacillus**, **Lentibacillus**, **Jeotgalibacillus**
    - **Bacillus**

Figure 2-1. Expected genera of isolates resulting from the enrichment process. Within the **Firmicutes**, aerobic endospore-formers are found almost exclusively within the Order **Bacillales**. Not all groups within this order form spores (gray). Enrichment under aerobic, cold, <2% salt, (pH) neutral conditions further restricts the phylogenetic associations of the isolates germinating from the surviving spores (black) because some groups require warmer conditions or higher salt concentrations, etc (orange). Under the given conditions, only isolates from the genera **Bacillus**, **Paenibacillus**, **Sporosarcina**, and **Brevibacillus** are expected.
2.3 Results

2.3.1 Characterization of isolates

The closest relative (by 16S rDNA comparison) and growth range of several isolates from different origins were examined (Table 2-1). Other phenotypic properties were also observed. Isolate *Sporosarcina* sp. CRE9 displayed motility in which the bacterial colonies move on the agar, leaving faint trails of cells behind as they grow outwards on centrally inoculated plates (Fig 2-2). Isolate *Sporosarcina* sp. CSA hydrolyzed more X-Gal when grown at 18°C than at 25°C and had no hydrolysis at 30°C, suggesting the presence of a β-galactosidase with a low thermal optimum of activity. Isolates *Paenibacillus* spp. C7 and CKG displayed more X-Gal hydrolysis at low temperatures (10 and 18°C) compared to other isolates. This could have been caused by either greater activity of the enzymes at these temperatures or greater production of β-galactosidase enzymes.

Figure 2-2. Motile colony pattern formation by isolate *Sporosarcina* sp. CRE9.

Figure 2-2. Motile colony pattern formation by isolate *Sporosarcina* sp. CRE9. Colony motility is described by “morphotype” (“B” (*Bacillus subtilis*), tip-splitting, chiral, vortex, or spiral-vortex) and handedness (preference for clockwise or counterclockwise rotation). Although it is capable of producing a chiral-like pattern (A), the isolate is of the “vortex” morphotype: vortices can be seen in the center of (B). CRE9 does not display a strong handedness, as shown by frequent switching of the direction of rotation (A). The growth has been stained with Coomassie blue for easy visualization.
Table 2-1 Characteristics of some X-Gal hydrolyzing isolates

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Growth range (°C)</th>
<th>Closest relative*, percent identity</th>
<th>Origin</th>
<th>37°C X-Gal hydrolysis?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRE9</td>
<td>2-25, not 30</td>
<td><em>Sporosarcina psychrophila</em>, 99.6</td>
<td>Rochester (NY)</td>
<td>NA**</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Sporosarcina globispora</em>, 99.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMM</td>
<td>2-37, not 45</td>
<td><em>Planococcus maritimus</em>, 99.9</td>
<td>Cheesequake bog (NJ)</td>
<td>No (30°C-yes)</td>
</tr>
<tr>
<td>CKG</td>
<td>2-37, not 45</td>
<td><em>Paenibacillus odorifer</em>, 98.5</td>
<td>Sewage treatment plant (PA)</td>
<td>Yes</td>
</tr>
<tr>
<td>CSA</td>
<td>2-30, not 37</td>
<td><em>Sporosarcina psychrophila</em>, 99.9</td>
<td>Cheesequake bog (NJ)</td>
<td>NA**</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Sporosarcina globispora</em>, 99.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>2-25, not 30</td>
<td>*Paenibacillus macquariensis, 97.6</td>
<td>Bear Meadows Bog (PA)</td>
<td>NA**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*Paenibacillus antarcticus, 97.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* as determined by analysis of 16S rDNA sequences
** NA – not applicable as organism does not grow at this temperature

Planococcus sp. CMM was not isolated following heat-treatment. It was selected based on a phenotype matching that of bacteria belonging to the genus Planococcus: orange pigmentation, tetrads of cocci cell morphology, and salt-marsh origin. Members of this genus are unable to form spores, but are part of the Bacillales family (Fig 2-1). In addition to belonging to this phylogenetic group of interest, there was also an opportunity for comparisons with another GHF 42 β-galactosidase whose gene had been cloned from a Planococcus isolate (33).

Three additional Paenibacillus spp. not isolated via my enrichment process were identified as belonging to the Paenibacillaceae through the phylogenetic work performed by Miteva et al. (23) (Table 2-2).

Table 2-2 Greenland ice-core isolates and their β-galactosidase thermal optima

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Closest relative*, percent identity</th>
<th>Growth range (°C)</th>
<th>B-galactosidase thermal optimum(s) for isolate (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GiC16</td>
<td><em>Paenibacillus pabuli</em>, 99.2</td>
<td>2 – 33, not 37</td>
<td>35 and 50</td>
</tr>
<tr>
<td>GiC1y</td>
<td>Same as GiCR21 via ERIC†</td>
<td>2-18, not 25</td>
<td>48</td>
</tr>
<tr>
<td>GiCR21</td>
<td><em>Paenibacillus amyloyticus</em>, 99.5</td>
<td>2-33, not 37</td>
<td>50</td>
</tr>
</tbody>
</table>

* as determined by analysis of 16S rDNA sequences
† Enterobacterial Repetitive Intergenic Consensus; see Miteva et al., 2004
2.3.2 Thermal dependency of β-galactosidase activities

Five fragments carrying genes encoding β-galactosidases were cloned into *E. coli* (Table 2-3) using genomic DNA from my five isolates. Clarified lysate containing heterologously-expressed enzyme was used to determine the thermal optima of these enzymes on the chromogenic substrate *o*-nitrophenyl-β-D-galactopyranoside (ONPG). Because the *E. coli* host contained a *lacZ* deletion, no background activity was observed in lysate from a vector-only control (pΔα18). These heterologously expressed β-galactosidases had optima lower than that of the β-galactosidase of *E. coli*, LacZ (55°C) (10) (Table 2-3). Thermal dependency data for the β-galactosidase from *Paenibacillus* sp. C7 reveals an exceptionally low optimum of 25°C.

<table>
<thead>
<tr>
<th>Isolate</th>
<th><em>E. coli</em> Transformant</th>
<th>Transformant 37°C X-Gal hydrolysis?</th>
<th>β-galactosidase thermal optimum (°C)</th>
<th>Most homologous β-galactosidase sequence*</th>
<th>% Identity, Homology*</th>
<th>GHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRE9</td>
<td>pCRE9-2</td>
<td>No</td>
<td>48</td>
<td><em>Bacillus halodurans</em></td>
<td>63, 76</td>
<td>2</td>
</tr>
<tr>
<td>CMM</td>
<td>pCMM-3</td>
<td>Yes</td>
<td>47</td>
<td><em>Planococcus</em> sp ‘SOS Orange’</td>
<td>91, 95</td>
<td>42</td>
</tr>
<tr>
<td>CKG</td>
<td>pCKG-6</td>
<td>Yes</td>
<td>43</td>
<td><em>Bacillus halodurans</em></td>
<td>68, 82</td>
<td>42</td>
</tr>
<tr>
<td>CSA</td>
<td>pCSA-5</td>
<td>No</td>
<td>36</td>
<td><em>Bacillus halodurans</em></td>
<td>69, 82</td>
<td>2</td>
</tr>
<tr>
<td>C7</td>
<td>pC7-4</td>
<td>No</td>
<td>25</td>
<td><em>Bacillus</em> sp. GL1</td>
<td>69, 82</td>
<td>3</td>
</tr>
<tr>
<td>GIC16</td>
<td>pGIC16-1</td>
<td>Yes</td>
<td>50</td>
<td><em>Fusarium oxysporum</em></td>
<td>49, 64</td>
<td>2</td>
</tr>
</tbody>
</table>

*over region sequenced as diagrammed in Fig 2-3*

The thermal profiles of β-galactosidase activity from three *Paenibacillus* spp. Greenland isolates obtained by Miteva et al. (23) were examined using clarified lysates from the isolates. *Paenibacillus* spp. GIC1y and GICR21 each showed a single maximum, while the profile of *Paenibacillus* sp. GIC16 showed two (Table 2-2). The thermal dependency shown by this isolate was interesting; therefore, a sixth β-galactosidase gene was cloned from *Paenibacillus* sp. GIC16. Several identical X-Gal
hydrolyzing transformants were obtained that displayed β-galactosidase activity with an optimum equivalent to the higher (50°C) one observed in the original isolate.

**2.3.3 Analysis of the cloned β-galactosidase genes and adjacent sequence**

Sections of the cloned fragments responsible for encoding the β-galactosidase activity expressed by the transformants were sequenced in order to identify the family of the glycoside hydrolase being expressed. Three genes belonged to GHF 2, the group to which the LacZ enzyme of *E. coli* belongs. Two genes belonged to GHF 42, and the remaining gene belonged to GHF 3 (Table 2-3). The most homologous sequences (as identified by BLAST (3)) for the sequenced regions of the β-galactosidase genes were identified (Table 2-3). The fragments cloned from *Sporosarcina* spp. CRE9 and CSA were not fully sequenced as the partial sequences showed the encoded β-galactosidases belonged to GHF 2. Analyses of the N-terminal translations using the Signalp WWW server (5) suggested that all of the β-galactosidases were intracellular.

Adjacent genes were examined for indications of the native function of these β-galactosidases. The sequenced regions of the β-galactosidase genes from *Sporosarcina* sp. CRE9 and *Sporosarcina* sp. CSA were most similar to a *Bacillus halodurans* ORF, BH2723, from GHF 2. **BH2723** is an annotation reference for the genome of *B. halodurans* (36) with adjacent genes typically having sequential numbers. Adjacent genes also had similarity to a possible operon in *B. halodurans* where the β-galactosidase gene is preceded by genes encoding a portion of an ATP-binding cassette (ABC) transporter (Fig 2-3). The *B. halodurans* genome annotates these genes as lactose permeases. Although the two *Sporosarcina* spp. β-galactosidases are both GHF 2 and are both in arrangements similar to that seen in *B. halodurans*, the β-galactosidases from
Sporosarcina sp. CRE9 and Sporosarcina sp. CSA are not identical; they share only 70% homology with each other. In a similar arrangement, the GHF 2 gene from Paenibacillus sp. GIC16 is preceded by a pair of ORFs with homology to sugar permeases. However, the closest homologs in B. halodurans are BH1865 and BH1866 (annotated simply as sugar permeases), not BH2724 and BH2725. Also, this β-galactosidase gene has highest identity and homology to hypothetical proteins in several species of fungi. The first bacterial β-galactosidase homolog returned by BLAST is EF2709 from Enterococcus faecalis (27) (Fig 2-3).

The GHF 42 genes from Paenibacillus sp. CKG and Planococcus sp. CMM are most similar to a B. halodurans ORF (BH3701) and the β-galactosidase gene of Planococcus sp. ‘SOS Orange’ (33). In both the Paenibacillus sp. CKG and B. halodurans sequences, the GHF 42 gene is preceded by sequence with homology to transcriptional regulator genes of the AraC/XylS family (Fig 2-3). In Planococcus sp. CMM the β-galactosidase gene is proceeded by a β-galactosidase belonging to GHF 35. The gene arrangement surrounding the Planococcus sp. ‘SOS Orange’ β-galactosidase gene is unknown. The only other instance of a GHF 35 gene occurring adjacent to a GHF 42 gene occurs in Carnobacterium maltaromaticum BA (12), in which there are also adjacent transporter genes that, like those in Paenibacillus sp. GIC16, are most similar to BH2725 and BH2724 (Fig 2-3). The sequences found in relation to Paenibacillus sp. C7 are discussed in the following chapter.
Figure 2-3. Homology of β-galactosidases and adjacent ORFs. The boxes represent the length and identity of the regions sequenced from cloned inserts (bolded titles); for reference, GHF 2 β-galactosidases are encoded by genes approximately 3 kb in length, and GHF 42, 2 kb. Similar arrangements are found for homologous genes belonging to fully sequenced genomes (not bolded). The gray boxes represent expected continuations of genes within unsequenced regions. The lengths of the unsequenced regions are sufficient to suggest the genes are not truncated. The GHF 35 gene of pCMM was not completely contained within the cloned fragment.
2.4 Discussion

2.4.1 Characterization of isolates

*Planococcus* sp. CMM, *Paenibacillus* sp. CKG, and *Paenibacillus* sp. C7 have 16S rDNA sequences that share greater than 97% identity with sequences from described species in the same genera. Although new species are defined by examining many other characteristics in addition to 16S rDNA, this level of identity indicates that it is unlikely that these isolates represent new species. The same conclusion cannot be made for *Sporosarcina* spp. CRE9 and CSA because of the species that their 16S rDNAs most closely match: *Sporosarcina psychrophila* and *Sporosarcina globispora*. These described species were originally isolated from soil and river water by Larkin and Stokes (19). Based on the “minor” phenotypic differences between the two species, *S. psychrophila* temporarily lost its nomenclatural standing (34). However, in spite of the two species possessing “effectively identical” 16S rDNA (15), low DNA-DNA relatedness and other differences justified the reinstatement of *S. psychrophila* (24). Therefore, identification within this particular phylogenetic cluster requires additional characterization in order to determine which of these species *Sporosarcina* spp. CRE9 and CSA belong to, or whether either represents a new species.

*Sporosarcina* sp. CRE9 exhibits an interesting phenomenon of colony motility previously observed in some *Bacillus* and *Paenibacillus* species. The pattern formation by *Sporosarcina* sp. CRE9’s resembles the “vortex” morphotype (6, 39). Members of the *Sporosarcina* genus (including *S. globispora, S. psychrophila* (19)) have not previously been observed to form motile microcolonies. Thus, this phenotype could be a strong
argument for classification of this isolate as a novel species, but work in this direction does not coincide with my intended goals.

Three of the five isolates obtained are considered psychrophilic since they are able to grow at 2°C, but not at 37°C (25), and all are related to isolates obtained from other cold environments. *Paenibacillus* sp. CKG is most closely related to a *Paenibacillus odorifer* strain isolated from chilled zucchini puree (8, 16), and *Paenibacillus* sp. C7 is most closely related to a pair of isolates from Antarctica and a sub-Antarctic island, *Paenibacillus antarcticus* and *Paenibacillus macquariensis*. Evidence of *Sporosarcina* and *Paenibacillus* species have been found in other cold environments including Siberian permafrost (4), Greenland (23, 35), and Antarctica (9, 29). 16S rDNA sequencing confirmed that isolate CMM belonged to the *Planococcus* genus, in agreement with its phenotype. Many planococci have been isolated from algae and algal mats (31), including those in Antarctica (2, 30), and in other marine environments (13, 42).

### 2.4.2 Thermal optima of β-galactosidases

The β-galactosidases encoded by genes cloned from *Sporosarcina* sp. CRE9 and *Sporosarcina* sp. CSA had optima of activity at 48°C and 36°C, respectively, when heterologously expressed (Table 2-2). This was unexpected since X-Gal hydrolysis on plates by these respective clones (which did occur at 18°C), did not occur at 37 and 25°C, respectively. Apparently, within the *E. coli* host, these enzymes are either unable to achieve or effectively maintain a conformation allowing activity at higher temperatures, or are for some reason the genes encoding them are transcribed or translated with a thermal bias. Explicit examination of the thermostability of these enzymes might also
provide insight into this discrepancy. The heterologously expressed β-galactosidase from isolate *Planococcus* sp. CMM had an optimum of 47°C. This is not significantly different from the 42°C optimum obtained by Dr. Sheridan in our research program from *Planococcus* sp. SOS Orange (33). The gradually declining activity observed at higher temperatures for the β-galactosidase from *Paenibacillus* sp. C7 (shown in Chapter 3) suggests that it is a cold-active enzyme whereas a sharp decline in activity at higher temperatures would have been more indicative of heat lability. The 25°C optimum of activity for the *Paenibacillus* sp. C7 β-galactosidase equivalent to the published optimum for a cold-active β-galactosidase from a *Pseudoaltermonas* sp. (26°C) (14) but higher than the impressively low optimum of BgaS from *Arthrobacter* sp. SB (18°C) (10).

The thermal-dependency profile of β-galactosidase activity in *Paenibacillus* sp. GIC16 (not shown), has two peaks of activity, at 35 and 50°C, suggesting the presence of two or more β-galactosidases. Based on overlapping the thermal dependency data from the isolate and the transformant (not shown), it can be concluded that the β-galactosidase expressed by the pGIC16-1 transformant is responsible for the higher of these two optima. Subsequent efforts were not successful in cloning the gene responsible for encoding the enzyme contributing to the cold-active portion of the activity.

### 2.4.3 Analysis of the cloned β-galactosidase genes and adjacent sequence

My original goal required isolates from a phylogenetically conserved group, certain families within the order *Bacillales*, some of which I intended to acquire through a specific enrichment process. In line with this goal, all the isolates belonged to the intended phylogenetic group. That some β-galactosidases did not belong to the desired family, GHF 42, was expected as a result of random cloning of genes encoding β-
galactosidases. However, it was anticipated that these other β-galactosidases would belong to groups already known for this activity. The insert from isolate *Paenibacillus* sp. C7 clearly encoded β-galactosidase activity, but surprisingly did not yield sequences with homology to GHF 1, 2, 35 or 42. The only sequence with GHF homology belonged to GHF 3, and was closest to a β-glucosidase from *Bacillus* sp. GL1 (17). This β-glucosidase did not show activity on a β-galactosidase substrate (17). Examination of the enzyme from *Paenibacillus* sp. C7 showed that it also had β-glucosidase activity, but the unusual presence of β-galactosidase activity made it worthy for further study (Chapter 3).

Adjacent gene sequences were analyzed for ORFs with functions that could be associated with possible functions for the adjacent β-galactosidase gene. Three different arrangements were observed. One of the arrangements showed transport proteins encoded adjacent to the β-galactosidases (Fig 2-3). Homology does not clearly indicate the exact substrate specificity of the transporters because the annotations of most homologous matches are putative assignments and not experimentally determined. However, the presence of these ORFs does suggest active transport of mono- or oligosaccharides. Conceptually, these transporters could be transporting either monosaccharide products resulting from extracellular β-galactosidase activity or oligosaccharidic substrates for intracellular β-galactosidases to act on. Analysis identifies the β-galactosidases as intracellular, indicating the latter, not the former.

The second arrangement had an associated ORF putatively encoding a homolog to the AraC transcriptional regulators, a widely distributed group of proteins that can function in regulating carbon metabolism. The third arrangement presents an adjacent β-galactosidase belonging to a different family (GHF 35), which would seem redundant if
both were for the hydrolysis of lactose. However, Coombs et al. studied a similar arrangement in a *Carnobacterium* sp. and suggested the enzymes acted in a synergistic fashion on an unknown poly- or oligosaccharide other than lactose (11). In spite of their differences, all three arrangements suggest an di- or oligo- saccharide substrate, as expected for a glycoside hydrolase. Further clues would have been suggested by the presence of adjacent lipases or proteases (suggesting a lipopolysaccharide or proteoglycan substrate) or the presence of an ORF encoding a polysaccharide lyase, or endo-acting glycoside hydrolase (giving a clearer indication of the likely source of the oligosaccharides). However, oligosaccharides from any of these sources could be released into the environment by the extracellular enzymes of other organisms, obviating the need a given microorganism to perform these preliminary degradation steps itself.
2.5 Materials and Methods

**Inocula.** Samples used in the enrichments for spore-formers were aseptically collected in winter from Cheesequake salt marsh, NJ; a sewage plant, PA; bodies of fresh water Rochester, NY; smoked salmon; and lake shores in Michigan and stored at 4°C. Aliquots of previous ongoing psychrophilic enrichments inoculated earlier using material from Jack’s Mt, PA; Bear Meadows Bog, PA; and Rochester, NY were also used.

**Selection and screening.** Selection for spore-forming bacteria was performed by exposing the samples to 70°C for 10 minutes in order to kill vegetative cells, but allow the survival of spores. This is a milder heat treatment than typically used for such a selection, but was used due to the reduced thermal resistance observed in spores from psychrophilic bacteria (18, 21). After heating, the samples were spread onto media and incubated at 2°C to enrich for psychrophiles. Nutrient agar medium was avoided in favor of TSA (tryptcase soy agar without dextrose) (Difco) and R2A (Difco) (28) media because previous results indicated that many environmental isolates grow poorly on it (not shown). Several authors have also noted poor growth on nutrient agar with regard to specific species of *Bacillus* (1, 20), *Paenibacillus* (37) and *Sporosarcina* (19). Cycloheximide was included in some media to reduce yeast and other fungal colonies from spores that could have survived the mild heat treatment. The resulting colonies were patched onto media containing the chromogen X-Gal (which turns blue when hydrolyzed by a β-galactosidase) and either kept at 2°C or moved to 10°C. Psychrophilic isolates from other laboratory projects identified as belonging to the families *Planococcaceae*, *Paenibacillaceae*, or *Bacillaceae* (Fig 2-1) were also examined.
**Cloning of β-galactosidase genes.** Chromosomal DNA was extracted from isolates by treating harvested cells with lysozyme (0.5% in Tris (10mM)-EDTA (0.5mM) buffer, pH 8) and then using the PureGene kit (Gentra Systems Inc. Minneapolis, MN) with the modification of lysing cells at 85°C for 10 minutes. The DNA was typically partially digested with the restriction enzyme *Pst*I, and used to create genomic libraries in the pΔα18 (*lacZ*, amp<sup>R</sup>) (40) vector using competent *E. coli* ER2585F’ (*lacY<sup>+</sup>, *lacZ*, *thi*, tet<sup>R</sup>) as a host. This host and vector combination lacks the ability to produce the native β-galactosidase (LacZ) of *E. coli*. The vector was made competent using the Z-competent kit (Zymo research) and transformed on ice. Transformants were spread on LB (Luria-Bertani) media, selected via ampicillin resistance, and screened for the hydrolysis of X-Gal (in the presence of IPTG (isopropyl-beta-D-thiogalactopyranoside) first at 37, and then at 18°C. Transformants that began hydrolyzing X-Gal only after the downward temperature shift, as indicated by the production of blue pigment, were expected to encode a cold-active β-galactosidase.

**Characterization of isolates.** The previously isolated chromosomal DNA was also used as a template for PCR amplification of 16S rRNA genes using Ready-To-Go beads (Amersham Pharmacia, Piscataway, NJ) and two universal primer pairs 8F with 907R, and 704F with 1492R. The overlapping products were sequenced at the Penn State Nucleic Acid Facility with an ABI Hitachi 3100 Genetic Analyzer. The resulting sequences were examined using BLAST searches of the National Center of Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov) and searches using the Ribosomal Database Project (RDP) (http://rdp.cme.msu.edu/html).
Thermal growth ranges of the isolates was explored by incubating TSA and/or R2A plates streaked with the isolate at 2, 10, 18, 25, 30 and 37°C and monitoring growth daily. Growth was defined as the formation of isolated colonies. Pattern formation by motile colonies was observed using the media and staining methods described by Ben-Jacob and colleagues (6, 7, 39). The photographs shown were representative of growth on low peptone medium (Fig 2-2A) and TSA (Fig 2-2B).

**Evaluation of thermal dependencies of β-galactosidase activity.** *E. coli* cells heterologously expressing β-galactosidases were grown in TB (terrific broth) (38) at 37°C until a OD$_{600}$ of 0.5, chilled to 18°C, induced with IPTG (100 µM, final), and incubated overnight. The cells were then harvested by centrifugation at 6,370 x g at 4°C for 11 minutes. The resulting cell pellet was resuspended in 3 mL of modified Z-buffer (without β-mercaptoethanol (22)) per gram of cell pellet. For preliminary assays the resuspended cells were lysed with chloroform and 0.1% SDS. For later assays the resuspended cells were disrupted with a single pass through a French pressure cell (18,000 lb/in$^2$), and centrifuged again (30,996 x g, 4°C, 30 min). A consistent volume between 10 – 50 µl of clarified lysate was used to initiate enzyme assays for thermal dependency. The reaction buffer for these consisted of 1.2 mL of modified Z-buffer with 2.2 mM *o*-nitrophenyl-β-D-galactopyranoside (ONPG) (Sigma) that was incubated at 25°C for 15 minutes prior to starting the reaction. The reactions were stopped with 0.5 mL of 0.5 M Na$_2$CO$_3$ and the change in absorbance at 420 nm, indicating release of the *o*-nitrophenyl group was measured. One unit of activity was defined as the release of 1 µmol of ONP per minute. Specific activity was expressed as units per milligram of protein (in clarified lysate) as determined using the Bio-Rad (Hercules, California) protein assay dye reagent protocol.
**Sequencing of cloned fragments.** DNA inserts were sequenced by using vector primers in combination with subcloning and primer walking at the Penn State Nucleic Acid Facility with an ABI Hitachi 3100 Genetic Analyzer. Compiled sequences were used in BLAST searches of the National Center of Biotechnology Information (NCBI) database ([http://www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) to identify the GHF of each β-galactosidase and identify the potential enzymes encoded by adjacent ORFs. The N-terminal region sequences (according to conceptual translations of the sequenced genes) of the β-galactosidases were analyzed using Signal pWWW ([http://www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/))(5) with neural networks trained on Gram-positive data to detect whether they lacked probable signal peptide cleavage sites, and thus were likely to be intracellular.
2.6 References


Chapter 3

Characterization of an unusual cold-active β-glucosidase belonging to family 3 of the glycoside hydrolases from the psychrophilic isolate Paenibacillus sp. C7
3.1 Summary

I enriched for spore-forming psychrophilic bacteria with β-galactosidase activity and one isolate, designated *Paenibacillus* sp. C7, was phylogenetically related to, but distinct from both *Paenibacillus macquariensis* and *Paenibacillus antarcticus*. Some *Escherichia coli* transformants obtained with genomic DNA from this isolate hydrolyzed X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside) only below 30°C, an indication of cold-active β-galactosidase activity. Sequencing of the cloned insert revealed an open reading frame encoding a 756-amino acid protein that, rather than belonging to a family typically known for β-galactosidase activity, belonged to glycoside hydrolase family 3, a family of β-glucosidases. Because of this unusual placement, the recombinant enzyme (BglY) was purified and characterized. Consistent with its classification, the enzyme had seven times higher activity with the glucoside substrate ONPGlu (*o*-nitrophenyl-β-D-glucopyranoside) than with the galactoside, ONPGal (*o*-nitrophenyl-β-D-galactopyranoside). In addition, the enzyme had, with ONPGlu, a thermal optimum around 30 to 35°C, activity over a broad pH range (5.5 to 10.9), and an especially low $K_m$ (< 0.003 mM). Further examination of substrate preference showed that the BglY enzyme also hydrolyzed other aryl-β-glucosides such as helicin, MUG (methylumbelliferyl-β-D-glucopyranoside), esculin, indoxyl-β-D-glucoside (a natural indigo precursor), and salicin, but had no activity with glucosidic disaccharides or lactose. These characteristics and substrate preferences make the BglY enzyme unique among the family 3 β-glucosidases. The hydrolysis of a variety of aryl-β-glucosides suggests that the enzyme may allow the organism to use these substrates in the environment and its low $K_m$ on indoxyl-β-D-glucoside may make it useful for producing
3.2 Introduction

Within the classic Enzyme Commission system (E.C.), β-glucosidases collectively have the designation EC 3.2.1.21. Under Henrissat’s classification, β-glucosidases are grouped in two glycoside hydrolase families (GHFs), 1 and 3, and the β-galactosidases (E.C. 3.2.1.23) in four families, 1, 2, 35 and 42. Although classification into these systems can be made by studying hydrolysis of synthetic chemicals or by sequencing the gene encoding the enzyme, respectively, neither proves the natural substrates or functions of the enzyme.

