ABSTRACT

Title of Dissertation: DEGRADATION OF PLANT CELL WALL POLYSACCHARIDES BY *SACCHAROPHAGUS DEGRADANS*

Larry E. Taylor II, Ph.D., 2005

Dissertation Directed By: Professor Ronald M. Weiner, Department of Marine and Estuarine Environmental Sciences

Saccharophagus degradans is an aerobic, Gram negative marine bacterium, isolated from decaying *Spartina alterniflora* in the Chesapeake Bay watershed. *S. degradans* can degrade and metabolize numerous complex polysaccharides, including the major components of the plant cell wall, cellulose, xylan and β-glucan. Genomic analyses reveal that *S. degradans* has 77 genes coding for enzymes that are predicted to participate in the degradation of plant cell wall polysaccharides. These include complete, functional, multienzyme systems for the depolymerization of cellulose, xylan, arabinan, β-1,3-glucans, β-1,4-glucans, and mannan. Most of the cellulases are modular, some of which contain novel combinations of catalytic and/or substrate binding modules. In addition to its well-predicted plant wall degrading systems, *S. degradans* encodes 19 proteins which contain a carbohydrate binding module, but lack an identifiable catalytic domain and 12 glycanases for which function cannot be predicted by sequence analysis. Many of the plant wall degrading enzymes contain lipoprotein signature sequences, indicating that they are likely attached to the cell surface, thereby maintaining their reactions near the cell and preventing loss of

enzyme or product to diffusion or competition. *S. degradans* is capable of using crystalline cellulose and intact plant matter as sole carbon and energy sources. Cellulose induces catalytic activity against all major plant cell wall polymers, suggesting a complex mechanism for coordinating the regulation of these multienzyme systems. In addition to its abundant carbohydrases, *S. degradans* encodes seven proteins with predicted molecular weights over 250,000 Daltons, one of which, CabA at 1,500,000 Daltons, is the largest known bacterial protein to date. These proteins contain calcium-binding repeat sequences suggesting a role in cell-to-surface adhesion or protein-to-protein interaction, perhaps as a means of surface-enzyme attachment. These studies establish *S. degradans* as the first marine bacterium with a complete and functional cellulase system with the further ability to degrade plants in monoculture.

DEGRADATION OF PLANT CELL WALL POLYSACCHARIDES BY *SACCHAROPHAGUS DEGRADANS*

By

Larry Edmund Taylor II

Dissertation 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
Doctor of Philosophy
2005

Advisory Committee:
Professor Ronald M Weiner, Chair
Professor Sammy Joseph (Dean's Representative)
Professor Richard Stewart
Professor Jocylene DiRuggiero
Professor Wenxia Song

Dedication

This Dissertation is dedicated to my Grandfather, Duane B. Williams, who taught me the crucial lessons of life through his caring words, and more importantly, by his daily example. Granddad, I will always be grateful to you for teaching me to value honesty, integrity, responsibility, and diligence. Thank you for teaching me what a man should be. I could not have accomplished this degree, or anything else worthwhile in my life, without the principles you instilled in me.

Acknowledgements

I would like to thank my mentor, Dr. Ronald Weiner, for his kindness, support and thoughtful guidance throughout this long and rigorous process. His boundless enthusiasm and total love for science have carried me through the more difficult stretches of my graduate education, influencing me, and my work, more than mere words can express. I would also like to thank my dissertation committee: Dr. Joseph, Dr. Stewart, Dr. Song, and Dr. DiRuggiero. Their suggestions have improved my work greatly and helped me to view my results from different perspectives.

I also wish to acknowledge some of the other people who have made my time at Maryland more enjoyable: Ruth Mausolf and Nancy Williams, thank you for always helping me out with such details as purchase orders and scheduling rooms for meetings, etc.; I also thank you both for your many kind words and deeds. Thank you to my former lab-mates, Dr. Nate Ekborg (Let's mix it up, Nate!), Dr. Ahmed AbdelHafez (Man, you've gotta take it easy!), and Christian DeRezende (¿Que?) for the fun times in the lab and for providing those moments of humor and insanity I needed to get through the tough times.

I also thank Tom Dunn (The pride of Carroll County) and John Russ (Go Pack!) for helping this displaced country boy to feel a bit more at home here. There aren't many folks like us in the DC suburbs, so it was great to find you guys. I would be wrong if I didn't mention my favorite members of the Stein lab: Ellen O'Connor (the soon to be Dr. Miss CBMG 2005), and Sam Bish (Mr. Tagliabue). Thank you both for your

friendship over the years. Whether complaining about failed experiments or having hallway discussions about the intricacies of fantasy football or just talking about life in general, you each greatly improved my time at Maryland.

Finally, I would like to thank my dear family: my Mother, my Grandparents, my Uncle, and my wonderful fiancé, Dr Karin Schmidova' M.D. (a "real" doctor), for all of their love, patience, and understanding. Without all of you, and your support, I would never have been able to realize this achievement. Thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA, or aa	amino acid
AZCL	azurine cross-linked
BCA	bicinchoninic acid (protein assay)
CAZy	Carbohydrate-active enzyme webserver/database
Cbh	cellobiohydrolase
CBM	carbohydrate binding module
CE	carbohydrate esterase
Cel	cellulase
CMC	carboxymethyl cellulose
CP	complex polysaccharide
DNSA	3,5-dinitrosalicylic acid
DTT	1,4-Dithiothreitol (Cleland's Reagent)
EDTA	ethylenediaminetetraacetic acid
FCL	fucose-binding lectin
FN3	fibronectin type 3 domain
GC	guanosine + cytosine content
GH	glycoside hydrolase
Gly	glycanase of uncertain specificity
GT	glycosyltransferase
His	histidine
HRP	horseradish peroxidase
IPTG	Isopropyl-beta-D-thiogalactopyranoside
JGI	Joint Genome Institute
KB	kilobases
kDa	kiloDaltons
LB	Luria-Bertani
LPB	Lipopotein box

MA	marine agar
MM	minimal medium
MW	molecular weight
MWCO	molecular weight cutoff
MS	mass spectrometry
NTA	nitrilotriacetic acid
ORF	open reading frame
ORNL	Oak Ridge National Laboratory
PAGE	polyacrylamide gel electrophoresis
PASC	phosphoric acid swollen cellulose
PCR	polymerase chain reaction
PIPES	Piperazine-1,4-bis(2-ethanesulfonic acid)
PKD	polycystic kidney disease domain
PL	polysaccharide lyase
PSL	polyserine linker
PTS	phosphotransferase system
SDS	sodium dodecylsulfate
SEM	scanning electron microscopy
SSS	polyserine linker
TEM	transmission electron microscopy
TMR	transmembrane region
Tris	tris (hydroxymethyl) aminomethane
TSP3	thrombospondin type 3 repeat
UDP	uridine diphosphate
VPC	viable plate count

Chapter 1: General Introduction

Complex polysaccharides in Marine Environments

In marine environments relatively little is known about how the annual production of over 25 billion tons of complex polysaccharides (CP) is recycled to usable carbon (92). These CP, including agar, chitin, alginate and cellulose, function as structural and/or energy-storage polymers in planktonic organisms, algal blooms, benthic invertebrates, as well as terrestrial and aquatic plants. Most of these CP are insoluble and relatively recalcitrant, requiring dedicated multienzyme systems for their degradation (10, 71). Marine bacteria are known to mediate the degradation of chitin (7, 118) and the so-called "marine snow" (1) that occurs in the pelagic zone. Recent studies have suggested that pelagic bacteria play a much larger role in recycling planktonic detritus to usable carbon than was thought previously (5, 17). While these findings highlight the important role of marine bacteria in the marine carbon cycle, much work is needed to identify the specific organisms and enzymatic processes involved (17).

Even less is known about the turnover of higher plant material within marine environments. Although bacterial decomposition of plant matter has been shown to be an important process in coastal nutrient cycling (79), the organisms involved and other details of the process are largely unknown. Answers to some of these questions may be found in a recently discovered group of marine bacteria that recycle complex polysaccharides (34, 39). One representative of this emerging group of degradative marine bacteria, *Saccharophagus degradans*, may prove to be a missing link in the

oceanic carbon cycle, being able to degrade at least 10 distinct types of complex polysaccharides which are present in coastal and marine environments.

Saccharophagus degradans

The marine bacterium originally designated strain 2-40 (2-40) was isolated in the mid 1980's from decaying salt marsh cord grass *Spartina alterniflora* in the Chesapeake Bay watershed (3). It is a pleomorphic, Gram-negative, aerobic, motile bacterium, which can degrade at least 10 different complex polysaccharides (CP), including agar, chitin, alginic acid, carboxymethylcellulose (CMC), β-glucan, laminarin, pectin, pullulan, starch and xylan (41). Strain 2-40 is regarded as a marine bacterium based on the following criteria: it has an absolute requirement for Sea salts, growing best at concentrations ranging from 2.5% to 3.5%, and it cannot tolerate concentrations below 1.1%. Furthermore, 16S rDNA analysis shows that 2-40 is a member of the gamma-subclass of the phylum *Proteobacteria*, related to such marine isolates as *Microbulbifer hydrolyticus* (43) and to *Teridinibacter sp.,*(34) cellulolytic nitrogen-fixing bacteria that are symbionts of shipworms. The proposed name of strain 2-40 was *Microbulbifer degradans* strain 2-40 (43), the name under which much of our group's work has been published. Recently, the taxonomy of strain 2-40 has been clarified with its placement in the novel genus, *Saccharophagus* (39). As such, *Saccharophagus degradans* forms a third genus in the newly emerging *Microbulbifer* / *Teridinibacter* / *Saccharophagus* group of marine carbohydrate degrading bacteria (39). Of the 23 bacteria in this group, only the genome of *S. degradans* has been sequenced.

The agarase, chitinase and alginase systems have been biochemically characterized. Zymograms indicate that all three systems are comprised of multiple depolymerases (21, 101, 112) and reducing-sugar assays reveal that their activities primarily reside with the cell fraction during logarithmic growth, while in later growth phases most the activity is found in the supernatant and cell-bound activity decreases dramatically (41, 101, 114). Immunoelectron microscopy probes suggest that some of the CP depolymerases are exocellular (112).

Because of its unique metabolic capabilities and the potential significance of its role in the marine carbon cycle, *S. degradans* was selected for genome sequencing by the United States Department of Energy's Joint Genome Institute (US DOE/JGI). The finished assembly of the genome, based on 20X sequence coverage, comprises 5.05MB in a single, circular chromosome. Automated annotations by the computational genomics division of the Oak Ridge National Laboratory (ORNL) are available on the World Wide Web at http://genome.ornl.gov/microbial/mdeg/.

The ORNL annotations and our own research team's analyses (N. Ekborg, M. Howard, S. Hutcheson, L. Taylor, R. Weiner. unpublished results) reveal that *S. degradans* encodes an unusually large number (>180) of enzymes for the degradation of many different CP. Among these are number of agarases, alginases and chitinases, correlating well with previous studies of these systems. Remarkably, the genome also contains an abundance of enzymes with predicted roles in the degradation of plant cell wall polymers, including a number of ORFs with homology to cellulases, xylanases, pectinases, and other glucanases and glucosidases. The carbohydrase genes of *S. degradans* occur throughout the genome and do not appear to be organized into operons;

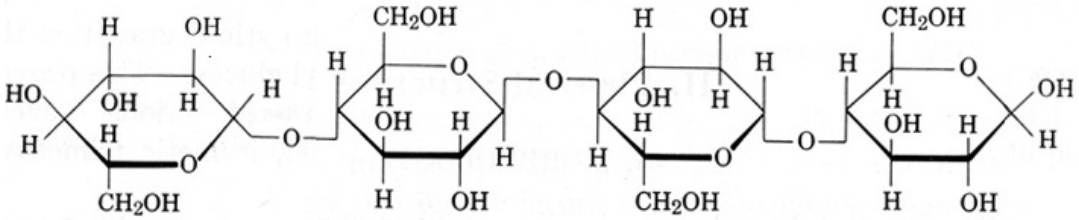
in the instances where polysaccharidase ORFs are clustered in the genome they often have different predicted functions. Given its extraordinary complement of carbohydrate degrading enzymes, it appears that *S. degradans* may play a significant role in the marine carbon cycle, mediating the breakdown and turnover of carbon sequestered in the complex polysaccharides of algae, invertebrates, and even plants in estuarine and coastal environments.

Characteristics of the Plant Cell Wall

The plant cell wall is both chemically and structurally complex, containing cellulose, hemicellulose (including xylans, arabinans and mannans), and pectin (20, 111). Cellulose is a homopolymer of β-1,4 linked D-glucose (Figure 1-1). Cellulose microfibrils act as the structural backbone of the plant cell wall. During cellulose biosynthesis, chains of poly-β-1,4-D-glucose self associate through hydrogen bonding and hydrophobic interactions to form cellulose microfibrils which further self-associate to form larger fibrils. Cellulose microfibrils are somewhat irregular and contain regions of varying crystallinity. The degree of crystallinity of cellulose fibrils depends on how tightly ordered the hydrogen bonding is between its component cellulose chains. Areas with less-ordered bonding, and therefore more accessible glucose chains, are referred to as amorphous regions (Figure 1-2). The relative crystallinity and fibril diameter are characteristic of the biological source of the cellulose (12, 73, 106). The irregularity of cellulose fibrils results in a great variety of altered bond angles and steric effects which hinder enzymatic access and subsequent degradation. Because of this variability,

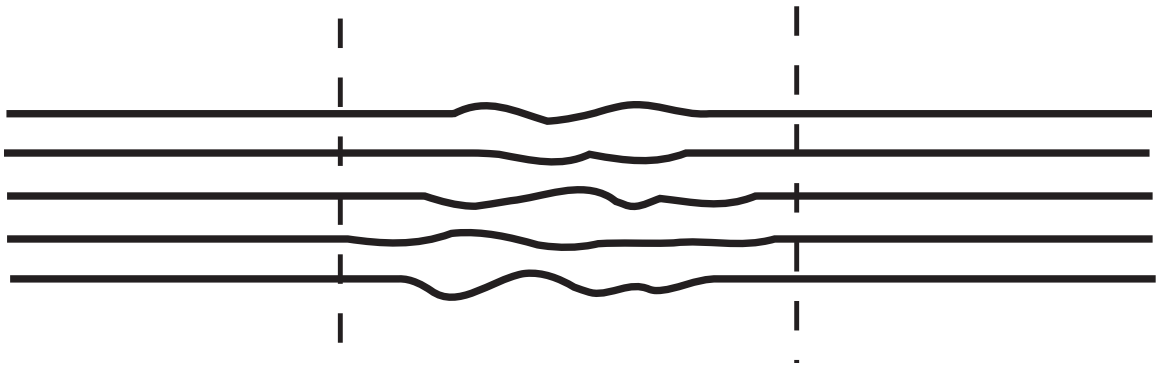
cellulose degradation requires a variety of enzymes, presumably with wide variations in the shape of the substrate-binding pockets and/or active sites (111).

Figure 1-1. Chemical structure of cellulose. The chemical structure of cellulose is shown with reducing and non-reducing sugar ends indicated. Natural cellulose chains are often many hundreds or a few thousand residues long.



Cellulose

Figure 1-2. Schematic representation of crystalline and amorphous cellulose. Crystalline cellulose consists of tightly packed cellulose chains with highly ordered cross-chain hydrogen bonding. In amorphous regions the chains are more disordered, and therefore more susceptible to enzymatic attack.



