ABSTRACT

Title of Document: ROLE OF SELECTED EXTRACELLULAR FUNCTION SIGMA FACTORS IN CARBOHYDRASE GENE EXPRESSION IN *SACCHAROPHAGUS DEGRADANS* 2-40

Corinn Ann Sinnott
Master of Science 2011

Directed By: Professor
Dr. Steven Hutcheson
Cell Biology and Molecular Genetics

This project explores the role of selected ECF sigma factors in regulating carbohydrase gene expression in *Saccharophagus degradans* 2-40. ECF sigma factors most likely to be involved in carbohydrase regulation were identified by scoring for the presence of high sequence similarity in cellulolytic but not in non-cellulolytic bacterial species. *C. japonicus*, *T. turnerae*, and *H. chejuensis* were determined to be the cellulolytic species most appropriate for evaluating transcriptional regulators for a role in carbohydrase regulation. Two sigma factors were identified in this survey. These sigma factor genes were cloned into a plasmid vector and expressed in *S. degradans.* Cellulase activity was detected for cell lysates of glucose-grown cultures of a strain expressing the sigma factor corresponding to gene 2407. It was concluded that sigma factor 2407 is involved in the regulation of endoglucanases in the *S. degradans* cellulolytic system.

ROLE OF SELECTED EXTRACELLULAR FUNCTION SIGMA FACTORS IN CARBOHYDRASE GENE EXPRESSION IN *SACCHAROPHAGUS DEGRADANS* 2-40

By

Corinn Ann Sinnott

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2011

Advisory Committee:

Professor Steven Hutcheson
Associate Professor Kevin McIver
Associate Professor Richard Stewart

Acknowledgements

Dr. Steven Hutcheson, Principal Investigator and Committee Chair
Dr. Kevin McIver, Committee Member
Dr. Richard Stewart, Committee Member

Dr. Haitao Zhang, Zymetis, Inc.
Dr. Young Hwan Moon
Dr. Brian Watson
Elizabeth Santos
Sabina Kaczanowska
Brandon Meyer
Alma Polo-Barrios

Table of Contents

List of Tables

List of Figures

List of Abbreviations

DNA	Deoxyribonucleic acid
GH5	Glycoside hydrolase family 5
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RNA	Ribonucleic acid
ECF	Extracellular function
BLAST	Basic Local Alignment Search Tool
PCR	Polymerase chain reaction
Tris-HCl	Tris (hydroxymethyl) aminomethane hydrochloride
M	Molar
C	Celsius
LB	Luria-Bertani medium
BCA	Bicinchoninic acid (protein assay)
Cel	Cellulase
DNS	3,5-dinitrosalicylic acid
IPTG	Isopropyl-beta-D-thiogalactopyranoside
Kb	Kilobases
kDa	Kilodaltons
PAGE	Polyacrylamide gel electrophoresis
CMC	Carboxymethyl cellulose

Chapter 1: Introduction to *Saccharophagus degradans* 2-40

Background on *Saccharophagus degradans* 2-40

Saccharophagus degradans strain 2-40 is a marine bacterium that was isolated from decaying marsh grass (*Spartina alterniflora*) in the Chesapeake Bay near Mathews, VA (Andrykovitch and Marx). This bacterium can be further characterized as a Gram-negative, motile, aerobic, gamma proteobacterium that requires salt in order to grow (Ensor, Stosz and Weiner , Ekborg et al.). *Saccharophagus degradans* 2-40 has the ability to degrade a variety of complex polysaccharides, including cellulose, xylan, pectin, agar, agarose, chitin, pullulan, starch, alginate, tequila bagasse, and laminarin, into their simple component sugars (Alva Munoz and Riley 2008, Ekborg et al. 2005). This degradation is accomplished through complex systems of enzymes that *S. degradans* produces when grown on one of these substrates as its major carbon source (Hutcheson, Zhang and Suvorov 2011).

The genome of *S. degradans* 2-40 has been fully sequenced and shows strong similarity to that of *Cellvibrio japonicus* as well as *Pseudomonas fluorescens*, from which *S. degradans* appears to have evolved (González and Weiner 2000, Weiner et al. 2008, DeBoy et al. 2008). This evolutionary relationship is indicated by the large number of genes shared between *S. degradans* and *P. fluorescens.* The *S. degradans* 2-40 genome contains over 180 homologs of carbohydrate degrading enzymes, which could contribute to its ability to degrade a variety of complex polysaccharides (Weiner et al. 2008).

The average distance between genes of *S. degradans* is 166 bp, while that of *E. coli* K-12 is 118 bp (Weiner et al. 2008, Blattner et al. 1997). The organizational structure of the genome of this organism is unusual in that many of the genes are not arranged into obvious operons, such as can be seen in other bacteria. In *E. coli*, most genes that belong to the same operon overlap and are therefore less than 0bp apart (Salgado et al. 2000). In *S. degradans,* genes that are of similar function and appear to be co-transcribed are often spaced further apart. For example, the *gcvH* and *gcvT* co-transcribed genes of *E coli* are 24bp apart, whereas in *S. degradans* these genes are separated by 211bp (Gama-Castro et al. 2010). This observation does not imply that operons do not exist within this system, but rather that the genes in an operon may be spaced farther apart than is typical of bacterial genomes. An operon analysis was conducted to assist in identifying some possible operons in this bacterium. The number of nucleotides between genes on the same strand of DNA and their predicted functions were used to assess whether genes could potentially belong to the same operon. Certain gaps between gene annotations could also indicate the presence of regulatory elements. Uncovering the unusual operon structure of this bacterium would lead to an increased understanding of how gene expression is regulated within this system and potentially provide a new model for genome organization.

The large number of apparent carbohydrase genes found in this bacterium is unusual, even amongst cellulolytic species. It is possible that *S. degradans* 2-40 may have used horizontal gene transfer mechanisms and duplication/expansion events to acquire these genes. The acquisition of many of these genes by horizontal gene transfer

events is supported by the observation that the codon usage and the third position nucleotide of the codons used in the carbohydrase genes often differ from those of "housekeeping" genes in this bacterium (Weiner et al. 2008, Hutcheson et al. 2011). Gene duplication events are evidenced by the presence of a few loose gene clusters of carbohydrases (Hutcheson et al. 2011, Weiner et al. 2008). Several enzyme systems have been identified in the *S. degradans* 2-40 genome that are comprised of multiple annotated carbohydrase genes (Ekborg et al. 2006, Howard et al. 2003, Taylor et al. 2006, Hutcheson et al. 2011). These enzyme systems include genes that are involved in the breakdown of chitin, agar, alginate, cellulose, xylan, pectin, mannan, laminarin, and α-glucans. Together these enzyme systems provide *S. degradans* with the ability to break down various complex polysaccharides, including whole plant material.

Characteristics of Plant Material

The plant cell wall is composed of various carbohydrate polymers including cellulose, hemicellulose, and pectin (Carpita and Gibeaut 1993, Carpita 1996). Cellulose is a polymer of β-1,4 linked D-glucose molecules that represents one of the primary structural and protective components of the plant cell wall (Warren 1996). Cellulose polymers are assembled into microfibrils that contain both amorphous and crystalline regions. Amorphous regions are more susceptible to enzymatic degradation, whereas the ordered structure of crystalline cellulose makes it more difficult for enzymes to degrade. Amorphous and crystalline cellulose regions form depending on the degree of order of the hydrogen bonds between individual cellulose chains. The mixture of amorphous and crystalline regions found in most cellulose in nature results in an

irregular structure that necessitates the action of several enzymes to degrade. In addition to the challenges of varied substrate crystallinity, carbohydrases must secure access to the substrate at their active site in order to carry out enzyme function and successfully degrade plant material. Substrate-binding modules can help to position enzymes such that their active site is in the proper orientation in order to achieve degradation.

The hemicellulose, which forms a complex with the cellulose, contributes to the stability of the cell wall structure. Hemicellulose is comprised of xylans, arabinans and mannans that together form a heterogeneous branched polymer (van Wyk 2001). Pectin encompasses another significant component of the plant cell wall. It too has a varied structure, containing polygalacturonic acid and rhamnogalacturonan, which may in turn have galactans, arabinans, and arabinogalactans attached at certain residues (Carpita 1996). The varied composition of both hemicellulose and pectin can necessitate a wide variety of enzymatic functions in order to achieve full degradation of these substrates. As for cellulose, substrate-binding modules in hemicellulose and pectinase systems can facilitate placement of enzymes such that the active site is able to come into contact with and carry out degradation of the substrate

Cellulolytic System

S. degradans 2-40 possesses an extensive system of enzymes that are involved in the degradation of cellulosic material. Many of the enzymes of the cellulolytic system have been well-characterized and this system is now believed to be comprised of

twelve GH5 domain containing β-1,4-endoglucanases (Cel5A, Cel5B, Cel5C, Cel5D, Cel5E, Cel5F, Cel5G, Cel5H, Cel5I, Cel5J, Gly5L, and Gly5M), at least three of which act processively to deliver cellobiose as their final product (Taylor et al. 2006, Watson et al. 2009, Hutcheson et al. 2011). It is believed that processive endoglucanases may compensate for the apparent lack of cellobiohydrolases in this system (Watson et al. 2009, Hutcheson et al. 2011). The cellulolytic system also contains one GH6 containing endoglucanase (Cel6A), two GH9 domain containing β-1,4-endoglucanases (Cel9A and Cel9B), five β-glucosidases (Bgl1A, Bgl1B, Bgl3C, Ced3A, and Ced3B), and one cellobiose phosphorylase (Cep94A) (Zhang et al. 2011, Hutcheson et al. 2011). Together these cellulases and accessory enzymes are responsible for the breakdown of cellulose into simpler sugars for metabolism by the bacterium.

Phylogenetic analysis of the *S. degradans* 2-40 GH5 domains has revealed two distinct clades: one containing Cel5G, Cel5H, Cel5J, and the carboxy terminal GH5 domain of Cel5A and the other clade containing Cel5B, Cel5C, Cel5D, Cel5E, Cel5F, and the amino terminal GH5 domain of Cel5A (Watson et al. 2009). Functional studies of these enzymes indicated that Cel5B, Cel5D, Cel5E, and Cel5F are classical endoglucanases, which cleave a substrate randomly to produce shorter chains. In contrast, Cel5G, Cel5H, and Cel5J act processively to degrade the substrate directly into dimer molecules. Phylogenetic analysis of the Cel5A GH5 domains indicated that the amino terminal GH5 is closely grouped with the classical endoglucanases, while the carboxy terminal GH5 is grouped with the processive endoglucanases

(Watson et al. 2009). Enzymes containing more than one catalytic domain are rare, making Cel5A and its mechanism of function of interest. It was proposed that these two catalytic domains could serve different functions in the degradation of cellulose.

Xylanolytic System

The ability of *S. degradans* 2-40 to successfully degrade xylan can be attributed to its array of various xylanolytic enzymes. Based on the annotated genome, this bacterium is predicted to contain seven endoxylanases, four xylosidases, one alpha-galactosidase, two alpha-glucuronidases, one beta-glucuronidase, eight arabinofuranosidases, two arabitan endo 1,5 alpha arabinosidases, three arabinogalactan endo 1,4 beta galactosidases, four acetoxylan esterases, one alpha glucosidase, and four carboxyl esterases (Hutcheson et al. 2011). The function of endoxylanases is to cleave off small chains of xylooligosaccharides from the polymer (Jain et al. 2009, Consortium 2011). Xylosidases function by releasing xylose monomers from the non-reducing end of xylan polymers. Galactosidases, glucoronidases, arabinofuranosidases, and glucosidases work to release their respective monomers from the substrate. The function of arabitan endo 1,5 alpha arabinosidases is to cleave short chains of arabinans from the polymer. The various activities of these enzymes comprise a system that allows *S. degradans* 2-40 to effectively metabolize xylan and grow with it as a sole carbon source.