The need for biochemical and physiological studies of microbial glycoside hydrolases is highlighted by the prevalence of open reading frames (ORFs) homologous to genes encoding GHF 1 and GHF 3 enzymes in a majority of analyzed bacterial genomes (10). Although the original interest in β-glucosidases focused on their participation in degrading the plant polymers cellulose and xylan (14), the importance of the interaction of β-glucosidases with plant-produced compounds goes beyond catabolic degradation and includes plant-phytopathogen interactions (6, 28, 31). There is also considerable interest in biotechnological applications of β-glucosidases relating to plant-based foods, for instance, converting phytoestrogen glucosides in fruits and vegetables to aglycone moieties, detoxification of cassava, aroma enhancement, and removing bitter compounds from citrus fruit juices or unripe olives (5). The β-glucosidase genes present in sequenced genomes could encode enzymes with any of these functions, or given the widespread occurrence of glucosylated compounds found in plants, additional unknown functions, some of which might contribute to plant-microbe signaling. Thus, the annotation of an ORF as a glycoside hydrolase alone provides little information about its
biochemical traits or physiological function. This gap in our knowledge illustrates the need to biochemically characterize these prevalent enzymes to provide the information needed to begin understanding their biological roles.

The Brenchley research group has been investigating psychrophilic microorganisms and their cold-active glycoside hydrolases (7, 9, 26, 36) with particular interest in their physiological roles. As part of this work, I enriched for and isolated numerous psychrophilic spore-forming bacteria to obtain phylogenetically related organisms with β-galactosidases active on the chromogen 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-Gal) at low temperatures. One transformant from a genomic library created from a Paenibacillus isolate, C7, hydrolyzed X-Gal, but sequence analysis grouped the enzyme responsible (BglY) in GHF 3, a family typically lacking appreciable β-galactosidase activity. Because of this interesting placement of an enzyme with β-galactosidase activity into GHF 3, I purified the enzyme and examined its thermal properties, pH dependence, substrate specificities, and kinetic characteristics.
3.3 Results

3.3.1 Characterization of the C7 isolate.

Numerous psychrophilic spore-forming isolates were obtained following a heat treatment to kill vegetative cells as described in Materials and Methods. The isolate designated as C7 was chosen for further study because it grew on lactose and hydrolyzed X-Gal at 2°C, an indicator of cold-active β-galactosidase activity. Microscopic characterization showed that the cells were Gram-negative rods during all growth stages and some contained ellipsoidal spores that formed terminally to subterminally in swollen sporangia. The C7 isolate developed colonies on TSA (trypticase soy agar without dextrose) at 25 and 18°C within 48 hours, at 10°C after five to six days, and at 2°C after two weeks, but no growth was observed at 30°C or 37°C. Colonies formed aerobically on TSA and R2A, but the isolate did not grow on nutrient agar.

Results of 16S rRNA gene phylogenetic analysis showed that isolate C7 robustly grouped within the *Paenibacillus* genus (Fig 3-1), a group of aerobic spore-formers (2) with species that produce polymer-degrading enzymes, such as xylanase (44), agarase (43), gelatinase (21), curdlanase (18), etc. Even though basal branches on some trees collapsed to polytomy, on average, the distance matrices yielded a 2.5 and 2.4% difference between isolate C7 and *Paenibacillus antarcticus* and *Paenibacillus macquariensis*, respectively (Fig 3-1).
Figure 3-1. Phylogenetic relationships of 16S rDNA sequences of isolate C7 and related Paenibacillus spp., based on a distance analysis (neighbor-joining algorithm with Jukes-Cantor model). Bootstrap values shown at the nodes were generated from 10,000 replicates. GenBank accession numbers are listed in materials and methods.

Further physiological characterization using API 50 test strips showed that, despite similarities, the phenotype of isolate C7 differed from those reported for P. antarcticus (27) and P. macquariensis (22) as well as for the related Paenibacillus spp. P. borealis (13), P. odorifer (4), and P. graminis (4) (data not shown). Some differences included the inability of isolate C7 to produce acid from glycogen, unlike P. macquariensis (13), or grow in media containing 3% NaCl, a characteristic of P. antarcticus (27). P. macquariensis has also been identified as Gram negative under all conditions (22), like isolate C7, whereas P. antarcticus was observed to be Gram variable (27). Based on my results, I designated the isolate C7 simply as Paenibacillus sp. C7
until future analyses determine whether it can be classified with either *P. antarcticus* or *P. macquariensis*, or as a new species.

### 3.3.2 Cloning of a gene encoding β-galactosidase activity.

Of the approximately 22,000 ampicillin-resistant transformants obtained from the genomic library from *Paenibacillus* sp. C7 cloned in pΔα18 and expressed in *Escherichia coli* ER2585F', all were white at 37°C after 16 h. However, when the plates were transferred to 18°C, three colonies became blue within 24 h indicating X-Gal hydrolysis at the lower temperature. The clarified lysates from these transformants hydrolyzed ONPGal (*o*-nitrophenyl-β-D-galactopyranoside) and ONPGlu (*o*-nitrophenyl-β-D-glucopyranoside), while clarified lysate from cells carrying the vector alone did not. The plasmids purified from these three transformants all contained 13.5 kb inserts with the same *Pst*I restriction patterns. Following subcloning of one of these, sequencing revealed that one open-reading frame, designated *bglY*, encoded a family 3 glycoside hydrolase.

### 3.3.3 Analysis of the *bglY* gene.

The gene *bglY* encoded the enzyme responsible for the activity on X-Gal and the expected translation of *bglY* had 68% identity to BglB, a GHF 3 enzyme from *Bacillus* sp. GL1 (17). Consistent with this homology, the BglY amino acid sequence possessed the conserved putative aspartate catalytic residue (20, 32) (position 246) in the motif commonly referred to as SDW, but represented in this enzyme as TDW. The assignment to GHF 3 was notable because I expected the enzyme to group with families known to
have β-galactosidase activities (GHF 2, 35, and 42), or with the one β-glucosidase family (GHF 1) frequently found to also have β-galactosidase activity.

The genes homologous to bglY belong to cluster F, one of several subgroups within GHF 3. This subgroup contains four enzymes as described by Cournoyer and Faure (10), two of which are known as AB’ enzymes because the second domain of the enzyme, B, is “truncated” compared to AB-type GHF 3 enzymes. My alignments (not shown) indicate that BglY is about 100 amino acids shorter than the AB enzymes, but not as truncated or condensed as the two in the AB’ group, BgxA of Erwinia chrysanthemi (45) and SalB of Azospirillum irakense (15).

3.3.4 Analyses of bglY and neighboring sequence regions.

Further analysis was performed on bglY and adjacent sequences to determine whether an operon existed that would provide clues to the enzyme’s function and relevant substrates. The sequence revealed two additional open-reading frames and a partial ORF. The NCBI (National Center of Biotechnology Information) database comparisons of the deduced amino acid sequences assigned the COG (Clusters of Orthologous Groups) (http://www.ncbi.nlm.nih.gov/COG/) (41) classifications 1082, 0673, and 2972 to orfA, orfB, and orfC, respectively (Fig 3-2). These correspond, in order, to IolE, sugar phosphate isomerases/epimerases; MviM, predicted dehydrogenases and related proteins, and, predicted signal transduction proteins with C-terminal ATPase domains (Fig 3-2). BglY was classified as COG 1472, BglX, beta-glucosidase-related glycosidases.
An examination of the fragment for potential regulatory regions showed that the sequence ANGGNGG, which resembles the “ideal” ribosome binding site in Bacillus subtilis (29), exists ten to eleven nucleotides before a putative translation start codon for each gene (Fig 3-2). The genes orfA and orfB may be expressed as an operon (Fig 3-2). However, the presence of putative rho-independent transcription terminators between the other genes and no association among their homologs suggests that bglY is not
cotranscribed with orfA and orfB, or with orfC. Thus, these nearby genes do not currently provide clues to the physiological role of BglY.

To determine whether any of the enzymes might be membrane associated or extracellular, I analyzed the sequences using the Dense Alignment Surface Method and the Signal pWWW with neural networks trained on Gram-positive data. The searches indicated that the protein encoded by orfC possesses a possible signal cleavage site and may be an integral membrane protein (as expected for a signal transduction protein). These same analyses, however, did not detect similar properties in orfA, orfB or bglY, indicating that their putative proteins are probably intracellular and not integral membrane proteins.

### 3.3.5 Enzyme purification.

The N-terminal six-His tagged BglY enzyme was purified for the purpose of examining its biochemical properties (Table 3-1). The enzyme was stable at 4°C but the specific activity gradually decreased over a month’s time during storage. SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) analysis of the recombinant BglY enzyme preparation showed that it was at least 95% pure and had an apparent molecular mass of 81 kDa (data not shown), which is comparable to the calculated molecular weight of 83,808 with the added His-tag.

| Table 3-1. Purification scheme for (His-tagged) BglY from *E. coli.* |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Step*           | Vol. (ml)   | Total Protein (mg) | Total Activity (U) | Sp. Act. (U mg⁻¹) | Yield (%) | Purification (fold) |
| Clarified lysate| 27.5        | 743          | 664          | 0.9          | 100        | 1.0          |
| Ammonium sulfate| 38.0        | 456          | 602          | 1.3          | 91         | 1.5          |
| Nickel-IDA      | 5.0         | 10           | 171          | 17.1         | 26         | 19.1         |
| Dialysis        | 5.0         | 10           | 254          | 25.4         | 38         | 28.4         |

* For purification procedure details, see Materials and Methods
3.3.6 Effects of temperature and pH on activity.

Because of its initial demonstration of $\beta$-galactosidase activity, the BglY enzyme was first assayed with ONPGal as a substrate but was also assayed with the glucosidic substrates X-Glc (5-bromo-4-chloro-3-indoyl-$\beta$-D-glucopyranoside) and ONPGlu because of its placement in GHF 3, a $\beta$-glucosidase assemblage. The enzyme hydrolyzed all of these substrates, and the activity with ONPGlu was seven times higher than with ONPGal (Table 3-2).

Table 3-2. Relative activity of the purified BglY enzyme on various chromogenic substrates as measured by ONP or PNP release at 25°C.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Relative activity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>o-nitrophenyl-$\beta$-D-glucopyranoside (ONPGlu)</td>
<td>100</td>
</tr>
<tr>
<td>p-nitrophenyl-$\beta$-D-glucopyranoside (PNPGlu)</td>
<td>56</td>
</tr>
<tr>
<td>o-nitrophenyl-$\beta$-D-galactopyranoside (ONPGal)</td>
<td>14</td>
</tr>
<tr>
<td>p-nitrophenyl-$\beta$-D-galactopyranoside (PNPGal)</td>
<td>3</td>
</tr>
<tr>
<td>o-nitrophenyl-$\beta$-D-fucopyranoside</td>
<td>4</td>
</tr>
<tr>
<td>o-nitrophenyl-$\beta$-D-xylopyranoside</td>
<td>0.7</td>
</tr>
<tr>
<td>p-nitrophenyl-$\alpha$-D-glucopyranoside</td>
<td>0.1</td>
</tr>
<tr>
<td>p-nitrophenyl-$\alpha$-D-galactopyranoside</td>
<td>&lt; 0.5</td>
</tr>
<tr>
<td>p-nitrophenyl-$\beta$-D-N-acetyl glucosaminide</td>
<td>&lt; 0.5</td>
</tr>
</tbody>
</table>

*Activity on ONPGlu taken as 100%, and corresponds to a specific activity of 21 U/mg.

The thermodependency of activity results showed that the highest specific activity with ONPGlu was around 30 to 35°C, whereas optimal activity with ONPGal was at 25°C (Fig 3-3) and was equal to 15% of the ONPGlu activity at 25°C. These thermal optima compare well with data obtained using clarified lysate containing heterologously expressed non-His-tagged BglY (data not shown). The purified enzyme demonstrated 5% of its activity at 0°C with both substrates. Thermal stability studies using ONPGlu showed that the BglY enzyme was stable at 25°C for over an hour (data not shown), but lost 23% activity after 10 minutes at 30°C and 85% after only 5 minutes at 40°C (Fig 3-4).
Fig 3-3. Thermal dependencies of activity of the purified BglY enzyme.

Figure 3-3. Thermal dependency of activity of the purified BglY enzyme with ONPGlu (–○–) and ONPGal (––□–). The specific activity corresponding to 100% was 18 U/mg with ONPGlu and 3 U/mg with ONPGal.

Fig 3-4. Thermostability of purified BglY.

Figure 3-4. Thermostability of purified BglY versus time of incubation at various temperatures: 25°C (□), 30°C (◊), 35°C (Δ), 40°C (○). The specific activity corresponding to the 100% value was 13 U/mg.
The enzyme was active over a broad pH range. The optimal activity was between pH 7 and 8 with Zm (Fig 3-5), with phosphate and PIPES (piperazine-N,N'-bis {2-ethanesulfonyl} acid) buffers providing roughly equivalent levels of activity at pH 7 (96%, 93%). Activity in the MOPS (morpholinepropanesulfonic acid) buffer was somewhat less (80%). The enzyme retained over 75% of the optimal activity between pH values 6.5 and 9 and at least 50% between pH values 6 and 10 and had residual activity at pH 10.9 and 5.5, but none at pH 5. Inclusion of 50 mM β-mercaptoethanol in the Zm buffer (modified Z-buffer) gave only 91% of the control activity. The enzyme was also stable (> 80% activity recovered after 24 h incubation in different buffers) throughout the pH range of pH 6 in Zm to 10.9 in carbonate buffer (data not shown). However, BglY lost activity in citric acid buffer at pH 6 or lower.

**Fig 3-5. Effects of pH on ONPGlu hydrolysis by purified BglY.**

![Graph showing the effects of pH on ONPGlu hydrolysis by purified BglY.](image)

Figure 3-5. Effects of pH on ONPGlu hydrolysis by purified BglY at 25°C using the following buffers: citric acid buffer (□), sodium acetate buffer (×), Zm (◇), PIPES (▲), MOPS buffer (+), Clark and Lubs buffer (Δ), carbonate buffer (○). Molarities and pH as described in methods. The specific activity corresponding to 100% was with ONPGlu in Zm, pH 7.5.
3.3.7 Effects of metal ions on activity.

The effects of metal ions on activity were first studied by dialyzing the enzyme in MOPS containing various metals and then assaying in the presence of the same metals. Compared to the activity determined in MOPS without metals the addition of 1 mM Mg$^{2+}$, 1 mM Ca$^{2+}$, 1 mM Mn$^{2+}$, 10 mM K$^+$, or 10 mM Na$^+$ had no effect. Assays with 1 mM Cu$^{2+}$, however, caused a 58% loss in activity. Only a slight activity loss (19%) was observed when the enzyme was assayed at 25°C after treatment with 50 mM EDTA at 0°C for 30 minutes. However, this same EDTA treatment at 25°C caused a 90% loss in activity that was partially restored with the addition of the cations Ca$^{2+}$, Mg$^{2+}$, and Mn$^{2+}$ (Table 3-3). A parallel control reaction demonstrated that the effects of the Sephadex column purification on the enzyme in the absence of EDTA treatment caused less than a 20% loss in activity. Incubating the treated enzyme with the metals at 25°C for 30 minutes prior to the assay did not affect the amount of activity recovered, nor did the addition of 10 mM KCl with 1 mM Ca$^{2+}$, Mg$^{2+}$, or Mn$^{2+}$.

**Table 3-3. Effects of ions on the activity of EDTA-treated BglY enzyme**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Concen (mM)</th>
<th>Relative activity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>EDTA, 30 min at 25°C</td>
<td>50</td>
<td>9</td>
</tr>
<tr>
<td>25°C EDTA treated, after column treatment:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MgCl$_2$</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>MgCl$_2$</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td>MgCl$_2$ and MnCl$_2$</td>
<td>1 each</td>
<td>48</td>
</tr>
<tr>
<td>MgCl$_2$ and CaCl$_2$</td>
<td>1 each</td>
<td>48</td>
</tr>
<tr>
<td>MnCl$_2$</td>
<td>0.1</td>
<td>44</td>
</tr>
<tr>
<td>MnCl$_2$</td>
<td>1</td>
<td>47</td>
</tr>
<tr>
<td>MnCl$_2$</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>MnCl$_2$ and CaCl$_2$</td>
<td>1 each</td>
<td>51</td>
</tr>
<tr>
<td>CaCl$_2$</td>
<td>0.1</td>
<td>44</td>
</tr>
<tr>
<td>CaCl$_2$</td>
<td>1</td>
<td>44</td>
</tr>
<tr>
<td>CaCl$_2$</td>
<td>10</td>
<td>47</td>
</tr>
<tr>
<td>CuCl$_2$</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>KCl</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>NaCl</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

* The specific activity at 100% is 14 U/mg
3.3.8 Substrate preference studies.

Synthetic chromogens were chosen based on the range of substrates hydrolyzed by other GHF 3 enzymes. The chromogenic substrate yielding the highest activity was ONPGlu. The enzyme was specific for the β-linkage (Table 3-2). The enzyme had greater activity on ONP substrates than PNP substrates even though studies with other GHF 3 enzymes more frequently report results using PNP chromogens. Although the enzyme possessed the greatest activity with the chromogenic β-glucoside substrates, it also had activity with ONPGal, and low activity on PNPGal (p-nitrophenyl-β-D-galactopyranoside) and o-nitrophenyl-β-D-fucopyranoside. Minimal activity (< 1 % of ONPGlu) was detected using o-nitrophenyl-β-D-xylopyranoside and p-nitrophenyl-β-D-N-acetyl glucosaminide.

I examined the hydrolysis of aryl-substrates with different structures (Fig 3-6) because the only cluster F enzyme with an identified function, SalB, is an aryl-β-glucosidase. BglY released glucose from both chromogenic, fluorogenic, and natural aryl-glucoside substrates (Table 3-4), but did not release significant amounts of glucose (less than 0.5 % of that released from ONPGlu) from any of the disaccharides, the cyanogenic substrate amygdalin, or a substrate with an alkyl aglycone, n-octyl-β-D-glucoside. The highest activity was with ONPGlu followed by helicin, a partially oxidized form of salicin. Even though salicin has a similar structure, activity was less than with helicin. Puerarin, a glucosylated flavonoid, was not significantly hydrolyzed nor was arbutin. The enzyme was also active with MUG (methylumbelliferyl-β-D-glucopyranoside), a synthetic coumaric substrate, and esculin, a natural coumaric substrate. Of special interest because of the possible application for dye production was
the hydrolysis of indican (indoxyl-β-D-glucoside), an indole glucoside, with concomitant
synthesis of indigo. The blue color developed as the reaction occurred at 25°C, but
intensified during inactivation of the enzyme at 65°C due to the oxidation and
dimerization of the intermediates to form indigo. Color formation during heating of a
control with indoxyl-β-D-glucoside but without enzyme was not significant.

Table 3-4. Relative activity of the BglY enzyme on various aryl
substrates as monitored by glucose release at 25°C.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Relative activity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONPGlu</td>
<td>100</td>
</tr>
<tr>
<td>Helicin</td>
<td>65</td>
</tr>
<tr>
<td>MUG</td>
<td>37</td>
</tr>
<tr>
<td>Escllin</td>
<td>26</td>
</tr>
<tr>
<td>Indican</td>
<td>21</td>
</tr>
<tr>
<td>Salicin</td>
<td>9</td>
</tr>
<tr>
<td>Arbutin</td>
<td>&lt; 0.5</td>
</tr>
<tr>
<td>Amygdalin</td>
<td>&lt; 0.5</td>
</tr>
<tr>
<td>Puerarin</td>
<td>&lt; 0.5</td>
</tr>
</tbody>
</table>

*Activity on ONPGlu taken as 100% which corresponds to a specific activity of 25 U/mg.

Fig 3-6. Structures for some of the substrates used in assays with purified BglY.

Figure 3-6. Structures for some of the substrates used in assays with purified BglY. (A) Phenolic substrates: ONPGlu, X = H, R = NO₂; helicin, X = H, R = CHO; salicin, X = H, R = CH₂OH; PNPG, X = NO₂, R = H; arbutin, X = OH, R = H; (B) Coumaric substrates: esculin, R = H; MUG, R = CH₃; (C) Indolyl substrates: Indican, X = Y = H, R = O-β-D-Glucose; X-Glc, X = Cl, Y = Br, R = O-β-D-Glucose; indole-3-acetic acid-glucoside, X = Y = H, R = CH₂CO-O-β-D-Glucose.
3.3.9 Kinetic studies.

The $K_m$ value for BglY was determined with PNPGlu ($p$-nitrophenyl-β-D-glucopyranoside) to be low (4.9 µM), and was 2 mM with ONPGal (Table 3-5) at 25°C. For comparison, the $K_m$ value with PNPGlu was also determined using clarified lysate containing non-His-tagged BglY and was found to be in the same range, 3.3 ± 0.6 µM, as that measured using the tagged enzyme. The kinetic constants for ONPGlu were estimated because the substrate concentrations used at these low $K_m$ values and the $o$-nitrophenol extinction coefficient were too low to accurately measure the initial velocity of product formation under the conditions used. However, my preliminary results indicate that the $K_m$ value for ONPGlu is also low (below 3 µM) since no significant increase in velocity occurs at greater substrate concentrations, suggesting saturation has been attained. Kinetics were also performed with indoxyl-β-D-glucoside by monitoring the increasing absorbance occurring at 668 nm due to indigo production (via spontaneous dimerization of the indoxyl product) at 25°C and the $K_m$ value was determined to be 0.2 mM.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (mM)</th>
<th>$V_{max}$ (µmol mg$^{-1}$ min$^{-1}$)</th>
<th>$k_{cat}$ (sec$^{-1}$)</th>
<th>$k_{cat}/K_m$ (mM$^{-1}$ sec$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNPGlu</td>
<td>0.0049 ± 0.0004</td>
<td>19.4 ± 0.6</td>
<td>27</td>
<td>5500</td>
</tr>
<tr>
<td>ONPGal</td>
<td>1.8 ± 0.5</td>
<td>5.9 ± 0.6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>ONPGlu</td>
<td>&lt; 0.003</td>
<td>~25</td>
<td>~35</td>
<td>n.c.*</td>
</tr>
<tr>
<td>Indican</td>
<td>0.2</td>
<td>u.d.$^{†}$</td>
<td>u.d.$^{†}$</td>
<td>u.d.$^{†}$</td>
</tr>
</tbody>
</table>

n.c.* Not calculated due to uncertainty in $K_m$ value
u.d.$^{†}$ Undetermined: The $V_{max}$ observed is not the true $V_{max}$ due to production of alternate products to indigo, which forms as a secondary reaction after the release of indoxyl from indican.
3.4 Discussion

3.4.1 Characterization of C7 and BglY and comparison to other GHF 3 enzymes

The C7 isolate is a spore-former that grows between 2 and 25°C and hydrolyzes X-Gal, consistent with the selection and screening protocols used for its isolation. The C7 isolate is clearly a member of the *Paenibacillus* genus, however, further characterization of the isolate may warrant its designation as a new species because it is phylogenetically at least 2.3% distant from both the *P. macquariensis* and *P. antarcticus* species.

Consistent with this isolate’s growth at low temperatures, the GHF 3 enzyme encoded by the cloned *bglY* gene is cold-active, with a thermal optimum of 30 to 35°C with ONPGlu and 5% activity at 0°C. This optimum is half the 60°C optimum of the mesophilic enzyme BgI from *Clostridium thermocellum*, an enzyme that has only about 5% at 20°C (34), with PNPGlu.

Interestingly, the thermal optimum of BgIY with ONPGal is lower than that found with ONPGlu. Although infrequently studied in GHF 3 enzymes, variations in thermal optima with different substrates were observed with the GHF 3 β-glucosidase from *C. thermocellum* (34). However, the divergence with that enzyme could have been due to the different assay conditions used, whereas the differences in thermal optima for BgIY with ONPGlu and ONPGal were observed using identical assay conditions. As frequently found with other cold-active enzymes, BgIY is heat-labile and loses activity at 30°C. However, the BgIY enzyme is extremely stable over a wide pH range, maintaining over 80% activity following incubation in buffers ranging from pH 6 to 10.9. This stability is surprising, as is the enzyme’s activity over a broad pH range.
Results of initial ion studies with BglY were similar to those found with other GHF 3 \(\beta\)-glucosidases with inhibition by copper, no activation by the addition of divalent ions (\(\text{Mg}^{+2}\), \(\text{Mn}^{+2}\), or \(\text{Ca}^{+2}\)), and no effect from EDTA treatment (17, 19, 30, 46). However, I also obtained results suggesting that EDTA had increased access to metal ions at a higher temperature, resulting in enzyme that could not be restored to full activity by reincorporation of divalent ions. This failure to restore activity may be due to an inability to obtain the proper conformation necessary for re-incorporation of metal ions that may be required for structural stability, but not catalytic activity. It is possible that this phenomenon might be observed with other GHF 3 enzymes if additional EDTA treatment conditions were used.

Not only does BglY have an unusual combination of temperature, pH, and ion characteristics, it is also atypical because it hydrolyzes both chromogenic \(\beta\)-galactosides and \(\beta\)-glucosides. Enzymes active on both \(\beta\)-galactosides and \(\beta\)-glucosides are frequently found in GHF 1, but not GHF 3. Activity with PNPGal has been reported for about two dozen microbial GHF 3 enzymes. With a single exception, this activity was less than 3% when compared with another appropriate substrate such as PNPGlu, PNPX (\((p\text{-nitrophenyl-})\beta\text{-D-xylopyranoside})\), or PNPNAG (\((p\text{-nitrophenyl-})\beta\text{-D-N-acetyl glucosaminide}\)). In contrast, both BglB from \textit{C. thermocellum} and BglY are unique in having twice this relative activity with PNPGal versus PNPGlu (34).

An additional biochemical feature of the BglY enzyme is its low \(K_m\) value of 4.9 \(\mu\text{M}\) with PNPGlu as the substrate. This \(K_m\) is 10-fold lower than that of the only cluster F enzyme with published kinetic analyses, SalB from \textit{A. irakense} (50 \(\mu\text{M}\), PNPGlu) (15). Enzymes in the subcluster C3, distant from cluster F (10), are the only other GHF 3
enzymes with $K_m$ values in this low range. One of these is from *Thermotoga maritima* with a $K_m$ of 3.9 µM on PNPGlu (16). The other is an enzyme from *Agrobacterium tumefaciens* with a $K_m$ value of 5 µM on PNPX (37). This enzyme also releases coniferyl alcohol from coniferin, an aryl-β-glucosidase, as part of *vir*-gene induction (6), and has a low $K_m$ (23 µM) on this substrate (47). The combination of the BglY enzyme’s ability to hydrolyze a variety of aryl glycosides and these low $K_m$ values makes it of biochemical interest.