crystalline amorphous crystalline

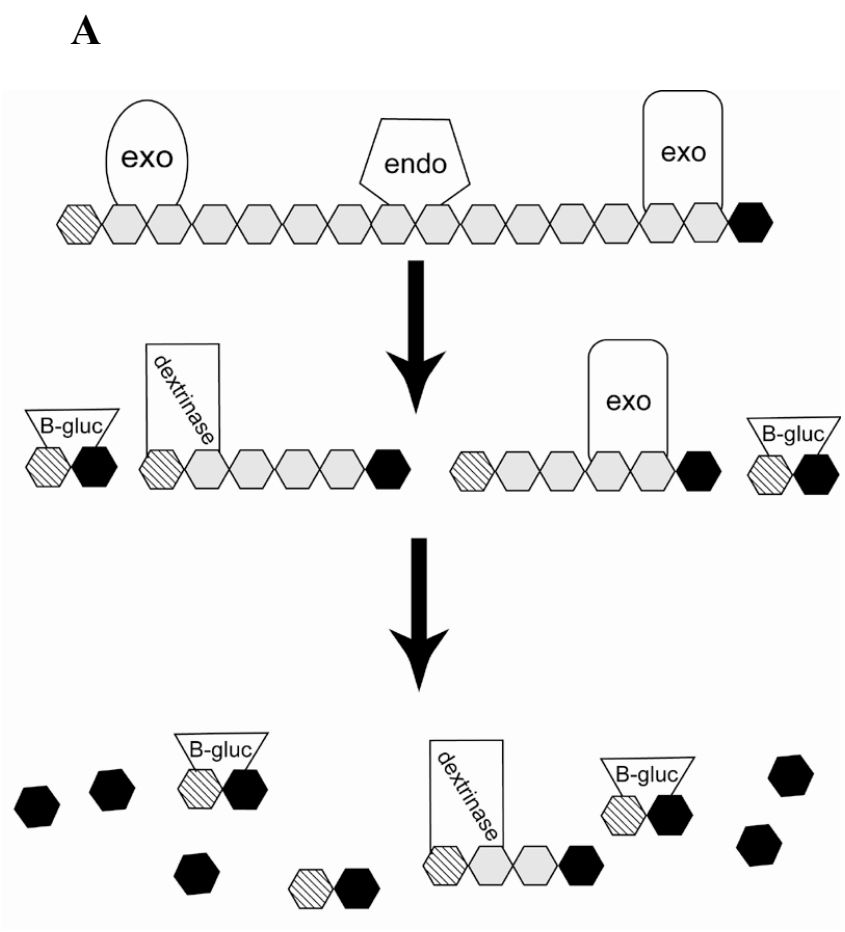
Cellulose degradation

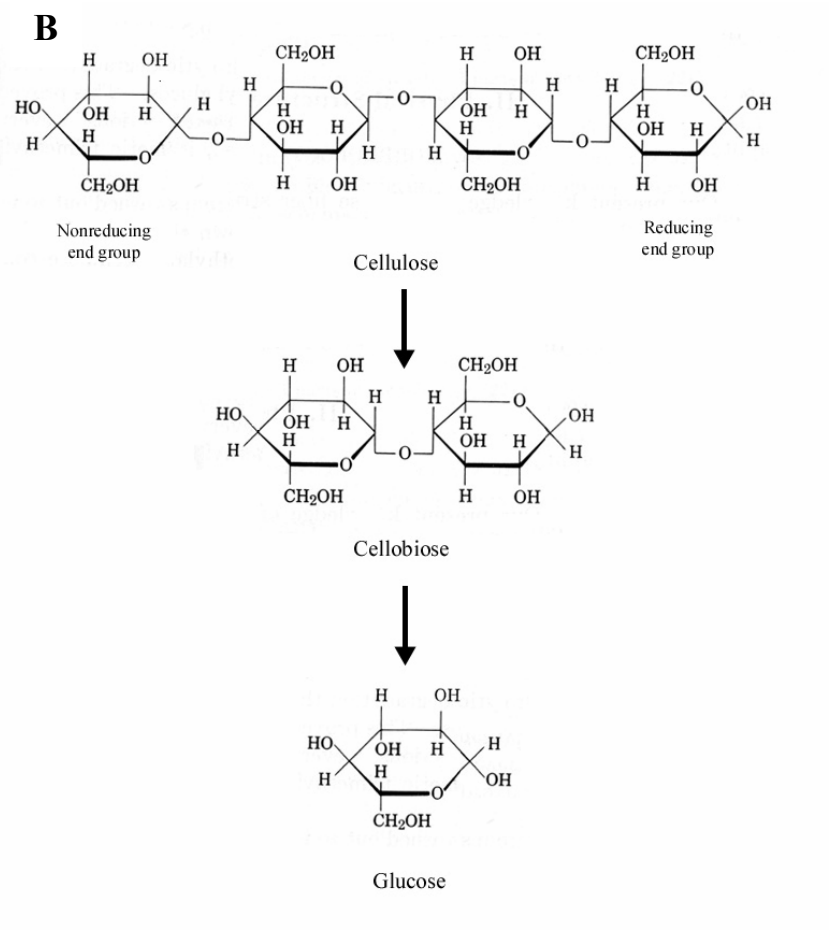
Despite its chemical simplicity, cellulose is one of the most recalcitrant complex polysaccharides (CP). It is the complex tertiary structure of cellulose that imparts its resistance to enzymatic attack. The enzymatic degradation of cellulose in terrestrial habitats is among the best-studied microbial processes, as evidenced by the numerous reviews on the subject (8, 10, 71, 73, 98, 106, 111). Cellulase systems are often referred to as "complete", meaning that they provide all enzymatic activities needed to degrade cellulose to glucose, or "incomplete". Organisms with incomplete cellulase systems can mediate some of the steps in cellulose degradation, but in order to mediate full cellulose degradation they must act in a consortium with other organisms which provide the missing capabilities.

The general model for cellulose depolymerization to glucose involves three distinct enzymatic activities (Figure 1-3). Endoglucanases cleave cellulose chains internally to generate shorter chains and increase the number of accessible ends, which are acted upon by exoglucanases. These exoglucanases are specific for either reducing ends or non-reducing ends and frequently liberate cellobiose, the dimer of cellulose (cellobiohydrolases). The accumulating cellobiose is cleaved to glucose by cellobiases (β-1,4-glucosidases). In many systems an additional type of enzyme is present: cellodextrinases are β-1,4-glucosidases which cleave glucose monomers from cellulose oligomers, but not from cellobiose.

Figure 1-3. Schematic representations of cellulose degradation.

A) Representation of enzymatic synergism between cellobiohydrolases (exoglucanases) and endoglucanases. Endoglucanases attack amorphous regions thereby generating shorter cellulose chains and increasing the number of available ends for cellobiohydrolases. The resulting cellulose oligomers (cellodextrins) are processed by cellodextrinases and β-glucosidases. Solid squares represent reducing ends and open squares represent non-reducing ends. B) Chemical representation of cellulose degradation to cellobiose and glucose. Reducing and non-reducing ends are indicated.


A
exo
endo
exo
B-gluc
dextrinase
exo
B-gluc
B-gluc
dextrinase
B-gluc


B
Nonreducing end group
Reducing end group
Cellulose
Cellobiose
Glucose

Because of the variable crystallinity and structural complexity of cellulose, and the enzymatic activities required for is degradation, organisms with "complete" cellulase systems synthesize a large number of endo and/or exo-acting β-1,4-glucanases. For example, *Cellulomonas fimi* and *Thermomonospora fusca* have each been shown to synthesize six cellulases while *Clostridium thermocellum* has as many as 15 or more (106). Presumably, the variations in the shape of the substrate-binding pockets and/or active sites of these numerous cellulases facilitate complete cellulose degradation (111). Organisms with complete cellulase systems are believed to be capable of efficiently using plant biomass as a carbon and energy source while mediating cellulose degradation. The ecological and evolutionary role of incomplete cellulose systems is less clear, although it is believed that many of these function as members of consortia (such as ruminal communities) which may collectively achieve total or near-total cellulose hydrolysis (71, 106).

Classification and nomenclature of Carbohydrate-Active Enzymes (CAZymes)

Enzymes have traditionally been classified by substrate specificity and reaction products. In the pre-genomic era, function was regarded as the most amenable (and perhaps most useful) basis for comparing enzymes and assays for various enzymatic activities have been well-developed for many years, resulting in the familiar EC classification scheme. Cellulases and other *O*-Glycosyl hydrolases, which act upon glycosidic bonds between two carbohydrate moieties (or a carbohydrate and non-carbohydrate moiety—as occurs in nitrophenol-glycoside derivatives) are designated as

EC 3.2.1.-, with the final number indicating the exact type of bond cleaved. According to this scheme an endo-acting cellulase (1,4-β-endoglucanase) is designated EC 3.2.1.4.

With the advent of widespread genome sequencing, ever-increasing amounts of sequence data have facilitated analyses and comparison of related genes and proteins on an unprecedented scale. This is particularly true for carbohydrases; it has become clear that classification of such enzymes according to reaction specificity, as is seen in the E.C. nomenclature scheme, is limited by the inability to convey sequence similarity. Furthermore, a growing number of carbohydrases have been crystallized and their 3-D structures solved. One of the major revelations of carbohydrase sequence and structure analyses is that there are discreet families of enzymes with related sequence, that contain conserved three-dimensional folds which can be predicted based on their amino acid sequence. Further, it has been shown that enzymes with the same three-dimensional fold exhibit the same stereospecificity of hydrolysis, even when they catalyze different reactions (27, 54). These findings form the basis of a sequence-based classification of carbohydrase modules which is available in the form of an internet database, the Carbohydrate-Active enZYme server (CAZy), at http://afmb.cnrs-mrs.fr/CAZY/index.html (26, 27).

CAZy defines four major classes of carbohydrases, based on the type of reaction catalyzed: Glycosyl Hydrolases (GH's), Glycosyltansferases (GT's), Polysaccharide Lyases (PL's), and Carbohydrate Esterases (CE's). GH's cleave glycosidic bonds through hydrolysis. This class includes many familiar polysaccharidases such as cellulases, xylanases, and agarases. GT's generally function in polysaccharide synthesis, catalyzing the formation of new glycosidic bonds through the transfer of a sugar molecule from an

activated carrier molecule, such as uridine diphosphate (UDP), to an acceptor molecule. While GT's often function in biosynthesis, there are examples where the mechanism is exploited for bond cleavage, as occurs in the phosphorolytic cleavage of cellobiose and cellodextrins (72). PL's utilize a β-elimination mechanism to mediate bond cleavage and are commonly involved in alginate and pectin depolymerization. CE's generally are deacetylases acting on *O*- or *N*- substituted polysaccharides. Common examples include xylan and chitin deacetylases.

Sequence-based families are designated by number within each class, as is seen with GH5: glycosyl hydrolase family 5. Members of GH5 hydrolyze β-1,4 bonds in a retaining fashion, using a double-displacement mechanism which results in retention of the original bond stereospecificity. Retention or inversion of anomeric configuration is a general characteristic of a given GH family (26, 52). Many examples of endocellulases, xylanases and mannanases belonging to GH5 have been reported, illustrating the variety of substrate specificity possible within a GH family. Also, GH5s are predominantly endohydrolases—cleaving chains of their respective substrates at random locations internal to the polymer chains. While true for GH5, this generalization does not hold for many other GH families. In addition to carbohydrases, the CAZy server defines numerous families of Carbohydrate Binding Modules (CBM). As with catalytic modules, CBM families are designated based on amino acid sequence similarity and conserved three-dimensional folds.

The CAZyme structural families have been incorporated into a new classification and nomenclature scheme, developed by Bernard Henrissat and colleagues (54). Traditional gene/protein nomenclature assigns an acronym indicating general function

and order of discovery; in this scheme an organism's cellulase genes are designated celA, celB, etc., regardless of their actual mechanism of action on cellulose. Some researchers have attempted to convey more information by naming cellulases as endoglucanases (engA, engB) or cellobiohydrolases (cbhA, cbhB), however this requires determination of function *in vitro* and still fails to convey relatedness of protein sequence and structure. CAZyme nomenclature retains the familiar acronym to indicate the functional system a gene belongs to and incorporates the family number designation. Capital letters after the family number indicate the order of report within a given organism system. An example is provided by two endoglucanases, CenA and CenB, of *Cellulomonas fimi*. In the old nomenclature nothing can be deduced from the names except order of discovery. Naming them Cel6A and Cel9A, respectively, makes it immediately clear that these two cellulases are unrelated in sequence, and so belong to different GH families (where Cel stands for cellulase, and 9 for glycosyl hydrolase family nine).

While this scheme does not distinguish between endo- and exo- activity, these designations are not absolute and can be included in discussion of an enzyme when relevant (i.e. the cellobiohydrolase Cel6A, the endoxylanase Xyn10B). Catalytic modules take precedence in naming carbohydrases; since many (or even most) carbohydrases contain at least one CBM, they are named for their enzymatic module. If more than one catalytic domain is present, they are named in order from N-terminus to C-terminus, i.e. cel9A-cel48A contains a GH9 at the amino-terminus and a GH48 at the carboxy-terminus. Both domains act against cellulose. There are, however, many examples of CBM modules occurring on proteins with no predicted carbohydrase module. In the absence of some other predicted functional domain (like a protease) these proteins are

named for the CBM module family. If there are multiple CBM families present, then naming is again from amino to carboxyl end, i.e. cbm2D-cbm10A (54). This nomenclature has been widely accepted and is used in the naming of all plant-wall active carbohydrases and related proteins considered as part of this study.

Statement of Purpose

The microbial degradation of plant cell walls in the marine environment is not well characterized. In fact, surprisingly little is known about how marine biomass is recycled to useable carbon. While the degradation of complex polysaccharides is generally mediated by consortia of microorganisms, analysis of the recently finished genome sequence of *S. degradans* suggests that it has the capabilities required to metabolize most or all major components of the plant cell wall in monoculture. This work was undertaken to test the following hypotheses: that *S. degradans* has a complete and functional cellulase system, that the cellulase system is complemented by additional systems for the degradation of plant polysaccharides, and that the genomically identified plant wall carbohydrase systems are coordinately regulated. In an extension of these, it is further hypothesized that *S. degradans* strain 2-40 is capable of growth on plant material as sole carbon and energy sources.

Chapter 2: Plant Wall Active Carbohydrases and Auxiliary Proteins of Saccharophagus degradans strain 2-40^T

Introduction

Previous studies had shown that *S. degradans* can grow on sole carbon source CMC, β-glucan, laminarin and xylan, among others (41). Although *S. degradans* degraded CMC, consisting of cellulose chains to which carboxymethyl side chains have been added to impart solubility, it still was not known whether it degraded "real world" cellulose (R. Weiner, personal communication) (106, 111), let alone thought to be capable of degrading intact plant material. This all changed with the sequencing of its genome.

Our research team's analyses (L. Taylor, N. Ekborg, M. Howard, S. Hutcheson, R. Weiner, unpublished results) of the genome sequence and of the annotations by ORNL revealed a plethora of carbohydrases, including agarases, alginases and chitinases—correlating well with the findings of previous studies of these systems (41, 57). Upon closer investigation, it became apparent that the genome also contains an abundance of open reading frames with homology to cellulases, xylanases, pectinases, glucanases and glucosidases. In all, over 180 open reading frames with a probable role in carbohydrate catabolism were identified in the draft genome. The deduced metabolic pathways illustrated on the *S. degradans* genome sequence website show that these predicted carbohydrases are complemented by the metabolic pathways required to utilize hexose and pentose monosaccharides, including a complete TCA cycle, the Embden-Meyerhof pathway, and the pentose phosphate pathway.

To begin to define the cellulase, xylanase and pectinase systems of *S. degradans*, genes were initially classified as belonging to one of those systems by BLAST homology. Ambiguous ORFs were tentatively assigned to the class of the best known hit. Other tools used to refine this tentative classification include Pfam (Protein families database of alignments and HMMs; http://www.sanger.ac.uk/Software/Pfam/) and SMART (Simple Modular Architecture Research Tool; http://smart.embl-heidelberg.de/) which use multiple alignments and hidden Markov models (statistical models of sequence consensus homology) to identify discreet modular domains within a protein sequence. These analyses were largely successful; however, a number of ORFs remained difficult to classify using the analysis tools that are freely available on the World Wide Web.

In order to obtain a more sophisticated domain analysis of *S. degradans* carbohydrases, a collaboration was initiated with an expert on the functional genomics of carbohydrase catalytic and accessory modules, Bernard Henrissat of the Architecture et Fonction des Macromolecules Biologques laboratory at the Centre National de la Recherche Scientifique (AFMB-CNRS), Marseille, France. His analyses significantly improved upon ONRL's annotation of the plant-wall degrading enzymes and confidently assigned many more ORFs to specific carbohydrase systems, i.e. cellulases, xylanases. These annotations also identified enzymes which could not be fully annotated, but which could be assigned to general classes of carbohydrate active proteins.

Materials and Methods

Genomic analyses

Every ORF of the draft genome was analyzed by Dr Henrissat's group on the CAZy ModO (Carbohydrase Active enZyme Modular Organization) server at AFMB-CNRS (27). Catalytic and substrate binding modules were identified and named according to the CAZy nomenclature scheme (54). The known characteristics of the GH and CBM families which are commonly involved in cellulose metabolism and also occur in the genome of *S. degradans* are summarized in Table 2-1.

Identified carbohydrase ORFs were also analyzed for lipoprotein signature sequences (lipoboxes) by the DOLOP web server (Database Of LipOProteins; http://www.mrc-lmb.cam.ac.uk/genomes/dolop/) and the Technical University of Denmark's CBS LipoP 1.0 server (http://www.cbs.dtu.dk/services/LipoP) which uses hidden Markov models (61) to identify candidate lipoproteins.

Bacterial strains, growth media and conditions

E. coli DH5α Electromax® (Invitrogen, Frederick, MD) was grown on LB agar or broth containing 75 µg/ml ampicillin at 37°C and Blue/white screened on LB/Amp/X-gal (0.2 mM). *E coli* BL-21(DE3)pLysS (Novagen, Madison, WI) was grown at 37°C on LB agar or broth containing 50µg/ml ampicillin and 30.6 µg/ml chloramphenicol. *S. degradans* strain 2-40 (ATCC43961^T) was maintained on ½ strength Marine Agar (½MA): 18.7 g/L Difco Marine Broth 2216® amended with 1.5% agar. For comparisons of enzyme production *S. degradans* was grown in minimal medium (MM) consisting of

(per L): 23 g Instant Ocean Sea salts (Aquarium systems, Mentor, OH), 1 g Yeast extract, 50 mM Tris buffer pH 7.4 and 0.5 g NH_4Cl. MM was supplemented with 0.2 % (w/v) of the desired carbon source. Carbon sources included substrates which *S. degradans* was known to utilize, glucose, barley glucan, laminarin, and xylan. Additionally, to determine if it is capable of metabolizing crystalline cellulose and/or plant material, glucose-grown cultures were inoculated into MM containing either 0.2 % Avicel or 0.2 % dried *Spartina* cut into ~2 cm segments.

Table 2-1. Properties of studied examples from the GH and CBM families occurring within the predicted cellulases, CBM proteins and glycanases of uncertain specificity of *S. degradans.*

Module[a]	Demonstrated function(s) of examples found in the literature	anomeric reaction conformation[b]	References[c]
GH1	β-glucosidase, β-galactosidase, β-mannosidase, others	r	
GH3	β-1,4-glucosidase, β-1,4-xylosidase, β-1,3-glucosidase, others	r	
GH5	β-1,4-endoglucanase, B-1,4-endoxylanase, β-1,4-endomannanase, licheninase, others	r	
GH6	non-reducing end cellobiohydrolase, β-1,4-endoglucanase	i	
GH9	β-1,4-endoglucanase, cellobiohydrolase, β-glucosidase	i	
GH16	β-1,3-endoglucanase, β-1,3(4)-endoglucanase, licheninase, β-agarase, others	r	
GH26	β-1,4-endomannanase, β-1,3-xylanase	r	
GH28	α-1,4-galacturonanase (polygalacturonase), Exo-polygalacturonase	i	
GH30	β-1,6-endoglucanase, glucosylceramidase	r	
GH43	β-1,4-xylosidase, α-arabinofuranosidase, α-endoarabinanase, β-1,4-endoxylanase	i	
GH94	cellobiose phosphorylase, cellodextrin phosphorylase, chitobiose phosphorylase	i	
CBM2a	binds crystalline cellulose	--	(100)
CBM6	binds glycan chains: cellulose, xylan, β-glucans	--	(19)
CBM10	binds crystalline substrates: cellulose, mannan	--	
CBM13	lectin like binding of mono and oligosaccharides, xylan	--	(82)
CBM16	found in: alginases, chitinases, mannanases. binding unknown	--	
CBM32	lectin like binding of small, soluble oligosaccharides, dp 1-3	--	

[a] GH, glycosyl hydrolase; CBM, carbohydrate binding module; numbers refer to families according to CAZy (http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html)

[b] r, retaining; i, inverting.