Pectinolytic System

S. degradans 2-40 has also been found to contain an extensive list of genes that are predicted to have a function in the degradation of pectin (Hutcheson et al. 2011). The annotated genome indicates the presence of twelve annotated pectate lyases, one rhamnogalacturonan lyase, three rhamnogalacturonan hydrolases, six beta-galactosidases, five pectinesterases, and one exopectate lyase. Pectate lyases are extracellular enzymes that cleave pectate resulting in the production of oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups on their non-reducing ends (Jain et al. 2009, Consortium 2011). Rhamnogalacturonan lyases break the alpha-1,4 glycosidic bonds between L-rhamnose and D-galacturonic acids in the chain of rhamnogalacturonan. The function of rhamnogalacturonan hydrolases is to cleave the alpha 1--2 linkages between rhamnosyl and galacturonosyl residues within the substrate. Beta-galactosidases function by hydrolyzing beta-galactosides into their monomeric constituents. Pectinesterases act to break down pectin into pectate and methanol. Finally, exopectate lyases remove 4-(4-deoxy-alpha-D-galact-4-enuronosyl)-D-galacturonate from the reducing end of pectate. Together, these enzymes would enable the bacterium to degrade pectin efficiently for use as a carbon source.

Chapter 2: Regulation of Carbohydrase Genes

Evidence for Regulation

Regulatory mechanisms exist for controlling the production of the complex carbohydrate degrading enzymes of *S. degradans* 2-40. This bacterium is capable of expressing complex carbohydrate degrading enzymes when grown with one of these substrates as its major carbon source (Ensor et al. 1999, Taylor et al. 2006). However, when glucose is present in the culture media, no complex carbohydrate degrading enzymes are produced by the organism, most likely to due to catabolite repression mechanisms contained in the genome of this bacterium (Weiner et al. 2008, Shin et al. 2010). This observation suggests that the production of these enzymes is being regulated by some mechanism in response to the carbon source available in the environment. When cellulose, an insoluble substrate, is used as the sole carbon source, cellulases are expressed.

Further evidence for regulation of these complex carbohydrate degrading enzymes has been obtained by measuring transcript levels of the cellulolytic system genes using qRT-PCR during growth on various substrates (Zhang and Hutcheson 2011). A large increase in the transcription of all except one of the genes for these degradative enzymes was observed during growth on media containing cellulose as its carbon source compared to when grown on glucose-containing media only. Three distinct patterns of gene expression could be observed within the cellulase system. The first pattern includes genes that are up regulated 5 -10 fold within four hours after the

introduction of cellulose growth media and then this level of expression is maintained for at least another 22 hours. The second pattern consists of the genes whose expression is highest between 4 and 10 hours after the introduction of cellulose and then decreases. The third pattern of gene expression includes the genes that reach their highest level of expression after 24 hours of growth on cellulose-containing media. The findings from the qRT-PCR analysis provide strong evidence that many of the cellulolytic system genes are regulated at the transcriptional level. Because different patterns of mRNA expression can be observed for different enzymes (Zhang and Hutcheson 2011), as many as three regulatory systems could be involved.

Possible Regulatory Mechanisms

The mechanism by which the *S. degradans* 2-40 genes for carbohydrate degrading enzymes are regulated is not clear from the sequence analysis alone. Many of the genes for carbohydrate degrading enzymes do not appear to be arranged in obvious operons with carbohydrate degrading enzymes nor with annotated regulatory genes. Many gene regulatory mechanisms have been identified in bacteria and could function in regulating carbohydrase gene expression. These regulators include such mechanisms as LuxR/LuxI, OmpR, LacI/GalR, AraC/XylS, and LysR family regulatory factors, along with catabolite repression control mechanisms and sigma factors. Similar sequences of all of these regulatory mechanisms, including sigma factors, have been identified in *S. degradans* 2-40 and it is possible that one or more of these regulatory mechanisms contributes to carbohydrase gene regulation in this bacterium.

First observed as regulators of expression of luminescence genes in *Vibrio fischeri*, LuxR/LuxI type regulators are now known to play a significant role in the regulation of many genes of diverse functions (Fuqua, Winans and Greenberg 1994, Fuqua, Winans and Greenberg 1996, Pompeani et al. 2008). Expression of the genes of the *lux* operon is activated by high levels of the autoinducer VAI-1, which is produced by the bacterium according to the concentration of cells in the environment. At high cell concentration, the amount of the VAI-1 is increased and expression of the *lux* operon allows the bacterium to luminesce. The protein LuxI is responsible for the production of VAI-1 and LuxR acts as a receptor for VAI-1. Once LuxR has bound to VAI-1 it is then able to recognize and bind to the *lux* promoter and activate transcription of the genes. LuR/LuxI type regulators have been discovered in many additional species of bacteria and have been found to function in the regulation of such genes as those for plasmid conjugation, virulence factors, and extracellular enzyme production. The latter function suggests that there is a possibility that they could be involved in the regulation of secreted carbohydrase genes. Sequences highly similar to LuxR have been found in *S. degradans* 2-40, suggesting that LuxR-type regulators may function to regulate gene expression in this bacterium (unpublished data).

Two-component systems are used by bacteria to detect changes in the environment through protein kinase-dependent signal transductions pathways (Gao and Stock 2009). Two component systems consist of a sensor molecule located in the inner membrane, which detects the environmental signal, and a response regulator that

affects an appropriate response. For example, OmpR is a transcriptional regulator that controls the expression of genes encoding porins in the outer membrane (Itou and Tanaka 2001). First, a sensor protein in the inner membrane detects the signal, in this case the osmolarity of the surrounding environment, and passes this signal on to OmpR by exercising kinase and phosphatase activity. OmpR is then able to activate the transcription of membrane porin genes through its binding to the gene promoter. Sequences of high similarity to OmpR family regulators have been found in many species of bacteria, including *S. degradans* 2-40 and their function in gene regulation in response to environmental signals makes them a potential candidate for regulation of carbohydrase genes (unpublished data).

LacI was one of the first identified gene regulators in bacteria and the LacI/GalR family of regulators to which it belongs is now known to play a critical role in controlling the transcription of a wide variety of bacterial genes (Lewis 2005, Swint-Kruse and Matthews 2009). LacI/GalR family regulators can function in both inducible and repressible systems. The characteristic example of such a regulator in an inducible system is LacI itself, which represses expression of the *lac* operon in the absence of lactose by binding to the gene operator and preventing transcription of the genes. When lactose becomes available, it acts as an inducer, binding LacI at an allosteric site which causes it to disassociate from the operator allowing transcription of the genes to occur (Wilson et al. 2007). LacI/GalR family regulators can also function in repressible systems, as exemplified by the purine repressor protein PurR (Swint-Kruse and Matthews 2009). The genes that PurR regulates are involved in

purine and pyrimidine metabolic pathways and are normally expressed in the bacterium. Only when a co-repressor molecule or co-factor protein is present and bound to PurR does it then bind the gene operator and block transcription of the genes. The widespread presence of sequences of high similarity to LacI in diverse species, including *S. degradans* 2-40, and their ability to regulate genes of various functions makes LacI type regulators another possible regulatory system for the carbohydrase genes of *S. degradans*.

AraC/XylS family regulators represent another common regulatory system in prokaryotes (Gallegos et al. 1997, Schleif 2010). AraC family regulators act as positive regulators to activate transcription of their target genes. AraC was the first described member of this family of regulators and it functions as a regulator of the L-arabinose operon in *E. coli*, which contains genes that function in the metabolism of arabinose. The dimerization of AraC plays an important role in its regulatory function. In the absence of arabinose, AraC is positioned with one monomer bound to one inducer site and the other bound to an operator site. This conformation results in a bending of the DNA that prevents transcription of the operon. When arabinose is present, it binds to AraC and alters the binding affinity of the protein such that the monomers shift location so that each is now bound to one of two inducer sites. This conformation releases the bending of the DNA, allowing transcription of the operon to proceed. AraC also regulates its own gene expression by binding to another operator site and blocking transcription of its own gene. AraC family regulators have since been characterized in many bacterial species, including *Pseudomonas* species,

and their presence in *S. degradans* 2-40 suggests a possible function in carbohydrase gene regulation (Gallegos et al. 1997, Potvin, Sanschagrin and Levesque 2008).

LysR family regulators represent a global regulatory system that is responsible for the regulation of genes of diverse functions, including genes for metabolism and quorum sensing (Maddocks and Oyston 2008, Schell 1993). LysR family regulators bind DNA and can act as either activators or repressors to regulate expression of genes. Co-inducer molecules have been found to play a significant role in this process and evidence indicates that these co-inducers are often some product of the pathway in which the LysR-regulated genes function, thereby establishing a feedback loop for expression of the genes. The ability of LysR family regulators to control the expression of genes of many different functions, to be acquired through horizontal gene transfer, their occasional location in areas of the genome other than that of the genes that they regulate, their established presence in *Pseudomonas* species, and the presence of similar sequences in *S. degradans* suggests a possible role for LysR family regulators in controlling gene expression in this bacterium (Schell 1993, Maddocks and Oyston 2008, Potvin et al. 2008).

Carbon catabolite repression is a term used to describe the phenomenon in which the expression of genes encoding proteins that are involved in the metabolism of a particular carbon source are repressed when a more preferable carbon source is present in the environment (Deutscher 2008, Görke and Stülke 2008). Once the preferred carbon source is exhausted, the bacteria will detect this deficiency and once

again begin to produce the enzymes necessary for degradation of the secondary carbon source. One of the most well-known examples of carbon catabolite repression is the glucose-lactose diauxic curve of *E. coli* in which expression of the *lac* operon, which contains genes for the metabolism of lactose, is repressed when the preferred carbon source, glucose, is present in the media along with lactose. Only when the glucose source is exhausted will the repression of the *lac* operon be lifted and the bacterium will be able to produce the necessary enzymes for the metabolism of lactose. Carbon catabolite repression represents a form of bacterial gene regulation that involves both global and operon specific mechanisms. This broad range of mechanism and the role of carbon catabolite repression in carbohydrase gene regulation build a compelling case for their possible function in the regulation of *S. degradans* 2-40 carbohydrase expression.

One way that carbon catabolite repression functions globally is through the transcriptional regulator known as cyclic AMP receptor protein (CRP) which interacts with cyclic AMP (cAMP) (Görke and Stülke 2008). When glucose levels are decreased, cAMP levels increase and bind CRP, which then acts as a positive inducer to activate transcription of genes involved in the metabolism of a particular carbon source. The other manner in which global catabolite repression can occur is through the mechanism of catabolite repression control (Crc). Crc is an RNA-binding protein that recognizes and binds to the 5' ends of mRNAs that encode regulatory proteins. This binding prevents the translation of the regulator and results in the regulation of the regulator at a post-transcriptional level. Crc has been well-studied in

Pseudomonas species and the genomic similarity between *Pseudomonas* and *S. degradans* 2-40 suggests that Crc may function to regulate genes in *S. degradans* as well (Görke and Stülke 2008, Potvin et al. 2008). Sequences of high similarity to both CRP and CRC have been found in the genome of *S. degradans* 2-40 and both may play a role in the regulation of the carbohydrase genes of this bacterium (unpublished data).

The regulation of bacterial genes involved in the degradation of cellulose, xylan, and pectin has not yet been well-characterized in many bacteria, including *S. degradans* 2-40 and its closely related species. In some bacterial species there is evidence that regulation of carbohydrase genes occurs, but the mechanisms governing this regulation have not been determined. For example, in *C. cellulovorans* and a *Streptomyces* sp., carbohydrase expression is high when the bacterium is grown on media containing the respective substrate and decreased when the substrate is absent, suggesting that regulatory mechanisms act to control carbohydrase gene expression (Han et al. 2004, Godden et al. 1989). However, the mechanism by which this regulation occurs has not been established.