### 3.4.2 Possible functions of BglY

In an attempt to define its physiological function, I compared BglY with other enzymes. The only other known GHF 3 enzyme from a *Paenibacillus* sp. (32) belongs to the distant subcluster C3, but the most homologous enzymes to BglY are those belonging to cluster F, such as BglB from *Bacillus* sp. GL1. (Although listed as a *Bacillus* sp., the 16S rDNA sequence most closely resembles sequences from other *Paenibacillus* spp. {data not shown}). The *Bacillus* sp. GL1 enzyme, BglB, has the highest identity to BglY and is well characterized, but its function is unknown. Within cluster F, SalB is the only enzyme with an identified function. This enzyme allows *A. irakense* to use salicin as a carbon source (15). The BglY enzyme also hydrolyzes salicin and *Paenibacillus* sp. C7 grows on this compound; it will be interesting to determine whether the BglY enzyme has a function similar to SalB. However, several factors suggest that the functions of both SalB and BglY may be more complex. For one, salicin might not be the only substrate of SalB (15, 38). Second, the *salB* gene in *A. irakense* is in an operon with a second GHF 3 gene and a transporter, with a LacI-type repressor nearby, whereas *bgly* does not appear as part of an operon. And third, BglY lacks the signal peptide found with SalB suggesting
the two enzymes could have different cellular locations. Examination of the genes adjacent to other cluster F enzymes does not reveal any consistent gene arrangements similar to those observed with salB, or bglY, nor to each other. Thus, no conserved functional associations can be made based on operon structure.

It is possible that the physiological role of the BglY enzyme is associated with its ability to degrade substrates with aryl-aglycones because the enzyme hydrolyzes ONPGlu, PNPGlu, helicin, MUG, esculin, and salicin but not disaccharides. Therefore, the \textit{in vivo} function of BglY might also involve hydrolysis of aryl-β-glucosidic compounds. Some possibilities could be plant metabolites such as coniferin, scopulin, mono and di-galloyl esters of glucose (precursors of tannin), indole-3-acetic acid-glucoside (Fig 3-6), and glucosylated flavonoids. The combination of the enzyme’s activity with several aryl-glucosides and its very low $K_m$ values suggests that it could degrade secondary plant metabolites with aryl-aglycone groups found in low concentrations. This could be environmentally relevant since \textit{Paenibacillus} spp. are frequently associated with plants and their roots (4).

The BglY enzyme’s biochemical properties contribute to our knowledge of β-glucosidases in general, and the understudied cluster F in particular. Although putative GHF 3 enzymes are encoded in many genomes, no specific function for BglY was indicated by annotations of homologous genes. Also, there are few GHF 3 enzymes that have been assayed using a wide suite of substrates for comparison. This highlights the need for continued characterization of enzymes to supplement the experimental knowledge used for genome annotation in order to improve the precision of predicted functions. The β-galactosidase activity observed with BglY using chromogenic substrates
is probably not relevant to lactose degradation, but is a consequence of the enzyme’s recognition of aryl groups and my data suggests that BglY has a role other than the degradation of cellulose, xylan, or cell-wall components most often associated with GHF 3 enzymes. The ability of BglY to cleave indican (indoxyll-β-D-glucoside) might have applications for the production of indigo dye from *Polygonum tinctorium*. The specific activity of purified BglY on indican is 300 times greater than that reported for an enzyme studied for possible industrial use, Novarom G (23), and the $K_m$ of BglY with indican is slightly lower than that reported for the β-glucosidase produced natively by *P. tinctorium* (25).
3.5 Materials and Methods

**Isolation and characterization of isolate C7.** Water and mud samples from Bear Meadows Natural Area, PA (N 40° 44’ W 077° 45’, 554 m elevation) were inoculated into a variety of liquid minimal media with lactose as the carbon source and incubated at temperatures between 2 and 10°C for several months with periodic transfers. Spore-forming bacteria were then selected by heating aliquots at 70°C for 10 min to kill vegetative cells. The treated aliquots were inoculated into trypticase soy broth without dextrose and incubated aerobically at 2°C for 10 days. This enrichment was plated on TSA (trypticase soy agar without dextrose) modified to contain sporulation salts (35). The resulting colonies were screened for cold-active glycoside hydrolases, as indicated by the cleavage of X-Gal (100 µg/ml) on TSA and R2A (33) at 2 and 10°C. Isolate C7 hydrolyzed X-Gal and was studied further.

API 50 CH strips (bioMerieux, Inc. Hazelwood, MO) were used to examine the oxidation of various substrates by isolate C7. The instructions were followed except that the incubation was at 25°C, and the inoculum was grown on TSA because isolate C7 did not grow on the recommended nutrient agar. The intensity of acidification was recorded at 24 h intervals for 120 h. The strips were also inoculated in combination with an M9 minimal medium (24) with phenol red (0.18 g L⁻¹). Growth on TSA with sodium chloride at final concentrations of 0.5, 1, 2, 3, 5, 7, and 10% was tested. Also, growth was monitored on TSA plates at 2, 10, 18, 25, 30, and 37°C.

**16S rRNA gene amplification and phylogenetic analysis.** Isolate C7 cells were treated with lysozyme and the PureGene kit (Gentra Systems Inc. Minneapolis, MN) used to
obtain genomic DNA with the modification of increasing the heating step to 85°C for 10 min to promote cell lysis. The 16S rRNA gene was amplified by PCR from the genomic DNA using Ready-To-Go beads (Amersham Pharmacia, Piscataway, NJ) and universal primers 8F with 907R and 704F with 1492R. Sequencing was performed at the Penn State Nucleic Acid Facility (NAF) on an ABI Hitachi 3100 Genetic Analyzer.

The 16S rRNA gene sequence (1,458 bp) was used to search the National Center of Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov) (via BLAST) (1) and the Ribosomal Database Project II (RDP II) (http://rdp.cme.msu.edu/html) (8). The 16S rDNA sequences of related organisms were initially aligned using Clustal W (BioEdit platform, Version 5.0.6; Department of Microbiology, North Carolina State University (http://www.mbio.ncsu.edu/Bioedit/bioedit.html)) and the sequence of isolate C7 was then aligned manually. Data for the following Paenibacillus spp. were used (GenBank accession numbers follow the name of each): P. polymyxna IAM 13419T (AB042063), P. durus LMG 14658T (AJ251195), P. stellifer IS1T (AJ316013), P. odorifer TOD45T (AJ223990), P. borealis KK19T (AJ011322), P. antarcticus LMG 22078T (AJ605292), P. macquariensis DSM 2T (AB073193), P. pabuli HSCC 492T (AB045094), P. amylolyticus JCM 9906T (AB073190), P. chibensis JCM 9905T (AB073194), P. ehimensis KCTC 3748T (AY116665), P. koreensis YC300T (AF130254), P. elgii SD17T (AY090110), P. lautus JCM 9073T (AB073188), P. glucanolyticus DSM 5162T (AB073189), P. lactis MB 1871T (AY257868), and P. campinasensis JCM 11200T (AB073187).

This alignment (1,469 nucleotide positions for 18 taxa) was imported into PAUP (Version 4.0b10; School of Computational Sciences and Information Technology, Florida
State University (http://paup.csit.fsu.edu/) in order to create bootstrapped (1,000 replicate) phylogenetic trees. The maximum likelihood method was performed using both the Heuristic and Fast algorithms. Distance analyses were performed using the same algorithms, in addition to the neighbor-joining algorithm (with all 10 models available in PAUP) to generate both trees and distance matrices. The Jukes-Cantor method with equal rates for variable sites was repeated with 10,000 replicates. The phylogenetic trees and distance matrices were compared and the trees found to be congruent.

**Glycoside hydrolase gene cloning.** Genomic DNA from isolate C7 was partially digested with PslI and ligated into vector pΔα18 (42), a derivative of pUC18 that lacks the E. coli lacZ alpha fragment. The constructs were transformed into competent E. coli ER2585F' (ΔlacZ lacY') cells that were plated onto Luria-Bertani medium supplemented with ampicillin (100 µg/ml), X-Gal (100 µg/ml), and IPTG (isopropyl-β-D-thiogalactoside) (100 µM). Plates were incubated at 37°C for 16 h and then transferred to 18°C. Three of the transformants screened positively for X-Gal hydrolysis and plasmid preparations from these were obtained using the Promega Wizard Plus SV Miniprep kit for further subcloning.

Subcloning in pΔα18 yielded a 4.7 kb PslI/EcoRI insert that conferred the ability to hydrolyze both X-Gal and X-Glc at 18°C but not at 37°C. Sequence from this and an adjoining 0.9 kb PslI fragment were used to perform BLAST searches of the NCBI database. The sequence was examined manually and using Signalp WWW server (3) for putative transcription and translation control regions including -35 and -10 regions similar to the E. coli σ70 promoter. The Dense Alignment Surface Method (http://www.sbc.su.se/~miklos/DAS) (11) was used to predict transmembrane regions.
Construction of an expression vector. The PCR-Script™ Amp-Cloning kit (Stratagene) was used to create an *Nhe*I site at the beginning of the bg/Y gene contained in the 4.7 kb fragment. The resulting fragment and vector pET28a were each digested with *Sal*I and *Nhe*I, ligated, and transformed into MC1061 (DE3) competent cells. The new construct, pETC7, contained the bg/Y gene with the coding region for a six-histidine tag at the N-terminus of the protein.

Enzyme purification. *E. coli* MC1061 (DE3) cells transformed with pETC7 were grown in 500 ml of terrific broth (40) with 30 µg/ml of kanamycin at 37°C until turbidity reached an optical density at 600nm of 0.4. The culture was then moved to 18°C, IPTG added (100 µM final concentration), and incubation continued for 15 h. The cells were harvested by centrifugation (6,370 x g, 4°C, 11 min). The cell pellet was resuspended with 3 ml/g of modified Z-buffer (Zm buffer) (100 mM sodium phosphate buffer, 10 mM KCl, 1 mM MgSO₄, pH 7), which does not contain β-mercaptoethanol (24), the cells disrupted with a single pass through a French pressure cell (18,000lb/in²), and centrifuged (30,996 x g, 4°C, 30 min). Saturated ammonium sulfate was added to the clarified lysate to a final concentration of 35% at 0°C and the mixture incubated for 30 minutes. After centrifugation (30,996 x g, 4°C, 30 min), the undialyzed supernatant was loaded onto a nickel-charged-IDA (iminodiacetic acid) column at 4°C.

The column wash solution was Zm buffer containing 300 mM NaCl and imidazole at three different concentrations. The column was first washed with buffer containing 5 mM imidazole, then 20 mM imidazole, and the tagged enzyme was eluted with buffer containing 150 mM imidazole. Fractions were collected and those with the
highest activity dialyzed overnight at 4°C in 1 L of Zm buffer and used for enzyme characterization.

SDS-PAGE was performed using a 10% polyacrylamide gel with Kaleidoscope Prestained Standards (Bio-Rad) on a Miniprotean II system (Bio-Rad).

**Enzyme characterization.** The specific activity measurements were performed in 1.2 ml of Zm buffer with 2.2 mM ONPGlu (Sigma), and incubated at 25°C for 15 min before starting the reaction with 10 µl of diluted purified enzyme. Reactions were stopped after 5 minutes with 0.5 ml of 0.5 M Na₂CO₃ and the release of o-nitrophenol immediately measured at 420 nm. One unit of activity was defined as the release of 1 µ mol of o-nitrophenol per minute. Specific activity was expressed as units per milligram of protein. Protein concentrations were determined using the Bio-Rad (Hercules, California) protein assay dye reagent protocol. All assays were performed at least in triplicate. The symbols on the graphs represent the average of the values obtained and the error bars the range of values.

The thermodependence of activity was assayed between 0 and 48°C for 5 min using either 2.2 mM ONPGlu or ONPGal. Thermostability assays were performed by incubating aliquots of enzyme at 25, 30, 35, or 40°C. Aliquots were removed at various times and assayed with ONPGlu as above at 25°C for 5 min.

The effect of pH on activity was determined with ONPGlu by assaying in the following buffers ranging from pH 5 to 10.9 in 0.5-pH unit increments: citric acid buffer (0.1 M) (pH 5 to 6), Zm buffer (0.1 M) (pH 6 to 8), Clark and Lubs pH 8.0 – 10.2 buffer (12) (0.05 M) (pH 8 to 10), and sodium carbonate buffer (0.025 M) (pH 10 to 10.9). The pH stability was examined by assaying with ONPGlu in Zm after incubating the enzyme
for 24 hours in each of the previously mentioned buffers at 4°C. Activity in
morpholinepropanesulfonic acid (MOPS), piperazine-N,N'-bis {2-ethanesulfonic} acid
(PIPES), and sodium phosphate buffer (all at 0.1 M) also were also tested at pH 7, and
sodium acetate buffer (0.1 M) at pH 5.5. Activity was measured as described above.

Possible metal ion requirements were examined in two ways. First, aliquots of
enzyme were dialyzed overnight at 4°C against 0.1 M MOPS (pH 7) or MOPS containing
either 1 mM MgCl₂, MnCl₂, CaCl₂, or CuCl₂; or 10 mM NaCl or KCl. Second, the
enzyme was treated with 50 mM EDTA for 30 minutes at either 0°C or 25°C, applied to a
Sephadex G-25 column (Sigma), and eluted with MOPS buffer. The EDTA-treated
enzyme was assayed at 25°C for 10 minutes in the MOPS buffer containing 2.2 mM
ONPGlu with and without the above ions in both 1 mM and 10 mM concentrations,
singly and in combinations. In a modification of this method, the enzyme was incubated
in the 0.1 M MOPS buffer at 25°C for 30 minutes before initiating the reaction with
ONPGlu.

Substrate preference was examined using chromogenic ONP and PNP substrates
at 2.2 mM at 25°C after 5 minutes, with one activity unit defined as the release of 1 µ mol
of o-nitrophenol or p-nitrophenol per minute. The substrates (Sigma) tested were
ONPGlu, PNPGlu, ONPGal, PNPGal, o-nitrophenyl-β-D-xylopyranoside, o-nitrophenyl-
β-D-fucopyranoside, p-nitrophenyl-β-D-N-acetyl glucosaminide (PNPNAG), p-
nitrophenyl-α-D-gluco pyranoside, and p-nitrophenyl-α-D-galactopyranoside.

The Sigma Diagnostic Glucose Kit was used to measure enzymatic glucose
release from non-chromogenic substrates. Substrates tested at 2.2 mM were the
disaccharides laminaribose, cellobiose, gentiobiose, sophorose, sucrose, and lactose, and
the glucosides amygdalin, arbutin, salicin, helcin, puerarin, 1-octyl-β-D-glucopyranoside, indoxyl-β-D-glucoside (indican), and esculin. Reactions subsequently used for the glucose assay were terminated by heating at 65°C for 10 min. The chromogenic substrate, ONPGlu, and the fluorogenic substrate MUG, were also tested using this kit. Appropriate controls showed that levels of ONPGlu hydrolysis measured by glucose release were similar to those measured by release of o-nitrophenol.

**Kinetic studies.** Kinetics studies were performed with freshly purified enzyme using PNPGLu concentrations from 0.75 µM to 40 µM, ONPGlu concentrations from 2 µM to 10 µM, and ONPGal concentrations from 0.4 mM to 7 mM. The absorbance at 420 nm was monitored for 7 min at 25°C. Kinetic studies were also performed using indoxyl-β-D-glucoside from 0.05 mM to 2 mM, monitoring the absorbance at 678 nm, at 25°C for 7 min. A standard curve was produced using synthetic indigo to determine an extinction coefficient of 1.99 mM⁻¹ cm⁻¹ at 678 nm. The resulting data were used to determine the $K_m$ using the Enzyme Kinetics computer program (39).

**Nucleotide sequence accession numbers.** The accession number for the 16S rRNA gene sequence from *Paenibacillus* sp. C7 is AY920751, and the accession number for the sequence that includes the $bgfY$ gene and surrounding open reading frames is AY923831.
3.6 References


Chapter 4

β-Galactosidases from glycoside hydrolase family 42: Biochemical and physiological perspectives, potential substrates, and detection via specifically designed primers
4.1 Summary

Even though there are at least four different families of β-galactosidases and some of these enzymes are found in microorganisms that do not live in locations where lactose would be expected to be found, this disaccharide is loosely assumed to be the substrate for all of these enzymes. I reexamined the physiological data available regarding GHF 42 and reviewed trends in the biochemical properties of enzymes belonging to this family. These data emphasize that lactose cannot be the natural substrate for at least some GHF 42 enzymes, and the physiological evidence available for the remainder does not clearly implicate these enzymes as the ones responsible for the observed lactose hydrolysis by their microorganisms. If the occasional activity on lactose observed for this group of enzymes is incidental, then GHF 42 enzymes are likely to have evolved to act on other substrate(s).

I hypothesized that a collection of several closely-related GHF 42 enzymes could be acquired and used to observe commonalities of growth and increased β-galactosidase activity on potential substrates to explore possible in vivo functions. Also, the cloned inserts could provide clues to possible substrates of the GHF 42 enzymes by the presence of certain genes adjacent to the genes encoding them. These would in turn guide in vitro studies with some of the enzymes encoded by the cloned genes. To obtain further examples of GHF 42-possessing isolates from a relatively small phylogenetic group (Chapter 2) I desired an additional screening method more specific than activity for GHF 42 β-galactosidases. Because no method was suggested by the biochemical properties of characterized GHF 42 enzymes, I compared GHF 42 gene and amino acid sequences and designed a pair of degenerate PCR amplification primers based on two conserved...
regions. I tested these primers on both plasmid and genomic control templates and showed that these primers were specific. I used the primers to: screen genomic DNA from environmental isolates before the creation of genomic libraries, determine which X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside)-hydrolyzing Escherichia coli transformants were carrying a GHF 42 gene, and sequence cloned DNA inserts. The phylogeny of the GHF 42 genes cloned in this way did not support the hypothesis that the GHF 42 enzymes would be closely-related. However, during alignments of GHF 42 for primers and analysis of cloned genes, I noticed the occasional occurrence of certain genes adjacent to GHF 42 genes that related to a source for one of the β-galactan substrates suggested by comparison with GHF 3 substrates and explored the possible significance of this association (Chapter 5).
4.2 Introduction

All known enzymes from GHF 42 and some enzymes from GHF 1, 2, and 35 are capable of β-galactosidase activity, but are distinguished from each other based on secondary structure. GHF 42 was recognized as distinct group in 1993 (12). Based on sequence, GHF 42 enzymes were presumed to have an (α/β)$_8$ barrel structure and belong to the 4/7 superfamily, named because of the locations of the proton donor and nucleophile on the β-strands numbered 4 and 7 (24). The first, and as yet only, structure of a GHF 42 β-galactosidase confirmed that GHF 42 enzymes share an (α/β)$_8$ domain in common with GHF 2, but that their overall tertiary structures differ (13). Thus far, the only general activity recognized in GHF 42 is β-galactosidase activity, whereas GHF 2 enzymes can be β-mannosidases or β-glucuronidases, and GHF 1 enzymes are more frequently β-glucosidases.

Lactose as a substrate for LacZ (GHF 2) is environmentally consistent given Escherichia coli’s occupancy of the human gut, a location where lactose is likely to be found, at least in infancy. However, several organisms with no known association with lactose-containing environments contain GHF 42 genes making a universal functional relationship to lactose implausible. Therefore, I compiled and examined the biochemical data for the GHF 42 β-galactosidases to determine if they had common features and whether lactose hydrolysis studies supported or contradicted the hypothesis of GHF 42 allowing growth on lactose. The results indicate that these enzymes are very likely to have some other function within their host organisms, and that other possible substrates should be explored to better understand the ecological significance of these enzymes.
I therefore wanted to obtain several closely-related isolates with GHF 42 enzymes, with the expectation that they would have a conserved function and could be used to identify consistent in vivo activities with potential β-galactosidase substrates. The ability of particular polysaccharides or sugars to increase β-galactosidase expression by the isolates would provide hints about nature of the in vivo substrate that could be tested by further biochemical characterization. Previously, I had cloned two GHF 42 β-galactosidase genes from bacteria isolated using an enrichment designed to target a relatively small phylogenetic group (Chapter 2). These isolates and (lacZ E. coli) transformants were screened by taking advantage of their ability to hydrolyze X-Gal. However, screening X-Gal (or with PNPG (p-nitrophenyl-β-D-galactopyranoside), or MUG (4-methylumbelliferyl-β-D-galactopyranoside)) cannot distinguish whether a β-galactosidase belongs to GHF 42 or one of the other β-galactosidase families. Also, none of the biochemical properties shared by GHF 42 suggest an alternative screening method to distinguish them from those belonging to other β-galactosidases families. Therefore, I designed degenerate primers specific for GHF 42 to develop a PCR-based screening method. I tested these primers on controls and then used them to identify GHF 42 isolates, and as part of the cloning process, resulting in acquisition of several more GHF 42-carrying isolates and cloned GHF 42 genes. I then examined the phylogeny of the resulting isolates and GHF 42 enzymes.
4.3 Results

4.3.1 Literature review: Biochemical characteristics of GHF 42

I searched the PubMed and CAZy (8) databases for sequences and publications describing GHF 42 enzymes and sequences and examined them for information suggesting possible functions. The characterized GHF 42 enzymes have come from diverse microorganisms, including halophiles, psychrophiles, and thermophiles, many of which are from environments not containing lactose. Not surprisingly, the enzymes themselves have a broad array of salt tolerances (7, 19, 40), thermal optima and stabilities, optimal pH values, and oligomeric states (Table 4-1). The solved GHF 42 structure has a trimeric form (13), which differs from the mono-, di- and tetramer forms observed for GHF 2 β-galactosidases. Other oligomeric states observed for GHF 42 enzymes may have been a result of purification methods (13, 34) or misinterpretation of borderline results swayed by the general rarity of trimers. Unlike GHF 2 enzymes, GHF 42 enzymes do not have specific catalytic divalent metal requirements as shown by a lack of effect by EDTA ((14, 27, 34, 40), although there is evidence for a metal ion serving a structural function (13). While divalent ions do not play an important activating role in GHF 42 comparable to that seen in GHF 2, they are not neutral in their effects. Cu\(^{2+}\) was inhibitory to all five GHF 42 enzymes tested, while Ni\(^{2+}\) and Zn\(^{2+}\) were also sometimes inhibitory (11, 19, 27, 34, 40). β-Mercaptoethanol, on the other hand, increased the activity of the three enzymes tested (11, 19, 34). All of the characterized GHF 42 enzymes, similar to the two I examined (Chapter 2), appear to be intracellular according to the Signalp WWW server (4).
## Table 4-1 Biochemical properties of characterized GHF 42 enzymes

<table>
<thead>
<tr>
<th>Host</th>
<th>Origin</th>
<th>Enzyme</th>
<th>Host Calculated monomer mass (kDa)</th>
<th>Measured Monomer Mass (kDa)</th>
<th>Multimer mass (kDa)</th>
<th>Substrate used for assays</th>
<th>Thermal optimum (°C)</th>
<th>t ½, thermal half-life</th>
<th>pH optimum</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arthrobacter psychrolactophilus</em> B7 ATCC 700733</td>
<td>Whey-treated PA farmland (winter)</td>
<td>BgaG</td>
<td>71</td>
<td>71</td>
<td>-</td>
<td>ONPG</td>
<td>45-50</td>
<td>~22 h</td>
<td>40°C</td>
<td>Gutshall, 1995</td>
</tr>
<tr>
<td><em>Bacillus licheniformis</em> ATCC 9800</td>
<td>Milk</td>
<td>LacB1</td>
<td>-</td>
<td>80</td>
<td>-</td>
<td>ONPG</td>
<td>45</td>
<td>-</td>
<td>5.7</td>
<td>Phan Tran, 1998</td>
</tr>
<tr>
<td><em>Bifidobacterium adolescentis</em> DSM 20083</td>
<td>Human intestine</td>
<td>Bgal II</td>
<td>78</td>
<td>89</td>
<td>350</td>
<td>PNPG</td>
<td>50</td>
<td>10 min</td>
<td>6</td>
<td>Van Laere, 2000; Hinz, 2004</td>
</tr>
<tr>
<td><em>Bifidobacterium longum bv. infantis</em> DSM 20088</td>
<td>Human intestine</td>
<td>INF1</td>
<td>77</td>
<td>73</td>
<td>140</td>
<td>ONPG</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Møller, 2001</td>
</tr>
<tr>
<td><em>Bifidobacterium longum bv. infantis</em> HL96</td>
<td>Human intestine</td>
<td>B-galIII</td>
<td>77</td>
<td>76</td>
<td>-</td>
<td>ONPG</td>
<td>40</td>
<td>-</td>
<td>8</td>
<td>Hung, 1998</td>
</tr>
<tr>
<td><em>Carnobacterium maltaromaticum</em> BA</td>
<td>Whey-treated PA farmland (winter)</td>
<td>BgaB</td>
<td>77</td>
<td>78</td>
<td>-</td>
<td>ONPG</td>
<td>30</td>
<td>35°C, 10 min</td>
<td>6-7**</td>
<td>Coombs, 1999</td>
</tr>
<tr>
<td><em>Clostridium cellulovorans</em> ATCC 35296</td>
<td>woody biomass digester</td>
<td>BgaA</td>
<td>76</td>
<td>78</td>
<td>170</td>
<td>PNPG &amp; PNPAp</td>
<td>30-40</td>
<td>-</td>
<td>6</td>
<td>Kosugi, 2002</td>
</tr>
<tr>
<td><em>Geobacillus kaustophilus</em> ATCC 8005</td>
<td>Pasteurized milk</td>
<td>B-gal I</td>
<td>78</td>
<td>70</td>
<td>-</td>
<td>ONPG</td>
<td>70</td>
<td>-</td>
<td>-</td>
<td>Hirata, 1984; Hirata, 1985; Panasik, 2002</td>
</tr>
<tr>
<td><em>Halofex aurecentense</em> DSM 14919</td>
<td>Spanish saltern pond</td>
<td>BgaH</td>
<td>74</td>
<td>78</td>
<td>180</td>
<td>ONPG</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Holmes, 1997</td>
</tr>
<tr>
<td><em>Planococcus sp. ‘SOS Orange’</em></td>
<td>Hypersaline pond</td>
<td>BgaA</td>
<td>77</td>
<td>75</td>
<td>155</td>
<td>ONPG</td>
<td>42</td>
<td>48°C, ~85 min</td>
<td>5.5</td>
<td>Sheridan, 2000</td>
</tr>
<tr>
<td><em>Thermus thermophilus</em> A4</td>
<td>Japanese hot spring</td>
<td>A4-B-gal</td>
<td>73</td>
<td>75</td>
<td>86</td>
<td>ONPG</td>
<td>≥90</td>
<td>90°C, 1 h</td>
<td>6.5</td>
<td>Ohtsu, 1998; Hidaka, 2002</td>
</tr>
<tr>
<td><em>Thermus sp. T2</em> ATCC 27737</td>
<td>Yellowstone hot spring</td>
<td>BgaA</td>
<td>73</td>
<td>-</td>
<td>-</td>
<td>ONPG</td>
<td>70</td>
<td>70°C, 1 h</td>
<td>80°C, 30 min</td>
<td>5 - 6</td>
</tr>
<tr>
<td><em>Thermus sp. IB-21</em> ATCC 43815</td>
<td>Icelandic hot spring</td>
<td>BgaA</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>PNPG</td>
<td>90</td>
<td>70°C, 74 h</td>
<td>80°C, 23 h</td>
<td>90°C, 4.7 h</td>
</tr>
</tbody>
</table>

ONPG = o-nitrophenyl-β-D-galactopyranoside, PNPG = p-nitrophenyl-β-D-galactopyranoside, PNPAp = p-nitrophenyl-α-L-arabinopyranoside