[c] data from CAZy database (http://afmb.cnrs-mrs.fr/CAZY/) unless otherwise indicated

To evaluate growth, cultures were incubated with shaking until the medium became visibly turbid. Aliquots of 100 µl were plated onto ½MA and LB. Recovery of agar-pitting colonies on ½MA and lack of growth on the corresponding LB plate was considered evidence of growth on the tested carbon source. Cultures were also examined under phase-contrast microscopy to observe characteristic cell morphology. All *S. degradans* cultures were grown at 27°C unless stated otherwise. All broth cultures, whether *E. coli* or *S. degradans*, were shaken at 200 rpm.

Cloning and expression of proteins in E. coli

Nucleotide sequences of predicted genes obtained from the DOE JGI's *Saccharophagus degradans* genome web server were used to design primers within the first and last 100 nucleotides of each ORF and 5′ restriction sites were added to the primers (Table 2-2) so as to permit in-frame cloning into pETBlue2 (Novagen Madison, WI).

All primers were purchased from Invitrogen (Frederick, MD). PCR reactions (50 µl) used standard parameters and conditions for tailed primers and Proof Pro® *Pfu* Polymerase (Continental Lab Products, San Diego, CA) and included 0.5 µl of *S. degradans* genomic DNA as the template. PCR products were cloned into pETBlue2, and the resulting plasmids were transformed into *E coli* DH5α by electroporation and blue/white screened on LB/amp/X-gal. Plasmids were recovered, singly digested and visualized by agarose electrophoresis for size confirmation.

The plasmids were transformed into the expression strain by heat shock. *E coli* BL-21(DE3)pLysS, transformants were selected on LB agar containing ampicillin and

chloramphenicol and incubated overnight at 37°C. Production of an appropriate-sized His-tagged protein was confirmed by comparing pre-induced and induced (1 mM IPTG) cell lysates in western blots using 1/5000 anti-HisTag® monoclonal primary antibody (Novagen) and 1/7500 goat anti-mouse HRP conjugated secondary antibody (BioRad, Hercules, CA). Blots were developed colorimetrically with the OPTI-4CN kit (BioRad).

Production and purification of recombinant proteins

Expression cultures were grown to A_{600} of 0.6 to 0.8 in 500 ml or 1 liter broths of LB containing ampicillin and chloramphenicol. The cultures were then induced by 1 mM IPTG incubated for four hours at 37°C or 16 hours at 20°C. Culture pellets were harvested by centrifugation (5000 x *g*, 20 min) and frozen overnight at -20°C. Cells were thawed on ice and suspended in 4ml urea lysis buffer (8 M Urea, 100 mM NaH_2PO_4, 25 mM Tris, pH 8.0) per gram wet pellet weight. Alternatively, cells were lysed with BugBuster HT® (Novagen) according to the manufacturer's instructions. The insoluble fraction from this procedure was dissolved in urea lysis buffer and processed identically to urea cell lysates. The samples were clarified by centrifugation at 15,000 *g*. The resulting supernatant was mixed with Nickel-NTA resin (QIAGEN, Valencia, CA) per the manufacturer's instructions. After 1 hour at 25°C, the slurries were washed twice with urea lysis buffer, pH 7.0.

Table 2-2. Primers used in this study.

Primer	Restriction site	Nucleotide sequence[a]
Cel5A-F	EcoRV	AGTCTTGATATCCATAACCGCCATGCAAAGCACT
Cel5A-R	XhoI	CCGCTCGAGGTTGAACATGCTGGGTGGTTCT
Gly3D-F	BamHI	CGGGATCCGAAGGCTACTGCGCAAACGAAT
Gly3D-R	AscI	AGGCGCGCCAGAACCAGGCCTGTGCCATTTA
Gly5K-F	AscI	AGGCGCGCCCGCAATCACACCACAATACCAA
Gly5K-R	ClaI	CCATCGATGACGTTATTAGCGCTGCCATTT
Gly5M-F	BamHI	CGGGATCCGGCACTGTTAACAAGCGCCAAGT
Gly5M-R	AscI	AGGCGCGCCGCCTCGGTTTGCCATGTTTGCCA
Gly9C-F	AscI	AGGCGCGCCAAACCGCGCCCATCAAATTAGC
Gly9C-R	ClaI	CCATCGATCTGATTAAATGGCGGTGAAGGCG
Gly43M-F	BamHI	CGGGATCCGACCGTAGCGCTAGACACAACAA
Gly43M-R	AscI	AGGCGCGCCCCTTCGGCGAACAGTTCACTTACA
Cbm2A-F	EcoRV	AGTCTTGATATCCAGTATGGGGACGTTGTCGTCT
Cbm2A-R	XhoI	CCGCTCGAGTGGCACCGTGGTGCTAATAAC
Cbm2B-F	AscI	AGGCGCGCCTTTCAGCGCTCTCGCTAGGTTT
Cbm2B-R	ClaI	CCATCGATTACACCAACGCCCACGGAATAA
Cbm2C-F	AscI	AGGCGCGCCCTATTTCGCTTGGTGCAATGA
Cbm2C-R	ClaI	CCATCGATCTTATCCTTCCCACCAGCATTC

[a] Underlines delineate restriction sites introduced to permit in-frame cloning into the pETBlue2 vector

Renaturation was performed on the column at 4°C using one column volume of renaturation buffer with decreasing urea concentrations (25 mM Tris pH 7.4, 500 mM NaCl, 20% glycerol, urea 6 M to 1 M in 1 M steps). The refolded proteins were eluted in 1 M urea, 25 mM Tris pH 7.4, 500 mM NaCl, 20% glycerol containing 250 mM imidazole. Void, wash and elution fractions were surveyed for HisTag® production in western blots as described above. Elution fractions containing the recombinant proteins were pooled and exchanged into Storage Buffer (20 mM Tris pH 7.4, 10 mM NaCl, 10% glycerol) using Centricon™ centrifugal ultrafiltration devices (Millipore). The enzyme preparations were then aliquoted and frozen at -80°C for use in activity assays. Final protein concentrations ranged from 500 µg/ml to 2200 µg/ml.

Zymograms

Gels were prepared as standard mini-format SDS-PAGE gels (67) with the appropriate CP substrate incorporated directly into the separating gel at the following final concentrations: 0.1% (HE-cellulose), 0.15% (barley β-glucan). Each lane contained ~5 µg of recombinant protein or ~10 µg of filtered supernatant from *S. degradans* cultures grown in MM containing 0.2% Avicel, Barley glucan, xylan or laminarin. After electrophoresis, gels were incubated 1 hr at room temperature in 80 ml of renaturing buffer (20 mM PIPES buffer pH 6.8, 2.5% Triton X-100, 2 mM DTT and 2.5 mM $CaCl_2$), after which they were held overnight at 4°C with gentle rocking. They were then equilibrated in 80 ml of 20 mM PIPES for 1 hour at room temperature, transferred fresh PIPES buffer and incubated for four hours at 37°C. Zymograms were stained for 30

minutes with 0.25% Congo red in dH_2O, and destained with 1 M NaCl until clear bands were visible against a stained background.

Nelson-Somogyi reducing-sugar assays

Purified proteins were assayed for activity using a modification of the Nelson-Somogyi reducing sugar method adapted for 96-well microtiter plates, using 50 ul reaction volumes (44). Briefly, 25 µl of the sample was added to 25 µl of substrate solution in a microtiter well, the plate was sealed with oleofin Microwell plate sealing film (Nunc, Rochester NY), vortexed and incubated at 37°C. Standards consisted of known concentrations of glucose or xylose in a volume of 50µl. Negative controls consisted of enzyme preparation that was autoclaved for 20 minutes to destroy activity. Test substrates included Avicel, CMC, phosphoric-acid swollen cellulose (PASC), Barley glucan, laminarin, and xylan dissolved or suspended at 1% (w/v) in 20mM PIPES pH 6.8 (Barley glucan and laminarin, 0.5%). Barley glucan, laminarin and xylan assays were incubated for 2 hours and Avicel, CMC and PASC assays were incubated 36 hours. Following incubation, assay reactions were stopped by freezing at -80°C and held at that temperature until analysis. Upon thawing, 75 µl of the copper reagent (four parts of 1:2:12:1.3 KNa tartarate: Na_2CO_3:$NaSO_4$:$NaHCO_3$ mixed with one part 1:9 $CuSO_4 \cdot 5H_2O$:Na_2SO_4) was added to the standards, reaction, and control tubes/wells and incubated 30 minutes at 80°C. After cooling to room temperature, 75 µl of the Arsenomolybdate reagent (25g ammonium molybdate in 450 ml DH_2O + 21 ml H_2SO_4 + 3 g $Na_2HAsO_4 \cdot 7H_2O$ dissolved in 25 ml DH_2O) was added and mixed briefly by vortex. Trial assays revealed that the substrates CMC, xylan and barley glucan formed a

precipitate which interfered upon addition of the assay reagents. As this precipitate interfered with A_{510} readings, assays using these substrates could not be performed directly in the microtiter plate wells. These assays were conducted in 0.6 ml siliconized microcentrifuge tubes. Reactions performed in 0.6ml tubes were centrifuged at 14,000 x *g* and 200 µl of the supernatant was transferred to a microplate well by pipette. Samples were assayed in triplicate, corrected for blank values, and levels estimated from a standard curve. Protein concentration of enzyme assay samples was measured in triplicate using the Pierce BCA protein assay according to the manufacture's instructions. Enzymatic activity was calculated, with one unit (U) defined as 1 µmol of reducing sugar released/minute and reported as specific activity in U/mg protein.

Exoglycosidase activity assays: pnp-derivatives

Purified proteins were assayed for activity against para nitrophenol derivatives of α-L-arabinofuranoside, -α-L-arabinopyranoside, -β-L-arabinopyranoside, -β-D-cellobioside, -α-D-glucopyranoside, -β-D-glucopyranoside, -α-D-xylopyranoside and -β-D-xylopyranoside. 25 µl of enzyme solution was added to 125 µl of 5mM substrate solution in 20 mM PIPES pH 6.8, incubated for 30 minutes at 37°C, and A_{405} was determined. After correcting for blank reactions, readings were compared to a p-nitrophenol standard curve and reported as specific activities in U/mg protein, with one unit (U) defined as 1µmol p-Np/minute.

Cellulose binding assays

In order to evaluate the function of CBM2 modules in cloned Cbm-only proteins a novel cellulose-binding assay was developed. Aliquots of rCbm2A, rCbm2B, and rCbm2C were adjusted to 500 µg/ml using Storage Buffer (20 mM Tris pH 7.4, 10 mM NaCl, 10% glycerol), and the control proteins, pepsin (37 kDa), porcine lipase (50 kDa), and ovalbumen (45 kDa) were prepared as 500 µg/ml solutions in the same buffer. A 20 µl drop of 10% (w/v) Avicel in Milli-Q water was added to 75 µl of 500 µg/ml bovine serum albumen (BSA) in 20 mM PIPES buffer pH 6.8 and intermittently vortexed for 15 minutes at 25°C. The Avicel pellet was washed twice in 400 µl of PIPES and resuspended in 20 µl of the test or control protein solution and intermittently vortexed 3 minutes at 25°C. The resulting pellets were given four ten-minute washes in 2 M urea at 56°C. The pellets were then suspended in 1X SDS treatment buffer and held at -80°C until analysis on 11% or 8% SDS-PAGE gels, depending on the MW of the tested protein. Staining with Coomassie blue revealed proteins remaining bound to the washed avicel. Specific binding attributed to CBM modules was defined as persistence of the test protein under conditions which removed all three negative control proteins.

Mass Spectrometry and proteomic analyses

Supernatants from avicel, CMC, and xylan-grown cultures were concentrated to ~25X by centrifugal ultrafiltration using Centricon™ or Microcon™ devices (Millipore). Protein concentrations were determined using the BCA protein assay (Pierce). Samples were exchanged into 100 mM Tris buffer, pH 8.5, containing 8 M urea and 10 mM DTT and incubated 2 hours at 37°C to denature and reduce the proteins. After reduction,

cysteine residues were alkylated by the addition of 1 M iodoacetate to a final concentration of 50 mM and incubated at 25°C for 30 minutes. The samples were exchanged into 50 mM Tris, 1mM $CaCl_2$, pH 8.5 using Microcon™ devices. The denatured, reduced, and alkylated samples were digested overnight at 37°C using proteomics grade trypsin (Promega) at a 1:50 enzyme to substrate ratio. Digestions were stopped by the addition of 99% formic acid to a final concentration of ~1% and analyzed by RPHPLC-MS/MS at the UMCP College of Life Sciences CORE Mass Spectrometry facility using a Waters 2960 HPLC linked to a Finnagin LCQ tandem Mass Spectrometer. Alternatively, to identify specific proteins from SDS-PAGE gels of avicel-grown supernatants, regions containing 120 kDa, 100 kDa, and 75 kDa proteins were excised from an 8% gel and sent to Stanford University's Mass Spectrometry facility (SUMS; http://mass-spec.stanford.edu), where they were processed similarly and analyzed using an ESI-Quadrupole-Time of Flight (Micromass Q-Tof) mass spectrometer. All peptide fragment masses were analyzed by the peptide analysis packages SEQUEST and MASCOT (37, 86), and compared to amino acid sequence translations of all gene models in the 2-40 draft genome and to the non-redundant Mass Spectrometry Database (ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam). Peptide identity matches were evaluated using the accepted thresholds of statistical significance specific to each program.

Results

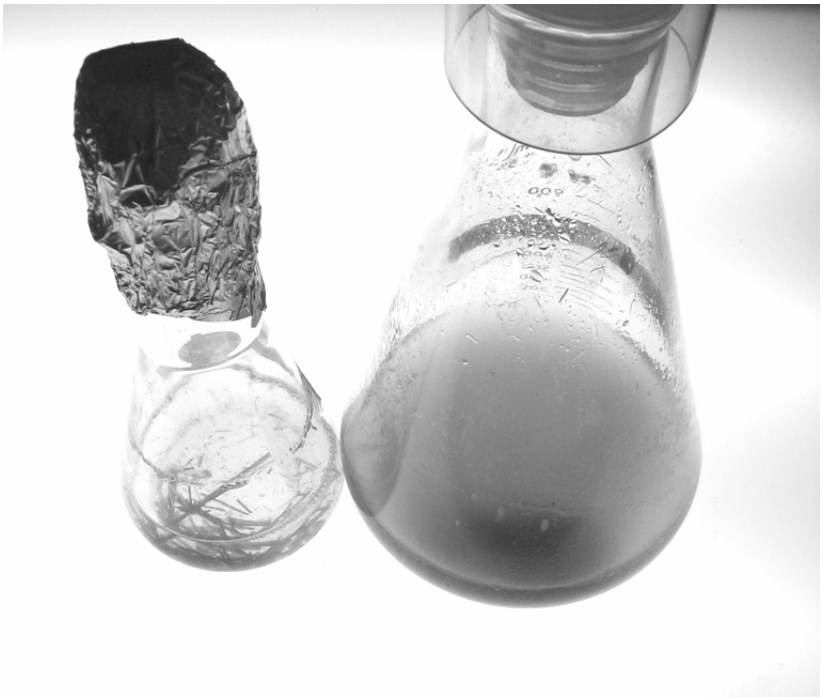
S. degradans uses Avicel and Spartina for growth and energy

Avicel-containing cultures became turbid within ca. 36 hours of inoculation. Numerous agar-pitting colonies formed on ½MA plates, while LB plates did not contain any visible colonies, even upon extended incubation at 27°C. Phase-contrast microscopy showed cells with typical *S. degradans* morphologies with no visible contaminants. These findings confirm growth on crystalline cellulose, indicating that 2-40 has a true cellulase system.

The cultures containing *Spartina* became visibly turbid after four days. As with Avicel cultures, analysis of marine agar plates and microscopic examinations confirmed culture purity. With extended incubation, pronounced disintegration of the *Spartina* segments was observed, indicating that 2-40 can utilize plant matter as sole carbon source and that growth is concurrent with physical disruption of plant leaf structure. Figure 2-1 shows an inoculated MM + *Spartina* culture and an uninoculated control flask of the same medium.

Figure 2-1. Growth of *S. degradans* 2-40 in *Spartina* minimal medium.

The uninoculated flask on the left was prepared identically to the inoculated flask on the right. Note intact *Spartina* and lack of turbidity in the uninoculated flask compared to the turbidity and visible deterioration of *Spartina* in the flask on the right. Photograph was taken on day 20.



Overview of the plant wall degrading enzymes

Since *S. degradans* appears to degrade plant matter in monoculture, it is predicted to have a broad spectrum of carbohydrases. CAZy ModO identified 222 ORFs containing carbohydrase and/or carbohydrate binding modules (CBM), a number second only to *Bacteroides thetaiotamicron,* an endosymbiont of the human intestinal tract which has 366 annotated carbohydrase-related genes (http://afmb.cnrs-mrs.fr/CAZY/index.html). In S. degradans there are 117 genes which have one or more glycosyl hydrolase (GH) modules. Additionally, there are 39 ORFs containing glycosyltransferase (GT) modules, 29 with polysaccharide lyase (PL), and 17 with carbohydrate esterase (CE) domains. Many carbohydrases carry at least one CBM, and most exhibit modular architecture.