Mechanisms of carbohydrase gene regulation have been determined in a few bacterial species. For example, it has been demonstrated that catabolite repression regulates cellulase expression in *C.cellulolyticum* and *C. flavigena* (Abdou et al. 2008, Herrera-Herrera et al. 2009). LysR homologs have been shown to repress expression of pectinolytic and cellulolytic enzymes in *Erwinia* species (Harris et al. 1998,

Hugouvieux-Cotte-Pattat et al. 1996). The sigma factor FliA has been shown to repress expression of pectate lyases in *D. dadantii* (Jahn, Willis and Charkowski 2008). It has also been shown that alternative sigma factors control expression of cellulosomal genes in *C. thermocellum* (Nataf et al. 2010). In *T. fusca* the LacI/GalR homolog CelR has been demonstrated to regulate its cellulase genes (Spiridonov and Wilson 1999). LuxR/LuxI homologs likely function in the regulation of pectinase and cellulase genes in *E. carotovora* and *P. aeruginosa* (Jones et al. 1993). It is clear from such findings that many mechanisms exist for regulating the expression of carbohydrases in various bacteria. Further research must be done to elucidate which mechanisms are the primary regulators of carbohydrase gene expression in complex-carbohydrate degrading bacteria, such as *S. degradans*.

Sigma Factor Structure and Function

Sigma factors represent one way by which bacterial genes can be regulated. Sigma factors function by recognizing and binding to RNA polymerase to form holoenzyme and recruiting holoenzyme to a specific gene promoter so that transcription can occur (Potvin et al. 2008). The structure of a sigma factor is made up of four distinct regions. The first region is involved in DNA binding and is the least conserved among sigma factors. The second region functions in the binding of RNA polymerase, recognizing the -10 region of the promoter, and in DNA melting. It is one of the most conserved regions amongst sigma factors. The third region recognizes RNA polymerase and is less conserved. The fourth region features a helix-turn-helix motif and recognizes the -35 region of the promoter. Similar to the second region, the

fourth region of the sigma factor structure is also highly conserved amongst bacterial species.

Types of Sigma Factors

Sigma factors can be divided into two main categories: sigma70 family and sigma54 family factors (Potvin et al. 2008). Sigma54 family factors include RpoN, which is involved in nitrogen metabolism, and evidence exists that they may have additional functions. Sigma 70 family factors can be further classified into four categories (Potvin et al. 2008, Paget and Helmann 2003). The first is the primary sigma factor, RpoD. Often referred to as the "housekeeping" sigma factor, RpoD is responsible for regulating the transcription of the majority of the genes in the bacterial cell, particularly those essential for survival. The second category of sigma70 factors are those closely related to RpoD, such as RpoS (Paget and Helmann 2003). These closely related sigma factors may function in the regulation of some genes, but are not usually necessary for survival of the bacterium. The third category of sigma70 factors are those that are more distantly related to RpoD, such as RpoH and FliA. This category includes sigma factors that effect heat shock response (RpoH) and morphological processes such as the formation of flagella (FliA) or endospores. Finally, the fourth category of sigma factors are known as extracellular function sigma factors (ECFs) (Potvin et al. 2008). The ECFs are the largest category of sigma factors and are highly divergent. The function of ECFs is to regulate gene transcription in response to environmental signals sensed by the bacterial cell.

Common examples of ECFs are RpoE (heat stress response ECF) and FecI (iron transport ECF).

Extracellular Function Sigma Factor Signalling

Extracellular function sigma factor signalling in gram negative bacteria begins with integral membrane proteins that detect particular substrates in the environment and then relay that signal to the cell machinery. Sigma factors themselves can be regulated by anti-sigma factor mechanisms. In Gram negative bacteria, expression of ECFs can be controlled by direct transcriptional regulation, Sigma E, or FecARI type signaling pathways.

Much sigma factor regulation occurs through direct transcriptional regulation. In this type of regulation, the transcription of a sigma factor is regulated by the presence of an inducer molecule that either negates the effect of a repressor protein or allows for the action of an activator protein (Potvin et al. 2008). In either case, the presence of the inducer molecule would enable the transcription of the gene encoding the sigma factor to occur, while in the absence of the inducer little or no transcription would occur. Inducers are known to play a critical role in facilitating the transcription of genes in many cellulolytic species. If a sigma factor is regulated at the transcriptional level then the amount of transcription of the sigma factor dependent gene will be proportional to the amount of sigma factor present.

The Sigma E signalling pathway is used by bacterial cells to detect heat stress and effect a response through sigma E-regulated genes (Brooks and Buchanan 2008). In

this pathway an integral inner membrane protein (DegS) is responsible for sensing a signal from the environment, in this case an unfolded outer membrane protein. DegS responds to this signal by becoming activated and using its protease domain to cleave off the anti-sigma factor RseB and part of a second integral membrane protein RseA. Under normal conditions, RseB acts as an anti-sigma factor, preventing the activation of a third integral membrane protein, RseP. With RseB removed, RseP is now able to cleave RseA at a second site, releasing the sigma factor-RseA complex. The protease ClpXP uses ATP to remove RseA and the sigma factor is now able to bind RNA polymerase and recruit it to the gene promoter.

The FecARI signalling pathway is used by bacterial cells to detect iron in the environment and activate the sigma factor FecI to allow for transcription of genes involved in iron uptake (Brooks and Buchanan 2008). In this signalling pathway, an integral outer membrane protein, FecA, binds iron-citrate in the environment. This binding acts as a signal that allows TonB to recognize a designated region of FecA. This recognition enables FecA to pass the signal on to the integral inner membrane protein FecR. FecR becomes activated and then activates the sigma factor, FecI. FecI is then able to bind the RNA polymerase and recruit it to the gene promoter.

Statement of Purpose

The purpose of this project is to determine if specific ECF sigma factors are involved in the regulation of the genes for carbohydrate-degrading enzymes in *Saccharophagus degradans* 2-40. Alternative sigma factors are one type of regulatory

system that could have been acquired by the bacterium through horizontal gene transfer along with the cellulase genes. A bioinformatic analysis was conducted on the sigma factors to predict which are the most likely to play a role in the regulation of carbohydrate degrading enzymes. It was predicted that if a sigma factor is involved in carbohydrase gene regulation then carbohydrase activity would be observed even when the transformed bacterium is grown on media containing glucose as its sole carbon source. It was also predicted that an increase in carbohydrase activity would be observed when the transformed bacterium is grown on media containing the appropriate complex carbohydrate as its sole carbon source.

Chapter 3: Bioinformatic Evaluation of Sigma Factors Involved in the Regulation of the Cellulolytic System

Introduction

The cellulolytic system is currently the most characterized complex-carbohydrate degrading enzyme system of *S. degradans* 2-40 (Hutcheson et al. 2011). The many genes of the cellulolytic system are dispersed in the genome, were apparently acquired by horizontal gene transfer, and are not obviously associated with known regulatory factors. For these reasons, the cellulolytic system of *S. degradans* 2-40 provides an interesting subject for bioinformatic analysis of carbohydrase systems.

Bioinformatic analysis of the cellulolytic system can potentially reveal critical information regarding evolutionary relationships, gene transfer, and gene regulation in this bacterium. As the genes of the cellulolytic system appeared to have been acquired by horizontal gene transfer, it is possible that regulatory systems for those genes were acquired from the source bacterium as well. The acquisition of these genes could be direct in that the source organism provided the genes for both the carbohydrases and the regulatory system or indirect in which another bacterium provided the genes for carbohydrases and regulatory system to many bacteria, including *S. degradans*. This shared ancestry should be evident through bioinformatics analyses of bacteria with similar cellulolytic systems. Such analysis could indicate possible species from which *S. degradans* acquired its various carbohydrase genes and potentially the regulatory mechanisms controlling carbohydrase expression.

Materials and Methods

BLAST (Basic Local Alignment Search Tool, NCBI) was used to identify sequences of high end-to end similarity to *S. degradans* cellulase and regulatory genes in other species of bacteria. The degree of similarity was determined by the statistical information provided by the program. The minimum coverage represents the length of similar regions between the two sequences. We have defined end-to-end high similarity as having a minimum score of 101, a minimum coverage of 77%, and a maximum e value of e^{-20} as determined by a BLAST search against a query sequence. The minimum coverage represents the length of similar regions between the two sequences. End to end similarity means that the genes have sequence similarity over almost all of their entire length and is used to indicate both an evolutionary and functional relationship between sequences.

Results

Bioinformatic Analysis of Cellulolytic Systems in other Bacteria

It was first necessary to establish which species have similar cellulolytic systems to that of *S. degradans* 2-40. Genes of the known cellulolytic system of the *S. degradans* 2-40 were analyzed using the BLAST tool to identify bacterial species with end-to-end homologs of high similarity. In this analysis a BLAST search was performed on both the nucleotide and amino acid sequence of each enzyme of the cellulolytic system and the sequences with the highest end-to-end similarity were selected for

further analysis. The object was to look for both content (the presence of high sequence similarity) and context (the same surrounding genes).

End to end similarity indicates that the two genes in question share sequence similarity over virtually the entire length of the sequence. An example can be seen in the results of the protein BLAST of *S. degradans* 2-40 Cel5A (Figure 1). These results depict the sequences with the highest similarity to Cel5A. The most similar sequence appears at the top and this is the sequence of *S. degradans* Cel5A itself. The next three sequences show end-to-end sequence similarity to Cel5A because their sequence coverage encompasses almost the entire length of the protein sequence. In these three results the small gap in the middle of the sequence is linked by a bar, indicating that these sequences are regions of the same protein. The remaining results only display similarity to certain regions of the Cel5A query sequence. These patterns correspond to the modulated structure of Cel5A that features two GH5 domains on either end of the protein separated by three CBM domains. Many proteins contain CBM or GH5 domains so it is not surprising that localized similarity to these regions may occur, however these do not represent an evolutionary relationship.