** (unpublished results, Coombs, 1999)
While it is useful to see the vast diversity of other biochemical properties of GHF 42 enzymes, the nature of the compounds these enzymes hydrolyze is more likely to provide clues to their function. Characterized GHF 42 β-galactosidases have been tested with a number of chromogenic substrates and maximal activity typically occurs with the substrates ONPG (o-nitrophenyl-β-D-galactopyranoside) and PNPG, with less than 10% relative activity on non-galactosidic chromogens (Table 4-2). The two exceptions occur with BgaH of *Haloferax lucentense* and BgaA of *Clostridium cellulovorans*. In the first, activity on PNP(L)F (p-nitrophenyl-β-L-fucopyranoside) is nearly half of that observed with ONPG (19). Detectable fucosidal activity has also been observed in other β-galactosidases in both GHF 42 and GHF 2. In the second, the enzyme has ten times greater activity on PNPAp (p-nitrophenyl a-L-arabinopyranoside) (27). The primary activity of GHF 42 enzymes on X-Gal, ONPG, PNPG, and MUG, but not on chromogenic compounds with other sugars indicates that galactose is the most relevant sugar to the enzymes’ function(s). Kinetic experiments have determined $K_m$, and less frequently $k_{cat}$ and/or $V_{max}$ for several of the GHF 42 β-galactosidases (Table 4-2). Using ONPG $K_m$ values ranging from 0.57 to 5.9 mM have been found. Comparative $K_m$ values on lactose have been higher in all cases (4.81, 19, 42 mM) than those found for the same enzymes with ONPG (0.57, 5.9, 0.41 mM) (11, 34, 25). $K_m$ values for LacZ (*E. coli*, GHF 2) are lower on both lactose and ONPG (1.4, 0.1 mM) (30). The lower $K_m$ value on lactose indicates LacZ binds this substrate more strongly than these GHF 42 enzymes. In total, the biochemical characteristics of the studied GHF 42 enzymes indicate a preference towards β-linked galactosidic substrates rather than other sugars, but provide no other clues pertaining to possible substrates found in the environment.
<table>
<thead>
<tr>
<th>Host</th>
<th>Enzyme</th>
<th>Substrates, % relative activity or:</th>
<th>K&lt;sub&gt;m&lt;/sub&gt;(mM)</th>
<th>[V&lt;sub&gt;max&lt;/sub&gt;] (umol min&lt;sup&gt;-1&lt;/sup&gt; mg&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>K&lt;sub&gt;v&lt;/sub&gt;(mM)</th>
<th>(s&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>[V&lt;sub&gt;max&lt;/sub&gt;] (umol min&lt;sup&gt;-1&lt;/sup&gt; mg&lt;sup&gt;-1&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arthrobacter psychrolactophilus</em> B7 ATCC 700733</td>
<td>BgaG</td>
<td>ONPG + PNPG 100 PNPG L-F 0.45 &lt;0.01; PNPGlu, PNPX, PNPM, PNPGar, PNPGlr, PNPA, PNPL, PNPC,</td>
<td>ONPG, 0.57 [254]</td>
<td>Lactose, 4.81 [3.97]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bifidobacterium adolescentis</em> DSM20083</td>
<td>Bgal II</td>
<td>+: PNPG, galactobiose, galactooligomers -: maltose, sucrose, cellobiose -: gal-β-1,6-gal; gal-α-1,4-gal, gal-β-1,4-mann</td>
<td>Galactobiose, 60 [1,219]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bifidobacterium infantis</em> DSM 20088</td>
<td>INF1</td>
<td>ONPG 100 &lt;10: ONPGlu, ONPX, ONPF, ONPGP</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bifidobacterium infantis</em> HL96</td>
<td>B-galIII</td>
<td>MUG + ONPG 100, Lactose 9.3, PNPG 2.4 -: ONPGlu, PNPGlu, PNPGlr -: sucrose, raffinose, melibiose, maltose</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Carnobacterium maltaromaticum</em> BA</td>
<td>BgaB</td>
<td>ONPG 100 PNPG 62 PNPG 0.2 -: PNPGlu, PNPGlu, PNPGlr -: sucrose, raffinose, melibiose, maltose</td>
<td>ONPG, 1.7 (588) [450]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium cellulovorans</em> ATCC 35296</td>
<td>BgaA</td>
<td>PNPAp 100 PNPG 6 PNPG 10  -: PNPGlu, PNPGlu, PNPGlu, PNPL, PNPC, PNPC, PNPA</td>
<td>PNPG, 6.06 [2.50]</td>
<td>PNPAp, 1.51 [10.4]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Geobacillus kaustophilus</em> ATCC8005</td>
<td>BgaB</td>
<td>ONPG 100 ONPGlu</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Halofex lucentense</em> DSM 1491</td>
<td>BgaH</td>
<td>Lactulose + ONPG 100 PNPG 43 -: ONPGlu, ONPX, PNPL, PNP, -: lactose, melibiose, cellobiose</td>
<td>ONPG, 0.87 [110-120]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Planoccus sp.</em> SOS Orange</td>
<td>BgaA</td>
<td>ONPG 100 PNPG 93 ONP 6 PNPG 4  -: ONPGlu, PNPGlu, PNPGlu, PNPL, PNPC</td>
<td>ONPG, 4.9 (603) [467]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Thermus thermophilus</em> A4</td>
<td>A4-B-gal</td>
<td>ONPG 100 lactose + PNPGlu, PNPGlu, PNPGlu</td>
<td>ONPG, 5.9 Lactose, 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Thermus sp.</em> T2 ATCC 27737</td>
<td>BgaA</td>
<td>ONPG 100 ONPGlu, PNPGlu, PNPGlu</td>
<td>ONPG, 1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Thermus sp.</em> IB-21 ATCC43815</td>
<td>BgaA</td>
<td>ONPG 100 ONPGlu, PNPGlu, PNPGlu Lactose 6</td>
<td>PNPG, 0.41 [418] Lactose, 42 (11.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References as per Table 1

ONPG o-nitrophenyl-β-D-galactopyranoside
PNPGlu p-nitrophenyl-β-D-glucopyranoside
PNPGlu o-nitrophenyl-β-D-galactopyranoside
PNP(2)F o-nitrophenyl-β-D-glucopyranoside
PNP(2)G o-nitrophenyl-β-D-galactopyranoside
PNP(Glu) p-nitrophenyl-β-D-galactopyranoside
PNP(Glu) o-nitrophenyl-β-D-galactopyranoside
PNP(L)F o-nitrophenyl-β-D-glucopyranoside
PNP(L)G o-nitrophenyl-β-D-galactopyranoside
PNPNA o-nitrophenyl-β-D-glucopyranoside
PNPL p-nitrophenyl-β-D-glucopyranoside
PNPL o-nitrophenyl-β-D-galactopyranoside
Gal galactose
Glu glucose
Fru fructose
Galactobiose Gal-β-1,4-Gal Lactose Gal-β-1,4-Glu Cellobiose Gal-β-1,4-Glu Maltose Glu-α-1,4-Glu
Lactulose Gal-β-1,4-Fru Melibiose Gal-α-1,6-Glu Sucrose Glu-β-1,2-Fru Raffinose Gal-α-1,6-glu-β-1,2-Fru
4.3.2 Literature review: Data regarding lactose as a substrate for GHF 42 enzymes

The habitat distribution of some microbes with β-galactosidase genes suggests that lactose is not the natural substrate of all GHF 42 enzymes. I reexamined published characterizations of GHF 42 enzymes to determine whether the biochemical and physiological data corroborate this implication. The data come from three types of experiments, those observing the hydrolysis of lactose by purified enzymes, induction of β-galactosidase activity by lactose, or growth using lactose as the sole carbon source. These types of experiments can refute lactose as the likely substrate for an enzyme, but only demonstration that a genetic knockout of a β-galactosidase gene affects growth on lactose can provide conclusive evidence.

Biochemical studies. Although several GHF 42 β-galactosidases have been characterized, examination of lactose hydrolysis is surprisingly weak, with some examples uncharacterized with this substrate (27, 32, 36). Two characterized GHF 42 β-galactosidases are, in fact, unable to cleave lactose (19, 43). Four other GHF 42 enzymes, three from *Thermus* species (25, 34, 44) and one from *Bifidobacterium longum bv. infantis* HL96 (20), hydrolyzed lactose *in vitro*. BgaG from *Arthrobacter psychrolactophilus* permitted *lacZ E. coli* transformants to grow on lactose minimal media when it was expressed at high levels (11). However, other experiments, described below, relating to induction of the GHF 42 enzymes in these organisms suggests these results should not necessarily be interpreted as indicating *in vivo* lactose hydrolysis by these enzymes.
**Induction studies.** The ability of an enzyme to hydrolyze lactose *in vitro* does not indicate that this is the function of the enzyme in its original host *in vivo*. If lactose is the substrate of a GHF 42 enzyme, then this β-galactosidase activity would be expected to be up-regulated during growth with this substrate (at least under conditions where catabolite repression was not occurring). However, microorganisms with multiple β-galactosidases are frequently found, as is the case for the microorganisms from which the first two characterized GHF 42 β-galactosidases originated (*Bacillus stearothermophilus* IAM11001 (15-17) (now known as *Geobacillus kaustophilus* IAM11001) (33, 37) and *A. psychrolactophilus* (11)). Therefore, while more than one β-galactosidase may be capable of hydrolyzing lactose *in vitro*, not all may be participating equally or even participating at all in this process *in vivo*, as can be shown by comparing induction or activity on lactose for the different β-galactosidases found within a single bacterium. Induction by lactose of the GHF 42 β-galactosidases of *Thermus* sp. T2 (5) and *Clostridium perfringens* (26) was not compared to the relative induction of other β-galactosidases that the *Thermus* sp. likely possesses and the *C. perfringens* does possess (41). The four examples (using *Thermus* sp. IB-21(25), *B. longum* bv. *infantis* HL96 (20-22), *G. kaustophilus* IAM11001 (17), and *A. psychrolactophilus* (11)) where the influence of lactose on the expressions or activities of distinct β-galactosidases were differentiated indicate a higher response from different β-galactosidases than the GHF 42 enzyme (GHFs 1, 2, 2, and 2 respectively). There is also at least one case of an organism possessing a GHF 42 enzyme where overall β-galactosidase activity decreased in the presence of lactose (6).
**Growth studies.** At a broader level of study than induction, growth via utilization can also be examined. If an organism possessing β-galactosidase activity is unable to use lactose as an only carbon source, then it is likely that this carbon source can be ruled out as being biologically relevant to those enzymes. Such is the case for *Haloferax lucentense* (19), *Bacillus subtilis* str. 168 (9), and two pathogenic *Leptospira interrogans* strains (1). Perhaps significantly, although *B. subtilis, B. licheniformis, Bacillus halodurans* all contain a GHF 42 gene, only the first does not respond to lactose, whereas *B. halodurans* (23) and *B. licheniformis* (39) do, probably by way of, respectively, a GHF 2 gene, and a GHF 1 gene not possessed by the other two species. Many other organisms possessing GHF 42 genes can utilize lactose. However, as mentioned above, attributing lactose hydrolysis as a function of a GHF 42 enzyme based solely on the growth of the microbe on this carbon source is premature. Many organisms have been observed to possess multiple β-galactosidase enzymes and organisms whose genomes have been sequenced have provided a great number of additional examples.

**Alternative substrates.** The trends above indicate that the biochemical and physiological data support the ecological indications that lactose is probably not the substrate for GHF 42 enzymes. The occurrence of enzymes with low-level lactase activity without correspondent *in vivo* lactase function may be a general consequence of having a structure compatible with β-galactosidase activity. Additionally, I could find no description of a knockout experiment demonstrating the necessity of a microorganism’s GHF 42 gene for growth on lactose as a sole carbon source. What then, are possible substrates for these enzymes? If the functions of GHF 3 enzymes are used as a model, then the GHF 42 substrates (oligosaccharides) could be arising from the degradation of β-
galactan polysaccharides. I reviewed three books on polysaccharides searching for possible sources for these degradation products (Fig 4-1). One of these, larch arabinogalactan, appears to have been directly tested *in vitro* in relation to the study of BgaA of *C. cellulovorans* and a very small amount (1.3%) of activity, versus that obtained with PNPAn, was detected (27). BgaH from *H. lucentense* was also studied with arabinogalactan (presumably from larch) but the enzyme was not active on this substrate, nor was *H. lucentense* able to grow with this substrate as the only carbon source (18). The experiment with *C. cellulovorans* does not indicate that degradation products from larch arabinogalactan are not potential substrates for this enzyme. Because BgaA appears to be an *intracellular* enzyme, it would *not* be expected to have high levels of activity on a polysaccharide that must be located *extracellularly*. Degradation products of a size that can be transported into the cell are a far more likely substrate. BgaH from *H. lucentense* could also act on degradation products, but lack of growth on the entire polysaccharide indicates that it would be dependent on another organism’s extracellular activities in order to have access to this substrate.

Comparison of the responses of several GHF 42-containing isolates to some of these substrates with regards to growth and β-galactosidase activity have not previously been performed and might implicate one of them, or provide clues to the structure of the actual GHF 42 *in vivo* target. From my analysis of the biochemical work and existence of probable substrates I concluded that the function of the GHF 42 enzymes is unknown. Thus, it was not possible to screen my isolates for growth on a specific compound to determine whether the X-Gal hydrolysis was due the activity of a GHF 42 β-galactosidase.
Figure 4-1 Sources for theoretical β-galactosidase substrates

<table>
<thead>
<tr>
<th>Possible Disaccharides</th>
</tr>
</thead>
<tbody>
<tr>
<td>β4</td>
</tr>
<tr>
<td>β1,4-galactobiose</td>
</tr>
<tr>
<td>β2</td>
</tr>
<tr>
<td>β1,2-galactobiose</td>
</tr>
<tr>
<td>β3</td>
</tr>
<tr>
<td>β1,3-galactobiose</td>
</tr>
<tr>
<td>β6</td>
</tr>
<tr>
<td>β1,6-galactobiose</td>
</tr>
</tbody>
</table>

Potential polysaccharidic sources for β-galactosidase disaccharide substrates above

- β1,4-galactan
- β1,2-galactan
- β1,3-galactan
- β1,6-galactan

Known polysaccharides containing linkages of the type indicated (as the backbone of a branched polysaccharide, or as branches on the polysaccharide):

- Citrus pectin
- Soybean pectin
- Picea sp. compression wood
- Gum arabic
- Larch arabinogalactan

Figure 4-1. Comparison with the polysaccharide β-glucan sources of GHF 3 β-glucosidase substrates suggests that GHF 42 β-galactosidases could be acting on the degradation products from β-galactan substrates. References to polysaccharides with β-galactan backbones or side chains were found in three review books (referenced below the figure). All are found in plants.

4.3.3 Primer design.

In order to obtain additional closely-related GHF 42-containing isolates from my enrichment (Chapter 2) and clone further examples of GHF-42 encoding genes, an additional screening method was desired that more specifically targeted this group of β-galactosidases. I aligned and examined all available GHF 42 sequences for consensus regions in order to identify areas appropriate for selecting specific primers for this group. However, no sufficiently selective primer sites of a suitable length were evident across the breadth of the alignment. Therefore, I reduced the sequence set to sequences from Gram-positive organisms, which encompasses the spore-formers I enriched for through a spore-selective heat treatment (described in Chapter 1). Within this smaller alignment, there was a clear region of homology in the amino acid sequence near the N-terminal...
portion of the sequences: G\textsubscript{G/A}DYNP\textsuperscript{E/D}QW. The corresponding nucleotide sequences were also fairly conserved, so the region was chosen as the forward primer, F42: 5’-GGNGGNGAYTAYAAYCCNGANCARTGG-3’. Further reduction of the sequences to those from Firmicutes (Low G+C Gram positives) was necessary to identify a potential reverse primer. The best such primer that was compatible with the forward primer, R42: DWEN(W/H/R/Y/M/D)WA, contained a string of 4 adjacent degeneracies 5’-GCCCA\textsubscript{VHRR}TTKTCCCATTC-3’. Numbering using the \textit{Geobacillus kautophilus} GHF 42 sequence, F42 covers amino acids 10-18, and R42, 399-405; using the \textit{Thermus thermophilus} A4 GHF 42 sequence, 3-11 and 405-411. Regions homologous to the portions of the GHF 42 β-galactosidase encoded by the primers occur on a β-sheet and α-helix of the \textit{Thermus thermophilus} A4 enzyme (Fig 4-2). The BLAST program was (2)

**Figure 4-2 Ribbon model structure showing locations of primer encoded regions**

![Ribbon model structure](image)

Figure 4-2 Ribbon model structure showing locations of primer encoded regions. The secondary sequences homologous to the regions conserved in GHF 42 enzymes from Firmicutes (Low G+C Gram positive) that were used to design primers are shown in light green on the structure of the GHF 42 enzyme from \textit{Thermus thermophilus} A4.
used to search the nonredundant database set for nucleotide sequences similar to the primers. The search yielded only sequences assigned to GHF 42, suggesting that the conserved regions were not common or conserved in other enzyme groups and that the primers were selective for GHF 42.

**4.3.4 Primer testing within vector background.**

Before the primers could be used for screening, it was necessary to demonstrate their utility using control experiments. I first tested the primers using single positive and negative control templates for amplification (Table 4-3A). The negative control was the vector used for creating genomic libraries, pΔα18 (a modified pUC18 derivative) (42), without insert. The positive control was a construct made from this vector carrying a GHF 42 gene (from *Paenibacillus* sp. CKG as per Chapter 2), pCKG-6. This template was chosen because the GHF 42 sequence was not part of the alignment used to design primers. The negative control did not yield a product, whereas pCKG-6 yielded the expected 1.2 kb product. The annealing temperature yielding the least background amplification, 61°C, was selected for further amplification efforts. A second positive control, pCMM-3, possessed a GHF 42 gene (from *Planococcus* sp. CMM as per Chapter 2) very similar to a sequence used in the alignment and also yielded a 1.2 kb product. A second negative control, pGIC16-1, carrying a GHF 2 gene (from *Paenibacillus* sp. GIC16 as per Chapter 2) but no GHF 42 sequence, yielded no product, as expected.
Table 4-3 PCR experiments and controls

<table>
<thead>
<tr>
<th>Source</th>
<th>GHF 42 presence/absence in control determined via:</th>
<th>Expected result</th>
<th>F42 PCR Product?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Controls using plasmid DNAs as templates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCKG-6</td>
<td>Sequencing of insert</td>
<td>Positive</td>
<td>Yes</td>
</tr>
<tr>
<td>pCMM-5</td>
<td>Sequencing of insert</td>
<td>Positive</td>
<td>Yes</td>
</tr>
<tr>
<td>pGIC16-1</td>
<td>Sequencing of insert</td>
<td>Negative</td>
<td>No</td>
</tr>
<tr>
<td>pΔα18</td>
<td>Vector without insert</td>
<td>Negative</td>
<td>No</td>
</tr>
<tr>
<td><strong>B. Controls using total genomic DNAs as templates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Paenibacillus</em> sp. CKG</td>
<td>Clone was positive</td>
<td>Positive</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em> str 168</td>
<td>Published genome</td>
<td>Positive</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Planococcus</em> sp. SOS Orange</td>
<td>Published paper</td>
<td>Positive</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Planococcus</em> sp. CMM</td>
<td>Clone was positive</td>
<td>Positive</td>
<td>Yes</td>
</tr>
<tr>
<td><em>E. coli</em> ER2585F’</td>
<td>Published genome</td>
<td>Negative</td>
<td>No</td>
</tr>
<tr>
<td><em>Sporosarcina</em> sp. CRE9</td>
<td>All clones negative</td>
<td>Negative</td>
<td>No</td>
</tr>
<tr>
<td><em>Sporosarcina</em> sp. CSA</td>
<td>All clones negative</td>
<td>Negative</td>
<td>No</td>
</tr>
</tbody>
</table>

4.3.5 Primer testing within genomic DNA background.

Although the primers were specific within the narrow pΔα18 background, it was necessary to test them with genomic DNA, which contains a greater number of potential false-priming sites. Two different methods for obtaining genomic DNA for PCR were also tested with these controls. Four genomic DNA templates were used as positive controls. Two were the isolates from which the positive vector controls were cloned (*Paenibacillus* sp. CKG, *Planococcus* sp. CMM). The third was from another isolate previously yielding a cloned GHF 42 gene, *Planococcus* sp. ‘SOS Orange’ (40). The fourth was *B. subtilis* sp 168, known to contain a gene belonging to GHF 42 via genome sequencing (29). DNA from the *E. coli* strain used to create genomic libraries, ER2585F’, was used as a negative control. Genomic DNA from two isolates that had not yielded any GHF 42-carrying transformants after extensive genomic library efforts (*Sporosarcina* sp. CRE9, *Sporosarcina* sp. CSA), were also tested. All positive controls yielded product of the correct size, and the negative control did not yield product (Table 4-3B). The two genomic DNAs (from *Sporosarcina* spp. CRE9 and CSA) that previously failed to yield
GHF 42 β-galactosidase genes in genomic libraries did not yield a 1.2 kb PCR product. These controls also demonstrated that the colony-based method was faster, easier, and less expensive than using the PureGene kit.

The PCR amplified fragment obtained from isolate *Paenibacillus* sp. CKG genomic DNA was ligated into a vector and sequenced using vector-based primers to confirm that the product contained GHF 42 sequence, as expected. The sequence compares favorably to that obtained from the cloned insert of pCKG-6 (Chapter 2).

### 4.3.6 Use of primers to screen genomic DNA from isolates.

The primers were used to guide the cloning process for specific isolation of GHF 42 genes (Fig 4-3). Using the colony-based method, genomic DNA was harvested from just over 100 psychrophilic isolates possessing β-galactosidase activity and were used as templates in PCR reactions. Nineteen of these reactions yielded PCR products of the expected size. Results of particular note came from analysis of isolate *Paenibacillus* sp. GIC16. Previously described in Chapter 2, this isolate appeared to possess at least two different β-galactosidase genes (Table 2-3). One of the responsible genes had been cloned (pGIC161) and sequencing indicated no GHF 42 homology. This construct was also used as a negative control, and did not yield a 1.2 kb product. However, genomic DNA from *Paenibacillus* sp. GIC16 produced a PCR product of the correct size and sequence, with highest homology (98% over 398 aa) to BgaA from *Bacillus circulans* (accession # L03424, unpublished). This indicated that a GHF 42 enzyme may have been responsible for the lower thermal optimum observed. It was desirable to clone this gene and confirm the cold-activity of the heterologously expressed enzyme but this was ultimately not pursued as it did not relate directly to attempts to discern the functions of GHF 42.
Figure 4-3. Primer-guided process for cloning GHF 42 genes. PCR amplification primers specific for GHF 42 genes belonging to a phylogenetically related group of microorganisms were first used to screen for isolates suitable for construction of genomic libraries. X-Gal hydrolyzing transformants from these were then screened with the same primers in order to eliminate those carrying genes belonging to GHFs other than 42.

4.3.7 Phylogeny of isolates

The 16S rDNA of the isolates was amplified and sequenced to confirm that the bacteria belonged to the expected spore-forming groups (Table 4-4). Most of the 16S rDNA sequences were greater than 97% identical to sequences in the NCBI database, as identified using BLAST. The isolates are closely grouped based on 16S rDNA results, with a majority of isolates belonging to the two genera, *Sporosarcina* and *Paenibacillus* (Fig 4-4). Additionally, the isolates form distinct clusters within each of these genera. Two of the isolates belong to the genera *Frigoribacterium* and *Microbacterium*, groups not known to form spores. Because a GHF 42 gene was cloned from the
Frigoribacterium isolate, genomic DNA from several Frigoribacterium isolates from Greenland (GIC6, GIC43, GIC64) were also screened and none were positive. Several other Greenland isolates belonging to the Paenibacillus genus (SO3-6, 1Y, and R21) were tested, and all were positive.

<table>
<thead>
<tr>
<th>Isolates</th>
<th>Location</th>
<th>F42 PCR product</th>
<th>Closest 16S rDNA result (BLAST)</th>
<th>Upper Growth range (plates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKG</td>
<td>PA</td>
<td>+</td>
<td>Paenibacillus odorifer</td>
<td>37°C or higher</td>
</tr>
<tr>
<td>CSW</td>
<td>PA</td>
<td>+</td>
<td>Paenibacillus odorifer</td>
<td>ND</td>
</tr>
<tr>
<td>CRE6</td>
<td>NY</td>
<td>+</td>
<td>Paenibacillus borealis</td>
<td>ND</td>
</tr>
<tr>
<td>C7</td>
<td>PA</td>
<td>-</td>
<td>Paenibacillus macquariensis / antarcticus</td>
<td>25°C, not 30°C</td>
</tr>
<tr>
<td>CRE4</td>
<td>NY</td>
<td>+</td>
<td>Paenibacillus macquariensis / antarcticus</td>
<td>30°C, not 37°C</td>
</tr>
<tr>
<td>CMC4</td>
<td>MI</td>
<td>+</td>
<td>Paenibacillus macquariensis / antarcticus</td>
<td>ND</td>
</tr>
<tr>
<td>CST</td>
<td>NJ</td>
<td>+</td>
<td>Paenibacillus macquariensis / antarcticus</td>
<td>25°C, not 30°C</td>
</tr>
<tr>
<td>CGG</td>
<td>NJ</td>
<td>+</td>
<td>ND* - Likely very close to above</td>
<td>30°C, not 37°C</td>
</tr>
<tr>
<td>CMG4</td>
<td>MI</td>
<td>+</td>
<td>S. psychrophila / S. globispora</td>
<td>25°C, not 30°C</td>
</tr>
<tr>
<td>CMM</td>
<td>NJ</td>
<td>+</td>
<td>Planococcus maritima</td>
<td>37°C or higher</td>
</tr>
<tr>
<td>COZ</td>
<td>PA</td>
<td>+</td>
<td>Microbacterium phyllosphaerae</td>
<td>37°C or higher</td>
</tr>
<tr>
<td>CMA2</td>
<td>MI</td>
<td>+</td>
<td>Frigoribacterium faeni</td>
<td>25°C, not 30°C</td>
</tr>
<tr>
<td>GIC6</td>
<td>GL</td>
<td>-</td>
<td>Frigoribacterium faeni</td>
<td>30°C, not 37°C&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>GIC43</td>
<td>GL</td>
<td>-</td>
<td>Frigoribacterium faeni</td>
<td>25°C, not 33 °C&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>GIC64</td>
<td>GL</td>
<td>-</td>
<td>Frigoribacterium faeni</td>
<td>30°C, not 37 °C&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>GIC16</td>
<td>GL</td>
<td>+</td>
<td>Paenibacillus amylolyticus</td>
<td>18°C, not 25°C&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>GICR21</td>
<td>GL</td>
<td>+</td>
<td>Paenibacillus amylolyticus / illinoisensis</td>
<td>33°C, not 37°&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>#23304</td>
<td>ATCC</td>
<td>-</td>
<td>Same as GIC1Y</td>
<td>33°C, not 37°&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>GICSO3-6</td>
<td>GL</td>
<td>+</td>
<td>Paenibacillus odorifer</td>
<td>37°C, higher?</td>
</tr>
<tr>
<td>CMC3</td>
<td>MI</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>CMA4</td>
<td>MI</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

* ND – not determined § Miteva et al., 2004 ‡ Larkin and Stokes, 1967
Figure 4-4 Composite 16S rDNA phylogenetic tree showing relationship of isolates with known species. This neighbor-joining tree shows that the isolates (red) are mainly Paenibacillus and Sporosarcina spp., as expected from the enrichment process used. Also shown are the two Actinobacteria isolates and the relevant GIC isolates of Miteva et al. (31).
4.3.8 Analysis of heterologously expressed GHF 42 β-galactosidases.