The ModO analysis predicts that 77 of these ORFs code for plant wall degradases, including apparent cellulase, xylanase, arabinoxylanase, β-mannanase, β-1,3-glucanase and pectinase systems. Each system appears to contain multiple depolymerases and accessory enzymes for removal of substituents and processing of oligomers to monomers. Most of these enzymes contain one or more carbohydrate binding modules (CBM). The confidently-predicted cellulases comprise an apparently complete system containing 13 cellulose depolymerases and 7 associated accessory enzymes from six different glycoside hydrolase (GH) families (Table 2-3).

Table 2-3. Predicted cellulases and accessory enzymes of *S. degradans*

Name[a]	Sde number[c]	Predicted function[d]	Modules[d,f]	amino acids[h]	MW[h]	MS[i]
Cel5A[b]	3003	endo 1,4-β-glucanase (EC 3.2.1.4)	GH5/CBM6/CBM6/CBM6/GH5	1,167	127.2	
Cel5B	2490	endo 1,4-β-glucanase	LPB/PSL(47)/CBM6/GH5	566	60.8	
Cel5C	0325	endo 1,4-β-glucanase	LPB/PSL(47)/GH5	451	49.1	
Cel5D	2636	endo 1,4-β-glucanase	CBM2/PSL(58)/CBM10/PSL(36)/GH5	621	65.9	
Cel5E	2929	endo 1,4-β-glucanase	CBM6/CBM6/GH5	673	72.6	
Cel5F	1572	endo 1,4-β-glucanase	GH5	365	42.0	
Cel5G	3239	endo 1,4-β-glucanase	GH5/PSL(21)/CBM6/PSL(32)/Y95	638	67.9	
Cel5H	3237	endo 1,4-β-glucanase	GH5/PSL(32)/CBM6/EPR(16)	630	66.9	av
Cel5I	3420	endo 1,4-β-glucanase	CBM2/PSL(33)/CBM10/PSL(58)/GH5	725	77.2	av
Cel5J	2494	endo 1,4-β-glucanase	GH5/CBM6/CBM6	610	65.2	
Cel6A	2272	cellobiohydrolase (EC 3.2.1.91)[e]	CBM2/PSL(43)/CBM2/PSL(85)/GH6	791	81.9	
Cel9A	0636	endo-1,4-β-glucanase	GH9	578	62.7	
Cel9B	0649	endo-1,4-β-glucanase	GH9/PSL(54)/CBM10/PSL(50)/CBM2	867	89.5	av
Ced3A	2497	cellodextrinase (EC 3.2.1.74)	LPB/GH3/PLP	1,072	116.0	av, cm, xn
Ced3B	0245	cellodextrinase	LPB/GH3	862	92.9	xn
Bgl1A	3603	cellobiase (EC 3.2.1.21)	GH1[g]	461	52.8	
Bgl1B	1394	cellobiase	GH1[g]	444	49.8	
Bgl3C	2674	cellobiase	LPB/GH3/UNK(511)	866	95.4	av, cm, xn
Cep94A	1318	cellobiose phosphorylase (EC 2.4.1.20)	GH94[g]	811	91.7	
Cep94B	0906	cellodextrin phosphorylase (EC 2.4.1.49)	GH94[g]	788	88.7	cm

[a] Acronyms: Cel , cellulase; Ced, cellodextrinases; Bgl, β-glucosidase; Cep, cellobiose/cellodextrin phosphorylase.

[b] *cel5A* was cloned, expressed as Cel5A:His_6, and assayed for activity.

[c] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[d] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[e] Cel6A is predicted to act on the non-reducing end of cellulose chains.

[f] Module abbreviations: CBM, carbohydrate binding module; GH, glycosyl hydrolase (numbers refer to families according to CAZy (http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html); EPR, glutamic acid-proline rich region; LPB, lipobox signature sequence; PLP, phospholipase-like domain; PSL, polyserine linker; UNK, unknown function. Numbers in parentheses indicate the length of a region in amino acid residues.

[g] No type II secretion signal sequence.

[h] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI gene model amino acid sequence translations including any predicted signal peptide.

[i] Protein identified by tandem mass spectrometry in concentrated supernatant from cultures grown in avicel (av), CMC (cm), or xylan (xn).

In addition to these 77 easily-annotated ORFs, there are another 35 with poorly-conserved CBM and/or GH modules which belong to families that are commonly involved in plant CP binding and/or degradation. To resolve the function of this group of genes requires further analyses. Nineteen of these sequences contain a CBM but have no identifiable carbohydrase domain, although one ORF includes a zinc-dependent protease module, suggesting a role as an accessory protease. The rest of these sequences lack any recognizable catalytic domains, and have thus been designated "CBM proteins" and given them the prefix "Cbm" (Table 2-4).

There are also 16 genes (Table 2-5) with GH modules of uncertain function or identity. These were termed "glycanases of uncertain specificity" and given the acronym "gly" pending an authoritative determination of their substrate. Several of these CBM proteins and glycanases of uncertain specificity were cloned, expressed, and subjected to functional analyses.

Table 2-4. Selected carbohydrate-binding module (CBM) proteins.

Name[a]	Sde number[c]	Modules[d,e]	amino acids[f]	MW[f]	MS[g]
Cbm2A[b]	1182	CBM2/EPR(48)/UNK(191)	371	38.0	
Cbm2B[b]	1183	CBM2/UNK(914)	1,042	112.1	av
Cbm2C[b]	0182	CBM2/PSL(58)/Y94/PSL(25)/UNK(577)	933	97.5	xn
Cbm2D-Cbm10A	0650	CBM2/PSL(43)/CBM10/PSL(101)/UNK(231)	558	56.3	
Cbm2E	0569	CBM2/PSL(18)/PSL(18)/UNK(471)	781	85.0	
Cbm2F	2939	CBM2/PSL(33)/PSL(17)/UNK(544)	787	84.8	
Cbm6A	0242	LPB/PSL(28)/CBM6/UNK(476)	681	72.5	
Cbm6B	0216	CBM6/CBM6/UNK(217)	630	67.2	
Cbm6C	3892	CBM6/UNK(216)	550	59.8	
Cbm6D	3260	CBM6/UNK(262)	500	54.6	
Cbm6E	1445	LPB/UNK(311)/CBM6	473	53.0	
Cbm6F	3927	UNK(311)/CBM6	465	50.4	
Cbm6G-Cbm16B	0112	UNK(620)/CBM6/CBM16/	1,024	111.4	
Cbm6H-Cbm32F	2795	UNK(514)CBM6/CBM32	912	97.5	
Cbm16A-Cbm32E	3272	CBM16/PTR(16)/CBM32/PTR(22)/UNK(232)	552	58.5	
Cbm32A	0478	CBM32/CBM32/UNK(251)/COG3488	1,028	111.9	xn
Cbm32B	3845	UNK(109)/CBM32/CBM32/CBM32	557	59.5	
Cbm32C	1503	ZDP/CBM32/FN3/FN3/CBM32	674	73.6	
Cbm32D	3709	CBM32/FCL/UNK(521)	818	88.8	

[a] Acronyms: Cbm, carbohydrate binding module.

[b] ORF was cloned, expressed as a HisTag® fusion protein and assayed for cellulose binding and enzyme activity.

[c] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[d] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[e] Module abbreviations: CBM, carbohydrate binding module; COG3488, thiol-oxidoreductase like domain; EPR, glutamic acid-proline rich region; FCL, similar to fucose-binding lectins in eel serum proteins; FN3, fibronectin type 3 module; LPB, lipobox signature sequence; PSL, polyserine linker; PTR, Proline-Threonine repeat; UNK, unknown function; Y94, novel domain identified in this protein and two xylanases by CAZyme analysis; ZDP, zinc-dependent protease domain.

[f] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI gene model amino acid sequence translations including any predicted signal peptide.

[g] Protein identified by tandem mass spectrometry in concentrated supernatant from cultures grown in Avicel (av) or xylan (xn).

Table 2-5. Glycanases of uncertain substrate specificity.

Name[a]	Sde number[d]	Predicted function[e]	Modules[e,f]	amino acids[g]	MW[g]	MS[h]
Gly3D[b]	0475	β-glycosidase (EC 3.2.1.-)	CBM32/CBM32/CBM32/GH3/CBM32	1,581	173.0	xn
Gly5K[b]	2993	endo-(1,3 or 1,4)-β-glucanase (EC 3.2.1.4/6)	GH5/CBM6/CBM6/CBM13	863	94.6	
Gly5L	2996	endo-(1,3 or 1,4)-β-glucanase	LQAC/GH5/CBM6	853	93.3	
Gly5M[b]	3023	endo-(1,3 or 1,4)-β-glucanase	LPB/EPR(34)/GH5/FN3/CBM6	869	94.9	
Gly5R[c]	1121	β-glycosidase or mannanase (EC 3.2.1.25)	TMR/TMR/GH5/FN3	523	59.2	
Gly5S	2285	β-glycosidase	GH5	383	42.8	xn
Gly9C[b]	0558	β-glycosidase	LPB/GH9	665	73.2	
Gly16H	2878	β-glycosidase	GH16/CBM4	968	107.2	
Gly26B	3022	β-glycosidase	GH26/CBM13/CBM13	644	71.6	
Gly30A	2992	β-glycosidase	LPB/GH30/CBM6/CBM6/CBM13	982	107.4	
Gly30B	2994	β-glycosidase	GH30	481	52.8	
Gly43M[b]	3317	β-glycosidase	LPB/GH43/LGL/LGL	1,174	127.2	

[a] Poorly-conserved modules from GH families which contain plant-wall depolymerases are designated by the generic prefix "gly" pending identification of substrate specificity.

[b] ORF was cloned and expressed as a HisTag® fusion protein and assayed for enzyme activity.

[c] Gly5R lacks an N-terminal secretion signal

[d] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[e] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[f] Module abbreviations: CBM, carbohydrate binding module; EPR, glutamic acid-proline rich region; FN3, fibronectin type 3 module; LPB, lipobox signature sequence; LQAC, putative alternate lipobox sequence with atypical second amino acid; TMR, transmembrane region.

[g] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI gene model amino acid sequence translations including any predicted signal peptide.

[h] Protein identified by tandem mass spectrometry in concentrated supernatant from cultures grown in xylan (xn).

Table 2-6. Summary of enzymes detected by mass spectrometry.

Name	Growth substrate[a]	amino acids[b]	MW[b]	Predicted function[c]	Modules[c]
Cbm2B	av	1,042	112.1	cbm only	CBM2/UNK(914)
Cbm2C	xn	933	97.5	cbm only	CBM2/PSL(58)/Y94/PSL(25)/UNK(577)
Cbm32A	xn	1,028	111.9	cbm only	CBM32/CBM32/UNK(251)/COG3488
Ced3A	av, cm, xn	1,072	116.0	cellodextrinase	LPB/GH3/PLP
Ced3B	xn	862	92.9	cellodextrinase	LPB/GH3
Cel5H	av	630	66.9	endocellulase	GH5/PSL(32)/CBM6/EPR(16)
Cel5I	av	725	77.2	endocellulase	CBM2/PSL(33)/CBM10/PSL(58)/GH5
Cel9B	av	867	89.5	endocellulase	GH9/PSL(54)/CBM10/PSL(50)/CBM2
Cep94B	cm	788	88.7	cellodextrin phosphorylase	GH94
Gly3D	xn	1,581	173.0	β-glycosidase	CBM32/CBM32/CBM32/GH3/CBM32
Xyl31A	xn	973	110.2	α-xylosidase	LPB/GH31
Xyl3A	av, cm, xn	893	97.6	β-xylosidase	LPB/GH3
Xyn10E	xn	670	75.2	β-xylanase	LPB/EPR(47)/GH10

[a] Protein was detected in supernatants of cultures grown in the following growth substrates: av, Avicel; cm, carboxymethyl cellulose; xn, xylan.

[b] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI gene model amino acid sequence translations including any predicted signal peptide.

[c] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

Identification of potential surface-associated enzymes

An unusually high percentage of *S. degradans* carbohydrases contain lipobox signature sequences, which are normally acylated and anchored to the outer membrane so that the enzymes function on the cell surface. In fact DOLOP analysis reveals that of the 112 predicted or potential plant-wall carbohydrases, 32 (28.5%) contain consensus lipobox sequences [LVI][ASTVI][GAS][C]. Five of these are cellulases and cellulase accessory enzymes (Table 2-3), five are glycanases of uncertain specificity (Table 2-5), and two are "CBM only" proteins (Table 2-4). There are also two genes (Gly5L, Table 2-5 and Man5Q, Table 3-5) with lipobox-like sequences which carry an atypical amino acid in the second position (LQAC); these two were identified as probable lipoproteins by LipoP 1.0.

Cellulases and related accessories.

Of the 13 predicted *S. degradans* cellulases, 12 annotate as β-1,4-endogulcanases and one as a cellobiohydrolase which acts on the non-reducing ends of cellulose chains (Table 2-3). The seven "cellulase accessory enzymes" process cellooligosaccharides and include two cellodextrinases, three cellobiases, a β-glucosidase, a cellodextrin phosphorylase and a cellobiose phosphorylase. All of these proteins include an N-terminal type-II secretion signal sequence except for two of the cellobiases and both phosphorylases. Two endoglucanases, Cel5B and Cel5C, the cellodextrinases Ced3A and Ced3B, and cellobiase Bgl3C contain lipobox motifs. Eight of the 13 cellulose depolymerases contain one or more extended regions comprised mainly of serine residues

(Table 2-3). These "polyserine linkers" (PSL) occur in an unusually high number of *S. degradans* carbohydrases, and are believed to function as flexible spacers or linkers (56).

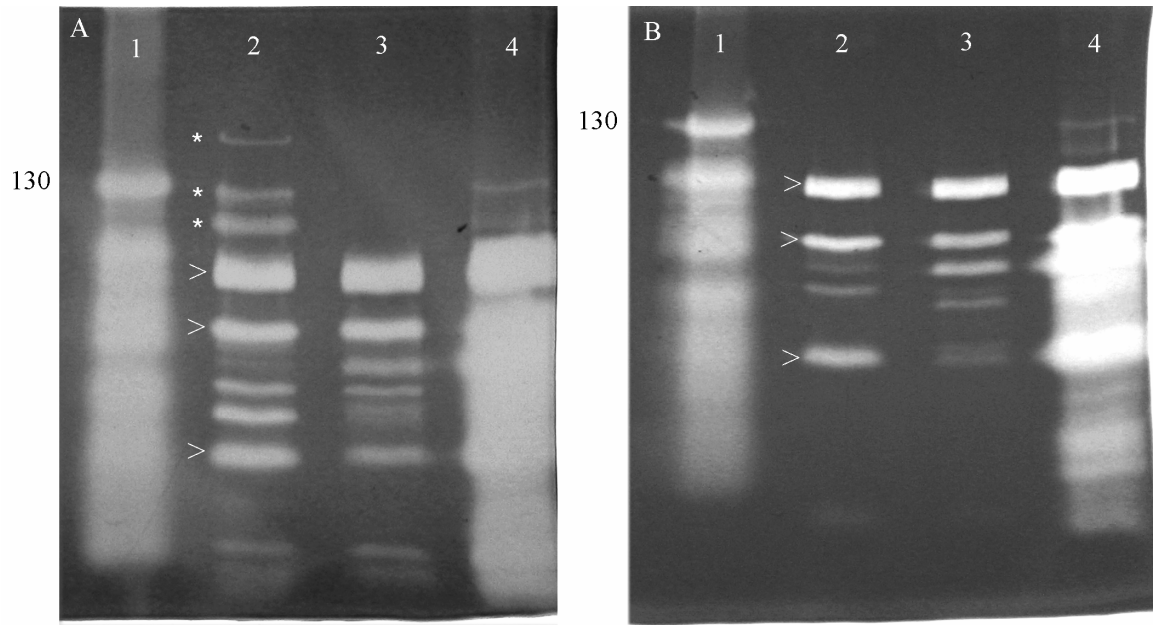
The activities of several predicted cellulases were confirmed using zymograms. Analysis of avicel-grown cultures in these zymograms revealed a number of protein bands with activity against HE-cellulose and barley β-glucan (Figure 2-2), showing that there are numerous endoglucanases induced during growth on cellulose. However, the number of predicted endoglucanases with molecular weights between 60-100 kDa made specific bands of activity difficult to correlate with genomically identified enzymes.

To identify individual carbohydrases, culture supernatants were analyzed by RP-HPLC MS/MS. Three cellulases and four cellulase accessory enzymes were repeatedly detected in cellulose and/or xylan grown culture supernatants (Table 2-6). The cellulases were Cel5H, Cel5I, and Cel9B, and the accessories included the cellodextrinases, Ced3A and Ced3B, as well as the secreted cellobiase, Bgl3C, and the cellodextrin phosphorylase, Cep94B.

Figure 2-2. Zymograms of Cel5A:His_6 and CP grown stationary-phase culture supernatants.

A, 7% SDS-PAGE zymogram containing 0.15% barley glucan. Lane 1) 5 µg Cel5A:His_6; lower MW bands are considered to be breakdown products 2) Laminarin-grown culture supernatant; 3) Barley glucan-grown culture supernatant. 4) Avicel-grown culture supernatant.

B, 8% SDS-PAGE zymogram containing 0.1% HE-Cellulose, Lanes as in panel A. Note high MW bands (*) in panel A, lane 2, showing three large β-1,3-endoglucanases produced in response to laminarin, but not cellulose or barley glucan. Also note the appearance of three conserved bands (arrows) in lanes 2-4 of both panels, suggesting that as many as three β-1,4-endoglucanases function as common members of laminarin, mixed glucan, and cellulose degradative pathways.


A
1
2
3
4
130
*
*
*
>
>
>
B
1
2
3
4
130
>
>
>

Domain analysis of cellulase catalytic and substrate-binding modules

Many of the cellulases of *S. degradans* are unlike those of other sequenced cellulolytic bacteria and contain a fascinating array of domains. When compared to other cellulolytic prokaryotes with sequenced genomes, *S. degradans* has an unusually high number of endoglucanases from glycoside hydrolase family 5 (Table 2-7).