Figure 1 Results of Protein BLAST of *S. degradans* 2-40 Cel5A

Cellulolytic system	pBLAST					
	Species	Accession #	Description	Max score	Coverage	e value
Cel5A	Hahella chejuensis	YP_435061.1	endoglucanase [Hahella chejuensis	496	87%	7.00E-138
	Cellvibrio mixtus	AAB61462.2	cellulase B [Cellvibrio mixtus]	452	89%	1.00E-124
	Pseudomonas sp. ND137	BAB79288.1	cellulase [Pseudomonas sp. ND137]	401	85%	3.00E-109

YP_528472.1	hypothetical protein Sde_3003 [Saccharophagus degradans 2-40] >gb|ABD82260.1| endoglucanase [Saccharophagus degradans 2-40]	2407	2407	100%	0.0
YP_435061.1	endoglucanase [Hahella chejuensis KCTC 2396] >gb|ABC30636.1| Endoglucanase [Hahella chejuensis KCTC 2396]	496	1057	87%	7e-138
YP_527966.1	hypothetical protein Sde_2494 [Saccharophagus degradans 2-40] >gb|ABD81754.1| putative endoglucanase [Saccharophagus degradans 2-40]	453	1018	89%	7e-125
AAB61462.2	cellulase B [Cellvibrio mixtus]	452	885	89%	1e-124
BAB79288.1	cellulase [Pseudomonas sp. ND137]	401	867	85%	3e-109
YP_528706.1	cellulase [Saccharophagus degradans 2-40] >gb|ABD82494.1| putative endoglucanase [Saccharophagus degradans 2-40]	290	647	82%	1e-75

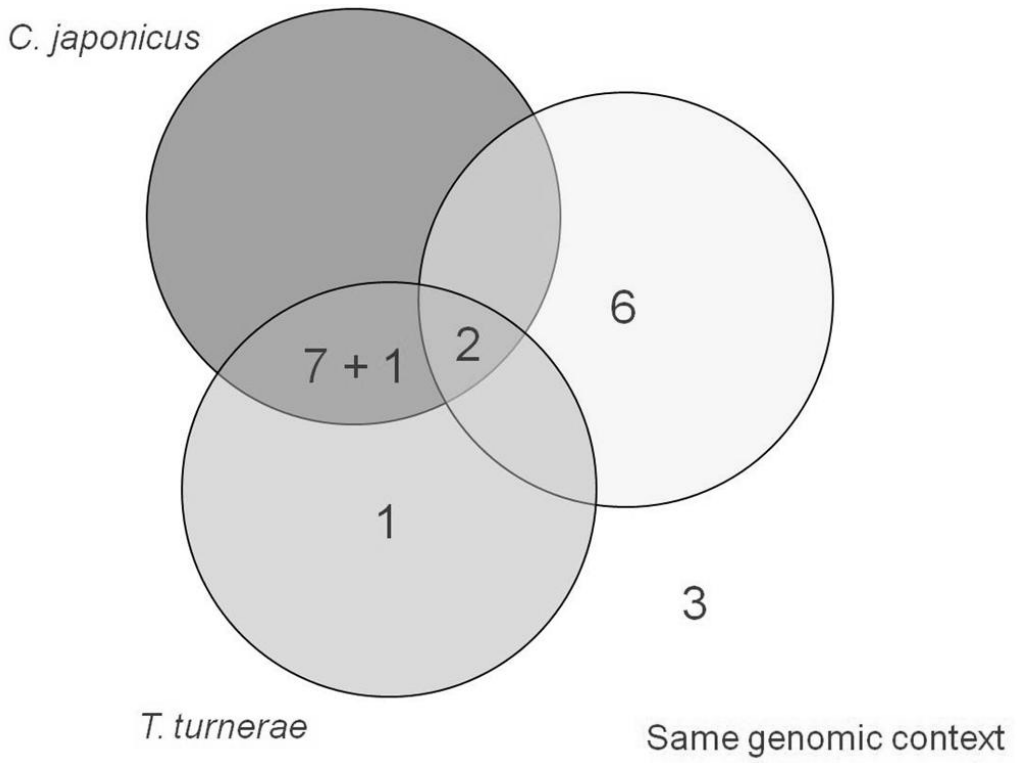
The use of the nucleotide BLAST tool was also explored as a method of detecting sequence similarity to the *S. degradans* 2-40 cellulolytic system. Nucleotide BLAST was not appropriate for this purpose because significant similarity was only detected for 3 out of the 13 cellulases and 2 of the associated accessory enzymes, with a maximum of two similar sequences returned. In contrast and as expected similarity analysis of the amino acid sequence produced similar sequence results for all cellulases except Cel5B, as well as all of the accessory enzymes. It was observed that the species with the most similar sequences to the accessory enzymes were often very different from the species with the most similar sequences to the cellulases.

The species most frequently carrying a full length high sequence similarity to the *S. degradans* 2-40 cellulases were *Cellvibrio japonicus*, *Teredinibacter turnerae*, *Hahella chejuensis*, *Cellvibrio mixtus*, and *Pseudomonas sp. ND137*. *C. mixtus* was omitted from further BLAST analysis due to its similarity to *C. japonicus* (DeBoy et al. 2008). The BLAST searches were repeated for all of the cellulases and accessory enzymes to determine if high scoring end to end similarity existed within these species at all, not just within the top results. Five patterns of sequence similarity were observed (Figure 2). Eight of the cellulolytic system enzymes (Cel5C, Cel5D, Cel5F, Cel5I, Cel9B, Ced3A, Cep94A, and Cep94B) had similar sequences in both *C. japonicus* and *T. turnerae* including one enzyme (Cep94B) that showed similarity in its genomic context. Similar sequences for two of the cellulolytic system enzymes (Cel9A and Ced3B) were found in all of the cellulolytic species examined. Six of the cellulolytic system enzymes (Cel5A, Cel5G, Cel5J, Cel6A, Bgl1A, and Bgl1B) had a

similar sequence in *H.chejuensis* only. Another cellulolytic system enzyme (Bgl3C) showed similarity only in *T. turnerae*. It was also found that three of the cellulolytic system enzymes (Cel5B, Cel5E, and Cel5H) did not have similar sequences in any of the species used in this analysis. None of the cellulases or accessory enzymes had end to end high sequence similarity in the non-cellulolytic *P. aeruginosa*.

Figure 2 Comparative Genomics of *S. degradans* 2-40 Cellulolytic System Enzymes

S. degradans 2-40 Cellulolytic System	*C. japonicus*	*T. turnerae*	*H. chejuensis*
Cel5A	-	-	+
Cel5B	-	-	-
Cel5C	+	+	-
Cel5D	+	+	-
Cel5E	-	-	-
Cel5F	+	+	-
Cel5G	-	-	+
Cel5H	-	-	-
Cel5I	+	+	-
Cel5J	-	-	+
Cel6A	-	-	+
Cel9A	+	+	+
Cel9B	+	+	-
Ced3A	+	+	-
Ced3B	+	+	+
Bgl1A	-	-	+
Bgl1B	-	-	+
Bgl3C	-	+	-
Cep94A	+	+	-
Cep94B	+	+	-



Based on the results of the bioinformatic analysis, it was determined that *Cellvibrio japonicus, Teredinibacter turnerae,* and *Hahella chejuensis* are the most appropriate species to predict regulators of cellulolytic enzymes. *Cellvibrio japonicus* was confirmed to have the cellulolytic system most similar to that of *S. degradans* 2-40. *Teredinibacter turnerae* is the species phylogenetically most closely related to *S. degradans* 2-40 (González and Weiner 2000). Finally, *Hahella chejuensis* is a cellulolytic bacterium from a marine estuary that also shows high similarity to *S. degradans* 2-40 in its cellulolytic system. It was postulated that *Pseudomonas sp. ND137* might also be used to score for high sequence similarity of regulators, but insufficient information about the genome of this organism prevented its use in further analysis.

Bioinformatic Analysis of Sigma Factors

Nineteen sigma factors were identified in the genome annotation of *S. degradans* 2-40. Five of these sigma factors (genes numbers 716, 3594, 2164, 1253, and 3179) show similarity to the known sigma factors RpoD, RpoH, RpoF, RpoS, and RpoN (Table 1). The other fourteen sigma factors are of unknown function and their annotations show similarity to regions 2 and/or 4 of sigma70 (Table 2). These fourteen sigma factors have been designated as candidate sigma factors for the regulation of carbohydrase genes. An operon analysis of these genes showed that eleven of these sigma factors could be part of an operon; however, no obvious carbohydrase regulatory function was indicated by the surrounding genes.

Table 1 Housekeeping Sigma factors of *S. degradans* 2-40 as Identified from Gene Annotations

Gene Number	Description	Possible Operon
716	RpoD	Possible with DNA primase
3594	RpoH	None
2164	RpoF	None
1253	RpoS	None
3179	RpoN	None

Table 2 Candidate Sigma factors for Carbohydrase Regulation as Identified from Gene Annotations

Gene Number	Description	Operon
95	Sigma 70 regions 2&4	Genes of unknown function
249	Sigma 70 regions 2&4	Genes of unknown function
297	Sigma 70 regions 2&4	Peptidase S8 & S53, unknown
799	Sigma 70 regions 2&4	Glutathione S-transferase
823	Sigma 70 region 2	Von Willebrand factor type A
1442	Sigma 70 regions 2&4	TonB dependent receptor, FecR
1647	Sigma 70 regions 2&4	None
2251	Sigma 70 regions 2&4	None
2365	Sigma 70 regions 2&4	Possible with genes of unknown function
2388	Sigma 70 regions 2&4	Genes of unknown function
2407	Sigma 70 regions 2&4	Genes of unknown function
2471	Sigma 70 regions 2&4	Anti-sigma factor
3322	Sigma 70 regions 2&4	None
3895	Sigma 70 regions 2&4	Genes of unknown function

Comparative genomics was performed on the annotated sigma factors of *S. degradans* 2-40 using the genomes of the bacterial species with the similar cellulolytic systems identified above. Sigma factors were evaluated for their likelihood of being involved in carbohydrase regulation based upon the presence or absence of high sequence similarity in cellulolytic but not in non-cellulolytic species. If a *S. degradans* 2-40 sigma factor was found to have high sequence similarity to sigma factors in non-cellulolytic species it would likely not be involved in carbohydrase regulation. In contrast, if it is determined that a *S. degradans* sigma factor has high sequence similarity only in cellulolytic species it could indicate that the sigma factor is a regulator of carbohydrase genes. Based on the results of the bioinformatic analysis of the cellulolytic system, *Cellvibrio japonicus*, *Teredinibacter turnerae*, and *Hahella chejuensis* were selected as the cellulolytic species to score for the presence or absence of an end to end high sequence similarity of *S. degradans* 2-40 sigma factors and *Pseudomonas aeruginosa* was used as the non-cellulolytic organism.

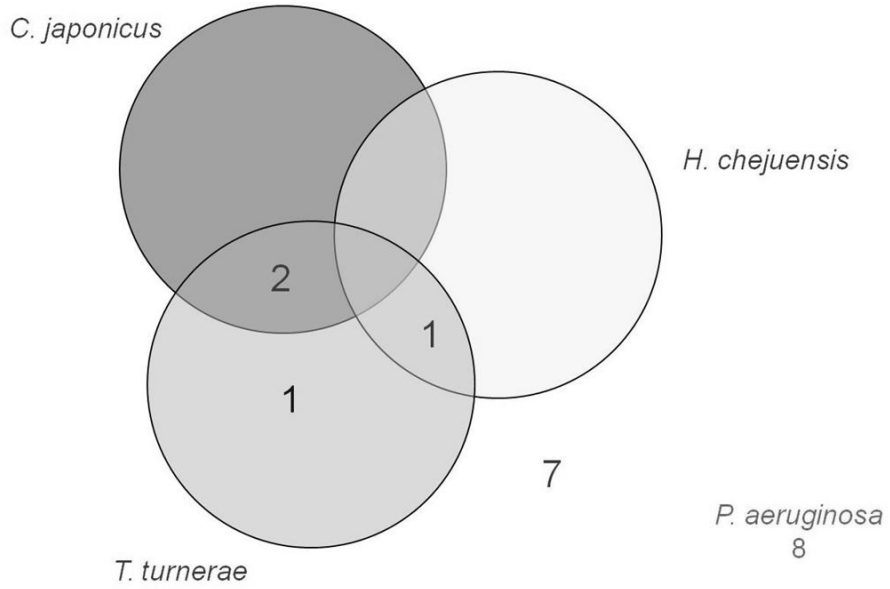
All of the annotated housekeeping sigma factors in *S. degradans* 2-40 had high sequence similarity in both the tested cellulolytic and non-cellulolytic species as expected (Figure 3). Seven of the candidate carbohydrase regulating ECF sigma factors did not have high sequence similarity in either the cellulolytic or the non-cellulolytic species. Four sigma factors were found to have high sequence similarity in cellulolytic species only. Eight sigma factors had highly similar sequences in the non-cellulolytic *P. aeruginosa*. It was concluded from this analysis that the sigma

factors for further study were genes 2471, 823, 2388, and 2407, which showed high sequence similarity to sigma factors in cellulolytic species only.