Working with those isolates identified as possessing GHF 42 genes, ten additional β-galactosidase genes belonging to the GHF 42 group were cloned using genomic libraries from the DNA of nine isolates. X-Gal hydrolyzing transformants with plasmids yielding no PCR product using the GHF 42 primers were also found and presumably contained genes for GHF 1, 2, or 35 β-galactosidases. These other clones are not discussed here. Sister clones were not uncommon, but the two transformants described for RE6-24 (Table 4-5) represent two different β-galactosidases.

Portions of some of the cloned GHF 42 genes were sequenced, and found to be homologous to other GHF 42 genes, although not necessarily to the ones expected (Table 4-5). It was possible to sequence plasmids carrying a GHF 42 gene using the degenerate primers by increasing the primer concentration from 1 µ M to 10 µ M, and changing the annealing temperature to 55°C. This was particularly advantageous with larger inserts because it allowed primer walking to begin within the gene of interest instead of from the vector ends. The forward GHF 42 primer yielded longer sections of readable sequence than the reverse for sequencing, probably because it is less degenerate and gives a higher signal to noise ratio.

The activity of the ten new and two previously cloned β-galactosidase genes was examined by plating the transformants (Table 4-5) at different temperatures to determine if any of the enzymes were cold-active. Although all of the enzymes were able to hydrolyze X-Gal at 37°C, some produced greater hydrolysis at lower temperatures, possibly indicating their optima may be close to 37°C.
<table>
<thead>
<tr>
<th>Isolate genus</th>
<th>Isolate</th>
<th>Blue colonies (total)</th>
<th>Selected plasmids</th>
<th>Thermal dependency†</th>
<th>Expected genus of most homologous β-galactosidase</th>
<th>Most homologous β-galactosidase via BLAST‡</th>
<th>As expected?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P CKG 2</td>
<td>18,500</td>
<td>pCKG-6</td>
<td>Optimum 43°C</td>
<td>B</td>
<td>B: BH3701 68, 82</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>P CSW 33</td>
<td>3600</td>
<td>pCSW-3</td>
<td>Bluer below 37°C</td>
<td>B</td>
<td>St: SCO7407 67, 78</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>P CRE6 7</td>
<td>6600</td>
<td>pCRE6-2, pCRE6-7</td>
<td>Blue at 37°C</td>
<td>B</td>
<td>ND*</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>P CRE4 7</td>
<td>6</td>
<td>pCRE4-2</td>
<td>Blue at 37°C</td>
<td>B</td>
<td>Pl: PlSOS 68, 81</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>P CMC4 1</td>
<td>1092</td>
<td>pCMC4-8</td>
<td>Blue at 37°C</td>
<td>B</td>
<td>St: SCO7407 66, 75</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>P CST 1</td>
<td>2500</td>
<td>pCST-1</td>
<td>Bluer below 37°C</td>
<td>B</td>
<td>St: SCO7407 63, 76</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>P CGG 1</td>
<td>4000</td>
<td>pCGG-1</td>
<td>Bluer below 37°C</td>
<td>B</td>
<td>ND</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>S CMG4 3</td>
<td>3900</td>
<td>pCMG4-9</td>
<td>Bluer below 37°C</td>
<td>Pl</td>
<td>St: SCO7407 70, 82</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Pl CMM 2</td>
<td>3000</td>
<td>pCMM-3</td>
<td>Optimum 47°C</td>
<td>Pl</td>
<td>Pl: PlSOS 91, 95</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>M COZ 13</td>
<td>9000</td>
<td>pCOZ-4</td>
<td>Blue at 37°C</td>
<td>A (or St)</td>
<td>ND</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>F CMA2 9</td>
<td></td>
<td>pCMA2-2</td>
<td>Blue at 37°C</td>
<td>A (or St)</td>
<td>ND</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

£: P – Paenibacillus, S – Sporosarcina, Pl – Planococcus, PlSOS – Planococcus sp. ‘SOS Orange’, M – Microbacterium, F – Frigoribacterium, B – Bacillus, BH – Bacillus halodurans, A – Arthrobacter, St. – Streptomyces, SCO, Streptomyces coelicolor
† - optimum determined as per Chapter 2; blueness refers to X-Gal hydrolysis on plates by selected transformants
*ND – Not determined
4.4 Discussion

4.4.1 Possible substrates for GHF 42 enzymes

The GHF 42 β-galactosidases have a wide range of biochemical properties, but the characteristics they share in common with each other differ from those shared by GHF 2 enzymes. These enzymes may have evolved to perform different functions. While lactose hydrolysis as a function for GHF 2 enzymes is supported, evidence supporting the same function for GHF 42 enzymes is weak. Therefore, it was of interest to determine whether microorganisms with closely-related β-galactosidases, which would presumably have a conserved function, might be using these enzymes to hydrolyze β-galactan portions of polysaccharides (Fig 4-1). Understanding the in vivo functions of β-galactosidases will help us to better appreciate their roles in the global carbon cycle, and better exploit them for molecular biology techniques, lactose degradation, and degradation of other polysaccharides found in food or industrial products.

4.4.2 Design and testing of GHF 42-specific primers

Degenerate primers specifically designed to amplify a region within GHF 42 genes have been constructed. I compared the regions homologous to the primers to the only solved structure for a GHF 42 β-galactosidase (13). Interestingly, although conserved, neither region is directly in the vicinity of the catalytic residues, or ones involved with substrate binding, metal binding, or subunit interactions. However, homology to this β-galactosidase indicates that the forward primer region encodes the first β-sheet of the (α/β)₈ barrel. Mutational studies have shown that this region is important for the flexibility and stability of GHF 42 enzymes by anchoring the barrel ring closed between the first β-sheet to the last α-helix (35). This may explain the importance
of the region, and thus its conservation.

The fortuitously wide spacing between the forward and reverse primers allows for amplification of DNA encoding all but the N-terminal most amino acids (approximately 3-18 residues) of domain 1, with a length of about 1.2 kb. This length is ideal for sequencing from both ends with an overlap, precluding the need for internal primers. The primers were tested both in the environment of vector background and with genomic DNA background and yielded a 1.2 kb product with those templates carrying a GHF 42 gene. The lack of PCR product with *E. coli* genomic DNA indicated that transformants could be tested using DNA extracted from β-galactosidase transformants without explicit extraction of plasmid DNA. The results from the positive and negative controls indicate this method can be used as intended, to screen β-galactosidase positive isolates to determine whether they possess a GHF 42 gene, and then to screen transformants expressing cloned inserts originating from these isolates. Screening transformants is a necessary step because organisms with β-galactosidase activity frequently possess more than one β-galactosidase gene, and therefore the cloned genes might not belong to GHF 42. This method is faster than sequencing to determine to whether a GHF 42 gene could be encoding the β-galactosidase activity, especially when the gene responsible is distant from the ends of a large insert, requiring time-consuming primer walking or subcloning.

It is important to consider that, due to variations in Family 42 genes outside those used in the alignment used for the creation of the primers, a negative result does not exclude the possible presence of a GHF 42 gene. Identification of GHF 42 genes within other specific groups, such as the high G+C Gram positive organisms, could be achieved by modifications to the current primers or creation of new primers for these specific
groups. However, primers able to identify GHF 42 genes across a wider phylogenetic range will be difficult to design because too much degeneracy will lead to the loss of specificity.

4.4.3 Phylogeny of isolates identified as possessing GHF 42 genes via PCR screen

In spite of possible limitations, the primers I designed were successful for screening isolates and transformants for the presence of GHF 42 genes. The presence of members of the genera *Frigoribacterium* and *Microbacterium* is surprising because they do not sporulate or have another currently known mechanism for heat-tolerance, and thus would not be expected to survive the spore-selection process. Their survival could be a result of the lower temperature used to select for spores. However, species belonging to both of these genera have been found alongside *Paenibacillus* species in both Siberian permafrost (3) and Greenland glacier ice (31), and can therefore survive certain types of extreme conditions. The remaining isolates are closely grouped based on 16S rDNA comparisons, which should be ideal for comparing similar β-galactosidase genes. It is interesting that most of the described species closely related to the isolates (Table 4-4) were themselves isolated from either Antarctica or plant-related habitats. The first indicates that the enrichment for psychrophiles was successful, while the second acknowledges the paenibacilli as impressive degraders of polymers and hints at a high potential for plant-produced polymers as possible substrates for GHF 42 enzymes. This is encouraging since all of the β-galactan sources listed on Fig 4-1 originate from plants.
4.4.4 Analysis of heterologously expressed GHF 42 β-galactosidases.

Even though many of the isolates were psychrophilic, no new transformants were obtained that hydrolyzed X-Gal only below 37°C. However, several probably have optima below that of LacZ (55°C) as indicated by greater hydrolysis of X-Gal below 37°C. The previously obtained GHF 42 enzymes from Paenibacillus sp. CKG and Planococcus sp. CMM (Chapter 2) were most identical to enzymes from B. halodurans and Planococcus sp. SOS Orange, respectively. The second match is ideal, and the first is reasonable given that no GHF 42 genes from Paenibacillus species are present in the NCBI databases. Unexpectedly, the phylogeny of some of the new enzymes encoded by the cloned DNA did not follow this pattern and are inconsistent with the phylogeny of the 16S rDNA sequences (Table 4-5). Several of the enzymes are most homologous to a GHF 42 gene from Streptomyces coelicolor, which belongs to a different phylum of bacteria than the Paenibacillus, Sporosarcina, and Planococcus spp. from which the enzymes originated.

The occurrence of GHF 42 genes within the Sporosarcina, Frigoribacterium, and Paenibacillus genera appears to be sporadic. Some species within each of the genera examined tested positive, while others tested negative. In the case of the Frigoribacterium species, it can be argued that the primers are not consistently detecting the GHF 42 genes present in these high G+C Gram-positive organisms. This same argument can also be used for the Paenibacillus β-galactosidases genes since even though the primers were designed for low G+C Gram positive GHF 42 genes, the Paenibacillus β-galactosidase genes are apparently not all closely related to each other.

In contrast to those isolates that may have possessed no GHF 42 genes, one
isolate possessed two different GHF 42 genes. The presence of two or more GHF 42 genes in the genome of *Paenibacillus* sp. CRE6, and in several microorganisms whose genomes have been sequenced (combined with analysis in Chapter 5) suggests a different function exists for each of the duplicate genes.

The resulting phylogeny of the cloned GHF 42 genes (discussed further in Chapter 5) indicates that the suggested substrate induction/growth experiments would not be likely to give congruent results, although they might still lead to answers implicating the nature of some GHF 42 β-galactosidase substrates. However, an adjacent gene with apparent homology to other genes known to encode extracellular GHs active on one of the alternative substrates was noticed during alignments. An extracellular enzyme capable of producing the type of oligosaccharides I expected the (intracellular) GHF 42 enzymes to act on, in a position suggestive of an operon, was very promising. A single example could easily be coincidental or not meaningful considering the wide range of enzyme activities that can be encompassed by a single GHF, but this one notable example suggested that another method might prevail for determination of a possible substrate for these enzymes. Therefore, an extensive analysis was made of the genes adjacent to GHF 42 genes to determine whether this relationship went beyond a single example, or if additional associations could be found (Chapter 5).
4.5 Materials and Methods

**Bacterial isolates.** Psychrophilic X-Gal-hydrolyzing bacterial strains were obtained from the enrichments described in Chapter 2 designed to isolate bacteria belonging to the *Bacillales* group. Additional isolates from other laboratory members’ enrichments were selected based on phylogenetic analysis of 16S rRNA sequences.

**Primer design.** Genes identified as belonging to GHF 42 by CAZY (http://afmb.cnrs-mrs.fr/CAZY/index.html)(8) were aligned both by hand, and by using ClustalW, within the software program Bioedit (Version 5.0.6; Department of Microbiology, North Carolina State University (http://www.mbio.ncsu.edu/Bioedit/bioedit.html)). Regions of high homology were identified by visual examination of alignments of all the available (35) sequences belonging to Family 42, or phylogenetically related subsets of this group. Degenerate primers were designed based on the nucleotide sequences of conserved regions, synthesized by IDT (Integrated DNA Technologies, Inc., IA), and were expected to yield a product of about 1.2 kb in length when PCR amplification occurred from a GHF 42 gene. The sequences of the primers are

F42: 5’-GGNGGNGAYTAYAAYCCNGANCARTGG-3’ and
R42: 5’-GCCCAVHRRTTKTCCCATTC-3’.

Locations homologous to the regions used to design the primers were examined on the *Thermus thermophilus* A4 β-galactosidase structure (13) by means of the Swiss-Pdbviewer (10) Deep View version 3.7 (http://www.expasy.org/spdbv), which was also used to create Figure 4-2.

**Vector background primer controls.** Plasmid DNA was harvested from the transformant plasmid pCKG-6, which encoded a GHF 42 gene from *Paenibacillus* sp.
CKG (Chapter 2), and used as a template in the PCR reaction. A gradient PCR
(Eppendorf, Mastercyler gradient PCR machine) was performed to determine optimal
amplification conditions (95°C 5 min, [95°C 1 min, 55°C +/- 10°C, 1 min, 72°C 1.5 min] X 32 cycles, 72°C 1 min) with twelve different annealing temperatures across the 20°C
gradient, using Ready-to-go PCR Beads (Amersham Pharmacia Biotech).

Plasmid DNA from additional transformants possessing genes already categorized
as belonging to particular GHFs was harvested in the same fashion as from pCKG-6.
These were used as template under the conditions identified as optimal to act as negative
and additional positive controls.

**Genomic background primer controls.** Genomic DNA was initially harvested using the
Puregene® Genomic DNA isolation kit (Gentra). As a colony-based method for
obtaining genomic DNA from Gram-positive bacteria, cells from one to two colonies
were suspended in lysozyme solution and incubated at 37°C for 20 min. Then, 5 µ l of
lysis buffer (0.25 M NaOH, 1.25% SDS) was added and incubated at 90°C for 15 min.
This mixture was diluted with 200 µ L molecular biology-grade water, and centrifuged
for 20 sec at 12,000xg. The supernatant was removed to a clean tube, and 1-3 µ L were
used for 25 µ L PCR reactions.

Genomic DNA from isolate *Paenibacillus* sp. CKG, *B. subtilis* strain 168,
*Planococcus* sp. ‘SOS Orange’, and the *Planococcus* sp. CMM were used as positive
controls, and *E. coli* was used as a negative control. Genomic DNAs from two isolates
from which only GHF 2 β-galactosidase genes had been cloned were also tested. The
PCR product resulting from using genomic DNA from isolate *Paenibacillus* sp. CKG as a
template was incorporated into a plasmid using PCR-Script™ Amp Cloning kit
(Stratagene) and sequenced by the Penn State Nucleic Acid Facility on an ABI Hitachi 3100 Genetic Analyzer.

**Use of primers to screen genomic DNA from isolates.** Genomic DNA was harvested from X-Gal hydrolyzing isolates with unknown β-galactosidase genes using the colony-based method, and the concentration of the templates compared by degree of UV fluorescence on an ethidium bromide stained agarose plate. The PCR conditions already identified were maintained, and either Takara Ex Taq polymerase (Takara Biomedicals) or Ready-to-go PCR Beads (Amersham) were used. Isolates with genomic DNA yielding a 1.2 kb product after amplification were identified as possessing a GHF 42 gene. The PCR product from isolate *Paenibacillus* sp. GIC16 genomic DNA template was inserted into the vector pPCRScript (Stratagene) and sequenced in the same method as described for the product from *Paenibacillus* sp. CKG, above.

**Preparation of genomic libraries and transformant screening.** The Puregene kit was used to extract DNA from cell material harvested from TSA or R2A (38) plates, with a preliminary lysozyme treatment, and increased heating to 85°C for 10 min to enhance lysis. This DNA was partially digested with *Pst*I for 10 to 40 min, and cleaned using the Qiagen PCR purification kit. No gel extraction was performed. This DNA was combined with pΔα18 (pUC18 without the LacZ alpha fragment)(42) that had also been treated with *Pst*I, and CIP (calf-intestinal phosphatase). Ligations were performed at room temperature using the FastLink DNA Ligation kit (Epicentre). Transformations used Z-competent (Zymo Research) *E. coli* ER2585F’ cells plated onto warm LB (Luria-Bertani) plates containing 100 µ g/ml ampicillin, 100 µ M IPTG (isopropyl-β-D-thiogalactoside) and 100 µ g/ml X-Gal, and incubated for no longer than 16 h at 37°C. Transformants able
to hydrolyze X-Gal at 37°C were restreaked to fresh plates with the same composition. The library plates, with colonies that had remained white at 37°C, were then moved to 18°C and reexamined within 3 days. Transformants becoming blue at 18°C were also restreaked.

Plasmid DNA was harvested from X-Gal hydrolyzing transformants. Multiple transformants from the same isolate DNA were compared using restriction digestion patterns of the plasmid DNA. Omitting sister clones, the plasmid DNA was then used as a template for PCR amplification using the GHF 42 primers. X-Gal hydrolyzing transformants identified as carrying a GHF 42 β-galactosidase gene were screened on plates for β-galactosidase activity across a range of temperatures (37, 30, 25, and 18°C).

**Phylogeny of isolates and β-galactosidase genes.** The 16S rRNA genes from the isolates whose genomic DNAs yielded clones carrying a GHF 42 gene were amplified by PCR. PCR was performed using Ready-To-Go beads (Amersham Pharmacia, Piscataway, NJ) and universal primers 8F and 1492R. Sequencing was performed at the Penn State Nucleic Acid Facility (NAF) on an ABI Hitachi 3100 Genetic Analyzer. Cloned inserts from selected transformants were also sequenced from plasmids, starting with the degenerate GHF 42 primers, and then by primer walking.

The 16S rRNA gene sequences and β-galactosidase gene sequences were used to search the National Center of Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov) via BLAST. The 16S rDNA sequences of related organisms were initially aligned using Clustal W (BioEdit platform, Version 5.0.6; Department of Microbiology, North Carolina State University (http://www.mbio.ncsu.edu/Bioedit/bioedit.html)) and the sequences from isolates were
aligned manually. Alignments were imported into MEGA (28) in order to create bootstrapped neighbor-joining phylogenetic trees (not shown) from which the closest phylogenetic relative among described species was determined for each isolate. The relationships between the isolates and representatives of both closely related and more distant phylogenetic groups were examined with a broader alignment, again in MEGA. Because the full-length 16S rRNA sequence was not obtained for all isolates, several different trees were created using different subsets of the full-length alignment. A composite diagram was then created to concisely display these data by overlapping the congruent regions of these trees. Although the general relationships were robustly maintained, the bootstraps are not shown because the numbers varied somewhat between the individual trees.
4.6 References


Chapter 5

β-Galactosidases from GHF 42:
Ecological and genomic perspectives relating to potential substrates
5.1 Summary

GHF 42 β-galactosidase genes are widespread but my analysis of biochemical data from previous studies (Chapter 4) indicates their function is almost certainly not lactose hydrolysis. My initial results based on analyzing cloned GHF 42 genes suggests that these genes possess a peculiar phylogeny and may have more than one function (Chapter 4). If the ability of organisms with GHF 42 genes to grow on lactose is due to the presence of other β-galactosidases, then what roles do the GHF 42 enzymes have? In order to obtain clues I used information available from the sequences adjacent to the genes encoding GHF 42 enzymes. I also analyzed the pattern of occurrence of GHF 42 genes both in terms of the ecological context of their hosts, and within their genomic context in order to find other clues to their *in vivo* function(s). A conserved gene assemblage was found for a subgroup of the sequences that was clearly suggestive of a specific primary substrate from plants, arabinogalactan type-I.
5.2 Introduction

Many organisms with no known association with the mammalian digestive tract possess GHF 42 genes making a universal functional relationship to lactose implausible. I decided to apply a bioinformatics approach to this problem by exploiting the information available from genome sequencing projects containing GHF 42 genes. I examined both the environmental background of the hosts of GHF 42 enzymes, as well as the proposed enzyme functions of the ORFs that occurred next to their genes in the genomes, in the hopes of finding conserved gene arrangements suggestive of function on specific substrate(s). I also examined the unusual apparent phylogeny of the GHF 42 enzymes.

The presence of specific genes adjacent to GHF 42 genes suggests that the encoded β-galactosidases act on pectic plant polysaccharide degradation products. This possibility suggests an important ecological function for these β-galactosidases in the carbon cycle. Although my research indicates that GHF 42 β-galactosidases are less likely to be useful for the dairy industry, they are still applicable for many biotechnological methods. Also, the natural substrate(s) of these enzymes are present in non-dairy foods, giving a biotechnological use for these enzymes through (indirect) modification of polysaccharides. This proposed function for GHF 42 enzymes may also present the opportunity for replicating a possible pathway for the evolution of lactase function from enzymes originally designed to function on other carbohydrates found in the diets of mammalian herbivores. This could be performed by mutations of a GHF 42 with poor lactose hydrolysis combined with a selective screen for improved growth on this substrate.
5.3 Results and Discussion

5.3.1 GHF 42 enzymes: distribution and phylogenetic relationships

In my analysis of GHF 42 enzymes described in Chapter 4 there were several characterized enzymes that had been isolated from environments where lactose would not be expected to be found. The first GHF 42 β-galactosidases studied were from isolates found in places where the presence of lactose was expected, such as milk (*Geobacillus stearothermophilus* ATCC (American Type Culture Collection) 8005 (13) and *Bacillus licheniformis* ATCC 9800) (29), whey-treated fields (*Arthrobacter psychrolactophilus* B7 (9), *Carnobacterium (piscicola) maltaromaticum* BA) (5), and the human intestine (*Bifidobacterium infantis* (15, 24) and *Bifidobacterium adolescentis* (40)), but other GHF 42 β-galactosidase genes were later cloned from extremophiles found in environments where lactose would not be expected to occur: hot springs (*Thermus* sp. A4 (28), *Thermus* sp. T2 (41), *Thermus* sp. IB2 (17)) and hypersaline environments (*Haloferax lucentense* (14) and *Planococcus* sp. ‘SOS Orange’) (34). Considering the sources of these β-galactosidases may be misleading because they are clearly biased by either our expectations (of where β-galactosidases can be found) or desires (for β-galactosidases with special extremophilic properties); additional undiscovered GHF 42 enzymes probably occur in prokaryotes of moderate (i.e. non-extremophilic) non-lactose environments. Alternatively, these enzymes may have been discovered, but gone unstudied and unreported in favor of working with the more promising (at least in terms of lactose hydrolysis) β-galactosidase clones belonging to GHF 2. This seems possible because of the many examples of GHF 2 enzymes having more efficient hydrolysis of lactose than GHF 42 enzymes (Chapter 4).
In order to discern how widespread examples of organisms with GHF 42 from “unexpected” habitats were, I examined their occurrence (as identified by CAZY (6) and/or searches using BLAST (1)) in fully sequenced genomes. There is still a bias because more genomes have been sequenced from microbes related to human, animal, and plant health (both beneficially and detrimentally), from extreme or unusual environments, and hard to culture phylogenetic groups than from “typical” environmental microbes. However, this bias is not directed at our preconceptions of β-galactosidase functions.

I first grouped over 200 prokaryotes with fully sequenced genomes into general habitats and then separated out those with GHF 42 genes (Fig 5-1). By restricting the data set to those microbes possessing GHF 42 genes, the relative proportion occurring in various habitats is modified (Fig 5-1). The (non-gastrointestinal) animal-associated habitat percentage is greatly reduced. This habitat is overrepresented in the GHF 42 portion of the diagram if one considers that the section of the pie symbolizes six bacterial strains, but only three different species (Leptospira interrogans, Yersinia pestis, and Yersinia pseudotuberculosis). The other sections do not contain multiple strains of a given species. This suggests that the GHF 42 genes are not involved in pathogenesis, and their enzymes do not act on an animal-produced compound. The proportion of GHF 42 possessing organisms found in aquatic (freshwater and marine) environments is also reduced, while the ratio of plant-associated microbes stays the same, and that in terrestrial (soil) environments and gastrointestinal habitats increases. The presence of these genes in microbes found in the digestive systems of animals (Lactobacillus, Bifidobacterium, and Bacteroides spp.) and in soil suggests these genes are involved in degrading a carbon source found in both the soil and in animal diets, possibly from plants.
Figure 5-1. Habitat of microorganisms with sequenced genomes (~250) compared with those containing GHF 42 genes. Bacteria and Archaea with sequenced genomes were assigned to one of five habitat categories (large pie). Opportunistic pathogens were assigned to their non-animal habitats. The smaller pie shows the habitat distribution within those prokaryotes possessing GHF 42 genes (about 9%).

I next compiled and aligned sequences known from research studies (including my sequences from Chapter 2 and 4) and sequencing projects from both completed and incomplete genomes. The aligned sequences were used to construct a phylogenetic tree showing the relationships between GHF 42 genes (Fig 5-2). Sequences containing N-termini were also examined using the Signalp WWW server (3) to determine whether, like all of the characterized GHF 42 enzymes, the additional predicted enzymes were also non-secerted. The only exceptions were a pair of ORFs from *Solibacter usitatus* Ellin6076, which each possess a signal peptide cleavage site.

Compared with a list of microorganisms with completed genomes this tree shows that not all of the major bacterial groups possess GHF 42 genes, which is not remarkable. However, the GHF 42 genes also do not occur consistently within those phylogenetic groups where there are examples; they occur in some members of groups, but not in other closely (or very closely) related members (Table 5-1). For instance, a GHF 42 from
Table 5-1 Inconsistent distribution of GHF 42 within genera

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species with GHF 42</th>
<th>Species lacking GHF 42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus</td>
<td>clausii halodurans licheniformis subtilis</td>
<td>anthracis cereus thuringiensis</td>
</tr>
<tr>
<td>Bacteroides</td>
<td>thetaotaomicron</td>
<td>fragilis</td>
</tr>
<tr>
<td>Burkholderia</td>
<td>vietnamensis G4</td>
<td>pseudomallei K96243</td>
</tr>
<tr>
<td>Clostridium</td>
<td>perfringens</td>
<td>tetani acetobutylicum</td>
</tr>
<tr>
<td>Deinococcus</td>
<td>geothermalis DSM 11300</td>
<td>radiodurans R1</td>
</tr>
<tr>
<td>Geobacillus</td>
<td>kaustophilus IAM11001 thermocatenulatus KNOUC105</td>
<td>kaustophilus HTA426</td>
</tr>
<tr>
<td>Pseudoalteromonas</td>
<td>atlantica T6c</td>
<td>haloplanktis TAC125</td>
</tr>
<tr>
<td>Silicibacter</td>
<td>sp. TM1040</td>
<td>pomeroyi DSS-3</td>
</tr>
<tr>
<td>Thermus</td>
<td>brockianus sp. T2 sp. A4 sp. IB21</td>
<td>thermophilus HB27 thermophilus HB8</td>
</tr>
</tbody>
</table>

Geobacillus kaustophilus IAM11001 isolated from pasteurized milk has been characterized (13), but the genome of *G. kaustophilus* HTA426 (from deep-sea sediment) (35) has no GHF 42 gene. This bolsters the hypothesis that these are not “house-keeping” genes, and also indicates that either frequent gene loss and/or horizontal gene transfer has occurred. Interestingly, the GHF 42 genes are apparently rare within the Archaea since none of the 24 fully sequenced genomes from this group possess these genes, although we have the *H. lucentense* (14) example. The tree also shows that fifteen microorganisms possess two GHF 42 genes, and there is one example each for three and four GHF 42 copies in a single genome. Therefore, at least three but probably less than eight (and certainly not twenty!) duplication events (either intra- or intergenomically mediated) have occurred. The multiple occurrences imply an advantage for possessing more than one GHF 42-encoding gene. A simple increase in expression level is unlikely because of the degree of difference in the homology of the duplicate sequences. This suggests that different GHF 42 enzymes have evolved different functions and that not all the GHF 42
enzymes are orthologous. Just as observed with my *Paenibacillus* sp. GHF 42 ORFs, ORFs from the gamma proteobacteria do not all cluster together. Thus, the analysis of the phylogenetic distribution of GHF 42 is consistent with my primer studies of my isolates. These confirmed results indicate that like the GHF 42 genes in total, the GHF 42 genes in my collection are probably not all orthologous. The phylogeny also indicates that it may be possible to make a case for one or more horizontal transfer events.