Twelve of the predicted cellulases are endoglucanases; ten have GH5 modules and the other two have GH9 modules (Table 2-3). There is also a GH6 non-reducing end cellobiohydrolase. Unlike two well-studied bacterial models of cellulose degradation, *Thermobifida fusca* and *Clostridium thermocellum* (59), the genome of *S. degradans* does not appear to encode a GH48 reducing-end cellobiohydrolase (Table 2-7).

Table 2-7. Distribution of candidate cellulases and accessory cellodextrinases, β-glucosidases and phosphorylases in the genomes of completely sequenced cellulolytic bacteria.

Family[a]	Sub-family[b]	*Saccharophagus degradans*[c]	*Cytophaga hutchinsonii*[c]	*Thermobifida fusca*[c]	*Clostridium thermocellum*[c]	*Thermotoga maritima*[d]
GH5		16	5	2	7	2
GH6	endo			1		
GH6	exo	1		1		
GH8			4		1	
GH9		3	7	2	14	
GH12						2
GH48				1	2	
GH74			1	1	1	1
GH1 (β-glucosidases)		2	1	2	2	1
GH3 (β-glucosidases)		5	5	1	1	3
GH94 (phosphorylases)		2			2	1

[a] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[b] GH family 6 includes endoglucanases and exo-acting cellobiohydrolases

[c] Data obtained by BLAST analysis on the NCBI microbial genomes server (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi)

[d] Data from the CAZy database (http://afmb.cnrs-mrs.fr/CAZY/)

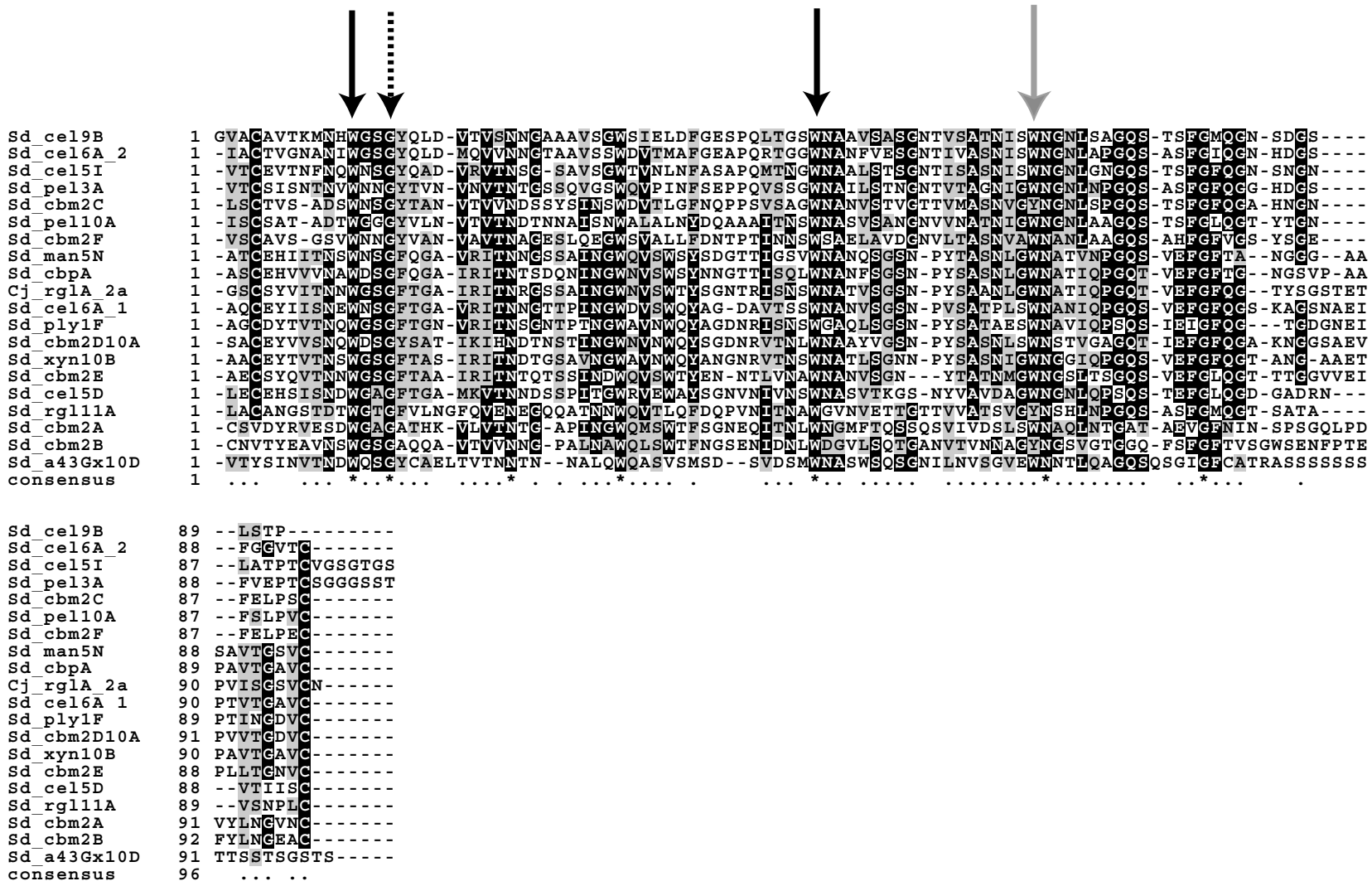
Most *S. degradans* cellulases contain one or more CBM belonging to families 2, 6 and 10 (Table 2-3). Sequence analysis using the Baylor College of Medicine's CLUSTALW web server (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) shows that all of the CBM2 modules in the genome are from subfamily 2a (Figure 2-3), indicating that they bind crystalline cellulose (18). Cellobiohydrolase Cel6A contains two such CBM2 modules, making it the first known GH6 protein with dual CBM2 modules. Three endoglucanases, Cel5D, Cel5I, and Cel9B, also have CBMs which are predicted to bind crystalline cellulose, each enzyme containing one CBM2a and one CBM10 (Table 2-3). Six other endoglucanases contain one or more CBM6 modules, and only three lack a CBM entirely (Table 2-3). The cellodextrinases, cellobiases and phosphorylases lack CBMs, in accordance with the soluble nature of their substrates. The substrate binding-specificities of the CBMs are considered important in assessing the binding sites, and thereby the function of the cellulase. For example, although the GH5 module of Cel5D is a predicted endoglucanase, the CBM2 and CBM10 both bind crystalline cellulose. Thus, Cel5D is thought to bind and act on or near crystalline regions (18).

Figure 2-3. ClustalW alignment of *S. degradans* CBM2 modules with the CBM2a module of rglA of *Cellvibrio japonicus.*

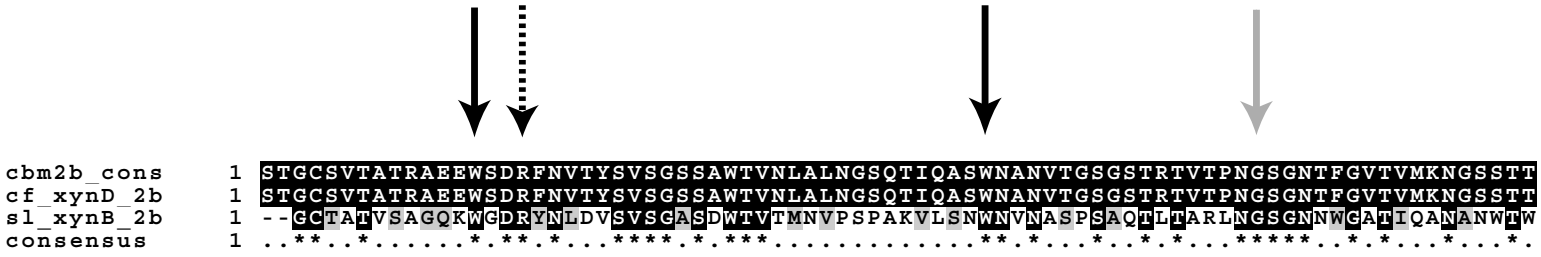
A, Solid black arrows indicate conserved aromatic residues involved in substrate binding in subfamilies a and b of CBM family 2. Dashed arrow indicates the position of a glycine residue conserved among CBM2a modules. This glycine is replaced by an arginine residue in CBM family 2b (see panel B). Shaded arrow indicates conserved aromatic residue in family 2a which also participates in binding crystalline cellulose. This residue is absent in CBM family 2b;

B, Solid black arrows indicate conserved aromatic residues involved in substrate binding in subfamilies a and b of CBM family 2. Dashed arrow indicates the position of the conserved arginine residue in CBM family 2b. Shaded arrow indicates the position of the conserved aromatic residue in family 2a which is absent in CBM family 2b.

A


Sd_cel9B 1 GVACAVTKMNHWGSGYQLD-VTVSNNGAAAVSGWSIELDFGESPQLTGSWNAAVSASGNTVSATNISWNGNLSAGQS-TSFGMQGN-SDGS----
Sd_cel6A_2 1 -IACTVGNANIWGSGYQLD-MQVVNNGTAAVSSWDVTMAFGEAPQRTGGWNANFVESGNTIVASNISWNGNLAPGQS-ASFGIQGN-HDGS----
Sd_cel5I 1 -VTCEVTNFNQWNSGYQAD-VRVTNSG-SAVSGWTVNLNFASAPQMTNGWNAALSTSGNTISASNISWNGNLGNGQS-TSFGFQGN-SNGN----
Sd_pel3A 1 -VTCSISNTNVWNNGYTVN-VNVTNTGSSQVGSWQVPINFSEPPQVSSGWNAILSTNGNTVTAGNIGWNGNLNPGQS-ASFGFQGG-HDGS----
Sd_cbm2C 1 -LSCTVS-ADSWNSGYTAN-VTVVNDSSYSINSWDVTLGFNQPPSVSAGWNANVSTVGTTVMASNVGYNGNLSPGQS-TSFGFQGA-HNGN----
Sd_pel10A 1 -ISCSAT-ADTWGGGYVLN-VTVTNDTNNAISNWALALNYDQAAAITNSWNASVSANGNVVNATNIGWNGNLAAGQS-TSFGLQGT-YTGN----
Sd_cbm2F 1 -VSCAVS-GSVWNNGYVAN-VAVTNAGESLQEGWSVALLFDNTPTINNSWSAELAVDGNVLTASNVAWNANLAAGQS-AHFGFVGS-YSGE----
Sd_man5N 1 -ATCEHIITNSWNSGFQGA-VRITNNGSSAINGWQVSWSYSDGTTIGSVWNANQSGSN-PYTASNLGWNATVNPGQS-VEFGFTA--NGGG--AA
Sd_cbpA 1 -ASCEHVVVNAWDSGFQGA-IRITNTSDQNINGWNVSWSYNNGTTISQLWNANFSGSN-PYSASNLGWNATIQPGQT-VEFGFTG--NGSVP-AA
Cj_rglA_2a 1 -GSCSYVITNNWGSGFTGA-IRITNRGSSAINGWNVSWTYSGNTRISNSWNATVSGSN-PYSAANLGWNATIQPGQT-VEFGFQG--TYSGSTET
Sd_cel6A_1 1 -AQCEYIISNEWNSGFTGA-VRITNNGTTPINGWDVSWQYAG-DAVTSSWNANVSGSN-PVSATPLSWNANIQPGQS-VEFGFQGS-KAGSNAEI
Sd_ply1F 1 -AGCDYTVTNQWGSGFTGN-VRITNSGNTPTNGWAVNWQYAGDNRISNSWGAQLSGSN-PYSATAESWNAVIQPSQS-IELGFQG---TGDGNEI
Sd_cbm2D10A 1 -SACEYVVSNQWDSGYSAT-IKIHNDTNSTINGWNVNWQYSGDNRVTNLWNAAYVGSN-PYSASNLSWNSTVGAGQT-IEFGFQGA-KNGGSAEV
Sd_xyn10B 1 -AACEYTVTNSWGSGFTAS-IRITNDTGSAVNGWAVNWQYANGNRVTNSWNATLSGNN-PYSASNIGWNGGIQPGQS-VEFGFQGT-ANG-AAET
Sd_cbm2E 1 -AECSYQVTNNWGSGFTAA-IRITNTQTSSINDWQVSWTYEN-NTLVNAWNANVSGN---YTATNMGWNGSLTSGQS-VEFGLQGT-TTGGVVEI
Sd_cel5D 1 -LECEHSISNDWGAGFTGA-MKVTNNDSSPITGWRVEWAYSGNVNIVNSWNASVTKGS-NYVAVDAGWNGNLQPSQS-TEFGLQGD-GADRN---
Sd_rgl11A 1 -LACANGSTDTWGTGFVLNGFQVENEGQQATNNWQVTLQFDQPVNITNAWGVNVETTGTTVVATSVGYNSHLNPGQS-ASFGMQGT-SATA----
Sd_cbm2A 1 -CSVDYRVESDWGAGATHK-VLVTNTG-APINGWQMSWTFSGNEQITNLWNGMFTQSSQSVIVDSLSWNAQLNTGAT-AEVGFNIN-SPSGQLPD
Sd_cbm2B 1 -CNVTYEAVNSWGSGAQQA-VTVVNNG-PALNAWQLSWTFNGSENIDNLWDGVLSQTGANVTVNNAGYNGSVGTGGQ-FSFGFTVSGWSENFPTE
Sd_a43Gx10D 1 -VTYSINVTNDWQSGYCAELTVTNNTN--NALQWQASVSMSD--SVDSMWNASWSQSGNILNVSGVEWNNTLQAGQSQSGIGFCATRASSSSSS
consensus 1 ... ... *..*... ....* . . ...*.... . ... *.. ....*........ ..*... .
Sd_cel9B 89 --LSTP---------
Sd_cel6A_2 88 --FGGVTC--------
Sd_cel5I 87 --LATPTCVGSGTGS
Sd_pel3A 88 --FVEPTCSGGGGSST
Sd_cbm2C 87 --FELPSC--------
Sd_pel10A 87 --FSLPVC--------
Sd_cbm2F 87 --FELPEC--------
Sd_man5N 88 SAVTGSVC--------
Sd_cbpA 89 PAVTGAVC--------
Cj_rglA_2a 90 PVISGSVCN-------
Sd_cel6A_1 90 PTVTGAVC--------
Sd_ply1F 89 PTINGDVC--------
Sd_cbm2D10A 91 PVVTGDVC--------
Sd_xyn10B 90 PAVTGAVC--------
Sd_cbm2E 88 PLLTGNVC--------
Sd_cel5D 88 --VTIISC--------
Sd_rgl11A 89 --VSNPLC--------
Sd_cbm2A 91 VYLNGVNC--------
Sd_cbm2B 92 FYLNGEAC--------
Sd_a43Gx10D 91 TTSSTSGSTS-----
consensus 96 ... ..

B


cbm2b_cons 1 STGCSVTATRAEEWSDRFNVTYSVSGSSAWTVNLALNGSQTIQASWNANVTGSGSTRTVTPNGSGNTFGVTVMKNGSSTT
cf_xynD_2b 1 STGCSVTATRAEEWSDRFNVTYSVSGSSAWTVNLALNGSQTIQASWNANVTGSGSTRTVTPNGSGNTFGVTVMKNGSSTT
sl_xynB_2b 1 --GCTATVSAGQKWGDRYNLDVSVSGASDWTVTMNVPSPAKVLSNWNVNASPSAQTLTARLNGSGNNWGATIQANANWTW
consensus 1 ..**..*.....*.**.*...****.*.***.............**.*...*..*.*...*****..*.*...*...*.

Cloning, expression and analysis of the predicted endoglucanase Cel5A

Cel5A was studied further because of its intriguing features. In addition to its three CBM6s, Cel5A contains dual GH5 modules, both of which annotate as endoglucanases although their sequences are dissimilar (37 % identity and 54 % similarity in amino acid sequence; 60.9 % identical nucleotide sequence). Cel5A was cloned, expressed and assayed for activity. Zymograms indicated that Cel5A:His_6 was active on HE-cellulose and barley β-glucan (Figure 2-2). Reducing sugar assays revealed that it had approximately 10 and 20 times more activity vs. barley β-glucan than vs. CMC and PASC, respectively (Table 2-8). No activity was detected against Avicel, xylan or laminarin. The activity of Cel5A:His_6 against pNP-β-cellobioside and absence of activity against pNP-β-glucopyranoside suggest that the enzyme does not cleave cello-oligos of less than three residues and corroborates the results from reducing-sugar assays. Thus it is concluded that Cel5A is a β-1,3(4) endoglucanase with secondary activity on amorphous cellulose chains.

Table 2-8. Activity assays of cloned enzymes

Substrate[a,b]	Cel5A:His_6	Gly5K:His_6
Barley β-Glucan	3.31 (1.50)[c]	N.D.[d]
CMC	0.34 (0.11)[c]	N.D.[d]
PASC	0.19 (0.10)[c]	N.D.[d]
p-NP-cellobiose	1.24 (0.05)[c]	N.D.[d]
p-NP-α-Arabinopyranoside	N.D.[d]	0.11 (0.004)[e]
p-NP-α-Arabinofuranoside	N.D.[d]	0.11 (0.02)[e]
p-NP-β-xylopyranoside	N.D.[d]	0.08 (0.003)[e]
p-NP-β-Glucopyranoside	N.D.[d]	0.10 (0.004)[e]

Approximately 5 ng of each protein was added to the indicated substrate solution and incubated as described in the materials and methods. Specific activity was calculated and expressed as U/mg protein where 1 U = 1 μmol reducing sugar or para-Nitrophenol produced per minute.

[a] CMC, carboxymethylcellulose; PASC, phosphoric acid swollen cellulose; p-NP, para-nitrophenol.

[b] Neither enzyme was active against avicel, laminarin, xylan, p-NP-β-arabinopyranoside, p-NP-α-glucopyranoside, or p-NP-α-xylopyranoside.