Figure 3 Comparative Genomics of *S. degradans* 2-40 Sigma Factors

Sigma Factor	Description	*P. aeruginosa*	*C. japonicus*	*T. turnerae*	*H. chejuensis*
716	RpoD	+	+	+	+
3594	RpoH	+	+	+	+
2164	RpoF	+	+	+	+
1253	RpoS	+	+	+	+
3179	RpoN	+	+	+	+

		BLASTp		
Sigma factor	*P. aeruginosa*	*C. japonicus*	*T. turnerae*	*H. chejuensis*
95	-	-	-	-
206	-	-	-	-
799	-	-	-	-
1647	-	-	-	-
3895	-	-	-	-
1442	-	-	-	-
2365	-	-	-	-
2471	-	-	+	+
2388	-	-	+	-
823	-	+	+	-
2407	-	+	+	-
2251	+	+	+	+
3322	+	+	-	+
297	+	-	-	-



Sigma Factor	Sequence Similiarity Species	Max Score	Coverage	e value
2407	C. japonicus	243	98%	7.00E-63
823	C. japonicus	198	84%	3.00E-49
2471	T. turnerae	166	84%	1.00E-39
	H. chejuensis	153	94%	1.00E-35
2388	T. turnerae	101	96%	4.00E-20

Of the four sigma factors that did not have a high sequence similarity in the non-cellulolytic species, but did have a high sequence similarity in the cellulolytic species, 3 patterns are observed: high sequence similarity in *T. turnerae*, high sequence similarity in *C. japonicus* and *T. turnerae*, and high sequence similarity in *H. chejuensis* and *T. turnerae* (Table 3). Considering that regulators of carbohydrase expression are likely to have been acquired by *S. degradans* 2-40 from the same species from which it acquired the carbohydrases themselves, the findings from the bioinformatic analysis of the sigma factors can then be compared to the bioinformatic analysis of the cellulolytic system. None of the cellulolytic system enzymes have high sequence similarity in *H. chejuensis* or *T. turnerae* only. Sigma factor 2388 has a high

sequence similarity in *T. turnerae* only; however both this sigma factor and its similar sequence appear to be in an operon with FecR indicating that it is likely a FecI homolog. Several cellulolytic system enzymes have high sequence similarity in both the genomes of *C. japonicus* and *T. turnerae*. The two sigma factors with high sequence similarity in *C. japonicus* and *T. turnerae* then become the most likely to act as extracellular function sigma factors that regulate carbohydrase expression.

Table 3 Comparison of Patterns of High Sequence Similarity in *S. degradans* Cellulolytic System Enzymes and Sigma Factors

Pattern of Similarity	Cellulolytic System	Sigma Factors
C. japonicus only	0	0
T. turnerae only	1	1
H. chejuensis only	6	0
C. japonicus and *T. turnerae*	8	2
C. japonicus and *H. chejuensis*	0	0
H. chejuensis and *T. turnerae*	0	1
C. japonicus, T. turnerae, and *H. chejuensis*	2	0
P. aeruginosa	0	8
None	3	7

Discussion and Hypothesis

BLAST analysis of the *S. degradans* 2-40 cellulolytic system enzymes was used to determine the bacterial species most appropriate for scoring for high sequence similarity of gene regulators to predict their likelihood of being involved in the regulation of *S. degradans* 2-40 carbohydrases. The object of the bioinformatic analysis was to determine sequences having high scoring end to end similarity to the *S. degradans* 2-40 cellulolytic system enzymes, which would indicate shared ancestry. It was found that *Cellvibrio japonicus*, *Teredinibacter turnerae*, and *Hahella chejuensis* are the most appropriate species for evaluating transcriptional regulators because *C. japonicus* has the cellulolytic system most similar to that of *S. degradans* 2-40, *T. turnerae* is the species phylogenetically most closely related to *S. degradans*, and *H. chejuensis* is a marine bacterium that contains many enzymes similar to those of the *S. degradans* cellulolytic system. These findings were then used to evaluate which sigma factors were most likely to be involved in the regulation of *S. degradans* 2-40 cellulases based upon high sequence similarity and comparative genomic patterns.

It was predicted that the sigma factors corresponding to genes 823 and 2407 were those that were most likely to regulate the *S. degradans* 2-40 cellulases and potentially other carbohydrase enzymes because similar sequences can be found in the same cellulolytic species as several of the cellulase enzymes. Both of the protein sequences of these annotated sigma factors have highly similar sequences in *C. japonicus* and *T. turnerae*. Several of the cellulolytic system enzymes including

Cel5C, Cel5D, Cel5F, Cel5I, Cel9B, Ced3A, Cep94A, and Cep94B also have highly similar sequences in *C. japonicus* and *T. turnerae*. None of the other patterns found of sigma factor sequence similarity matched those observed for the cellulolytic system, other than a probable FecI homolog. The hypothesis is that the sigma factors corresponding to genes 823 and 2407 of *S. degradans* 2-40 are involved in the regulation of carbohydrase gene expression in this bacterium.

Chapter 4: Molecular Cloning of Sigma Factors and Assays for Carbohydrase Activity

Introduction

The bioinformatic analyses predicted that sigma factors corresponding to genes 823 and 2407 were those most likely to act as regulators of expression of the *S. degradans* cellulolytic system genes. Both of these sigma factors display the same pattern of sequence similarity as many of the cellulolytic system enzymes. In order to test this hypothesis each sigma factor gene was cloned into an expression vector. The recombinant plasmid was transformed into *S. degradans* 2-40 and expression of the sigma factor was induced. Carbohydrase activity of the cell lysate was then measured to determine whether the sigma factor had an effect on the expression of carbohydrases. It was predicted that if the sigma factor is involved in regulation of carbohydrase genes then carbohydrase activity would be observed even when the bacteria was grown on glucose-containing media. It was also predicted that if the sigma factor regulates carbohydrase gene expression then an increase in carbohydrase activity would be observed when the bacteria was grown on media containing a complex carbohydrate substrate. The complex carbohydrate substrates tested in this study include cellulose, xylan, and pectin.

Materials and Methods

Strains and Growth Media

S. degradans 2-40 (ATCC 43961) was grown at 27C in minimal medium broth containing per liter 2.3% Instant Ocean (Aquarium Systems), 0.2% yeast extract, 0.5% tryptone, 0.1% ammonium chloride, and 16.7 mM Tris-HCl, pH 8.6 (Sigma) supplemented with 0.2% glucose, Avicel, Birchwood xylan, or citrus pectin. For agar, half-strength marine media 2216 (Difco) was used. *E. coli* DH5α (Invitrogen) was grown at 37C in modified Luria-Bertani broth or agar containing 0.5% yeast extract (Difco), and 1% sodium chloride. Streptomycin was added to the media at a concentration of 50 ug/mL for agar or 100 ug/mL for broth cultures.

Cloning of Sigma Factors 823 and 2407

Primers were designed to amplify each selected sigma factor gene (823 and 2407) and the surrounding region in order to include the ribosome binding site (Appendix 1). *Xba*I and *Hin*dIII sites were added to the primers so that the resulting PCR products would contain these restriction sites at their 5' and 3' ends respectively. Amplified fragments were ligated into the IncQ plasmid pMMB503EH (gift from Michael Bagdasarian, Michigan State University). IncQ plasmids have a broad host range and it has been demonstrated that these plasmids can be successfully transformed and stably replicated in *S. degradans* 2-40, whereas many other plasmids cannot. The pMMB503EH plasmid features single *Xba*I and *Hin*dIII restriction sites, the *tac* promoter for expression of the inserted gene, and *lacI* for strong regulation of the *tac* promoter. pMMB503EH also contains streptomycin resistance genes, *mob* genes that allow for plasmid conjugation, and *rep* genes that control plasmid replication. pMMB503EH was isolated from an overnight culture of *E. coli* using alkaline lysis

followed by purification with a DNA Clean and Concentrator kit (Zymo). It was demonstrated that pMMB503EH could be successfully transformed into *S. degradans* 2-40 via electroporation and could therefore be used in this experiment. The ligation mixtures were first transformed into DH5α and colony PCR and sequencing (Genewiz) were used to confirm the construct. Next each construct was transformed in to *S. degradans* 2-40 through electroporation. Colony PCR using primers designed to amplify the region of pMMB503EH containing the insert were used to confirm the presence of the construct.

Sigma Factor Expression in S. degradans 2-40

The design of the experiment was to induce expression of the sigma factor in *S. degradans* 2-40 grown on various carbon sources and assay for a change in carbohydrase activity compared to an uninduced culture. In order to express the cloned genes, cultures of *S. degradans* 2-40 containing each plasmid-insert construct were grown to $OD_{600} \geq 2.00$ in marine media containing 0.2% glucose, Avicel, Birchwood xylan, or citrus pectin as a carbon source. Expression of each cloned sigma factor gene was induced by addition of 1mM IPTG to half of the culture and incubation at 27C for 14 hours. The cells were collected by centrifugation and resuspended in 50mM sodium phosphate, 300mM sodium chloride, 10mM imidazole, pH8.0 buffer and 0.5mM phenyl methyl sulfonyl fluoride. Cells were lysed in a mini Bead Beater (BioSpec) and the lysate was cleared by centrifugation. The cell lysate was quantified using a microBCA kit (Pierce).

Assays for Cellulolytic and Xylanolytic Activity

Assays were performed in 1M sodium citrate buffer, pH 6.5 at 27C for various time points. Assays for cellulolytic and xylanolytic activity were conducted by the addition of cleared cell lysate to 1mg of substrate (Avicel or Birchwood xylan) or 1% CMC in a total volume of 400uL. Uninduced samples and substrate with no enzyme added were used as controls. The reaction was stopped by boiling for 3 minutes. For Avicel and Birchwood xylan assays, the samples were centrifuged and transferred to new tubes to remove the substrate. DNS assays were used to detect enzyme activity compared to a standard curve (Ghose 1987, Ghose and Bisaria 1987). Carbohydrase activity was detected by measuring the absorbance at OD_{540} compared to a cellobiose or xylose standard curve.

Assay for Pectinolytic Activity

Reactions of the cleared lysate and 0.5% polygalacturonic acid substrate were performed as described above. Pectinolytic activity was assayed according to the protocol of Worthington Biochemical Corporation (Corporation 1993).

SDS-PAGE and Zymogram

SDS-PAGE was performed to ensure expression of the induced sigma factor. Gels were visualized with Sypro Ruby Protein Gel Stain (Molecular Probes). Zymogram analysis of the cleared lysates was performed according to the protocol of Zhang and Hutcheson 2011. After refolding, gels were washed twice with 20mM PIPES buffer,

pH 6.8 then transferred to PIPES buffer, and incubated for 24 hours at 37C before staining with 0.25% Congo Red.

qRT-PCR

Quantitative RT-PCR analysis was performed jointly with Zymetis, Inc. *S. degradans* cultures were grown on glucose media to mid-log phase then induced with 1mM IPTG for 5 hrs. at 27C. Sample preparation and qRT-PCR was performed according to the procedures of Zhang and Hutcheson 2011.

Results

Cloning of Genes 823 and 2407

It was found that colonies of DH5α transformed with the ligated plasmid and PCR-amplified insert could only be obtained with the use of LB media from which tryptone was excluded. Tryptone is a complex media ingredient typically obtained from bovine milk and it is therefore possible that it contains small amounts of lactose that could induce expression of the inserted gene. Because sigma factors play a large role in gene regulation, their expression, particularly in a foreign system, could affect the expression of various genes and prove lethal to the bacterial cell. The finding that DH5α transformed with the ligation mixture containing pMMB503EH and either gene 823 or 2407 could only be grown on LB media lacking tryptone suggests that expression of these genes is lethal to this strain of *E. coli*. Once colonies were obtained, colony PCR and sequencing were used to confirm the construct prior to transformation into *S. degradans*.