5.3.2 Examination of GHF 42 gene arrangements

My next endeavor was to see if by taking advantage of the vast data from the rapidly-growing number of fully-sequenced genomes I could identify adjacent genes with suggestive related functions. I first searched for ORFs homologous to known GHF 42 enzymes. Several “new” GHF 42 ORFs were detected in draft genomes that were not yet integrated into the CAZy database. I then examined the adjacent ORF annotations within both these new sequences and those identified by CAZy. However, the annotation methods varied, and “hypothetical protein” designations, although accurate, are not very informative. Additionally, other (non-genomic) studies rarely analyzed sequence adjacent to the GHF 42 gene of interest. To determine possible functions of adjacent genes in a consistent manner, I focused on the conserved COG (Clusters of Orthologous Groups) (37, 38) designations returned by protein-protein BLAST searches using the conceptual translations. Actual BLAST results were examined when no COG homologies were identified, with careful attention to the possible occurrence of familiar annotation loci (numbers similar to those identifying the GHF 42 ORFs in other organisms). Where sequence information was available, I analyzed at least three ORFs up- and down-stream of the GHF 42 genes (not shown).
Figure 5-2. Phylogenetic relationships between GHF 42 enzymes based on neighbor-joining. Bootstrap values shown at the nodes were generated from 1,000 replicates. The accession locus tag of the enzyme (or name, for those genes not identified by genome sequencing) precedes the microorganism name. Enzymes originating from different phylogenetic groups are color coded as per the group names on the right of the figure. Enzymes with published characterization are circled. Arrows show the expected locations of fragmentary sequences, as based on BLAST search results. Sequences resulting from work in Chapters 2 and 4 are shown in bold.
Figure 5-2 Phylogenetic relationships between GHF 42 enzymes

Key
- Unclad near ribulose-5-phosphate carboxylase oxygenase
- [Published character sets]
- Homeologous incomplete sequence
- Cloned from dissertation incomplete isolate

1. Firmicutes
   - Bacilli
   - Lactobacillales
   - Clostridia

2. Gamma
   - Proteobacteria

3. Thermotogae

4. Actinobacteria
   - Bacillus sp. CSM
   - Paenibacillus sp. CM4
   - Paenibacillus sp. CSE
   - Planococcus sp. CMG4

5. Other

6. Beta
   - Proteobacteria

7. Deinococci

8. Alpha
   - Proteobacteria

* Probably alpha proteobacteria, based on homology of other genes in contig
With a large (and growing) number of GHF 42 genes, this effort involved enormous amounts of raw data. A major difficulty was organizing the data so that any patterns could be observed. To do this I developed a spreadsheet format after it became apparent that organizing linear text would not permit efficient analysis. I then selected a color-scheme for general functions and constructed a diagram of each gene arrangement so that it could be compared to others to detect any shared patterns. Patterns were most easily detected by sorting the color-coded physical printouts of the gene arrangements. Fig 5-3 shows shortened versions of these arrangements overlaid on a phylogenetic tree of the GHF 42 enzymes. This diagram emphasizes the variety of patterns that occur even among closely related genes and the difficulty in detecting meaningful relationships.

The most consistent co-occurring genes are a series of three genes putatively encoding ATP-binding cassette (ABC) transporter genes (Fig 5-3, orange genes) with homology to COG2182 (MalE)/COG1653 (UgpB); COG3833 (MalG)/COG0395 (UgpE); and COG4209, (LplB)/COG1175 (UgpA). The first, MalE/UgpB are periplasmic binding proteins for maltodextrin and sugars, while the second and third are permeases for maltodextrin/sugar and oligosaccharides/sugar. The Yersinia and Pectobacterium spp. also share the presence of maltoporin-like transporters (LamB) (cd01346 & pfam02264) in their arrangements. Some arrangements appear to encode complete ABC transport systems, as they each also possess a COG3839 (MalK) gene, which encodes an ATPase compatible with ABC transport systems for sugars. The other arrangements lack the necessary ATPase part of the ABC transport system, but this does not mean that they are nonfunctional or irrelevant: the relevant ATPase genes could be coded for elsewhere in
Figure 5-3. Gene arrangements of GHF 42 enzymes overlaid on phylogenetic tree. The putative functions of genes nearby GHF 42 genes were examined using CDsearch and BLAST and are color coded as per the key at the right of the figure. The alignments are centered on the (blue) GHF 42 genes. The order of the arrangements corresponds to the order in which they occur on the preceding phylogenetic tree (Fig 5-2). The black lines “cross out” sequences occurring on Fig 5-2 for which there is insufficient sequence available for analysis. The jagged lines represent the end of a sequence - either the end of a cloned insert or the end of the contig of an incomplete genome sequence.
Figure 5-3 Gene arrangements of GHF 42 enzymes overlaid on phylogenetic tree

Key
Polysaccharide degradation
- GHF 42
- GHF 53
- GHF w/ B-galactosidase activity (2 or 35)
- GHF w/ alpha-galactosidase activity (4 or 36)
- GHF #
- Carbohydrate esterase family #
- Polysaccharide lyase family #

Regulation
- PurR family, Lae I-like
- Other transcriptional regulator
- 2-component response regulator

ABC Transport System
- UgpB/MalE, periplasmic binding-component
- UgpA/LplB, permease
- UgpE/MalG, permease
- MalK, ATPase
- As above, but w/ homology to different substrate specificity

Other Transporters
- Permeases of the major facilitator superfamily
- Maltoporin
- Other transport protein

Sugar metabolism
- Galactokinase
- UDP-glucose-4-epimerase
- Galactose-1-phosphate uridyl transferase
- Galactose mutarotase
- Possible other relationship to sugar metabolism

Other
- Metal-dependent hydrolase
- Transposonase
- Other
- End of sequence
- No adjacent sequence
the respective genomes. ATPases that work with multiple systems, like MsmX of *Streptococcus mutans* (30), could easily be responsible.

Maltodextrins are oligosaccharides of α-1,4-linked glucose but the transporter homology is probably more indicative of an oligosaccharidic substrate for both them and adjacent genes, rather than one of this composition. As in Chapter 2, the presence of transporters for di- or oligosaccharides is consistent with the expected *products from* or *substrates for* most glycosyl hydrolases, with the transport of substrates being more likely for the intracellular GHF 42 enzymes. Several arrangements also have galactose metabolism genes (GalK, GalT, GalM, and/or GalE), which is not unexpected if galactose is ultimately a product of the action of the β-galactosidases.

Although the data compiled for Fig 5-3 showed many different gene arrangements, it proved to be extremely useful. Careful inspection showed that a dozen of the gene arrangements have a conserved pattern clearly indicating a possible substrate (Fig 5-4). Often just preceding or just following the GHF 42 genes are genes of COG 3867 (light green). This COG encodes enzymes belonging to GHF 53, whose members are arabinogalactan endo-1,4-β-galactosidases (EC 3.2.1.89). These enzymes hydrolyze the β-1,4-galactan backbone of arabinogalactan, a pectic substance found in soybeans and citrus fruit, into oligomers. More generally, these can be termed β-galactanases. This pectic polysaccharide could be the target for a pathway involving the proteins encoded by these genes and was suggested as a potential substrate source for GHF 42 by earlier comparisons with GHF 3 substrates (Fig 4-1). The chemical structure of a disaccharide from this polysaccharide, “galactobiose” (Fig 5-5, A) is quite similar to lactose (Fig 5-5, B). GHF 42 β-galactosidase activity on the galactooligosaccharides
released by the action of β-galactanases would be consistent with the proximity of these two genes and the evidence provided by the transporter homology, the specificity of GHF 42 enzymes for the galactose that composes these oligosaccharides, and the presence of nearby galactose metabolism genes for further processing of this sugar.

Additional examples of this gene arrangement may exist where sequencing has not extended much beyond the β-galactosidase gene. For instance, partial sequence from *Thermotoga neapolitana* is extensively homologous to the arrangement from *Thermotoga maritima* (Fig 5-2) and it seems likely that further sequencing would reveal similar genes. It is also possible that a galactanase gene exists elsewhere in the genome for those organisms whose genomes have not been completed.

A β-galactan-oligomeric substrate is also consistent with my predictions based on genome-habitat analysis (Fig 5-1): it is not related to house-keeping in bacteria, pathogenicity in animals, and is not synthesized by animals. Found in plants, this substrate is also more likely to be environmentally accessible to a larger variety of microbes than lactose. The vastness of the field of plant glycomics and the difficulties in purifying pectic substances means that the exact distribution of this substrate is unknown, although it is likely to be widespread. The subset of bacteria possessing β-galactosidase genes adjacent to galactanase genes is not restricted to a particular origin – microbes from soil, plants, and the human gut are represented alongside the pathogenic *Yersinia* spp.. However, both the soil and the human gut have access to the arabinogalactan produced by plants through the deposition of plant detritus and herbivory, respectively. That there is only one representative from the halotolerant-philic/thermophilic group may be because the sequences from the other species representing this type of environment
Figure 5-4. Gene arrangements containing GHF 42 and GHF 53.
Some of the arrangements with GHF 42 genes (blue) seen in Fig 5-3 possess a nearby GHF 53 gene (light green), as well as ABC transporter genes (light orange) and transcriptional regulators similar to LacI (red).

Figure 5-5. Chemical structures of proposed and known β-galactosidase substrates.

(A) “galactobiose” (galactose-β-1,4-galactose) (B) lactose (galactose-β-1,4-glucose) and (C) ONPG (o-nitrophenyl-β-1,4-galactose).
(mostly the *Thermus* spp.) are too short for analysis. The significance of the distinctly different arrangement of the *Thermotoga maritima* genes is also unclear, but could be influenced by the decreased likelihood of finding arabinogalactan from plants in anaerobic marine mud, or an adaptation of the operon for yet another function. Interestingly, those GHF 42 genes clearly associated with GHF 53 genes (boxed on Fig 5-2) are not strongly clustered phylogenetically, but occur in several major groups of bacteria (*Firmicutes, Gamma Proteobacteria, Thermotogae, and Actinobacteria*).

### 5.3.3 GHF 53 associations with GHF 42 and GHF 2

I was curious to learn the extent of the GHF 53 association with GHF 42, and to determine whether the phylogeny of GHF 53 was similar to that presented by the GHF 42 enzymes. In order to do this, I constructed a phylogenetic tree of galactanase sequences using the genes listed by CAZY as belonging to GHF 53 and homologous genes from incomplete genomes revealed by BLAST searches (Fig 5-6). A more complete tree (not shown) revealed that the fungal sequences formed a distinct branch. About half of the bacterial GHF 53 enzymes are associated with GHF 42 genes (Fig 5-6, boxed enzymes), and most of these are more related to each other than to other GHF 53 genes without this association. I then examined the arrangements surrounding the unfamiliar GHF 53 genes (those from bacteria that were not associated with GHF 42 genes). A surprising relationship was revealed: the GHF 53 genes from *Xanthomonas* spp., *Bacteroides thetaiotamicron*, and *Microbulbifer degradans* are associated with β-galactosidases from GHF 2 (Fig 5-6, circled). This suggests that the GHF 2 enzymes may also have a non-lactose hydrolysis function on a substrate that probably predated the evolution of lactose production by mammals.
Figure 5-6. Phylogenetic relationships between GHF 53 enzymes. 
Phylogenetic relationships between GHF 53 enzymes based on neighbor-joining. Bootstrap values shown at the nodes were generated from 1,000 replicates. The accession locus tag of the enzyme (or name, for those genes not identified by genome sequencing) precedes the microorganism name. Enzymes originating from different phylogenetic groups are color coded as per the group names on the right of the figure. Enzymes with associated GHF 42 genes are boxed, and those with GHF 2 genes, circled.
5.3.4 GHF 53 and β-galactosidase synergy

My analysis of these gene arrangements suggests the action of a GHF 53 enzyme is required to release galactan-oligomeric substrates for GHF 42 enzymes, yielding a hypothetical function for these enzymes, and the transporters encoded nearby (Fig 5-7). In this proposed pathway the GHF 53 enzyme acts extracellularly to release galactooligomers that are then transported into the cell by the ABC transporter, and hydrolyzed into their galactose components by the GHF 42 enzyme.

What do we know about GHF 53, and what details can it tell us about the possible actions of GHF 42 enzymes on the proposed substrate? GHF 53 shares structural and functional similarities with the four GHFs with β-galactosidase activity and is also a member of 4/7 superfamily. The only activity reported for this family is arabinogalactan endo-1,4-beta-galactosidase activity, also known as EC 3.2.1.89. These enzymes hydrolyze the β-1,4-galactan backbone of arabinogalactan into galactooligomers in an endo- fashion and structures for GHF 53 enzymes from both fungi and a bacterium (21, 31, 32) have been solved. Plant galactanases and β-galactosidases have frequently been studied in the context of cell-wall modification and ripening, but these galactanases are exo-, not endo- acting, and both of these enzyme activities in plants are found in GHF 35, not GHF 53 and GHF 42. Only sequences from bacteria and fungi are found in GHF 53.

There are a few examples of galactanases studied in bacteria. Several of these galactanases are from *Bacillus* spp. that yielded galactotetramers (Gal⁴) and galactotrimers (Gal³) as products (20, 31, 39, 44). In comparison, most fungal galactans (4, 7, 22, 26, 32, 43), and one known bacterial enzyme (25), yield galactobiose (Gal²). The production of Gal⁴ and Gal³ by most bacterial galactanases, but Gal² by fungal
Figure 5-7. Hypothesized functionality indicated by GHF 42/53 gene arrangements.

Starting outside the cell, an unknown secreted enzyme cleaves the arabinose from the arabinogalactan I, yielding galactan. The actions of a GHF 53 enzyme release galactooligomers such as galactotetraose, from the galactan. Using an ABC (ATP-binding cassette) transport system, the oligomers enter the cell. A homolog of MalE is the binding protein, with MalG and LplB forming a heterotetramer permease, and a MalK homolog encoding the ATPase that drives the transporter. Once within the cell, the galactooligomers release the LacI repression of the operon and are degraded by the GHF 42 β-galactosidase, releasing galactose for the cell to use. Presumably, the arabinose is likewise transported into the cell for use as a carbon source, and may also contribute to the regulation of this proposed system.
galactanases leads to interesting hypotheses regarding the nature and occurrence of β-galactosidases. First, it is clear that a GHF 42 enzyme would be somewhat redundant in fungi because their galactanases can already hydrolyze Gal$_{4}$ and Gal$_{3}$. Second, since the fungi still require a Gal$_{2}$ degrading enzyme, GHF 2 enzymes may serve this purpose. The association between GHF 53 and GHF 2 in a few bacteria, as described above, provides further support for the possibility of GHF 53 products acting as GHF 2 substrates. Availability of Gal$_{2}$ in the environment by the extracellular action of fungal galactanases could also help explain the occurrence of GHF 2 in other microorganisms that occur in environments lacking lactose. About one of every three of the fully-sequenced microbial genomes contains at least one GHF 2 gene, and there are examples, as with GHF 42, of the host organisms being isolated from habitats not expected to contain lactose. Alternatively, these GHF 2 enzymes could be acting on Gal$_{2}$ released by another GH not yet recognized for this function, or could be β-mannosidases or β-glucuronidases instead of β-galactosidases. Thirdly, if some of the GHF 42 enzymes do not act on Gal$_{2}$ (but do act on Gal$_{3}$ or Gal$_{4}$), then GHF 2 enzymes in the same organism might complete the hydrolysis.

The process of Gal$_{4}$ degradation by GHF 42 almost certainly leads to (temporary) production of Gal$_{2}$ prior to complete hydrolysis to galactose. This might mean that the different quaternary structures observed for GHF 42 (monomers and trimers) are related to its action on a range of substrates. As a monomer, the structure of the β-galactosidase from *Thermus thermophilus* A4 shows a cleft-type active site suitable for degradation of a oligosaccharidic substrate (10), such as Gal$_{4}$. However, the A4-β-galactosidase can also form a trimer with a pocket-type active site more suitable for hydrolysis of smaller
substrates such as Gal$_{2}$ or lactose. The LacZ β-galactosidase also has pocket-type active site, but this is created by the presence of additional domains rather than multimerization (10). The GHF 42 enzyme may exist in both forms (probably with one dominating) in vivo with the monomer form hydrolyzing Gal$_{4}$ into Gal$_{2}$ and the trimer form completing the hydrolysis of Gal$_{2}$. It is also possible that some or all of the GHF 42 enzymes act purely in an exo- fashion.

5.3.5 Galactan-galactosidase relationships reported in literature

As mentioned in Chapter 4, efforts regarding relationships between β-galactosidases and galactans were made using larch arabinogalactan. However, this arabinogalactan is of type-II and has a β-1,3 linked backbone, not a β-1,4 linked backbone and the two enzymes were tested on whole polysaccharide rather than on β-1,3-galactooligomers. These two enzymes also did not have adjacent GHF 53 genes. In fact, none of the characterized enzymes (Table 4-1, circled enzymes on Fig 5-2) has adjacent sequence data encoding a GHF 53 gene either because the sequence is too short, or another gene arrangement is present. Because the gene arrangements vary so widely within a given genus, and because this arrangement does not appear to be tightly conserved with respect to 16S rDNA phylogeny, I cannot predict whether any of those GHF 42 sequences lacking adjacent sequence data would be likely to have an adjacent GHF 53 ORF.

A search of the literature reveals that hydrolysis of galactooligomers by β-galactosidases was suggested by Nakano et al. (27). They examined the kinetic response of several enzymes to galactooligomers, and achieved $K_m$ values ranging from 4.5 to 19.4 mM from several fungal β-galactosidases (which almost certainly do not belong to GHF
42) with Gal\textsuperscript{2}. Two of the fungal β-galactosidases were unable to act on higher oligomers and the other three had higher K\textsubscript{m} values on Gal\textsuperscript{3} and Gal\textsuperscript{4}, in line with the optimization towards Gal\textsuperscript{2} production by fungal galactanases. For four of the five fungal enzymes, the K\textsubscript{m} values at least doubled when lactose was the substrate. This supports the above hypothesis that the GHF 2 β-galactosidases of fungi are degrading Gal\textsuperscript{2} produced by fungal GHF 53 (or other unknown) enzymes rather than lactose. Nakano et al. (27) also tested LacZ from \textit{E. coli} on these substrates (ranging from dimers to tetramers), but this GHF 2 enzyme was unable to hydrolyze these substrates, indicating LacZ is not bifunctional in this manner.

Unfortunately, these observations of Nakano et al. went unnoticed in the context of biochemical studies of bacterial β-galactosidases. More recently the kinetics of \textit{Bifidobacterium adolescentis} BgalII were examined on galactooligomeric substrates and oligomers with additional galactoside moieties linked β-1,4 to lactose (known as galactooligosaccharides, GaOS or GOS or transgalactooligosaccharides, TOS) (40), who did not cite the work by Nakano et al. This enzyme did not hydrolyze lactose, but yielded K\textsubscript{m} values from 2.2 to 6.4 mM for these other substrates. Of the galactooligomers, the enzyme had the lowest K\textsubscript{m} for Gal\textsuperscript{2}, followed by Gal\textsuperscript{4}, and lastly Gal\textsuperscript{3}. Further studies showed that the enzyme was specific for β-1,4 linkages as the enzyme had no activity on galactose units linked α-1,4 or β-1,6 and had only weak activity on galactose linked β-1,3 (11). \textit{Bifidobacterium longum} has been observed to grow on galactan in pure-culture, but apparently required large amounts of the polysaccharide to compensate for the inability to use most of the arabinogalactan molecule (8). β-galactosidase activity increased during growth on this substrate (8), but the genes encoding the observed activity were not
identified \((B. \textit{longum})\) has a GHF 2 gene as well as two GHF 42 genes) nor was the \(\beta\)-galactosidase activity shown to be necessary for growth.

An \textit{in vivo} substrate for an enzyme would be expected to allow growth of the host microorganism, cause upregulation of that enzyme, and have associations with other genetically-encoded functions in the same microorganism. GHF 42 possessing microorganisms do not all grow on lactose, expression of their GHF 42 enzymes does not necessarily increase in response to lactose, and the transporter genes associated with the GHF 42 genes are not homologous to those used by \textit{E. coli} for the transport of lactose. GHF 42 enzymes do have activity on galactooligomers \((11)\), \(\beta\)-galactosidase activity increases during growth on galactan, and my observations regarding gene arrangements support an association with galactanases. This suggests that further proof of GHF 42 enzyme participation in arabinogalactan type-I degradation is worth pursuing.

\textbf{5.3.6 Other possible functions}

As mentioned earlier, the occurrences of duplicate GHF 42 genes in some organisms suggest that these enzymes might have more than one function. Unfortunately, no other relationships suggestive of function appear as consistently as the GHF 53 association shown in Figure 5-4. Perhaps this should not be surprising given that the sporadic occurrence of the GHF 42 enzymes already indicates a tumultuous history whereby such clues are easily lost. The presence of transposon genes (Fig 5-3, gray) may indicate the basis for some of incongruities observed in both the gene arrangements and overall phylogeny. Alpha-galactosidase genes occur in many of the arrangements, suggesting additional galactose units may be present via alpha linkages in the
polysaccharidic substrate(s). The GHF 35 association with GHF 42 in *Carnobacterium maltaromaticum* BA (5) (two β-galactosidase genes adjacent to each other) seems significant, but does not appear to encode a similar association as GHF53-GHF 42 for several reasons. First, GHF 35 are only known to act in a exo-fashion, not an endo-fasion, so oligosaccharide products would not be expected. Secondly, unlike many of the GHF 35 enzymes for which a function has been found, this GHF 35 appears to be intracellular, and therefore would not be expected to act on galactan directly. However, GHF 35 could be cooperatively acting on the galactooligomers intracellularly, since the β-1,3 linkages bacterial GHF 35 enzymes have been demonstrated to act on (16, 36, 42) have recently been found to occur infrequently in some galactans (12).

Genes that could encode endo-acting glycoside hydrolases like GHF 53 are present in some other arrangements, but again, not consistently. Endo-acting functions are known for several of the GHFs represented, but only GHF 5 (near GHF 42 genes SAV1027, SCO7407, and Lmes03001880 (Fig 5.3)) is known to be active in this way on a β-galactan. GHF 5 has a single example from a fungus (*Trichoderma viride*) which acts as a endo-1,6-β-galactanase on β-(1,3)(1,6)-galactan from a green alga (*Prototheca zopfii*) (18). No equivalent activity has been observed in a bacterial GHF 5 enzyme. Exploration of associations between the actions of GHF 42 and other GHs will require further studies of novel GHs, and a willingness to hypothesize reactions downstream of the products yielded by a given GH during *in vitro* characterization.
5.4 Conclusion

The physiological and biochemical data available for GHF 42 enzymes suggests that their natural substrate *in vivo* is not lactose (Chapter 4). Understanding the *in vivo* functions of β-galactosidases will help us to better appreciate their roles in the global carbon cycle, better exploit them as tools for molecular biology, and develop them for degradation of polysaccharides found in foods or industrial products. The alternative substrate(s) are probably polysaccharidic in origin and not restricted in occurrence to extreme environments. The phylogenetic history of GHF 42 enzymes suggests that they can have more than one function, and that frequent gene loss, duplication, and perhaps horizontal transfer have occurred, perhaps mediated by transposons. A relationship between β-galactanase and GHF 42 genes, together with part of an ABC binding cassette transporter system was found in several genomes. This strongly supports exploration of these galactooligomers as a substrate for some GHF 42 enzymes. Although previous hints implying this potential function exist, my analysis of gene arrangements and patterns is the first evidence of a widespread function for this enzyme group.

The substrates of other GHF 42 enzymes remain unknown, as it is unlikely that all of those without the conserved gene arrangement are acting on galactooligomers from galactan. Another possible substrate source are the polysaccharides found in algal cell walls. Algae seem likely to be found in the widest range of habitats occupied by the extremophilic and “normal” organisms, but the limited information available regarding cell-wall composition does not support this hypothesis. Hypotheses regarding additional putative substrates will likely be revealed by further genome sequences, better knowledge of the types and presence of polysaccharides present in the environment, and continued
biochemical characterization. The gene arrangements also suggest that some GHF 2 β-galactosidases may also act on galactooligomers in vivo, implying additional substrates beyond this and lactose may also exist for this well-studied enzyme group.

My examination of sequence data and gene organization led to detection of an association between GHF 42 and GHF 53 genes that implies a relationship between the enzymes that they encode. Subsequent review of the literature supports this association, but indicates that it has not been explicitly explored, and that further experimentation is necessary to confirm the functional link between these two GHs in vivo. Based on these results I propose that some GHF 42 enzymes play a role in the degradation of arabinogalactan type-I downstream from the actions of GHF 53 and the ABC transporter at least partially encoded nearby (Fig 5-7). Growth of a bacterium on galactan as a sole carbon source and simultaneous induction of a GHF 42 β-galactosidase can provide supporting evidence for this hypothesis. However, the best demonstration would involve knocking out the GHF 42 enzyme and observing a change in the ability of the microorganism to grow on galactan. The hypothesis that GHF 42 enzymes are involved in galactan degradation will be tested with mutant studies in B. subtilis, a mesophilic microorganism that possess two GHF 42 genes (Chapter 6).
5.5 Materials and methods

**Genome analysis.** I compared the list of microorganisms with fully sequenced prokaryotic genomes ([http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi](http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi)) to the list of organisms known to possess GHF 42 genes (CAZy). I analyzed the general habitat and specific isolation source of the microorganisms with and without these genes according to the isolation source as indicated by the NCBI Entrez Genome Project database ([http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj)) or culture collection databases with reference to the exact strain used such as the DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (German Collection of Microorganisms and Cell Cultures)) ([http://www.dsmz.de/](http://www.dsmz.de/)) and the ATCC (American Type Culture Collection) ([http://www.atcc.org/catalog/all/allIndex.cfm](http://www.atcc.org/catalog/all/allIndex.cfm)). I also examined the genes and nucleotide sequences proximal to GHF 42 genes found in both genomes and resulting from smaller sequencing projects, currently represented by over 50 sequences. When possible, I analyzed at least three ORFs upstream and downstream. The orientation and size of the ORFs were also considered. The theoretical amino acid sequences were used to search the NCBI's Conserved Domain Database (CDD) (23) ([http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi](http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)) using CDsearch in parallel with searches of the protein database using BLAST (1). Frequently, significant alignments with conserved domains were detected (in the SMART (33), pfam (2) and/or COG databases (37, 38)), and the matches with the lowest E values (i.e. least likelihood of being random) were recorded. When none of these were identified, or the alignment did not match a majority of the conserved domain, I examined the BLAST results and the general trend of putative function recorded instead, when possible. I then assigned
different colors to genes with different functions, producing color-coded alignment diagrams whereby conserved patterns could more easily be discerned. Thirty to fifty amino acids of the N-terminal sequences of each GHF 42 were examined using the Signalp WWW server (3) to determine whether or not they were probably intracellular.