[c] (U/mg protein) $x10^2$ as glucose reducing equivalents, determined by Nelson-Somogyi reducing sugar assay (20)

[d] N.D., Not detected.

[e] (U/mg protein)$x10^2$ as nitrophenol released from p-NP derivative. Assays were run in triplicate; standard deviation ($x10^2$) is shown parenthetically.

CBM proteins

Analysis of the proteins which incorporate a CBM, but not an annotated carbohydrase domain, reveals that they all have a type II secretion signal and most have extended regions of unknown sequence large enough to accommodate binding and/or catalytic modules. Many of these CBM proteins contain unusual sequence features. In Cbm2C, Cbm2D-Cbm10A, Cbm2E, Cbm2F, and Cbm6A, a number of binding modules are separated from the rest of the protein by PSLs (Table 2-4). Other proteins with unusual features include Cbm2A, with a glutamate-proline repeat (EPR), and Cbm2B containing a 914 amino acid region which has no homolog in the non-redundant database of the NCBI at the time of this writing. Cbm2C and two predicted xylanases contain a novel domain, designated Y94, which is thus far unique to *S. degradans*. RP-HPLC MS/MS analyses detected Cbm2C xylan-grown culture supernatants and Cbm2B in avicel-grown cultures (Table 2-6).

Cbm6A and Cbm6E have lipobox motifs. Two other CBM6 proteins include an unusual second binding module; Cbm6G-Cbm16B and Cbm6H-Cbm32F contain a family 16 and family 32 CBM, respectively. There are four other ORFs with CBM32s, three of which have multiple copies of the binding module. Cbm32A was detected by MS/MS in xylan-grown supernatant and contains tandem CBM32s followed by a 251 amino acid unknown region and a thiol oxidoreductase-like module (Table 2-4). Cbm32D has an N-terminal CBM32 followed by a novel domain, designated FCL, which has homology to fucose-binding lectins that are found in serum proteins of the Japanese eel, *Anguilla japonica.* An INTERPRO database search

(http://www.ebi.ac.uk/interpro/displayIproEntry?ac=IPR006585) did not identify any prokaryotic examples of this domain except for *S. degradans*.

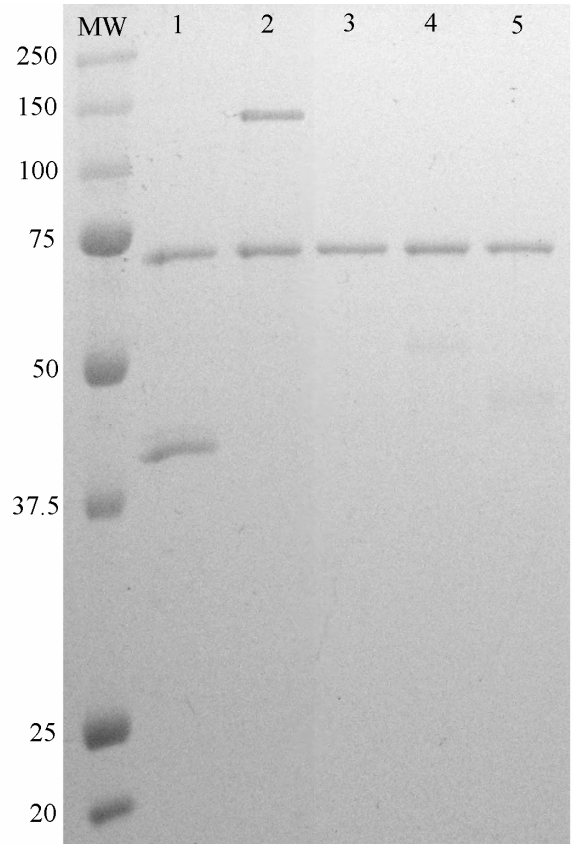
In order to assess whether the cloned CBM proteins were successfully renatured following expression and purification, and to verify the predicted binding of their CBM2a modules, a novel cellulose-binding assay was developed, in which Avicel was pre-incubated with BSA, blocking nonspecific protein binding. Cbm2A:His_6, Cbm2B:His_6 and Cbm2C:His_6 were subjected to this Avicel-binding assay, examined in zymograms and checked for enzyme activity. All three of these fusion proteins specifically bound to Avicel (Figure 2-4), and none exhibited detectable enzymatic activity.

Figure 2-4. Cbm2-containing proteins specifically bind to Avicel.

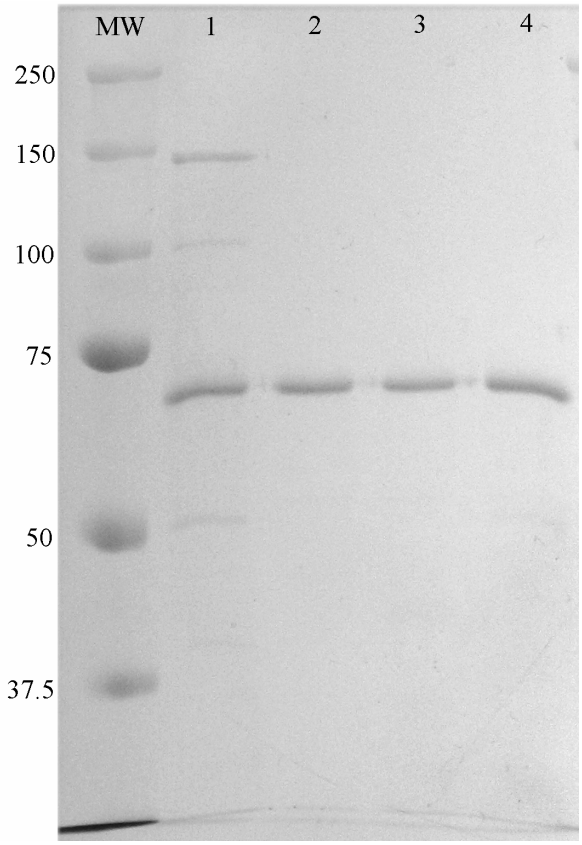
A) 11% SDS-PAGE gel. Sample proteins were incubated with BSA-treated Avicel as described in materials and methods. Arrows indicate test protein bands. Lanes: 1- Cbm2A:His_6; 2- Cbm2B:His_6; 3- Pepsin; Porcine Lipase and Ovalbumen controls not shown.

B) 8% SDS-PAGE gel prepared as in panel A. Lanes: 1- rCbm2C; 2- Pepsin; 3- Porcine Lipase; Ovalbumen control not shown. For reasons yet to be determined a number of cloned 2-40 proteins, including Cbm2B and Cbm2C, consistently run higher in SDS-PAGE then their predicted MW.

A



B



Glycanases of uncertain function

The glycanases for which precise substrate specificity cannot be predicted by sequence analyses belong to five different well-characterized GH families (Table 2-5), although many deviate from the typical consensus sequence. For example, the catalytic domains of Gly5K, Gly5L and Gly5M annotate as possible β-1,3 or β-1,4 endoglucanases, but are not well-enough conserved to permit more specific prediction of their substrate specificity. Similarly, Gly5R is distantly related to known β-mannanases, and the sequence of Gly9C is only 25 % conserved with known GH9 cellulases.

There are also glycanases with modules from less-well characterized families, such as Gly26B, Gly30A and Gly30B. Approximately half of the glycanases contain one or more CBM from families 6, 13, or 32, suggesting that these enzymes bind to glycan chains or oligosaccharides. While CBM families 6 and 13 are well-characterized, relatively little is known about CBM family 32, quadruply represented in Gly3D. Unlike the cellulases and CBM proteins, none of the glycanases contain PSLs, however there are other unusual motifs. Gly5M contains a 34 amino acid glutamate-proline repeat (EPR), and Gly43M carries two modules (designated LGL; Table 2-5) with homology (BLASTP; 2e-12, 2e-18) to concanavalin A-like lectins, which are known to bind to glucose and mannose (87).

To determine the function of some of these glycanases, Gly3D, Gly5K, Gy5M, Gly9C and Gly43M were cloned, expressed and assayed for activity (Table 2-5). Amino acid sequence comparisons showed that the active sites of Gly5K and Gly5L were so similar to each other as to share common substrate specificity. Accordingly, Gly5K was the chosen representative for cloning and functional analyses. Gly3D:His_6, Gly5M:His_6,

Gly9C:His_6, and Gly43M:his_6 were not active in the chosen zymograms or enzyme assays, and their substrate specificities remain to be determined, although Gly3D:his_6 was detected by MS/MS in xylan-grown supernatants (Table 2-6), implying a role in plant wall degradation. On the other hand, Gly5K:His_6 was moderately active against p-NP conjugated α-L-arabinofuranoside, α-L-arabinopyranoside, β-D-glucopyranoside, and -β-D-xylopyranoside (Table 2-8), suggesting that Gly5K acts on arabinoxylan oligomers.

Discussion

Enzyme complement and domain analysis

The degradation of complex polysaccharides often requires consortia of enzymes with diverse and complementary activities. Genomic and functional analyses reveal that *S. degradans* has ten distinct CP-degrading systems (41) and more carbohydrases and accessory proteins than any marine bacterium studied thus far. To our knowledge, *S. degradans* is the first marine bacterium shown to have complete systems for metabolizing every major polysaccharide component of the plant cell wall, including cellulose, common hemicelluloses (xylans, arabinans, glucans and β-mannans), and pectin. This is particularly remarkable in the context of its other complete systems which metabolize other polysaccharides commonly found in the marine environment: agar, alginate and chitin (41, 57).

It is also noteworthy that the genome of *S. degradans* contains more cellulases than many other cellulolytic organisms, including the cellulolytic fungus *Trichoderma reesei,* which has seven cellulases (two cellobiohydrolases and five endoglucanases (42).

Examples of cellulase systems from genomically-sequenced bacteria include the soil organisms *Cytophaga hutchinsonii, Clostridum thermocellum,* and *Thermobifida fusca*, and the marine bacterium, *Thermotoga maritima* (Table 2-7). Like *Thermobifida fusca* and *Clostridium thermocellum*, *S. degradans* has a non-reducing end cellobiohydrolase (106, 116). However, unlike these*, S. degradans* does not have a GH48 reducing-end cellobiohydrolase.

One reason for the high number of cellulases encoded by *S. degradans* may be that many of the enzymes act synergistically. In such cases particular enzymes may not exhibit high specific activity when acting alone, but may be far more effective in combination(s). This may explain why the observed activities of Cel5A:His_6 are about 1/100 of those observed for GH5 cellulases from fungal species such as *Orpinomyces joyonii*, in which two such cellulases were shown to have specific activities of around 3.2 U/mg against β-glucan (90).

Studies of organisms such as *T. fusca* and *T. reesei* have strongly suggested that the synergistic activities of endo and exo-acting cellobiohydrolases combined with endoglucanases are essential for efficient cellulose utilization (6, 111). This perceived requirement is often extended to cellulolytic microorganisms in general. However, accumulating evidence suggests that there may be other operative mechanisms, as well. For example, the Gram-negative gliding soil bacterium, *Cytophaga hutchinsonii,* is regarded as an efficient cellulolytic bacterium (75) despite the absence of either a GH6 or GH48 cellobiohydrolase (Table 2-7). Furthermore, a recent study of *Clostridium thermocellum* indicates that during growth on Avicel, cellotetrose is the major oligomer imported into the cytoplasm, where further metabolism occurs by phosphorolytic

cleavage (117). The cellulase system of *C. thermocellum* is reported to contain 12 endoglucanases, as compared to only two cellobiohydrolases. Thus, it appears that the combined activities of the abundant endoglucanases produce the bulk of assimilated carbon during cellulose metabolism (117). In such a system, cellobiohydrolases may function more in promoting the breakdown of particularly recalcitrant crystalline regions than in the production of cellobiose as a primary energy source. *S. degradans* also has an abundance of endoglucanases and only one cellobiohydrolase. This disparity, and the presence of cellodextrin and cellobiose phosphorylases (Table 2-7), suggests that *S. degradans* cellulose metabolism proceeds, at least in part, through the uptake of endoglucanase-derived oligomers (n>2) and that phosphorolytic cleavage is also an operative mechanism.

Most *S. degradans* cellulase components appear to be freely secreted, as indicated by type II secretion signals. However, two cellulases, the cellodextrinases, and a cellobiase have lipobox sequences (Table 2-3), suggesting that they are anchored to the cell surface (31, 88, 99). Within the cellulase system, the lipoprotein enzymes are specific for single glycan chains and cello-oligomers. Thus, it appears that the extracellular enzymes which process intermediate-length glucan chains into cello-oligos and glucose are maintained on the cell surface, thereby minimizing substrate and enzyme loss through diffusion (12, 95, 106).

Many carbohydrases of *S. degradans* have features which are unique and others that are unusual among prokaryotes. In addition to well-predicted plant wall carbohydrases, there are numerous ORFs with binding and/or catalytic modules for which precise substrate specificity cannot be predicted due to insufficient relatedness to a well

characterized enzyme. It has long been noticed that the sequence-based families of glycoside hydrolases often contain enzymes of varying substrate specificity (51), making predictions of substrate specificity difficult for distantly related members (28, 53).

S. degradans encodes many such genes. Although GH5 is among the best studied CAZyme families, it has five GH5 modules for which substrate specificity cannot be predicted by sequence analysis. Of the two GH5 proteins which were cloned and assayed in this study, Gly5M:His_6 was not active on any tested substrate, and Gly5K:His_6 was moderately active on arabinoside, xyloside, and glucoside p-NP conjugates, but not on laminarin, β-glucan, xylan or three kinds of cellulose. These findings suggest that the substrates of these GH5 glycanases, much like their sequences, are unlike those of previously characterized GH5 cellulases. This is also the likely case for Gly9C and Gly43M—both of which were inactive in reducing-sugar and p-Np assays. Thus, further study of *S. degradans* carbohydrases should extend our knowledge of even well-studied GH families.

Although it is possible that one or more of the expressed recombinant proteins failed to refold properly, evidence suggests that many of them retained or recovered functionality. For example, Cel5A:His_6 was catalytically active and Cbm2A:His_6, Cbm2B:His_6, and Cbm2C:His_6 exhibited specific binding to Avicel. Including the detected activities of Gly5K:His_6, these results show that four of the nine expressed proteins exhibited their predicted function and a fifth was active against arabinoside and xyloside p-Np linkages. Therefore it is probable that Gly3D:His_6, Gly5M:His_6, Gly9C:His_6, and Gly43M:His_6 were expressed in a functional state and that they act on untested substrates.

There are also a number of *S. degradans* genes with CBM and/or GH modules from families which count only a few characterized members and others which contain apparently novel catalytic or binding domains. Genes from less-characterized families include those with GH26 and GH30 modules. Although most studied GH26 modules are β-mannanases, Gly26B is only a distantly relative. The specificities of the proteins with GH30 modules are also uncertain. Those GH30 modules which have been functionally characterized are eukaryotic in origin, and are either β-1,6-glucanases or β-glucoceramidases (27). While Gly30B does not contain a CBM, Gly30A has two CBM6s and a CBM13, suggesting binding to individual glycan chains and oligosaccharides. These predicted binding affinities are consistent with a β-1,6-glucanase activity. Functional characterization of Gly30A and/or Gly30B should enhance understanding of this GH family.

S. degradans encodes four distinct modules which might be novel CBMs. These include the fucolectin-like module (FCL) in Cbm32D (Table 2-4) and the concanavalinA lectin-like modules of Gly43M (Table 2-5). Two other candidate CBMs are the conserved modules designated Y94 (Figure 2-5) and Y95 (Figure 2-6). Y94 modules are about 130 amino acids in length and are found in Cbm2C, which was identified by Mass Spectrometry in supernatant from *S. degradans* grown with xylan as the sole carbon source. Y94 modules also appear in two predicted xylanases, which together with the MS/MS data, suggests a possible xylan-binding role. Y95 modules are about 95 amino acids in length and occur in Cel5G (Table 2-3) as well as in two poorly-conserved polysaccharide lyases.

Alternatively, it is possible that one or more of these unknown modules are involved in specific attachment of proteins to the bacterial cell surface as exemplified by the dockerin-cohesin interaction in *C. thermocellum* (12). Surface attachments in Gram negative bacteria are not solely mediated by lipobox sequences (9, 62, 104).

Figure 2-5. ClustalW alignment of conserved domains designated Y94

Conserved Y94 domains of unknown function in the *S. degradans* proteins Xyn10A, Cbm2C, and Xyl/Arb43G-Xyn10D.