Assays for Carbohydrase Activity

To determine if sigma factors 823 and 2407 are involved in carbohydrase gene regulation, *S. degradans* 2-40 strains containing these sigma factors were grown on media containing glucose, Avicel, Birchwood xylan, or citrus pectin as a carbon source. Expression of the sigma factor was induced with IPTG. The cleared cell lysate was allowed to react with a complex carbohydrate substrate and the resulting carbohydrase activity was measured. No change in cellulolytic, xylanolytic, or pectinolytic activity was detected when the strains were grown on Avicel, Birchwood xylan, or citrus pectin respectively as compared to wild-type *S. degradans* 2-40 and *S. degradans* containing pMMB503EH (Table 4). No change in carbohydrase activity was detected for the strain expressing sigma factor 823 grown with glucose as a carbon source. Carbohydrase activity was detected for glucose cultures of *S. degradans* 2-40 expressing sigma factor 2407 (Table 4 and Figure 4).

Figure 4 Carbohydrase Activity of Cell Lysate of *S. degradans* Expressing Sigma Factor 2407 Grown on Glucose Media

Cultures of *S. degradans* containing the sigma factor 2407 construct were grown with glucose as a carbon source. IPTG was used to induce protein expression. The cell lysate of the culture was assayed for activity on CMC, Avicel, Birchwood xylan, and polygalacturonic acid. All assays were performed in triplicate.

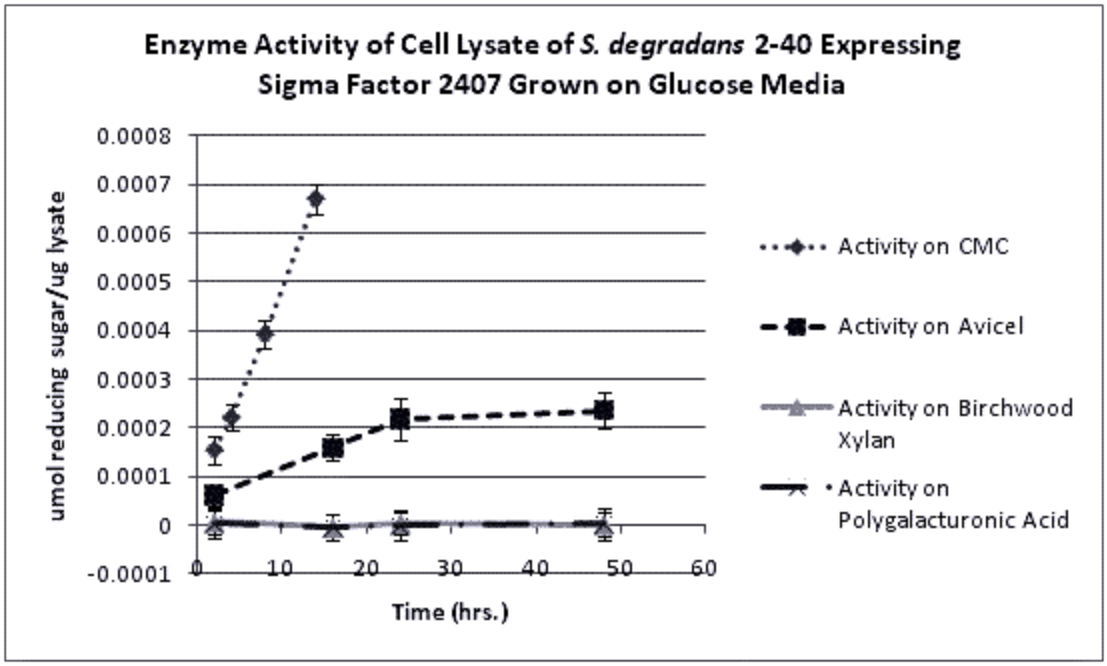


Table 4 Carbohydrase Activity of *S. degradans* Strains Grown on Various Carbon Sources

This table shows the carbohydrase activity detected from the cell lysate of the indicated *S. degradans* strain grown on the indicated carbon source. The left column indicates the assay performed as well as the carbon source used. Activity is reported as μmol product/minute/μg cell lysate. All assays were performed in triplicate. The standard deviation for all assays was less than 5.42E-07.

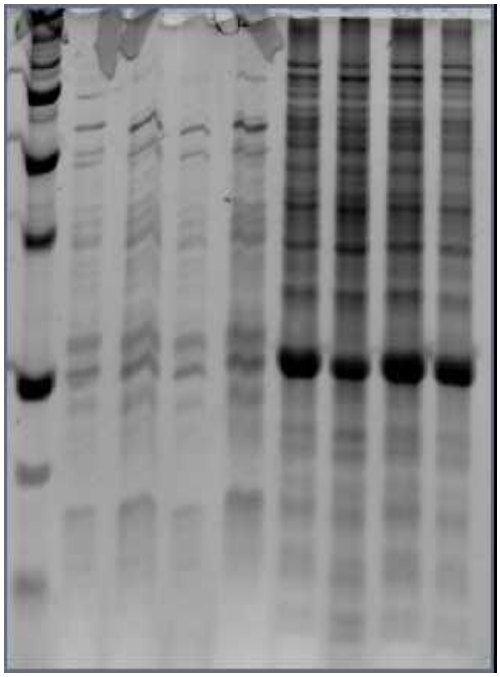
	Strain							
Assay	Wild-type Induced	Wild-type Uninduced	pMMB503EH Induced	pMMB503EH Uninduced	823 Induced	823 Uninduced	2407 Induced	2407 Uninduced
CMC assay of glucose culture	3.38E-07	3.45E-07	3.64E-07	3.69E-07	3.29E-07	3.39E-07	1.13E-06	8.78E-07
Avicel assay of glucose culture	1.84E-07	1.88E-07	1.88E-07	1.92E-07	1.94E-07	1.68E-07	3.35E-07	2.17E-07
Xylan assay of glucose culture	1.11E-07	1.19E-07	1.22E-07	1.23E-07	1.21E-07	1.09E-07	1.28E-07	1.10E-07
Polygalacturonic acid assay of glucose culture	1.05E-07	1.06E-07	1.07E-07	1.06E-07	1.13E-07	1.06E-07	1.05E-07	1.08E-07
CMC assay of Avicel culture	1.99E-06	1.95E-06	1.83E-06	2.05E-06	2.02E-06	1.98E-06	1.80E-06	2.00E-06
Avicel assay of Avicel culture	5.16E-07	4.97E-07	4.65E-07	5.28E-07	5.12E-07	5.01E-07	4.57E-07	5.10E-07
Xylan assay of Xylan Culture	2.77E-07	3.05E-07	3.02E-07	2.86E-07	2.94E-07	2.80E-07	2.87E-07	2.82E-07
Polygalacturonic acid assay of pectin culture	2.44E-07	2.81E-07	2.75E-07	2.33E-07	2.45E-07	2.62E-07	2.88E-07	2.43E-07

SDS-PAGE and Zymogram

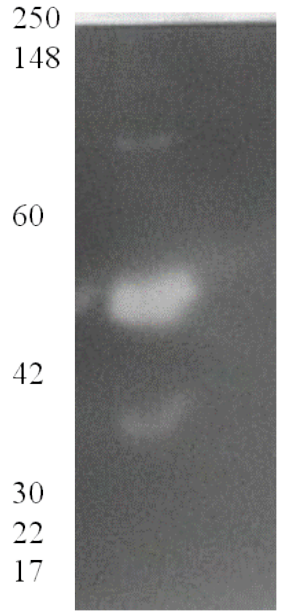
SDS-PAGE was performed on the cell lysate following induction to ensure expression of the cloned sigma factor genes. Expression of each sigma factor can be visualized as a ~26kDa band (Figure 5A). Significant protein expression was evident even in uninduced cultures of strains containing genes 823 and 2407. Zymogram analysis was used to verify carbohydrase activity and attempt to identify the cellulases being regulated by sigma factor 2407. Activity was observed as three clear bands (Figure 5B).

Figure 5 SDS-PAGE and Zymogram of Cell Lysate of Glucose Grown Cultures

A. SDS-PAGE of cell lysates of induced and uninduced *S. degradans* cultures grown on glucose media. Lanes are 1. Precision Plus Protein standard (Bio-rad) 2. Wild-type induced 3. Wild-type Uninduced 4. pMMB503EH induced 5. pMMB503EH Uninduced 6. Sigma factor 2407 induced 7. Sigma factor 2407 uninduced 8. Sigma factor 823 induced 9. Sigma factor 823 uninduced



B. Zymogram of cell lysate of glucose culture of *S. degradans* expressing sigma factor 2407. Molecular mass of MultiMark Multi-colored Standard (Invitrogen) is reported in kilodaltons. Lane is 1. 40 ug of cell lysate of Sigma factor 2407 induced



qRT-PCR

Due to the number of cellulases and potential degradation products that could be responsible for the bands observed in the zymogram, qRT-PCR was used to definitively identify the cellulases being regulated by sigma factor 2407. Transcript levels of known cellulase genes were measured for cultures of sigma factor 2407 and pMMB503EH strains following induction. Transcript levels measured from the sigma factor 2407 strain are reported relative to those of pMMB503EH strain. Transcript levels of cellulase genes were normalized to either those of housekeeping genes *folA*, *spoR*, or *16s* (Table 5). The results indicate a decrease in transcript levels of the tested cellulase genes in the sigma factor 2407 strain compared to the pMMB503EH strain.

Table 5 Relative Transcript Levels of Known Cellulase Genes of Induced Cultures of *S. degradans* Expressing Sigma Factor 2407

Gene	Ratio of Expression in Strain 2407 Relative to pMMB503EH Strain Normalized to *16s*	Ratio of Expression in Strain 2407 Relative to pMMB503EH Strain Normalized to *folA*	Ratio of Expression in Strain 2407 Relative to pMMB503EH Strain Normalized to *spoR*
cel5A	0.05	0.26	0.02
cel5B	0.15	0.77	0.07
cel5C	0.02	0.10	0.01
cel5D	0.03	0.16	0.02
cel5E	0.18	0.90	0.09
cel5F	0.07	0.37	0.04
cel5G	0.02	0.12	0.01
cel5H	0.11	0.53	0.05
cel5I	0.05	0.25	0.02
cel5J	0.04	0.22	0.02
cel9A	0.06	0.29	0.03
cel9B	0.03	0.16	0.02
cel6A	0.06	0.31	0.03
16S	1.00	5.01	0.48
spoR	2.08	10.45	1.00
folA	0.20	1.00	0.10
gly5K	0.08	0.42	0.04
gly5L	0.03	0.17	0.02
gly5M	0.03	0.13	0.01
gly5R	0.01	0.06	0.01
gly5S	0.08	0.40	0.04
gly9C	0.11	0.55	0.05

Discussion

Bioinformatic analysis demonstrated that sigma factors 823 and 2407 were those most likely involved in carbohydrase regulation because similar sequences were found in the same cellulolytic species of bacteria as many of the *S. degradans* cellulolytic system enzymes. Sigma factors 823 and 2407 were cloned into an expression vector and expressed in *S. degradans.* Cellulolytic, xylanolytic, and pectinolytic activity was assayed from the resulting cell lysate compared to wild-type and plasmid-only controls. No change in carbohydrase activity was detected for *S. degradans* cultures expressing sigma factor 823 compared to control cultures when glucose, Avicel, Birchwood xylan, or citrus pectin was used as a carbon source. Activity on CMC and Avicel was detected for glucose cultures of *S. degradans* expressing sigma factor 2407. This finding supports the hypothesis that sigma factor 2407 does function in the regulation of cellulase genes and is even able to up regulate the expression of these genes in the presence of catabolite repression. No change in carbohydrase activity was detected when *S. degradans* expressing sigma factor 2407 was grown on Avicel, Birchwood xylan, or citrus pectin. The lack of an increase in cellulase activity for Avicel-grown cultures of *S. degradans* expressing sigma factor 2407 can be explained by the presence of a multitude of other cellulases in this organism whose activity could be masking the effect of the enzymes being up regulated by sigma factor 2407. No change in xylanolytic or pectinolytic activity was detected for cultures grown on any of the tested media.

qRT-PCR was used to identify the cellulase genes being up regulated by sigma factor 2407. None of the tested cellulase genes showed increased transcript levels during expression of sigma factor 2407 and instead transcript levels of these genes decreased. It was concluded that sigma factor 2407 does not appear to function in the regulation of known cellulase genes.