**Phylogenetic trees.** The amino acid translations of GHF 42 and GHF 53 genes identified using CAZY and BLAST were initially aligned using Clustal W (BioEdit platform, Version 5.0.6; Department of Microbiology, North Carolina State University (http://www.mbio.ncsu.edu/Bioedit/bioedit.html)) and visually inspected to correct errors in the alignments. The sequences were trimmed to avoid bias in the highly divergent N and C-terminal regions. The alignments covered regions homologous to BSU34130 (*Bacillus subtilis*) from amino acid position 16 to 684 for GHF 42; and BSU34120 from amino acid position 53 to 337 for GHF 53. Bootstrapped neighbor-joining phylogenetic trees were produced using MEGA (19), using the complete deletion, amino: Poisson correction model, and uniform rates among sites, for 1000 replicates, using 64238 as the random seed number. Fragmentary sequences were not included in the alignment, but were instead positioned on the tree according to the highest homology indicated by BLAST.
5.6 References


Chapter 6

A natural function for a Glycoside Hydrolase Family 42 β-galactosidase of *Bacillus subtilis*
6.1 Summary

My examination of the gene arrangements surrounding GHF 42 genes in several organisms led me to hypothesize that degradation products from arabinogalactan type-I could be their natural substrate (Chapter 5). Arabinogalactan type-I is a pectic substance from plants, and as such, would be available in the habitats where many organisms with genes encoding GHF 42 enzymes are found. *Bacillus subtilis*, the best-characterized Gram-positive organism, is ideal for testing whether the GHF 42 enzyme can hydrolyze this substrate. This organism has a well-studied and easily manipulated genetic system, a sequenced genome, does not have competing β-galactosidases in GHF 2 (the family to which LacZ to *Escherichia coli* belongs) or GHF 35, and does not use lactose as a sole carbon source. The *B. subtilis* genome contains two non-adjacent GHF 42 encoding genes, *lacA* and *yesZ*. Adjacent to *lacA* is a gene *galA* that encodes a ORF homologous to arabinogalactan endo-1,4-β-galactosidases from GHF 53, hypothetically allowing *B. subtilis* to produce the galactooligomers, which LacA can then hydrolyze. The *lacA* gene together with *galA* and nearby ABC transporter genes represents an example of the conserved homologous arrangement observed in other genomes (Chapter 5, Fig 5-4) (gene arrangement in *B. subtilis* is shown at the top of Fig 6-1).

I used *B. subtilis* to determine whether the GHF 42 LacA could be involved in arabinogalactan type-I utilization. I constructed plasmids in *E. coli* containing fragments of the *B. subtilis* genome that included *lacA*, *yesZ*, or *galA* and studied the effects of the heterologously expressed proteins on the physiology of *E. coli*. I then inserted a chloramphenicol resistance (Cm<sup>R</sup>) cassette into the coding regions of each of these genes, transformed *E. coli* cells, and obtained constructs where each of the genes was
independently interrupted. The Cm\textsuperscript{R} constructs were transferred to \textit{B. subtilis} where recombinants with independent insertions in the \textit{lacA}, \textit{yesZ}, or \textit{galA} genes were obtained. I then examined the altered responses of these mutants to induction by, and growth on, a variety of carbon sources. In this way, I provide evidence that LacA of \textit{B. subtilis} strain 168 functions as a galactooligomerase contributing to the utilization of galactan backbone of arabinogalactan type-I as a carbon source. The function of the other \(\beta\)-galactosidase, YesZ, is not clear. The presence of similar gene arrangements in several other organisms suggests that other GHF 42, and perhaps GHF 2, \(\beta\)-galactosidases encoded adjacent to GHF 53 genes have similar functions to LacA.
6.2 Introduction

The genome sequence *B. subtilis* contains two sequences encoding β-galactosidases, *lacA* (BSU34130) and *yesZ* (BSU07080), both belonging to GHF 42, and in spite of the presence of these β-galactosidase genes, *B. subtilis* does not use lactose as a sole carbon source (4). The presence of two different genes each in a different context within the genome suggests that they have separate functions. The arrangement of genes surrounding *lacA*, and homologous arrangements in other genomes, suggested that function of these GHF 42 enzymes could be the further degradation of products yielded by the activity of GHF 53 enzymes on arabinogalactan type-I substrates (Chapter 5, Fig 5-7). GHF 53 activity in *B. subtilis* was confirmed by the work of Labavitch et al. in 1976 (12), although the gene encoding the enzyme was not identified. The functions of the β-galactosidases of *B. subtilis* have not been studied in detail probably due to the infrequency with which their expression has been observed. The β-galactosidase activity of LacA in *B. subtilis* was observed because of interference with LacZ-reporter studies (5, 7) and *lacA* was consequently mutated to eliminate this problem (8). Even less is known about YesZ, which, prior to the sequencing of the genome, was known only because of expression generated by a transposon-promoter experiment (24). Much later, Errington returned to *lacA* (identified in the genome as *yvfN*) and confirmed that it was regulated by the repressor *lacR* (4), which the nascent genome of *B. subtilis* had revealed was nearby. They also noticed the potential for a polycistronic operon comprised of *lacA*, three genes similar to those found to encode maltose or maltodextrin transport systems (*yvfK, L and M*), and a gene (*yvfO*, also known as *galA*) encoding an ORF with homology to an arabinogalactan-endo-1,4,β-galactosidase (4). This is the same as the
Surprisingly, in spite of several clues, a functional connection between LacA and GalA has not yet been explored. Labavitch et al. (12) reported induction of β-galactosidase activity in *B. subtilis* using arabinogalactan from soybean (which contains arabinogalactan type-I) and although they studied the products yielded by the *B. subtilis* galactanase, they never mention further degradation of these by the induced β-galactosidase. Two decades later Daniel et al. (4) reported that they were unable to induce expression of the endogenous β-galactosidase on sugars (but admitted that they had not tested any plant exudates), apparently ignoring the potential relevance of the adjacent arabinogalactan-endo-1,4,β-galactosidase gene.

Confirmation of the ability of *B. subtilis* to grow on galactan as a sole carbon source with a consequent increase in β-galactosidase activity would support my proposal of the following pathway: low levels of galactanase (GalA) release galactooligomers (gal<sub>4</sub>) (the proposed inducer), from galactan, these gal<sub>4</sub>s release repression (derepression) of LacR allowing upregulation of galactanase (GalA) (which releases gal<sub>4</sub> from galactan if it is present) and upregulation of β-galactosidase activity (just LacA, not YesZ), which then acts on gal<sub>4</sub>, yielding galactose as a carbon source. The hypothesis that galA (*yvfO*) and lacA (*yvfN*) respectively produce an extracellular arabinogalactan-endo-1,4-β-galactosidase and an intracellular β-galactosidase that act consecutively on arabinogalactan type-I, with LacA degrading the products yielded by the action of GalA, can be tested by disrupting this proposed pathway at different steps by gene inactivation. This would be expected to have the effects on X-Gal hydrolysis outlined in Table 6-1.
Therefore, I created \textit{B. subtilis} mutants in which \textit{lacA}, \textit{galA}, and \textit{yesZ} were each independently inactivated. I tested to confirm that the \textit{lacA::Cm}^R mutant, but not the \textit{yesZ::Cm}^R mutant, no longer expressed β-galactosidase activity in the presence of this substrate, and was affected in its ability to grow on galactan. I also tested the \textit{galA::Cm}^R mutant to determine if its activity (production of galactooligomers) was necessary for increased expression of LacA as proposed.

### Table 6-1 Expected effects of mutations on the proposed pathway in the presence or absence of galactan

<table>
<thead>
<tr>
<th>Galactan present?</th>
<th>Release of inducer (gal^4)</th>
<th>De-repression</th>
<th>GaLA expression</th>
<th>LacA expression (X-Gal hydrolysis)</th>
<th>Hydrolysis of gal^4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>Yes</td>
<td>→</td>
<td>→</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>Wild-type</td>
<td>No</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ΔlacA</td>
<td>Yes</td>
<td>→</td>
<td>→</td>
<td>→</td>
<td>X</td>
</tr>
<tr>
<td>ΔgalA</td>
<td>Yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ΔyesZ</td>
<td>Yes</td>
<td>→</td>
<td>→</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>ΔlacR</td>
<td>Yes</td>
<td>→</td>
<td>→</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>ΔlacR&amp;ΔgalA</td>
<td>Yes</td>
<td>X</td>
<td>(→)</td>
<td>→</td>
<td>→</td>
</tr>
</tbody>
</table>

¥ cannot occur if either galactan or galactooligomers are absent
→ step occurs, pathway proceeds
X step does not occur, normal pathway blocked
(→) no repression by LacR occurring, step bypassed, pathway not blocked
6.3 Results

6.3.1 Production of β-galactosidase activity in *B. subtilis*.

A variety of simple sugars or polysaccharides were placed in 48-well plates with *B. subtilis* cells to efficiently screen β-galactosidase activity levels, using ONPG, to determine whether increased β-galactosidase activity resulted from the presence of galactan or other carbon sources (Table 6-2). Of the sugars, only gentiobiose (glucose-β-1,6 glucose) caused slight increase (A\textsubscript{420} 0.35) in β-galactosidase activity above that found for the glucose control (A\textsubscript{420} 0.30). For the polysaccharides, cellulose was used as a control. The average amount of background β-galactosidase activity was greater on polysaccharides (A\textsubscript{420} 0.40 - 0.70) than simple sugars. Soy flour and galactan clearly increased β-galactosidase activity (A\textsubscript{420} \geq 1.5). The ONPG hydrolysis was above background levels, but less than observed for soy flour and galactan, with gum arabic, polygalacturonic acid, apple and citrus pectin, phytone, xylan, and gellan gum (A\textsubscript{420} 0.71 – 1.35). Since this confirmed stimulation of β-galactosidase activity by galactan, I then interrupted the β-galactosidase genes to see which, or if both, was responsible, and whether either affected growth on galactan as a sole carbon source.

Table 6-2 Some polysaccharides tested for upregulation of β-galactosidase activity

<table>
<thead>
<tr>
<th>Polysaccharide</th>
<th>Backbone</th>
<th>Linkage</th>
<th>Increased ONPG hydrolysis?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starch</td>
<td>Glucan</td>
<td>α-1,4</td>
<td>-</td>
</tr>
<tr>
<td>Dextran</td>
<td>Glucan</td>
<td>α -1,6</td>
<td>-</td>
</tr>
<tr>
<td>Cellulose</td>
<td>Glucan</td>
<td>β -1,4</td>
<td>-</td>
</tr>
<tr>
<td>Laminarin</td>
<td>Glucan</td>
<td>β -1,4</td>
<td>-</td>
</tr>
<tr>
<td>Polygalacturonic acid</td>
<td>Galacturonan</td>
<td>α -1,4</td>
<td>+</td>
</tr>
<tr>
<td>Apple pectin</td>
<td>As above, but with branches</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Citrus pectin</td>
<td>As above, but with branches</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Xylan</td>
<td>Xylan</td>
<td>β -1,4</td>
<td>+</td>
</tr>
<tr>
<td>Locust bean gum</td>
<td>Galactan</td>
<td>α -1,6</td>
<td>-</td>
</tr>
<tr>
<td>Arabinogalactan type II</td>
<td>Galactan</td>
<td>β -1,3</td>
<td>-</td>
</tr>
<tr>
<td>Gum arabic</td>
<td>Galactan</td>
<td>β -1,3/1,6</td>
<td>+</td>
</tr>
<tr>
<td>Arabinogalactan type I</td>
<td>Galactan</td>
<td>β -1,4</td>
<td>++</td>
</tr>
</tbody>
</table>

- no upregulation, + modest upregulation, ++ strong upregulation

Additional sugars and polysaccharides were tested and are described in Materials & Methods
6.3.2 Construction of vectors and knockouts

Genomic DNA from *B. subtilis* was used to construct genomic libraries in *E. coli* and the transformants were screened for X-Gal hydrolysis. The restriction patterns of the cloned fragments in X-Gal hydrolyzing transformants were compared to those expected as per the *B. subtilis* genome. In this way, the genes encoding LacA (GHF 42 β-galactosidase) and GalA (GHF 53 arabinogalactan-endo-1,4-β-galactosidase) were cloned; a gene encoding a second GHF 42 β-galactosidase in *B. subtilis*, yesZ, was separately cloned. Subclones were created from each of these larger inserts (pYvf, pYvfK, and pYesZ) (Fig 6-1 A & B) and were ultimately used to create the constructs

![Figure 6-1 Plasmid inserts from Bacillus subtilis genome](image)

Figure 6-1. Plasmid inserts from *Bacillus subtilis* genome. Fragments of the *Bacillus subtilis* genome encoding LacA (Panel A) and YesZ (Panel B) were targeted by creation of genomic libraries. The fragments were subcloned using the restriction endonuclease sites indicated. Other restriction sites (red, arrows) were used to insert a CAT cassette in the given subclones in order to disrupt the genes.
placA::Cm\textsuperscript{R}, pgalA::Cm\textsuperscript{R}, and pyesZ::Cm\textsuperscript{R}, which carry the CAT (chloramphenicol acetyltransferase) cassette interrupting the lacA, galA, and yesZ genes. In *E. coli*, disruption of the plasmid-borne lacA or yesZ genes yielded colonies that did not hydrolyze X-Gal, and disruption of galA alone did not affect X-Gal hydrolysis by the preceding but intact lacA gene in the pYvfs construct. *E. coli* expressing GalA alone did not hydrolyze X-Gal.

### 6.3.3 Physiological effects on *E. coli*.

The process of creating in *E. coli* the constructs used for the interruption of genes in *B. subtilis* provided the opportunity to determine whether the expression of the enzymes encoded by the galA, lacA, and yesZ genes could effect changes in the physiology of *E. coli*. The specific physiologies of interest were the ability to grow on lactose or galactan as sole carbon sources. Wild-type *E. coli* grows very poorly on galactan, but grows well on lactose using its native GHF 2 β-galactosidase, LacZ. The *E. coli* strain ER2585F', which contains a deletion of lacZ, only hydrolyzed X-Gal when it possessed a plasmid carrying lacA, yesZ, or lacZ (Table 6-3), confirming expression of these β-galactosidases. The GHF 42 β-galactosidases encoded by lacA (pLacA) and yesZ (pGalAYesZg) were each expressed in (lacZ) *E. coli* to see whether they could replace lacZ in allowing growth on lactose. *E. coli* expressing LacZ was able to grow on lactose minimal media, but those expressing LacA or YesZ were not immediately able to do so (Table 6-3). The *E. coli* expressing LacA eventually exhibited some growth on lactose, but it was still less than that displayed by *E. coli* expressing LacZ.
Table 6-3 Physiological effects of lacZ, lacA, galA and yesZ, expression in E. coli

<table>
<thead>
<tr>
<th>Enzymes expressed:</th>
<th>Hydrolysis of*:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X-Gal</td>
</tr>
<tr>
<td>pΔα18</td>
<td>-</td>
</tr>
<tr>
<td>pLacZ</td>
<td>+</td>
</tr>
<tr>
<td>pLacA</td>
<td>-</td>
</tr>
<tr>
<td>pGalA</td>
<td>-</td>
</tr>
<tr>
<td>pYvfs (pGalALacA)</td>
<td>-</td>
</tr>
<tr>
<td>PGalAYesZg</td>
<td>-</td>
</tr>
<tr>
<td>pGalALacZ</td>
<td>+</td>
</tr>
</tbody>
</table>

*Hydrolysis of X-Gal determined by formation of blue product.
Hydrolysis of other compounds determined by growth on minimal media.

Of greater interest was determining whether expression of both GalA and LacA would confer on E. coli the ability to grow on galactan. Plasmids carrying galA together with yesZ or lacZ were also constructed (not shown) to see whether or not these other β-galactosidases would be functionally equivalent to the LacA β-galactosidase. The only E. coli strain that exhibited growth on galactan minimal media equivalent to that seen on the positive controls (galactose) was the one containing both the galactanase gene (galA) and lacA; the yesZ and lacZ encoded enzymes were not functional substitutes (Table 6-3). E. coli with either the galactanase gene (galA) or the lacA gene alone did not grow because these enzymes must act consecutively in order to produce a substrate E. coli can use as a sole carbon source (galactose).

6.3.4 Creation of B. subtilis mutants.

Transformation of B. subtilis with placA::Cm<sup>R</sup>, pgalA::Cm<sup>R</sup>, or pyesZΔCm<sup>R</sup> led to Cm<sup>R</sup> colonies, indicating that marker-replacement double-crossover recombination had
occurred as none of these vectors can replicate within \textit{B. subtilis}. Genomic DNA from these Cm\textsuperscript{R} recombinants yielded the expected products when amplified using primers (see materials and methods) where at least one of the primer pair annealed outside the cloned regions of DNA; the wild-type Cm\textsuperscript{S} colonies yielded products smaller by 1.4 kb, the difference indicating the presence of the Cm\textsuperscript{R} cassette and disruption of the gene of interest.

\textbf{6.3.5 Comparison of wild-type and mutants}

Wild-type \textit{B. subtilis} is unable to grow on galactose as a sole carbon source because it relies on the arabinose-transporter (which is expressed in the presence of arabinose, but not expressed in response to galactose) to import this sugar in its monosaccharide form (11). Also, in its natural state, arabinogalactan possesses arabinose containing side chains. For these reasons arabinose was used in combination with other sugars to determine whether the typically present arabinose might influence expression of \(\beta\)-galactosidase activity in response to galactan, and to determine whether galactose as a monosaccharide (instead of the hypothesized galacooligomers) effected \(\beta\)-galactosidase activity. Wild-type \textit{B. subtilis} and mutant strains were grown on Schaeffer media containing X-Gal and either arabinose, galactose, galactan, or a combination of these substrates. This medium is typically used to observe processes occurring during the stationary phase of growth such as sporulation and expression of extracellular enzymes such as \(\alpha\)-amylase – and hypothetically GalA and LacA which are similarly breaking down a larger polysaccharide. Expression of LacA but not YesZ was observed with galactan, and functional GalA was required for expression of LacA (Table 6-4). Lack of an effect of \textit{galA} interruption on LacA expression in \textit{E. coli} constructs (as observed by
X-Gal hydrolysis) suggests that this is not a polar effect. The addition of arabinose enhanced the intensity of the blue color observed from both the wild-type and mutant *B. subtilis* strains in the presence of galactan, but when used alone did not cause increased β-galactosidase activity in wild-type. Galactose combined with arabinose also did not increase β-galactosidase activity. These results suggest that galactooligomers, not full-length galactan, or galactan hydrolyzed to galactose, is responsible for the observed upregulation of the LacA β-galactosidase.

**Table 6-4 B. subtilis induced hydrolysis of X-Gal by various substrates**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Arabinose</th>
<th>Arabinose +galactose</th>
<th>Galactan</th>
<th>Arabinose +galactan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>LacA::CmR</td>
<td>-</td>
<td>NT*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>YesZ::CmR</td>
<td>-</td>
<td>NT*</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>GalA::CmR</td>
<td>-</td>
<td>NT*</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*NT – not tested

Wild-type *B. subtilis* and CmR insertion strains were plated on M9 minimal agar containing 0.2% galactan or glucose to observe effects on growth. *B. subtilis* grew better on galactan than glucose, but the effect of the lacA::CmR and galA::CmR mutations on growth was subtle (Table 6-5). The data sheet for the commercially-available galactan used indicates that it is only 83% galactose (13). I assumed that 90% of the galactose was present as galactan, and constructed an M9 medium with a sugar mixture equivalent to the amount and type of contaminating carbon sources. The levels of growth by all three mutants were identical and demonstrated that this carbon source could have contributed to the unexpected level of growth of the lacA::CmR and galA::CmR mutants on the galactan.
Table 6-5 Growth of wild-type and insertion strains on minimal media plates

<table>
<thead>
<tr>
<th></th>
<th>wt</th>
<th>lacA::Cm\textsuperscript{R}</th>
<th>yesZ::Cm\textsuperscript{R}</th>
<th>galA::Cm\textsuperscript{R}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Galactan</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Sugar mix</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

To better observe the effects of the mutations on β-galactosidase activity, the 48-well ONPG plate assay was repeated using the mutants and a decreased number of substrates. Interruption of lacA reduced the β-galactosidase activity observed with polygalacturonic acid, citrus pectin, apple pectin, galactan, and soy flour to a level observed with controls. An equivalent decrease in activity was not detected for the lacA::Cm\textsuperscript{R} and yesZ::Cm\textsuperscript{R} mutants on gentiobiose, phytone, xylan, and gellan gum. The galA::Cm\textsuperscript{R} mutant had reduced activity with soy flour and galactan (compared to wild-type and the yesZ::Cm\textsuperscript{R} mutant), but expressed more activity than the lacA::Cm\textsuperscript{R} mutant. The reduction of activity was less pronounced on the galactan than the soy flour.

6.3.6 LacA β-galactosidase purification

Because LacA may have a different function than other β-galactosidases, and no GHF 42 enzyme known to be encoded adjacent to a GHF 53 gene has been characterized, it was of interest to biochemically characterize the enzyme. Enzyme from clarified lysate was initially examined (data not shown). Then, N-terminal six-His tagged LacA enzyme was purified using Nickel-charged IDA column material (Table 6-6). The purified protein had a specific activity of 51 units/mg.
Table 6-6. Purification scheme for (His-tagged) LacA from *E. coli*.

<table>
<thead>
<tr>
<th>Step*</th>
<th>Vol. (ml)</th>
<th>Total Protein (mg)</th>
<th>Total Activity (U)</th>
<th>Sp. Act. (U mg⁻¹)</th>
<th>Yield (%)</th>
<th>Purification (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarified lysate</td>
<td>25</td>
<td>362</td>
<td>8950</td>
<td>25</td>
<td>100</td>
<td>1.0</td>
</tr>
<tr>
<td>Nickel-IDA</td>
<td>12</td>
<td>62</td>
<td>3390</td>
<td>54</td>
<td>39</td>
<td>2.2</td>
</tr>
</tbody>
</table>

* For purification procedure details, see Materials and Methods

6.3.7 LacA β-galactosidase characterization.

His-tagged LacA had the highest activity at pH 6.0 to 6.5 (Fig 6-2) and 50°C (Fig 6-3), and was thermolabile above this temperature (Fig 6-4). These results agree with those obtained from untagged LacA (data not shown). None of the metal ions increased enzyme activity, and Zn²⁺, Co²⁺ and Cu²⁺ were inhibitory (Table 6-7). Activity on PNPG was about equal to that observed with ONPG, and trace activity (4%) was observed with PNPAp and ONP-fucose. Most of the other substrates tested yielded no detectable activity (< 0.1% of that observed with ONPG) (Table 6-8).

Fig 6-2. Effects of pH on ONPG hydrolysis by purified LacA.

![Figure 6-2](image)

Figure 6-2. Effects of pH on ONPG hydrolysis by purified LacA at 40°C using the following buffers: citric acid buffer (□), Zm (○), MOPS buffer (△). Molarities and pH as described in methods. The specific activity corresponding to the 100% value was 47 U/mg.
Figure 6-3. Thermal dependency of activity of purified LacA enzyme with ONPG.

![Graph showing thermal dependency of activity of purified LacA enzyme with ONPG.]

Fig 6-3. Thermal dependence of activity of purified LacA. The specific activity corresponding to 100% was 39 U/mg.

Fig 6-4. Thermostability of purified LacA.

![Graph showing thermostability of purified LacA versus time of incubation at various temperatures: 40°C (□), 50°C (◊), 55°C (Δ), 60°C (○). The specific activity corresponding to the 100% value was 39 U/mg.]

Figure 6-4. Thermostability of purified LacA versus time of incubation at various temperatures: 40°C (□), 50°C (◊), 55°C (Δ), 60°C (○). The specific activity corresponding to the 100% value was 39 U/mg.
Table 6-7 Effects of ions on the activity of LacA enzyme.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Relative activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOPS pH 7, no additions</td>
<td>100</td>
</tr>
<tr>
<td>MOPS w/</td>
<td></td>
</tr>
<tr>
<td>1mM Mg^{++}</td>
<td>110</td>
</tr>
<tr>
<td>1mM Ca^{++}</td>
<td>100</td>
</tr>
<tr>
<td>1mM Mn^{++}</td>
<td>88</td>
</tr>
<tr>
<td>1mM Co^{++}</td>
<td>75</td>
</tr>
<tr>
<td>1mM Zn^{++}</td>
<td>33</td>
</tr>
<tr>
<td>1mM Ni^{++}</td>
<td>29</td>
</tr>
<tr>
<td>1mM Cu^{++}</td>
<td>3</td>
</tr>
<tr>
<td>10mM Na^{+}</td>
<td>104</td>
</tr>
<tr>
<td>10mM K^{+}</td>
<td>101</td>
</tr>
</tbody>
</table>

*The specific activity at 100% is 48 U/mg

Table 6-8. Relative activity of the purified LacA enzyme on various chromogenic substrates as measured by ONP or PNP release at 40°C.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Relative activity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o$-nitrophenyl-$\beta$-D-galactopyranoside (ONPG)</td>
<td>100</td>
</tr>
<tr>
<td>$p$-nitrophenyl-$\beta$-D-galactopyranoside (PNPG)</td>
<td>104</td>
</tr>
<tr>
<td>$p$-nitrophenyl-$\alpha$-L-arabinopyranoside</td>
<td>4</td>
</tr>
<tr>
<td>$o$-nitrophenyl-$\beta$-D-fucopyranoside</td>
<td>4</td>
</tr>
<tr>
<td>$p$-nitrophenyl-$\alpha$-D-galactopyranoside</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>$p$-nitrophenyl-$\beta$-D-glucopyranoside (PNPGlu)</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>$o$-nitrophenyl-$\beta$-D-xylopyranoside</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>$p$-nitrophenyl-$\beta$-D-mannopyranoside</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>$p$-nitrophenyl-$\alpha$-L-arabinofuranoside</td>
<td>&lt;0.1%</td>
</tr>
</tbody>
</table>

*Activity on ONPG taken as 100%, and corresponds to a specific activity of 46 U/mg.
6.4 Discussion

A conserved arrangement of the genes encoding GHF 42, GHF 53, and several ABC transporter genes was observed in several organisms and I proposed that the encoded intracellular GHF 42 galactosidases functioned to degrade oligomers released from galactan by the extracellular GHF 53 enzyme. I first looked for increased β-galactosidase activity in wild-type *B. subtilis* in response to the pre-substrate arabinogalactan type-I (β-1,4-linked galactose backbone) but not the β-galactan pre-substrate found in larch arabinogalactan type-II or gum arabic (β-1,3(/6)-linked galactose backbones). As previously observed (4), easily available mono- and di-saccharides did not result in increased β-galactosidase expression. The greater background observed on polysaccharides was consistent with general *Bacillus* spp. expression of multiple glycosyl hydrolases (including β-galactosidases) under low (accessible) carbon conditions (i.e. pre-sporulation / stationary phase). The highest expression was seen with β-1,4 galactan as found in arabinogalactan type-I, or sources of this polysaccharide such as soy flour, and apple and citrus pectins, with less occurring as a result of the β-1,3/ β-1,6 galactan of gum arabic, and no increase due to β-1,3 galactan of larch arabinogalactan type-II.