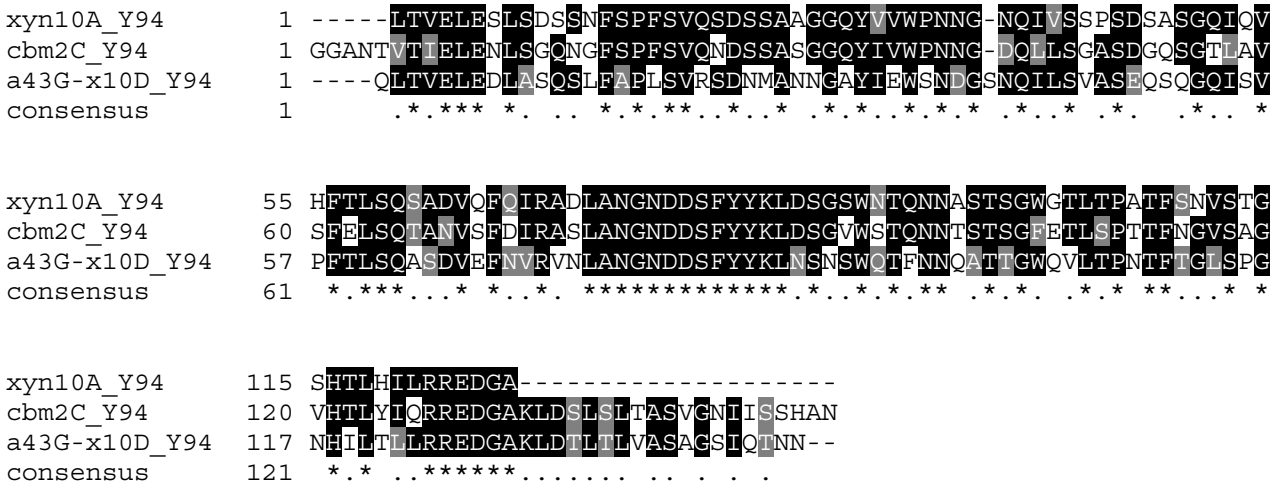
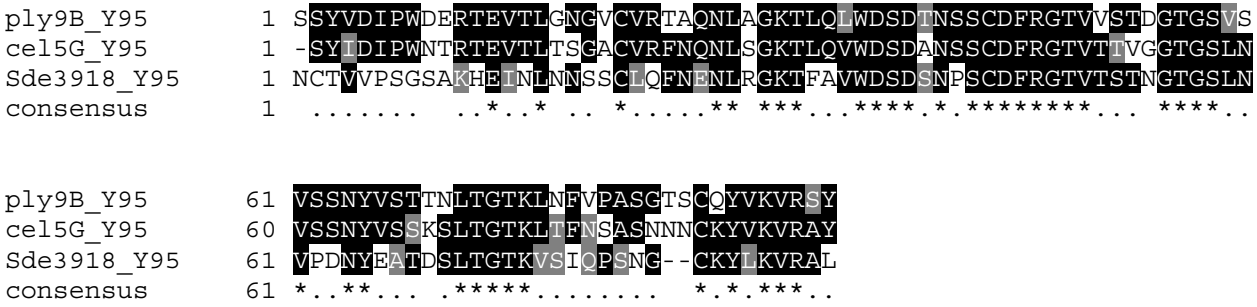

xyn10A_Y94 1 -----LTVELESLSDSSNFSPFSVQSDSSAAGGQYVVWPNNG-NQIVSSPSDSASGQIQV
cbm2C_Y94 1 GGANTVTIELENLSGQNGFSPFSVQNDSSASGGQYIVWPNNG-DQLLSGASDGQSGTLAV
a43G-x10D_Y94 1 ----QLTVELEDLASQSLFAPLSVRSDNMANNGAYIEWSNDGSNQILSVASEQSQGQISV
consensus 1 .*.*** *. .. *.*.**..*..* .*.*..*.*.* .*..* .*. .*.. *
xyn10A_Y94 55 HFTLSQSADVQFQIRADLANGNDDSFYYKLDSGSWNTQNNASTSGWGTLTPATFSNVSTG
cbm2C_Y94 60 SFELSQTANVSFDIRASLANGNDDSFYYKLDSGVWSTQNNTSTSGFETLSPTTFNGVSAG
a43G-x10D_Y94 57 PFTLSQASDVEFNVRVNLANGNDDSFYYKLNSNSWQTFNNQATTGWQVLTPNTFTGLSPG
consensus 61 *.***...* *..*. **********.*..*.*.** .*.*. .*.* **...* *
xyn10A_Y94 115 SHTLHILRREDGA--------------------
cbm2C_Y94 120 VHTLYIQRREDGAKLDSLSLTASVGNIISSHAN
a43G-x10D_Y94 117 NHILTLLRREDGAKLDTLTLVASAGSIQTNN--
consensus 121 *.* ..******..... .. . . .

Figure 2-6. ClustalW alignment of conserved domains designated Y95.

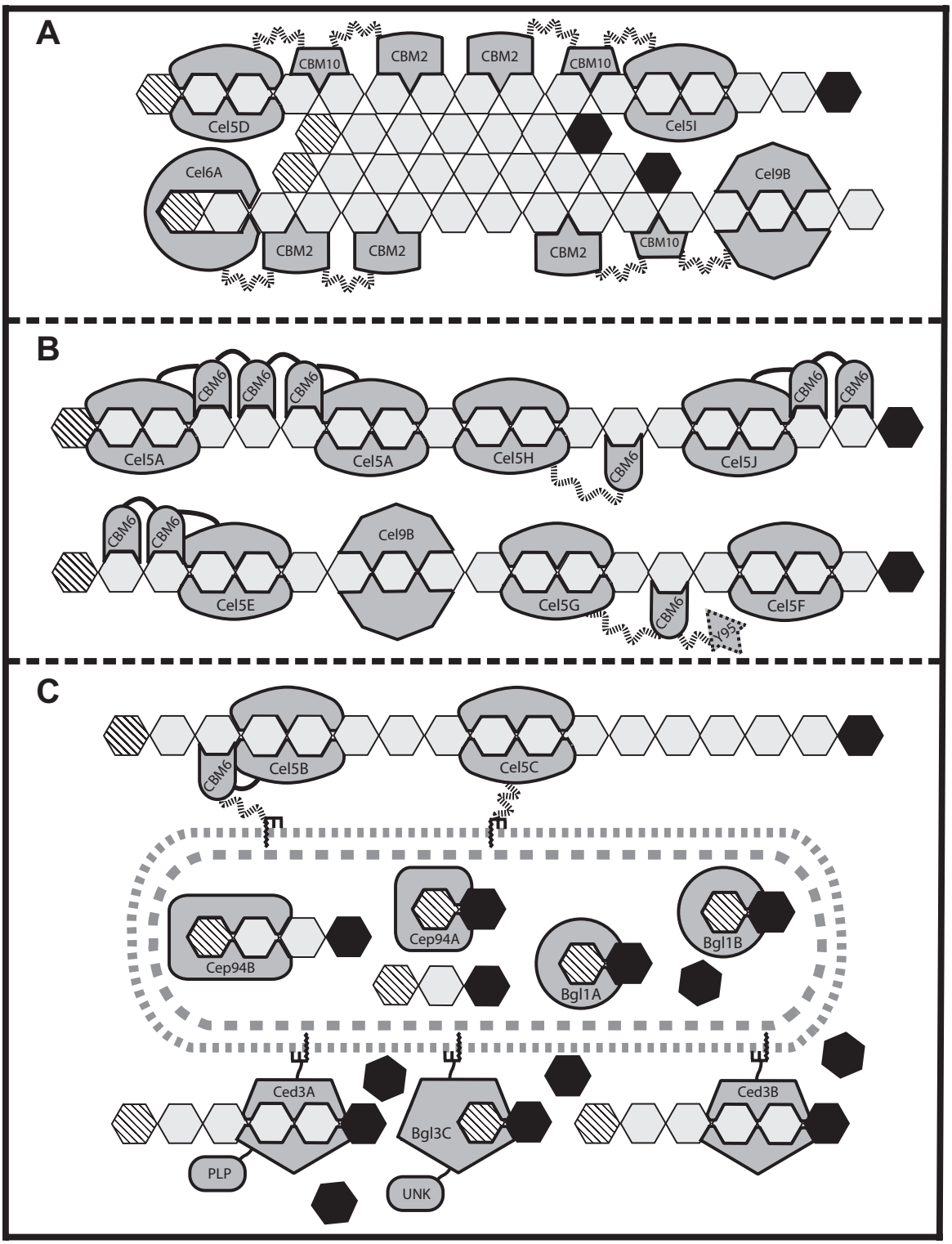
Conserved Y95 domains of unknown function in the *S. degradans* proteins Ply9B, Cel5G, and Sde3918, a putative alginate lyase


ply9B_Y95 1 SSYVDIPWDERTEVTLGNGVCVRTAQNLAGKTLQLWDSDTNSSCDFRGTVVSTDGTGSVS
cel5G_Y95 1 -SYIDIPWNTRTEVTLTSGACVRFNQNLSGKTLQVWDSDANSSCDFRGTVTIVGGTGSLN
Sde3918_Y95 1 NCTVVPSGSAKHEINLNNSSCLQFNENLRGKTFAVWDSDSNPSCDFRGTVTSTNGTGSLN
consensus 1 ....*..* .. *.....** ***...****.*.********... ****..
ply9B_Y95 61 VSSNYVSTTNLTGTKLNFVPASGTSCQYVKVRSY
cel5G_Y95 60 VSSNYVSSKSLTGTKLIFNSASNNNCKYVKVRAY
Sde3918_Y95 61 VPDNYEATDSLTGTKVSIQPSNG--CKYLKVRAL
consensus 61 *..**... .*****........ *.*.***..

Functional analysis of the S. degradans cellulase system

The general characteristics of the cellulase system of *S. degradans* can be inferred from the analysis of its proteome with the CAZy database tools. Annotations and analyses of the GH and CBM modules within an ORF allow prediction of its catalytic and binding activities. Furthermore, analyses of lipoprotein or secretion signal sequences, if present, suggest where the gene product will perform its expected function. From this, the predicted location of each enzyme relative to the cell and its likely substrate are presented in Figure 2-7.

Figure 2-7. **Representation of cellulose degradation by *S. degradans*.** The diagram shows the predicted cellulases of *S. degradans* participating in three distinct phases of cellulose degradation, based on their expected location relative to the cell, as well as their enzymatic and substrate binding specificities. In panel **A**, a region of crystalline cellulose located distal from the cell is attacked by the non-reducing end cellobiohydrolase, Cel6A, and endoglucanases Cel5D, Cel5I and Cel9B. Panel **B** shows the secreted cellulases which contain CBM6 modules (or lack a CBM) acting on freely accessible cellulose chains at an intermediate distance from the cell surface. In panel **C**, the lipoprotein-anchored components are shown processing cello-oligomers. Endoglucanases Cel5B and Cel5C likely bind and degrade single cellulose chains at the cell surface, while the cellodextrinases cleave cellodextrins to generate glucose and (n-1) cello-oligomers. Extracellular cellobiose is cleaved to glucose by Bgl3C. Additionally, within the cell cytoplasm, the cellobiose and cellodextrin phosphorylases are shown bound to their substrates, as are the cytoplasmic cellobiases, Bgl1A and Bgl1B. Internal glucose residues of cellulose chains are represented by open hexagons, while the non-reducing ends are indicated by diagonal lined hexagons, with reducing ends shown as solid filled hexes. The bacterial inner and outer membranes are indicated by shaded dashed lines. Enzyme names (i.e. Cel9B) are indicated within catalytic module, while CBM family type is indicated within CBM. Wavy, dashed lines represent polyserine linkers (PSLs), and solid lines represent non-polyserine linker regions. Lipoprotein anchors are shown embedded in the outer membrane with the lipid moiety represented by the zigzag line. Cel5H, Cel5I, Cel9B, Ced3A, Ced3B, Bgl3C and Cep94B were detected by RPHPLC-MS/MS in cellulose and/or xylan-grown culture supernatants.


A
CBM10
CBM2
CBM2
CBM10
Cel5D
Cel5I
Cel6A
Cel9B
CBM2
CBM2
CBM2
CBM10
B
CBM6
CBM6
CBM6
Cel5A
Cel5A
Cel5H
CBM6
Cel5J
CBM6
CBM6
CBM6
CBM6
Cel5E
Cel9B
Cel5G
CBM6
Cel5F
Cel95
C
CBM6
Cel5B
Cel5C
Cep94B
Cep94A
Bgl1A
Bgl1B
Ced3A
Bgl3C
Ced3B
PLP
UNK

As the roles of the CBM-only proteins remain unclear, they were not included in Figure 2-7 even though three of them were detected by MS/MS analyses of Avicel or xylan grown culture supernatants (Table 2-6) and cellulose-binding assays confirmed that Cbm2A, Cbm2B and Cbm2C specifically bind to Avicel. Perhaps one or more of the many CBM proteins enhance cellulose degradation through a "disruption" mechanism similar to that recently shown for the non-hydrolytic, chitin-binding protein CBP21 of *Serratia marcescens* (108). Interestingly, and like *S. marcescens*, the *S. degradans* chitinase system also contains a CBM33 containing chitin-binding protein that is believed to assist in the degradation of chitin (57). To this end, various CBM proteins of *S. degradans* may be bacterial equivalents of plant expansins (24) or the swollenin protein of *T. reesei* (94). These proteins also contain CBMs yet lack catalytic modules, and there is evidence that swollenins act to disrupt the surface of crystalline cellulose, thereby increasing the availability of cellulose chains for enzymatic attack (94).

The model calls for initiation of cellulose degradation by the freely secreted endoglucanases which bind insoluble cellulose, namely Cel5D, Cel5I, and Cel9B (Figure 2-7, A). These endoglucanases attack amorphous regions within insoluble cellulose, thereby generating shorter cellulose chains and increasing the availability of non-reducing ends for attack by cellobiohydrolase Cel6A. Although a reducing-end cellobiohydrolase has not been found by sequence analysis, the cellulase system of *S. degradans* is quite functional, permitting growth on Avicel and even filter paper as sole carbon sources. Therefore, it appears that its numerous endoglucanases act synergistically to compensate for the lacking reducing-end cellobiohydrolase in a manner similar to that discussed above for *C. thermocellum*. Alternatively it is possible, if unlikely, that there is

a cellobiohydrolase which was not identified in the analyses of the recently closed genome. Such an enzyme may eventually be found in one of the glycanases of uncertain substrate specificity or among the unknown sequence within a CBM-only protein.

Continuing with the cellulose degradation model, there are five freely secreted endoglucanases with CBM6s and two which do not have a CBM. Because of their CBM specificity (or lack of CBM) these enzymes are expected to act on more accessible forms of cellulose such as individual chains or extended amorphous regions (Figure 2-7, B). Six of these "secondary" endoglucanases are from GH family 5, the other is a GH9. These enzymes cleave cellulose to shorter, random-length chains and cellodextrins. While it is possible that one or more of the endoglucanases in this group may generate oligomers which are acted on by other endoglucanases, no such functional hierarchy can be inferred by sequence analysis.

Some of the cellulose chain fragments and cellodextrins generated by *S. degradans* enzymes are likely processed at the cell-surface by the lipoproteins: endoglucanases Cel5B and Cel5C, the cellodextrinases Ced3A and Ced3B, and the cellobiase Bgl3C (Fig. 3*C*). Cel5B and Cel5C hydrolyze individual cellulose chains to cellodextrins, which are cleaved to glucose and cello-oligos by Ced3A and Ced3B. Cellobiose is hydrolyzed to glucose at the cell surface by Bgl3C and then imported to the cytoplasm. Alternatively, short cellodextrins and cellobiose may be imported into the cytoplasm for further metabolism by cellobiases and/or phosphorylases (Figure 2-7, C).

The presence of the cellobiases and phosphorylases suggests that in addition to the surface-bound enzymes there are also two distinct cytoplasmic pathways for cello-oligo metabolism, both of which rely on energy-dependent import of cellobiose and/or

cellodextrins. In the hydrolytic pathway, cellobiose is cleaved to glucose by either of the two cellobiases, Bgl1A and Bgl1B. In the phosphorolytic pathway, cellobiose and cellodextrins are cleaved into glucose-1-phosphate and glucose or cellodextrin (n-1) by the cellobiose or cellodextrin phosphorylase (Figure 2-7, C). Although it is impossible to predict the extent to which each pathway contributes to cellulose metabolism, the energy savings afforded by phosphorolytic cleavage may be advantageous under certain conditions (98). The detection of Bgl3C in sole carbon source avicel, CMC, and xylan grown cultures by MS/MS, and the detection of Cep94B in CMC grown cultures, suggests that, in *S. degradans,* the surface-bound hydrolysis of cellobiose and the intracellular phosphorolytic cleavage of cellodextrins may occur simultaneously.

Another intriguing finding is illustrated in Table 2-6. Certain accessory proteins from the cellulase and xylanase systems appear to be induced during growth on cellulose and xylan. Cellodextrinase Ced3A was detected in supernatants of cultures grown in Avicel, CMC, and xylan, while Ced3B was detected in xylan grown cultures. Similarly, the β-xylosidase, Xyl3A, was detected in Avicel and CMC grown cultures in addition to xylan grown cultures. These findings suggest that certain components of the cellulase and xylanase systems may be induced by either substrate. Thus, it appears that the cellulase and xylanase systems may be coordinately regulated.

Collectively these data support the notion that *S. degradans* synthesizes a complete and functional cellulase system. Until now reports on the degradation of cellulose in marine environments have been descriptive. The discovery and molecular biochemical characterization, beginning with *S. degradans*, of an emerging group of

cellulolytic marine bacteria promises a more complete understanding of the marine carbon cycle.

Chapter 3: The Hemicellulolytic and Pectinolytic Systems of *Saccharophagus degradans* strain 2-40

Introduction

Saccharophagus degradans strain 2-40 is a representative of an emerging group of marine complex polysaccharide (CP) degrading bacteria. It is unique in its metabolic versatility, being able to degrade at least 10 distinct CP from diverse algal, plant and invertebrate sources. Because of its remarkable carbohydrate-degrading abilities, *S. degradans* was selected for genome sequencing by the US Department of Energy's Joint Genome Institute (JGI). The genome has recently been completed and more than 180 open reading frames have been identified that encode carbohydrases. Over half of these are likely to act on plant cell-wall polymers.

Previous work had shown that *S. degradans* 2-40 is capable of growth on non-cellulosic plant wall components, including xylan, β-glucan, mannan, and pectin (41). Furthermore, *S. degradans* is the first marine bacterium shown to grow on and completely degrade plant cell walls (chapter 2, this work). These findings are supported by recent genomic analyses which indicate that *S. degradans* 2-40 contains a number of ORFs with predicted activity against hemicellulose and pectin, As with the cellulases, there are numerous related hemicellulases and pectinases, many of which annotate to have complementary activities. In most cases, these annotations are of high confidence, allowing the research to focus on the expression of enzymes through zymogram or MS/MS analyses. The experimental portions of this work included the cloning of two laminarinases. Both were expressed and subjected to activity assays. The purpose of this

section is to describe the genomically predicted hemicellulose and pectin degrading enzymes, organizing them into putative systems, where possible. Analyses of these systems provide insight into how *S. degradans* degrades plant-wall polysaccharides, thereby contributing to the understanding of carbon recycling in the marine environment.