Chapter 5: General Discussion and Future Research

S. degradans is a marine bacterium that employs a variety of enzymes, including cellulases, xylanases, and pectinases, to accomplish degradation of whole plant material. Regulation of these carbohydrase enzymes is evidenced by their expression when the bacterium is grown on media containing the respective carbohydrate, but not on glucose-containing media. In order to investigate the potential role of alternative sigma factors in regulation of these carbohydrase genes a bioinformatic evaluation was performed to predict the sigma factors most likely to be involved. Because it is likely that this organism acquired carbohydrase regulating mechanisms from the same species from which it acquired its carbohydrases through horizontal gene transfer, sigma factors were evaluated for end to end high sequence similarity in cellulolytic and non-cellulolytic species.

The cellulolytic species established as the most appropriate for comparative genomics of transcription factors were *Cellvibrio japonicus*, which was confirmed to have the most similar cellulolytic system to *S. degradans*, *Teredinibacter turnerae*, the most closely related species, and *Hahella chejuensis*, a marine bacteria with many highly similar sequences to *S. degradans* cellulases. The non-cellulolytic but closely related *Pseudomonas aeruginosa* was used to eliminate sigma factors involved in regulation of non-carbohydrase genes. It was hypothesized that sigma factors corresponding to genes 823 and 2407 were those most likely to be involved in carbohydrase gene regulation in *S. degradans* due to their shared pattern of sequence similarity with many of the cellulolytic system enzymes. Expression of these sigma factors in *S.*

degradans revealed that sigma factor 2407 induces cellulase activity even when the bacterium is grown on glucose media. This finding is significant because it represents the first regulator of carbohydrase genes identified in this bacterium.

Quantitative RT-PCR analysis of an induced glucose-grown culture of the sigma factor 2407 strain indicated a decrease rather than an increase in transcript levels of known cellulase genes compared to the plasmid-only strain. Several possibilities exist to explain this observation. One possibility is that sigma factor 2407 does up regulate expression of one or more of the tested carbohydrase genes, but that the increase in cellulase transcript levels was too slight to be detected by qRT-PCR. Zymograms, particularly those that use barley β-glucan as a substrate are extremely sensitive to cellulase activity therefore it is possible that cellulase activity could be indicated in a zymogram, but not in qRT-PCR. Another possibility is that sigma factor 2407 acts to up regulate expression of a second transcription factor and that that factor is then responsible for up regulating expression of cellulase genes. In this scenario, up regulation of cellulase genes may not be observed in qRT-PCR analysis. Another possibility is that sigma factor 2407 up regulates expression of one or more cellulases other than the known cellulases whose transcript levels were measured in this study. As stated previously, the *S. degradans* genome contains 180 open reading frames that are predicted to encode carbohydrases. It is therefore not surprising that a transcriptional regulator could be involved in the regulation of one or more cellulase genes that have not been previously described. Further research must be performed in

order to determine the identity of the cellulase genes up regulated by sigma factor 2407.

Once the cellulases regulated by sigma factor 2407 have been identified, future research would include determining the mechanism by which this sigma factor regulates their expression. Identification of the promoters of these genes would indicate the site at which sigma factor 2407 binds and recruits RNA polymerase to facilitate transcription of the genes. It would also be of interest to investigate the mechanism by which sigma factor 2407 itself is regulated, whether by direct transcriptional regulation or the Sigma E or FecARI signalling pathways. There are many homologs of TonB-dependent receptors in this system, but few anti-sigma factors so there is some question as to how either of these signalling pathways would function in this bacterium.

From the *S. degradans* 2-40 genome annotation, sigma factor 2407 appears to be in an operon with an unknown protein containing transmembrane helices and a signal sequence. BLAST of the amino acid sequence indicates that this protein is an ornithine carbamoyltransferase and that it has a lower scoring similar sequence in *C. japonicus* that is located in an operon with an ECF sigma factor, unknown proteins, and a murein polymerase. The ECF sigma factor in this operon is the same sigma factor that was determined to have high sequence similarity to sigma factor 2407 of *S. degradans*. No similar sequence was found in *T. turnerae*. It is also possible that this operon contains another unknown protein containing transmembrane helices and an

AraC type protein containing a helix-turn-helix domain due to the unusual gene organization in this bacterium. The unknown protein has a lower scoring similar sequence in *C. japonicus* that appears to be in the same operon as the other unknown protein with the ECF sigma factor. In short, sigma factor 2407 of *S. degradans* not only has a high scoring end to end sequence similarity to an ECF sigma factor of *C. japonicus*, but also these similar sequences appear in the same context in their respective genomes. This contextual evidence strongly suggests an evolutionary relationship by which *S. degradans* 2-40 either acquired this gene cluster from *C. japonicus* directly or that both species acquired these genes from another source organism. The membrane-associated proteins and signal sequence could suggest that regulation of the sigma factor occurs at the membrane, such as occurs in both the Sigma E and FecARI signalling pathways. The AraC type regulator does not have end to end sequence similarity in *C. japonicus* or *T. turnerae*, but does in various distantly related, non-cellulolytic organisms, suggesting it may not function in the regulation of sigma factor 2407 and, with 191bp separating it from the other genes, may not be part of this operon. Further research is necessary in order to suggest a model for the regulation of sigma factor 2407.

Future research could be to examine the potential role of other sigma factors involved in regulating carbohydrase gene expression in *S. degradans* 2-40. Of the nineteen annotated sigma factors present in the *S. degradans* genome, fourteen are of unknown function. It was found that sigma factors 95, 206, 799, 1647, 2365, 1442, and 3895 did not have high sequence similarity in either the cellulolytic or the non-cellulolytic

species used for scoring. In addition, besides the sigma factors 823 and 2407 examined in this study, two other sigma factors 2471 and 2388 were found to have high sequence similarity in the cellulolytic species used for scoring, and not the non-cellulolytic species. The pattern of similarity for sigma factor 2471 did not correlate with those of the cellulolytic system enzymes and sigma factor 2388 is likely a FecI homolog. It is possible that the sigma factors that are unique to *S. degradans* 2-40 could still play a role in carbohydrase gene regulation in this bacterium. It would be of interest to determine the patterns of sequence similarity for additional carbohydrases, such as those of the xylanolytic and pectinolytic systems, in order to assist in predicting which transcription factors are most likely to be involved in the regulation of these enzyme systems.

More future research would include using the bioinformatic methods employed in this study to search for high sequence similarity of other known gene regulators, such as LacI, AraC, LysR, OmpR, and LuxR and evaluate them for potential roles in carbohydrase gene expression. These regulators could be scored for presence or absence of high sequence similarity in cellulolytic and non-cellulolytic species, as was performed in this study, in order to evaluate their likelihood of being involved in carbohydrase gene regulation. The genes for these regulators could then be cloned into a plasmid vector, expressed in *S. degradans* 2-40, and the cell lysate could be assayed for carbohydrase activity. Determining the role of other regulators in the expression of carbohydrases in *S. degradans* 2-40 would allow researchers to

manipulate their expression and lead to a better understanding of the overall gene regulation and evolution of this bacterium.

Examining the role of carbon catabolite repression in *S. degradans* 2-40 would require a different approach due to its negative regulation in carbohydrase expression. Studying the role of carbon catabolite repression in *S. degradans* 2-40 could be accomplished by mutating or preventing the expression of *crp* and *crc* genes and observing carbohydrase activity. Because attempts at transposon mutagenesis and site-directed mutagenesis have not been successful in this bacterium, gene suppression could be used to prevent translation of *crp* and *crc* genes. High sequence similarities for CRP and Crc have already been found in this bacterium and may function prominently in the regulation of carbohydrase gene expression.

Further research could also include additional bioinformatic analyses on this organism to elucidate evolutionary relationships amongst its related species. Such analyses would work to establish the pathway by which this bacterium evolved from a pseudomonad ancestor to the distinct species *S. degradans* 2-40. Further in-depth bioinformatic analyses could also help reveal the source organisms from which *S. degradans* 2-40 acquired its carbohydrase genes, likely through horizontal gene transfer mechanisms. Determining the paths of vertical and horizontal gene transfer from which this organism acquired its genes would lead to a greater understanding of its phylogenetic relationships with other species of bacteria.

Appendices

Appendix 1 Primers Used for PCR amplification and Colony PCR

Name	Restriction Site	Nucleotide Sequence[a]
823_F	*Xba*I	ATACGCTCTAGACATGGTGGAACGATACTGGGA
823_R	*Hin*dIII	ATACGCAAGCTTATCAGTCATGATGATTGCCCC
2407_F	*Xba*I	ATACGCTCTAGATATCGGCCAAGCAAGGGTAAC
2407_R	*Hin*dIII	ATACGCAAGCTTTGCTAGCTGCTAACTGTAAGC
pMMB503EH_F		TCGGCTCGTATAATGTGTGG
pMMB503EH_R		GCCAGGCAAATTCTGTTTTATC

[a] Restriction sites underlined if added

Appendix 2 Primers Used for Quantitative RT-PCR

Name	5'Primer	3'Primer
cel5A	AACATTGCCGAAGCCATTAC	AGCGCGAATAGCAGGAATAA
cel5B	TATGCAGTCAGGCGAGTACG	CATCTACACTCAGCGCCAAA
cel5C	AGAAACCGAAGCACAAATGG	GTCCACAATAGCGAGGTGGT
cel5D	AGCTCGAGCACCAGCTCTAC	TCGCTTCTAGGGAGTTTCCA
cel5E	ATGGCGGCAGTTTTAATTTG	ATTCCAGCCACCAGTAATCG
cel5F	TATTCGCCTGCCTATTCACC	GTGTCGTACATGGGGTGTTG
cel5G	CCAAGGTCTTCCACCCTACA	AGAATGCGGTTGCCATCTAC
cel5H	CCTACTTGGTCGCAGGATGT	GTGCCCGCATAAAAGTGAAT
cel5I	CTCAAGGTGTGGGTGATGTG	GGAGGCTCTGCATCTAAACG
cel5J	CGGTATGAACGTAGGCTGGT	TGGCTGGCTACTCGGTAAGT
cel6A	TGTGATCCAAACGCACCTAA	TTTGGAATGCTTCTGGGAAC
cel9A	AAAGGGTTGGTACGATGCTG	GTACGCAGCCATTAGGGTGT
cel9B	GCTTGTGCGGTAACCAAAAT	CAGTAAGCTGTGGCGATTCA
gly5K	CAGAAGTACAAGCGGCTTCC	GTTTCGCCGTACTCACCAAT
gly5L	CGCAGGTATACCCGATCAGT	CAAGGTTAAAGCCTGCTTCG
gly5M	GGGCTTACAAACCGAAAACA	ACTCGCTAAGCGGTAGTGGA
gly5R	CGTGATGCGGTAGTTTGCTA	GAACGTTACGGCCAAACCTA
gly5S	GCAAAGGCGTAGACGAAAAG	ACCAAGCGTACACCCTTCAC
gly9C	GCAATGCGGCGTTAGATTAT	GCCGTAGCGAGAAGTTTTTG
folA	CGAGGGCTACGAAGATCAAG	ATAACTGCGCACCACCTACC
spoR	TGCCACTATGGTGGAAAACA	TGCAGCAGGTATGAGTTTGC
16S	CAGCTCGTGTCGTGAGATGT	CACCGGCAGTCTCCTTAGAG

Bibliography

Abdou, L., C. Boileau, P. de Philip, S. Pagès, H. P. Fiérobe & C. Tardif (2008) Transcriptional regulation of the Clostridium cellulolyticum cip-cel operon: a complex mechanism involving a catabolite-responsive element. *J Bacteriol,* 190**,** 1499-506.