Increased expression in the presence of polygalacturonic acid may be occurring because this polysaccharide represents the main backbone of pectin, a complex plant substance which includes branches consisting of β-1,4-galactan.

In order to further explore my hypothesis, two β-galactosidase genes belonging to GHF 42 and a galactanase gene belonging to GHF 53 were cloned from *B. subtilis* into *E. coli* in order to create constructs where these genes were interrupted by a selectable marker, a chloramphenicol resistance cassette. Homologous recombination between these
constructs and the *B. subtilis* genome yielded strains where these genes were interrupted. Cloning of the genes also allowed their effects on *E. coli* to be tested. Neither YesZ nor LacA effectively allowed *E. coli* to grow on lactose, adding further support to the hypothesis that lactose is not the natural substrate for either of these β-galactosidases. However, the combination of GalA and LacA did allow *E. coli* to grow on galactan. Controls expressing GalA with other β-galactosidases (YesZ or LacZ) did not grow, suggesting that degradation of galactooligomers is a property of only certain β-galactosidases.

Not only has the LacA GHF 42 β-galactosidase been demonstrated to have an *in vivo* function, it is encoded in a specific operon arrangement that has not been observed for any of the GHF 42 genes whose encoded β-galactosidases have been biochemically characterized. Therefore, the enzyme heterologously expressed in *E. coli* was purified and examined. In spite of having a possibly different physiological role than other GHF 42 enzymes previously examined, its biochemical properties are not greatly different from other GHF 42 enzymes (compare with Chapter 4, Table 4-1).

In order to confirm the suggested physiological role, wild-type *B. subtilis* was tested to confirm that growth was possible using galactan as the sole carbon source. Then, mutations in these genes were constructed in *B. subtilis* to determine which genes were required for growth on galactan. The genes *lac* and *yes* were separately interrupted and expression patterns by these mutant strains showed that the presence of galactan was causing increased expression of LacA, and not YesZ, only interruption of *lac* (and not *yes*) affected growth on this substrate. The absence of GalA also affected growth on galactan, and decreased the expression of LacA in the presence of this substance. The
failure of the combination of arabinose and galactose to increase $\beta$-galactosidase activity combined with the requirement for GalA for $\beta$-galactosidase expression suggests that galactooligomers may be the natural inducer that releases the repressor LacR.

The mutants were also used to explore the higher $\beta$-galactosidase activities observed for substrates other than galactan and soy flour. Activity on gellan gum, gentiobiose, and xylan, was unexpected as none of these contains galactose. The observed activity on these substrates even when the $\beta$-galactosidase genes had been mutated indicates that accumulated side activities of enzymes other than $\beta$-galactosidases are likely responsible for the observed activity. In contrast, increased expression in response to phytone was expected due to the soybean source of this material (and previous observations with another enzyme (3)), but the lacA::Cm$^R$ mutation curiously failed to affect the activity. The background activity and decreased effect on $\beta$-galactosidase activity by the galA::Cm$^R$ mutant as compared to the lacA::Cm$^R$ mutant further explains the level of growth seen by the lacA::Cm$^R$ and galA::Cm$^R$ mutants on galactan minimal media. Given the poor growth of wild-type *E. coli* on galactan, its enzymes apparently lack this level of background activity.

In all, these results indicate that LacA and GalA are involved in the ability of *B. subtilis* to degrade galactan. Furthermore, *E. coli* strains carrying these two genes also gain the ability to grow on this substrate. It is unknown what specific or nonspecific transport system *E. coli* is using to import galactan oligomers; LacY is a possibility. These are the first results to clearly prove a reasonable *in vivo* function for a GHF 42 $\beta$-galactosidase, and similar gene arrangements in other organisms suggest that this will not be the only example of this function. A function other than lactose hydrolysis for GHF 42
genes has implications for molecular biological implementations of these β-galactosidases and for the search for lactases with unusual properties.

Additional questions are provoked by these results. One of these is whether the preceding genes encoding ABC transporter proteins (yvfK, L, and M) do form an operon with \( \text{lacA} \) and \( \text{galA} \). Evidence for this would support the hypothesized function of YvfK, L and M in the transport of galactooligomers into the cell. This is of particular interest because a previous experiment in \( \text{B. subtilis} \), suggested that the YvfK (CycB) binding protein was involved in transporting cyclodextrin, and had no functional relationship to the GHF 42 enzyme (10).

Another question regards the absence of an ATPase for the ABC transporter system adjacent to the other ABC subunits. A similar gene arrangement (neighboring extracellular and intracellular GHs genes together with genes encoding an ABC binding protein and two permeases) exists in \( \text{B. subtilis} \) for arabinan (a polysaccharide with a backbone consisting of \( \alpha\)-1,5-linked arabinose) degradation that also lacks an ABC transporter ATPase (17, 19). Some other microorganisms with GHF 53 and GHF 42 genes do have ATPase genes encoded nearby that have homology with MalK genes. In yet another similar gene arrangement observed in \( \text{Alicyclobacillus} \) sp., MalK was able to complete the functionality (21). Performing a BLAST (1) search using a GHF 42 associated MalK sequence against the translated genome of \( \text{B. subtilis} \) yields the homologs YurJ (BSU32550) and MsmX (BSU38810). A search in PubMed regarding MsmX reveals that in \( \text{Streptococcus mutans} \) this ATPase is involved in the transport of more than one sugar, and is referred to as the multiple sugar-binding transport ATP-binding protein (18). This description alone sounds promising, but YurJ is the ATPase
which Quentin et al (16) have predicted interacts with the \( \gamma \nu f \) transporter gene sets in \( B. \ subtilis \). Experiments monitoring the effects of a \( \gamma v f K \) mutant on growth on cyclodextrin as a sole carbon source, or microarray data demonstrating regulation of \( \gamma v f K \) and the ATPase during growth with and without galactan might answer these questions. Another question regards the function of \( \text{YesZ} \). Although \( \text{YesZ} \) and \( \text{LacA} \) are both found in \( B. \ subtilis \) and both belong to GHF 42, they only share 43% homology, and clearly have different functions. Expression of \( \text{YesZ} \) was not observed, although nearby genes suggest that rhamnogalacturan type I may be a relevant substrate worth investigating.
6.5 Materials and Methods

**Initial screen for inducing sugars or polysaccharides.** A semiquantitative screen using 48-well plates was used to quickly analyze many carbon sources for increased β-galactosidase activity in wild-type *B. subtilis*. A homogenous solution of wild-type *B. subtilis* was made by washing cells harvested from TSA plates in M9 medium (15) containing no carbon. This was used to inoculate 200 μL wells containing Spizizen’s minimal media (SMM) (9) (per l: 2 g (NH4)2SO4, 14 g K2HPO4, 6 g KH2PO4, 1 g Na3citrate·2H2O, 0.2 g MgSO4·7H2O containing 0.2% of one of the following sugars: glucose, galactose, ribose, xylose, mannose, fucose, fructose, sorbose, rhamnose, sucrose, trehalose, salicin, cellobiose, melibiose, maltose, raffinose, lactose, sorbitol, gentiobiose, N-Acetyl-D-Galactosamine, N-Acetyl-D-glucosamine, D-galacturonic acid, D-glucuronic acid, lyxose, or lactobionic acid; or 0.2% one of the following polysaccharides: cellulose, dextran I, dextran II, dextran sulfate, starch, inulin, glycogen, xanthan gum, laminarin, locust gum, chitin, arabinogalactan type-II (from larch), gum arabic, polygalacturonic acid, apple pectin, citrus pectin, phytone, xylan, gellan gum, soy flour, or galactan. After 24 h of incubation at a 25° angle at 30°C, ONPG (o-nitrophenyl-β-D-galactopyranoside) was added to 2.2 mM, and the plates returned to 30°C for another 24 h before adding 0.5 M Na2CO3 to intensify the color. The upper half of the liquid was removed to fresh 96-well plates to avoid precipitates and read in a Molecular Devices Thermo Max microplate reader at 420 nm at room temperature. The results were interpreted as no effect ($A_{420} < 0.7$), modest increase in activity ($A_{420} > 0.7$), and clear increase in activity ($A_{420} \geq 1.5$).
**Construction of vectors.** *Bacillus subtilis* str. 168 (ATCC 33234) was acquired from the ATCC and grown on TSA medium. Genomic DNA was obtained from cells treated with lysozyme by using the PureGene kit (Gentra Systems Inc. Minneapolis, MN), with the modification of increasing the heating step to 85°C for 10 min to promote cell lysis. This DNA was digested with *Bgl*II or *Pml*I, and the pΔα18 plasmid (23) DNA with *Bam*HI or *Sma*I. Using gel extraction, genomic fragments from 6.6 to 9.5 kb were purified away from smaller and larger fragments. The restricted plasmid was treated with calf-intestinal phosphatase (CIP), and the genomic DNA fragments and vector were ligated (Fast-link, Epicentre, Madison, WI). These constructs were used to transform *E. coli* ER2585F’(*ΔlacZ lacY*’), creating a genomic library. The transformants were screened for 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-Gal) (100 µg/ml) hydrolysis on LB with ampicillin (100 µg/ml) and IPTG (isopropyl-β-D-thiogalactoside) (100 µM) at 37°C.

Transformants from the *Bgl*II library, pYvf and pYes, and *Pml*I library, pYvfK, were subcloned yielding constructs pYvfs, pYess, and pYvfKs. The pYvfs plasmid was used to produce two further subclones pLacA and pGalA, in order to allow the β-galactosidase and the galactanase, respectively, to be expressed separately. On plasmids the genes *lacA*, *galA*, and *yesZ* were not oriented in the same direction as the *lacZ* promoter present on the vector. Plasmids expressing both GalA (using pGalA) and either YesZ (using fragment pYesZsG, Fig 6-1, Panel B) or LacZ (*Escherichia coli*) were also created, yielding pGalAYesZsG and pGalALacZ. The *lacZ* gene was taken from a pET-28a(+) construct (2) and placed into pΔα18 & pGalALacZ. This process involved a modification to create an *Xba*I site into a non-coding 5’ region using QuikChange™ Site-
Directed Mutagenesis kit (Stratagene), and resulted in a product without the N-terminal 6X-His-tag encoded by pET28a(+) (Table 6-9).

<table>
<thead>
<tr>
<th>Name</th>
<th>Primer pair sequences (Forward on top, Reverse on bottom)</th>
<th>Purpose(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAX</td>
<td>GCCACAATGACTGGGAAAG GACACCTACAGCTTTTTC</td>
<td>Detection of Cm(^{R}) insertion in lacA Detection of Cm(^{R}) insertion in galA</td>
</tr>
<tr>
<td>YES</td>
<td>CGACCGTGGAATATGAAC GCTGTCCCTGATAACCATTGG</td>
<td>Detection of Cm(^{R}) insertion in yesZ</td>
</tr>
<tr>
<td>ZX</td>
<td>CAGCAAGCGGTCTAGAGCAGCGCGGCCAGCTAGACGCCGCTG</td>
<td>“Quickchange” addition of XbaI site as part lacZ manipulation</td>
</tr>
<tr>
<td>LACM</td>
<td>GCCAAGAAGAACAAGGAGGAGGACATATGAGCTCAAGAGCTTTG CAAGCTTTGACATATGCTCCTCTTCCTTACAGCC</td>
<td>“Quickchange” to create NdeI site for pETLacM creation</td>
</tr>
</tbody>
</table>

**Studies in E. coli.** The plasmids pΔα18, pLacA, pGalA, pYesZs, pLacZ, pYvfs, pGalAYesZsG, and pGalALacZ and were maintained in *E. coli* ER2585F' on LB with ampicillin (100 µg/ml). The *E. coli* strains carrying these plasmids were plated on solid M9 minimal media containing 0.2% galactose as a sole carbon source, ampicillin (100 µg/ml), and vitamins, and on LB containing X-Gal (100 µg/ml) and ampicillin (100 µg/ml). Cell material from each of the *E. coli* strains was taken from these M9Galactose plates, washed and resuspended in M9 without carbon, and equal aliquots were streaked to M9 plates containing ampicillin (100 µg/ml), vitamins, and: no carbon, 0.2% galactan, 0.2% galactose, or 0.2% lactose. The plates were incubated at 30°C for 5 days, and visually inspected. Growth was considered positive if colony size was greater than or equal to that of pΔα18 on galactose. Negative results had colony sizes less than or equal to the colony size of pΔα18 (without insert) on lactose.

**Creation of B. subtilis mutants.** The pYvfs plasmid was treated with *Srf*I or *Pml*I, and pYess with *Sma*I to create blunt-end restriction sites within the *lac*A, *gal*A, and yesZ genes, respectively. These linearized plasmids were then treated with calf intestinal phosphatase (CIP) and a *Hinc*II or *Hinc*II/*Sma*I fragment of ptrpBGLI-PLK (14) carrying
the Cm$^R$ determinant of pC194 (6) was ligated into these sites, respectively creating the constructs placA::Cm$^R$, pgalA::Cm$^R$, and pyesZ::Cm$^R$.

*E. coli* containing these interrupted-gene constructs were plated on media containing IPTG and X-Gal (concentrations as above). Screening for loss of β-galactosidase activity allowed detection of placA::Cm$^R$ and pyesZ::Cm$^R$, while pgalA::Cm$^R$ was screened examining the plasmids for a correctly modified restriction pattern.

These plasmids carrying interrupted genes were transformed into a *B. subtilis* strain 168, referred to as wild-type in this chapter, previously made prototrophic (for tryptophan) by transformation with *B. subtilis* W168 genomic DNA (Babitzke, personal communication). The cells were made competent via the two-step transformation procedure described in Molecular Biological Methods for *Bacillus* (9), but omitting the freezing step. Recombinants were selected on TSA with chloramphenicol (5 u g ml$^{-1}$).

PCR using Ready-To-Go beads (Amersham Pharmacia, Piscataway, NJ) and primers pairs where at least one primer annealed to a region outside of the cloned insert (Table 6-9) were used to confirm insertion of the Cm$^R$ cassette within the desired regions.

**Comparison of wild-type and mutant expression and growth patterns.** Expression of β-galactosidase activity was also observed on plates containing Schaffer media (20) (per l: 8 g nutrient broth, 1 g KCl, 0.12 g MgSO$_4$, 0.002 g MnCl$_2$·4H$_2$O, 0.5 mM CaCl$_2$, 1 μ M FeSO$_4$, pH 7) containing X-Gal (100 μ g/ml), with or without arabinose (0.05 %), and with 0.2% galactose or galactan (from lupin) (Megazyme, Ireland). The phenotypic and physiological properties of the wild-type and three mutant types were compared on the same media.
The wild-type and mutant *B. subtilis* strains were also plated on M9 minimal agar with 0.2% galactan, glucose, or 0.05% sugar mixture as the sole carbon source. The sugar mixture was based on the data supplied for the galactan by the company, Megazyme, which indicated that 17% of the galactan substrate was not galactose (13). Based on the assumption that 90% of the galactose was present as galactan, the media was made to contain 0.017% galactose (10% of that found in the galactan), 0.006% arabinose, 0.01% rhamnose, 0.004% xylose, 0.01% galacturonic acid, and 0.004% mannose (in place of the unidentified remaining percentage).

Using the mutants as well as the wild-type, the ONPG plate assay was repeated as previously using a reduced number of substrates: glucose, arabinose, xylose, mannose, rhamnose, lactose, starch, cellulose, arabinogalactan type-II (from larch), gentiobiose, PGA, citrus pectin, apple pectin, phytone, xylan, gellan gum, galactan, and soy flour.

**LacA β-galactosidase purification.** Using the QuikChange™ Site-Directed Mutagenesis kit (Stratagene), the start site of lacA was modified to form an *NdeI* site, allowing insertion and ligation of an *NdeI-SacI* fragment carrying lacA into the vector pET28a+.

This construct, pETLacM, transformed into *E. coli* MC1061 (DE3) competent cells, allows inducible expression of the LacA protein modified to possess a 6 x histidine tag at the N-terminus of the protein. This strain was grown in TB (22) with 30 µg/ml of kanamycin at 37°C until an optical density of 0.4 was attained. This culture was then cooled to 18°C, IPTG added (100 µM final concentration), and incubated at 18°C for 15 hours. The cell pellet resulting from centrifugation (6,370 x g, 4°C, 11 minutes) was resuspended (with inclusion of a Complete EDTA-free protease inhibitor cocktail tablet (Roche)) at 3 ml/g of Za buffer. Za buffer is a modification of Z-buffer (15) without β-
mercaptoethanol, consisting of 0.1 M phosphate buffer pH 6.5 with 10 mM KCl and 1 mM MgSO₄. A single pass through a French pressure cell (18,000 lb/in²) was used to disrupt the cells. Centrifugation (30,996 x g, 4°C, 30 min) yielded clarified lysate that was applied to nickel-IDA beads at 4°C. The 6x histidine-tagged β-galactosidase was then purified using a batch elution process using centrifugation (1,239 x g, 10°C, 5 min). The material was triple washed with an initial wash (Zm buffer (Za buffer at pH 7), 300 mM NaCl, 5 mM imidazole), and a secondary wash (Zm buffer, 300 mM NaCl, 20 mM imidazole), prior to elution (with Zm buffer, 300 mM NaCl, 150 mM imidazole). Aliquots of elution supernatant were dialyzed overnight at 4°C in 1 l of Za buffer, followed by a second 2 h dialysis in the same volume of buffer, and used for enzyme characterization.

**LacA β-galactosidase purification.** The specific activity measurements were performed in 1.2 ml of Za buffer with 2.2 mM ONPG, and incubated at 40°C for 15 min before starting the reaction with 10 µl of diluted purified enzyme. Reactions were stopped after 5 min with 0.5 ml of 0.5 M Na₂CO₃ and the release of o-nitrophenol measured at 420 nm. One unit of activity was defined as the release of 1 µmol of ONP per minute. Specific activity was expressed as units per milligram of protein. Protein concentrations were determined using the Bio-Rad (Hercules, California) protein assay dye reagent protocol. All assays were performed at least in triplicate. The symbols on the graphs represent the average of the values obtained and the error bars the range of values.

The following characterizations all used 2.2 mM ONPG, and measured activity as described above. Thermodependence of activity was assayed between 27 and 60°C for 5 minutes. Thermostability assays were performed by incubating aliquots of enzyme at 40,
50, 55, or 60°C. Aliquots were removed at various times and assayed at 40°C for 5 minutes. The optimal pH values were determined by assaying in citric acid buffers (0.1 M) (pH 5 to 6), and Zn buffer (0.1 M) (pH 6 to 8). Activity in sodium phosphate buffer (0.1 M) was also tested at pH 7. Activation or inhibition by metal ions was examined in a solution of 0.1 M MOPS (pH 7) without additions, or containing either 1 mM MgCl$_2$, MnCl$_2$, CaCl$_2$, or CuCl$_2$, or 10 mM NaCl or KCl.

Substrate preference was examined using chromogenic ONP and PNP substrates at 2.2 mM at 40°C after 5 minutes, with one activity unit defined as the release of 1 µmol of o-nitrophenol or p-nitrophenol per minute. The substrates (Sigma) tested were ONPG, PNPG ($p$-nitrophenyl-$\beta$-D-galactopyranoside), $p$-nitrophenyl-\(\alpha\)-D-galactopyranoside, $p$-nitrophenyl-\(\alpha\)-L-arabinopyranoside (PNAP), $p$-nitrophenyl-$\alpha$-L-arabinofuranoside (PNAPf), $o$-nitrophenyl-$\beta$-D-glucopyranoside, $p$-nitrophenyl-$\beta$-D-glucopyranoside, $p$-nitrophenyl-\(\alpha\)-D-glucopyranoside, $o$-nitrophenyl-$\beta$-D-fucopyranoside, $p$-nitrophenyl-$\beta$-D-mannopyranoside and $o$-nitrophenyl-$\beta$-D-xylopyranoside. Although testing with $p$-nitrophenyl-$\beta$-D-galactotrioside or galactotetraside would be ideal, these compounds are not yet commercially available as the need for them has only recently been discerned.
6.6 References


Chapter 7

Summary
My dissertation research employed microbiology, phylogenetics, microbial physiology, bioinformatics, microbial genetics, biochemistry, and environmental microbiology to examine different aspects of bacterial glycoside hydrolase functions in the environment. By combining the results from these different facets, I have identified the potential functions for two glycoside hydrolases in separate families. Both of these glycoside hydrolases, BglY and LacA, have β-galactosidase activity, but they are very different enzymes and are unusual within the context of other currently characterized enzymes.

The BglY enzyme represents an unusual example within a large, frequently studied group of enzymes (GHF 3), whereas the LacA enzyme appears to be a typical enzyme of a smaller and less studied group (GHF 42). The methodology for determining the functions of enzymes in these families differs. My characterization of BglY was a matter of identifying a function from a pool of known possible functions; but discerning a function for LacA required me to use a genomic analysis method. My use of this analysis method also indicated probable functions for additional genes that differed from the functions predicted by the genome annotations (e.g. galactooligomer transport instead of maltodextrin transport).

Amidst these differences these enzymes also have common aspects. Both BglY and LacA are exceptions to the traditional assumptions about the function of X-Gal hydrolyzing enzymes because neither function in lactose hydrolysis. The deep-seated nature of this assumption regarding β-galactosidases and lactose hydrolysis is revealed by the name LacA was historically given even though experimental evidence for a lactose-related function was lacking. Both of BglY and LacA act on substrates originating in plants rather than animals, and in addition to being unusual for enzymes with β-galactosidase activity, they also do not have functions traditionally associated with other members in their respective GHFs. The GHF 3 enzyme group, to which BglY belongs, is typically associated with degradation of the glucan-polysaccharide cellulose, but BglY has a low $K_m$ with aryl-glucoside substrates suggesting it may have a functional role involving signaling rather than catabolism. Likewise, the LacA enzyme does not function in the assumed role of the GHF 42 group, lactose-hydrolysis, but instead, as shown in
Chapter 6, the LacA enzyme fits the general role traditionally expected for BglY – contributing to catabolism of a β-1,4 linked polysaccharide. Just as bacteria use β-glucosidases, β-mannosidases and β-xylosidases to complete degradation of the oligosaccharides (produced by the activity of cellulases, mannases and xylanases) from cellulose, mannan, and xylan into glucose, mannose, and xylose, β-galactosidases contribute to the process of completely hydrolyzing galactooligomers (produced by the activity of galactanases) from galactan into galactose.

BglY and LacA would each have been disregarded by researchers expecting these enzymes to contribute to cellulose or lactose-hydrolysis, respectively. However, my clarification of their functional activities gives them different potential roles in biotechnology. Aryl-glucosidases have a variety of applications expounded in Chapter 3, such as improving the smell or taste of foods, or enhancing the healthfulness of foods by detoxifying harmful compounds or increasing the bioavailability of beneficial substances. The GHF 42 galactooligomerases may be useful for the modification of pectic substances, which could be useful for modifying the clarity, viscosity, or gelling characteristics of food stuffs.

Prior to my work represented in this thesis, there were no in vivo demonstrations of a function for a GHF 42 enzyme. The association of the functions of GHF 42 and GHF 53 may previously have gone unnoticed because of tight regulation, as observed in B. subtilis. This would have prevented expression of β-galactosidase activity under typical laboratory conditions, making B. subtilis and organisms like it appear to be devoid of β-galactosidases genes. The GHF 42 genes encoding β-galactosidase activity are not expressed when the B. subtilis is grown on TSA, even with IPTG. B. subtilis does not grow on lactose, and galactan is not commonly tested as a carbon source. Thus, most of the GHF 42 genes that have been cloned have either been constitutively expressed (and probably do not have the GHF 53 associated arrangement), or have been cloned because of interest created by the activity of another (more easily expressed) β-galactosidase, such as those from GHF 2. Also, because neither the relationship between GHF 42 enzymes nor the association of some of them with GHF 53 is not tightly phylogenetically
constrained (as shown by Chapters 4 and 5), it is unlikely that small-scale comparisons that currently accompany traditional cloning projects would have uncovered the relationship and predicted the galactooligomerase function. Thus, a full-scale analysis of all the available data combined with a wider viewpoint of glycoside hydrolase functions outside of the β-galactosidases was essential for discerning a function for GHF 42. Further research in the field of plant glycomics combined with continued examination of the wealth of data being revealed by genome sequencing projects will doubtless yield more information regarding “alternative” roles for glycoside hydrolases in the future.
Curriculum Vitae

Stephanie Shipkowski
770 Toftrees Ave. Apt 330, State College, PA 16803
(814) 237-6972 sshipkowski@gmail.com

Education:
1999 – 2006 Ph.D. Candidate at The Pennsylvania State University, Department of Biochemistry, Microbiology, and Molecular Biology (BMMB)
Advisor: Jean E Brenchley, Ph.D.
Participant in Biogeochemical Research Initiative for Education program

1995 – 1999 B.A. Biochemistry, and Biology, Douglass College, Rutgers University, Highest Honors, Scholars Program, Douglass Honors Independent Research

Teaching Experience:
2001 – 2004 Instructor for Teaching Assistants for Biochemistry, Microbiology, and Molecular Biology Department at The Pennsylvania State University (PSU)
2003 Instructor for Microbiology / Molecular Biology section of Introduction to Biogeochemical Analysis 597C
2000 – 2001 Teaching Assistant for Biochemistry, Microbiology, and Molecular Biology Department at PSU
1995 – 1999 Led and assisted in various science themed educational programs for elementary and middle school students through Project Outreach and Girl Scouts

Publications and Posters (S. Shipkowski and J.E. Brenchley):
2006 (Publication in preparation) Bioinformatic evidence that some glycoside hydrolase family 42 ß-galactosidases function in hydrolysis of arabinogalactan type-I oligomers, with supporting studies using Bacillus subtilis.


2003 Poster: Blue Plate Specials: Cold-active beta- galactosidases from Bacillaceae spp. 103rd General Meeting. American Society for Microbiology, Washington DC.

2002 – 2005 Posters at Environmental Chemistry Student Symposia (ECSS), PSU

Honors:
2005, 2003 1st place Graduate Poster in Biochemistry and Microbiology Session, ECSS
2003 Recipient of an ASM 103rd General Meeting student travel grant
2000 Recipient of Biogeochemical Research Initiative for Education (BRIE) funding
1999 – 2001 Braddock Graduate Fellowship, awarded by Eberly College of Science
1999 – 2001 Roberts Graduate Fellowship, awarded by Eberly College of Science
1999 Life Sciences Consortium Scholar, awarded by Life Sciences Consortium of PSU
1999 Ruth E. Salny Fellowship, awarded by the Douglass Associate Alumnae
1997 Golden Key National Honor Society
1995 National Merit Scholarship
1995 Senior Girl Scout Gold Award