Alva Munoz, L. E. & M. R. Riley (2008) Utilization of cellulosic waste from tequila bagasse and production of polyhydroxyalkanoate (PHA) bioplastics by Saccharophagus degradans. *Biotechnol Bioeng,* 100**,** 882-8.

Andrykovitch, G. & I. Marx (1988) Isolation of a new polysaccharide-digesting bacterium from a salt marsh. *Appl Environ Microbiol,* 54**,** 1061-2.

Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau & Y. Shao (1997) The complete genome sequence of Escherichia coli K-12. *Science,* 277**,** 1453-62.

Brooks, B. E. & S. K. Buchanan (2008) Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochim Biophys Acta,* 1778**,** 1930-45.

Carpita, N. C. (1996) STRUCTURE AND BIOGENESIS OF THE CELL WALLS OF GRASSES. *Annu Rev Plant Physiol Plant Mol Biol,* 47**,** 445-476.

Carpita, N. C. & D. M. Gibeaut (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J,* 3**,** 1-30.

Consortium, U. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res,* 39**,** D214-9.

Corporation, W. B. 1993. *Worthington Enzyme Manual*. Lakewood.

DeBoy, R. T., E. F. Mongodin, D. E. Fouts, L. E. Tailford, H. Khouri, J. B. Emerson, Y. Mohamoud, K. Watkins, B. Henrissat, H. J. Gilbert & K. E. Nelson (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus. *J Bacteriol,* 190**,** 5455-63.

Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol,* 11**,** 87-93.

Ekborg, N. A., J. M. Gonzalez, M. B. Howard, L. E. Taylor, S. W. Hutcheson & R. M. Weiner (2005) Saccharophagus degradans gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides. *Int J Syst Evol Microbiol,* 55**,** 1545-9.

Ekborg, N. A., L. E. Taylor, A. G. Longmire, B. Henrissat, R. M. Weiner & S. W. Hutcheson (2006) Genomic and proteomic analyses of the agarolytic system expressed by Saccharophagus degradans 2-40. *Appl Environ Microbiol,* 72**,** 3396-405.

Ensor, L., S. Stosz & R. Weiner (1999) Expression of multiple complex polysaccharide-degrading enzyme systems by marine bacterium strain 2-40. *J Ind Microbiol Biotechnol,* 23**,** 123-126.

Fuqua, C., S. C. Winans & E. P. Greenberg (1996) Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annu Rev Microbiol,* 50**,** 727-51.

Fuqua, W. C., S. C. Winans & E. P. Greenberg (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol,* 176**,** 269-75.

Gallegos, M. T., R. Schleif, A. Bairoch, K. Hofmann & J. L. Ramos (1997) Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev,* 61**,** 393-410.

Gama-Castro, S., H. Salgado, M. Peralta-Gil, A. Santos-Zavaleta, L. Muniz-Rascado, H. Solano-Lira, V. Jimenez-Jacinto, V. Weiss, J. S. Garcia-Sotelo, A. Lopez-Fuentes, L. Porron-Sotelo, S. Alquicira-Hernandez, A. Medina-Rivera, I. Martinez-Flores, K. Alquicira-Hernandez, R. Martinez-Adame, C. Bonavides-Martinez, J. Miranda-Rios, A. M. Huerta, A. Mendoza-Vargas, L. Collado-Torres, B. Taboada, L. Vega-Alvarado, M. Olvera, L. Olvera, R. Grande, E. Morett & J. Collado-Vides (2010) **RegulonDB (version 7.0): transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units)**. *Nucleic Acids Research*.

Gao, R. & A. M. Stock (2009) Biological insights from structures of two-component proteins. *Annu Rev Microbiol,* 63**,** 133-54.

Ghose, T. K. (1987) Measurement of Cellulase Activities. *Pure and Applied Chemistry,* 59**,** 256-268.

Ghose, T. K. & V. S. Bisaria (1987) Measurement of Hemicellulase Activities Part 1: Xylanases. *Pure and Applied Chemistry,* 59**,** 1739-1752.

Godden, B., T. Legon, P. Helvenstein & M. Penninckx (1989) Regulation of the production of hemicellulolytic and cellulolytic enzymes by a Streptomyces sp. growing on lignocellulose. *J Gen Microbiol,* 135**,** 285-92.

González, J. M. & R. M. Weiner (2000) Phylogenetic characterization of marine bacterium strain 2-40, a degrader of complex polysaccharides. *Int J Syst Evol Microbiol,* 50 Pt 2**,** 831-4.

Görke, B. & J. Stülke (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol,* 6**,** 613-24.

Han, S. O., H. Y. Cho, H. Yukawa, M. Inui & R. H. Doi (2004) Regulation of expression of cellulosomes and noncellulosomal (hemi)cellulolytic enzymes in Clostridium cellulovorans during growth on different carbon sources. *J Bacteriol,* 186**,** 4218-27.

Harris, S. J., Y. L. Shih, S. D. Bentley & G. P. Salmond (1998) The hexA gene of Erwinia carotovora encodes a LysR homologue and regulates motility and the expression of multiple virulence determinants. *Mol Microbiol,* 28**,** 705-17.

Herrera-Herrera, J. A., O. Pérez-Avalos, L. M. Salgado & T. Ponce-Noyola (2009) Cyclic AMP regulates the biosynthesis of cellobiohydrolase in Cellulomonas flavigena growing in sugar cane bagasse. *Arch Microbiol,* 191**,** 745-50.

Howard, M. B., N. A. Ekborg, L. E. Taylor, R. M. Weiner & S. W. Hutcheson (2003) Genomic analysis and initial characterization of the chitinolytic system of Microbulbifer degradans strain 2-40. *J Bacteriol,* 185**,** 3352-60.

Hugouvieux-Cotte-Pattat, N., G. Condemine, W. Nasser & S. Reverchon (1996) Regulation of pectinolysis in Erwinia chrysanthemi. *Annu Rev Microbiol,* 50**,** 213-57.

Hutcheson, S. W., H. Zhang & M. Suvorov (2011) Carbohydrase Systems Degrading Polymers of Marine Origin in *Saccharophagus degradans*. *Marine Drugs,* 9.

Itou, H. & I. Tanaka (2001) The OmpR-family of proteins: insight into the tertiary structure and functions of two-component regulator proteins. *J Biochem,* 129**,** 343-50.

Jahn, C. E., D. K. Willis & A. O. Charkowski (2008) The flagellar sigma factor fliA is required for Dickeya dadantii virulence. *Mol Plant Microbe Interact,* 21**,** 1431-42.

Jain, E., A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B. E. Suzek, M. J. Martin, P. McGarvey & E. Gasteiger (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics,* 10**,** 136.

Jones, S., B. Yu, N. J. Bainton, M. Birdsall, B. W. Bycroft, S. R. Chhabra, A. J. Cox, P. Golby, P. J. Reeves & S. Stephens (1993) The lux autoinducer regulates the production of exoenzyme virulence determinants in Erwinia carotovora and Pseudomonas aeruginosa. *EMBO J,* 12**,** 2477-82.

Lewis, M. (2005) The lac repressor. *C R Biol,* 328**,** 521-48.

Maddocks, S. E. & P. C. Oyston (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology,* 154**,** 3609-23.

Nataf, Y., L. Bahari, H. Kahel-Raifer, I. Borovok, R. Lamed, E. A. Bayer, A. L. Sonenshein & Y. Shoham (2010) Clostridium thermocellum cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc Natl Acad Sci U S A,* 107**,** 18646-51.

Paget, M. S. & J. D. Helmann (2003) The sigma70 family of sigma factors. *Genome Biol,* 4**,** 203.

Pompeani, A. J., J. J. Irgon, M. F. Berger, M. L. Bulyk, N. S. Wingreen & B. L. Bassler (2008) The Vibrio harveyi master quorum-sensing regulator, LuxR, a TetR-type protein is both an activator and a repressor: DNA recognition and binding specificity at target promoters. *Mol Microbiol,* 70**,** 76-88.

Potvin, E., F. Sanschagrin & R. C. Levesque (2008) Sigma factors in Pseudomonas aeruginosa. *FEMS Microbiol Rev,* 32**,** 38-55.

Salgado, H., G. Moreno-Hagelsieb, T. F. Smith & J. Collado-Vides (2000) Operons in Escherichia coli: genomic analyses and predictions. *Proc Natl Acad Sci U S A,* 97**,** 6652-7.

Schell, M. A. (1993) Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol,* 47**,** 597-626.

Schleif, R. (2010) AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS Microbiol Rev,* 34**,** 779-96.

Shin, M. H., d. Y. Lee, K. Skogerson, G. Wohlgemuth, I. G. Choi, O. Fiehn & K. H. Kim (2010) Global metabolic profiling of plant cell wall polysaccharide degradation by Saccharophagus degradans. *Biotechnol Bioeng,* 105**,** 477-88.

Spiridonov, N. A. & D. B. Wilson (1999) Characterization and cloning of celR, a transcriptional regulator of cellulase genes from Thermomonospora fusca. *J Biol Chem,* 274**,** 13127-32.

Swint-Kruse, L. & K. S. Matthews (2009) Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol,* 12**,** 129-37.

Taylor, L. E., B. Henrissat, P. M. Coutinho, N. A. Ekborg, S. W. Hutcheson & R. M. Weiner (2006) Complete cellulase system in the marine bacterium Saccharophagus degradans strain 2-40T. *J Bacteriol,* 188**,** 3849-61.

van Wyk, J. P. (2001) Biotechnology and the utilization of biowaste as a resource for bioproduct development. *Trends Biotechnol,* 19**,** 172-7.

Warren, R. A. (1996) Microbial hydrolysis of polysaccharides. *Annu Rev Microbiol,* 50**,** 183-212.

Watson, B. J., H. Zhang, A. G. Longmire, Y. H. Moon & S. W. Hutcheson (2009) Processive endoglucanases mediate degradation of cellulose by Saccharophagus degradans. *J Bacteriol,* 191**,** 5697-705.

Weiner, R. M., L. E. Taylor, B. Henrissat, L. Hauser, M. Land, P. M. Coutinho, C. Rancurel, E. H. Saunders, A. G. Longmire, H. Zhang, E. A. Bayer, H. J. Gilbert, F. Larimer, I. B. Zhulin, N. A. Ekborg, R. Lamed, P. M. Richardson, I. Borovok & S. Hutcheson (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, Saccharophagus degradans strain 2-40 T. *PLoS Genet,* 4**,** e1000087.

Wilson, C. J., H. Zhan, L. Swint-Kruse & K. S. Matthews (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci,* 64**,** 3-16.

Zhang, H. & S. W. Hutcheson (2011) Complex expression of the cellulolytic transcriptome of Saccharophagus degradans.

Zhang, H., Y. H. Moon, B. J. Watson, M. Suvorov, E. Santos, C. A. Sinnott & S. W. Hutcheson (2011) Hydrolytic and phosphorolytic metabolism of cellobiose by the marine aerobic bacterium Saccharophagus degradans 2-40(T). *J Ind Microbiol Biotechnol*.