The plant cell wall is both chemically and structurally complex, containing cellulose, hemicellulose (including xylans, arabinans and mannans), and pectin (20, 111). Cellulose microfibrils, which act as the structural backbone of the plant cell wall, are embedded in a matrix of hemicelluloses (including xylans, arabinans and mannans), pectins (galacturonans and rhamnogalactans), and β-1,3 and β-1,4 glucans. Hemicellulose polymers are often substituted with arabinose, galactose, glucuronic acid and/or xylose residues, yielding arabinoxylans, glucuronoarabinoxylans, galactomannans, xyloglucans, and others (64, 73, 106, 111).

The particular polysaccharides that comprise hemicellulose are species and tissue dependent. The cell walls of the leaves of *Arabidopsis thaliana* are well characterized. Their major hemicellulose components are xyloglucans (XG) and glucuronoarabinoxylans (GAX) (115). Xyloglucans, which comprise ~20 % of the cell wall in dicots, have a branched β-1,4-glucan backbone on which as many as 75 % of the residues are substituted with an α-D-xylose residue at the O-6 position (77, 115). The xylose substituents may also be substituted at the C-2 position by α-L-fucose or β-D-galactose (often acetylated) (115). GAX comprises 5 % of the cell wall of *A. thaliana* and has a β-1,4-xylan backbone, in which up to 25 % of the xylose residues may be substituted at O-2 with α-L-arabinose, β-D-glucuronic acid or 4-O-methyl glucuronic acid (115). Other components may include arabinan, a polymer of α-1,5-linked arabinose

residues and galactoglucomannan (64, 69, 83). Galactoglucomannan has a backbone comprised of β-1,4-linked D-mannose and D-glucose residues which are substituted with D-galactose residues.

Acetylated galactoglucomannan is the major hemicellulose of softwood; other mannans, including unsubstituted mannan, glucomannan and galactomannan function as energy storage polymers (55, 69). Additionally, the cell walls of cereals and other grasses contain mixed-linkage (β-1,3(4)) glucans (115).

There are three major classes of pectins: Homogalacturonan, Rhamnogalacturonan-I (RG-I), and Rhamnogalcturonan-II (RG-II). Homogalacturonan is commonly referred to as pectin, or smooth pectin. It is a 1,4-linked chain of α-D-galacturonic acid residues, many of which are methylated. De-methylated pectin is known as pectate (30, 103). RG-I is a heteropolymer, with a backbone consisting of[-α-D-galacturonic acid-(1,2)-α-L-rhamnose-(1,4)-], which is substituted with linear and branched oligomers containing arabinose, galactose and/or glucuronic acid residues (115). RG-II has a homogalacturonan backbone, but is highly substituted, carrying as many as 30 different modified glycosyl residues (30, 76, 115).

As indicated above, non-cellulosic CP has great complexity and numerous glycosyl bond combinations so that its degradation requires enzyme systems that rival cellulase systems in count and complexity. Because of such heterogeneity and complexity, plant cell wall degradation often requires consortia of microorganisms (71, 106), however the findings reported here show that *Saccharophagus degradans*, alone, can degrade all of the components of the plant cell wall.

Materials and Methods

Genomic analyses

The *S. degradans* draft genome was analyzed by Dr Henrissat's group on the CAZy ModO (Carbohydrase Active enZyme Modular Organization) server at AFMB-CNRS (27). Catalytic and substrate binding modules were identified and named according to the CAZy nomenclature scheme (54). Open reading frames were also analyzed for presence of N-terminal secretion sequences using signalP (http://www.cbs.dtu.dk/services/SignalP/), for lipoprotein signature sequences using DOLOP (http://www.mrc-lmb.cam.ac.uk/genomes/dolop/) and lipoP (http://www.cbs.dtu.dk/services/LipoP/), and for transmembrane helices using TMpred (http://www.ch.embnet.org/software/TMPRED_form.html). Molecular weights were calculated using the ProtParam tool (http://us.expasy.org/tools/protparam.html).

Cloning and expression

Of the hemicellulases, two laminarinases were chosen for cloning. PCR was performed as previously described using the following primer sequences: Lam16A-AscI-F, AGGCGCGCCGAAAGTTAGCGTTGGCTTCAGGC; Lam16A-ClaI-R, CCATCGATATCGCGCTCGCTATTGCTTA; Lam16D-BamHI-F, CGGGATCCCTTAAGCCTATCTCTGCCCGTGT

All primers were purchased from Invitrogen (Frederick, MD). PCR reactions (50 µl) used standard parameters and conditions for tailed primers and Proof Pro® *Pfu* Polymerase (Continental Lab Products, San Diego, CA) and included 0.5 µl of *S. degradans* genomic DNA as the template. PCR products were cloned into pETBlue2, and

the resulting plasmids were transformed into *E coli* DH5α by electroporation and blue/white screened on LB/amp/X-gal. Plasmids were recovered, singly digested and visualized by agarose electrophoresis for size confirmation.

The plasmids were transformed into the expression strain by heat shock. *E coli* BL-21(DE3)pLysS, transformants were selected on LB agar containing ampicillin and chloramphenicol and incubated overnight at 37°C. Production of an appropriate-sized His-tagged protein was confirmed by comparing pre-induced and induced (1 mM IPTG) cell lysates in western blots using 1/5000 anti-HisTag® monoclonal primary antibody (Novagen, Madison, WI) and 1/7500 goat anti-mouse HRP conjugated secondary antibody (BioRad, Hercules, CA). Blots were developed colorimetrically with the OPTI-4CN kit (BioRad).

Production and purification of recombinant proteins

Expression cultures were grown to A_{600} of 0.6 to 0.8 in 500 ml or 1 liter broths of LB containing ampicillin and chloramphenicol. The cultures were then induced by 1mM IPTG incubated for four hours at 37°C or 16 hours at 20°C. Culture pellets were harvested by centrifugation (5000 x *g*, 20 min) and frozen overnight at -20°C. Cells were thawed on ice and suspended in 4ml urea lysis buffer (8 M Urea, 100 mM NaH_2PO_4, 25 mM Tris, pH 8.0) per gram wet pellet weight. The samples were clarified by centrifugation at 15,000g. The resulting supernatant was mixed with Nickel-NTA resin (QIAGEN, Valencia, CA) per the manufacturer's instructions. After 1 hour at 25°C, the slurries were washed twice with urea lysis buffer, pH 7.0.

Renaturation was performed on the column at 4°C using one column volume of renaturation buffer with decreasing urea concentrations (25 mM Tris pH 7.4, 500 mM NaCl, 20% glycerol, urea 6 M to 1 M in 1 M steps). The refolded proteins were eluted in 1 M urea, 25 mM Tris pH 7.4, 500 mM NaCl, 20% glycerol containing 250 mM imidazole. Void, wash and elution fractions were surveyed for HisTag® production in western blots as described above. Elution fractions containing the recombinant proteins were pooled and exchanged into Storage Buffer (20 mM Tris pH 7.4, 10 mM NaCl, 10% glycerol) using Centricon™ centrifugal ultrafiltration devices (Millipore, Bedford, MA). The enzyme preparations were then aliquoted and frozen at -80°C for use in activity assays. Final protein concentrations ranged from 500µg/ml to 2200µg/ml.

Zymograms

Gels were prepared as standard mini-format 8% SDS-PAGE gels (67) with 0.15 % Birchwood xylan incorporated directly into the separating gel. Each lane contained ~5 µg of recombinant protein or ~10 µg of filtered supernatant proteins from *S. degradans* cultures grown in MM containing 0.2 % xylan. After electrophoresis, gels were incubated 1 hr at room temperature in 80ml of renaturing buffer (20 mM PIPES buffer pH 6.8, 2.5 % Triton X-100, 2 mM DTT and 2.5 mM $CaCl_2$), after which they were held overnight at 4°C with gentle rocking. They were then equilibrated in 80 ml of 20 mM PIPES for 1 hour at room temperature, transferred fresh PIPES buffer and incubated for four hours at 37°C. Zymograms were stained for 30 minutes with 0.25 % Congo red in dH_2O, and destained with 1 M NaCl until clear bands were visible against a stained background.

Nelson-Somogyi reducing-sugar assays

Purified proteins were assayed for activity using a modification of the Nelson-Somogyi reducing sugar method adapted for 96-well microtiter plates, using 50 ul reaction volumes at 37°C (44). Standards consisted of glucose or xylose and autoclaved enzyme preparation were used as negative controls. Test substrates included Avicel, CMC, phosphoric-acid swollen cellulose (PASC), Barley glucan, laminarin, and xylan dissolved or suspended at 1 % (w/v) in 20 mM PIPES pH 6.8 (Barley glucan and laminarin, 0.5 %). Barley glucan, laminarin and xylan assays were incubated for 2 hours and Avicel, CMC and PASC assays were incubated 36 hours. Assay reactions were stopped by freezing at -80°C and kept frozen until analysis. Upon thawing, 75 µl of the copper reagent (four parts of 1:2:12:1.3 KNa tartarate: Na_2CO_3:$NaSO_4$:$NaHCO_3$ mixed with one part 1:9 $CuSO_4 \cdot 5H_2O$:Na_2SO_4) was added to the standards, reaction, and control tubes/wells and incubated 30 minutes at 80°C. After cooling to room temperature, 75 µl of the Arsenomolybdate reagent (25 g ammonium molybdate in 450 ml DH_2O + 21 ml H_2SO_4 + 3 g $Na_2HAsO_4 \cdot 7H_2O$ dissolved in 25 ml DH_2O) was added and mixed briefly by vortex. Trial assays revealed that the substrates CMC, xylan and barley glucan formed a precipitate which interfered upon addition of the assay reagents. As this precipitate interfered with A_{510} readings, assays using these substrates could not be performed directly in the microtiter plate wells. These assays were conducted in 0.6 ml siliconized microcentrifuge tubes. Reactions performed in 0.6 ml tubes were centrifuged at 14,000 x *g* and 200µl of the supernatant was transferred to a microplate well by pipette. Samples were assayed in triplicate, corrected for blank values, and levels estimated from a standard curve. Protein concentration of enzyme assay samples was measured in triplicate

using the Pierce BCA protein assay according to the manufacture's instructions. Enzymatic activity was calculated, with one unit (U) defined as 1 μmol of reducing sugar released/minute and reported as specific activity in U/mg protein.

Exoglycosidase activity assays: pnp-derivatives

Purified proteins were assayed for activity against para nitrophenol derivatives of α-L-arabinofuranoside, -α-L-arabinopyranoside, -β-L-arabinopyranoside, -β-D-cellobioside, -α-D-glucopyranoside, -β-D-glucopyranoside, -α-D-xylopyranoside and -β-D-xylopyranoside. 25 μl of enzyme solution was added to 125 μl of 5 mM substrate solution in 20mM PIPES pH 6.8, incubated for 30 minutes at 37°C, and A_{405} was determined. After correcting for blank reactions, readings were compared to a p-nitrophenol standard curve and reported as specific activities in U/mg protein, with one unit (U) defined as 1μmol p-Np/minute.

Mass Spectrometry and proteomic analyses

Supernatants from avicel, CMC, and xylan-grown cultures were concentrated to ~25X by centrifugal ultrafiltration using Centricon™ or Microcon™ devices (Millipore). Protein concentrations were determined using the BCA protein assay (Pierce). Samples were exchanged into 100 mM Tris buffer, pH 8.5, containing 8 M urea and 10 mM DTT and incubated 2 hours at 37°C to denature and reduce the proteins. After reduction, cysteine residues were alkylated by the addition of 1 M iodoacetate to a final concentration of 50mM and incubated at 25°C for 30 minutes. The samples were

exchanged into 50 mM Tris, 1 mM $CaCl_2$, pH 8.5 using Microcon™ devices. The denatured, reduced, and alkylated samples were digested overnight at 37°C using proteomics grade trypsin (Promega) at a 1:50 enzyme to substrate ratio. Digestions were stopped by the addition of 99% formic acid to a final concentration of 1 % and analyzed by RPHPLC-MS/MS at the UMCP College of Life Sciences CORE Mass Spectrometry facility using a Waters 2960 HPLC linked to a Finnagin LCQ tandem Mass Spectrometer. Alternatively, to identify specific proteins from SDS-PAGE gels of avicel-grown supernatants, regions containing 120 kDa, 100 kDa, and 75 kDa proteins were excised from an 8% gel and sent to Stanford University's Mass Spectrometry facility (SUMS; http://mass-spec.stanford.edu), where they were processed similarly and analyzed using an ESI-Quadrupole-Time of Flight (Micromass Q-Tof) mass spectrometer. All peptide fragment masses were analyzed by the peptide analysis packages SEQUEST and MASCOT (37, 86), and compared to amino acid sequence translations of all gene models in the 2-40 draft genome and to the non-redundant Mass Spectrometry Database (ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam). Peptide identity matches were evaluated using the accepted thresholds of statistical significance specific to each program.

Results

Overview of hemicellulose and pectin-degrading enzymes

The cellulase system of *S. degradans* is complemented by 57 genes which comprise five complete systems for the degradation and metabolism of the hemicellulose and pectin components commonly found in plant cell walls. These systems contain an

extensive battery of depolymerases and accessory enzymes predicted to act on heterogeneously decorated xylans and arabinans, β-1,3-glucanases (laminarinases), mannanases, and pectinases. These hemicellulose and pectin-degrading enzymes, for which ModO analysis resulted in confident predictions of function, were organized into putative systems (Tables 3-1, 3-2, 3-3, 3-4 and 3-5).

Identification of potential surface-associated enzymes

An unusually high percentage (28.5 %) of *S. degradans* carbohydrases contain lipobox signature sequences (Chapter 2). The hemicellulases and pectinases are no exception. Lipobox sequences are known to function in the localization of proteins to the outer membrane of a number of gram negative bacteria (80, 88). Lipoboxes occur in three depolymerases among the xylan and arabinan-degrading enzymes. These include an endo-xylanase, (Xyn10E), an exo-arabinosidase (Arb43B), and an endo-arabinanase (Arb43E). Six accessory enzymes in this group also carry lipoboxes: the β-xylosidase Xyl3A, all four predicted xylosidase/arabinosidases, and the α-xylosidase, Xyl31A, which is predicted to remove α-linked xylose substituents from xyloglucan chains (Tables 3-1 and 3-2).

Within the laminarinases, three depolymerases, Lam16F, Lam16G, and Lam81A, contain lipobox sequences (Table 3-3). Another predicted laminarinase, Lam16D, contains a predicted C-terminal transmembrane region, as identified by the TMpredict server (http://www.ch.embnet.org/software/TMPRED_form.html). The lipoprotein motif is particularly common in the pectinase system; seven of the 17 predicted pectinases or pectinase accessories have a lipobox sequence (Table 3-4). Five of these predicted lipoproteins are pectate depolymerases (Pel1D, Pel1E, Pel9A, Pel10B, and Pgl28A). The

other two are pectin methylesterases (Pes8A and Pes8B), which de-methylate pectin to pectate, thereby allowing depolymerization by pectate lyases (105). There is also a conserved lipobox sequence in the mannanase Man26A, and a possible lipobox as identified by LipoP 1.0, but with an atypical amino acid in the second position (LQAC) in another mannanase, Man5Q.

Table 3-1. Genomically predicted xylanases, xylosidases and related accessory proteins

Name[a]	Sde number[b]	Predicted function[c]	Modules[c,d]	amino acids[e]	MW[e]	MS[f]
Xyn10A	0181	β-1,4-xylanase	GH10/PSL(29)/Y94/PSL(21)/CBM5	574	61.7	
Xyn10B	2934	β-1,4-xylanase	CBM2/PSL(42)/CBM10/PSL(48)/GH10	619	65.0	
Xyn10C	2633	β-1,4-xylanase	GH10	374	42.3	
Xyl/arb43G-Xyn10D	3612	β-xylosidase/α-L-arabino furanosidase, β-1,4-xylanase	GH43/CBM6/Y94/PSL(24)/CBM2/PSL(22)/CBM22/GH10	1,186	129.6	
Xyn10E	0323	β-1,4-xylanase	LPB/EPR(47)/GH10	670	75.2	xn
Xyn11A	0701	β-1,4-xylanase	GH11	275	30.4	
Xyn11B-axe4A	3061	β-1,4-xylanase, acetylxylanesterase	GH11/PSL(10)/PSL(19)/CE4/CBM10	767	80.8	
Axe2A	2370	acetylxylanesterase	CE2	360	40.9	
Xyl31A	2500	α-xylosidase	LPB/GH31	973	110.2	xn
Xyl3A	1487	β-xylosidase	LPB/GH3	893	97.6	av cm xn
Xyl43L	0946	α-xylosidase	GH43	317	36.1	
Xyl/arb43H	0598	bifunctional β-xylosidase/α-L-arabinofuranosidase	LPB/GH43	577	63.8	
Xyl/arb43I	1655	bifunctional β-xylosidase/α-L-arabinofuranosidase	LPB/GH43	566	62.6	
Xyl/arb43J	0789	bifunctional β-xylosidase/α-L-arabinofuranosidase	LPB/GH43	317	35.1	
Xyl/arb43K	0822	bifunctional β-xylosidase/α-L-arabinofuranosidase	LPB/GH43	385	42.6	
Agu67A	1025	α-1,2-glucuronidase	GH67	738	82.0	

[a] Acronyms: xyn, xylan depolymerase,; xyl, xylosidase; agu, α-glucuronidase; axe, acetylxylanesterase; arb, arabinofuranosidase; xyl/arb, bifunctional enzyme mediating exo-cleavage of β-xylose and α-arabinose residues.

[b] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[c] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[d] Module abbreviations: CBM, carbohydrate binding module; CE, carbohydrate esterase; EPR, glutamic acid-proline rich region; GH, glycosyl hydrolase; LPB, lipobox signature sequence; PSL, polyserine linker; Y94, novel domain thus far unique to *S. degradans*

[e] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI amino acid sequence translations including any predicted signal peptide.

[f] Protein identified by tandem mass spectrometry in concentrated supernatant from cultures grown in Avicel (av) carboxymethylcellulose (cm) or xylan (xn).

Table 3-2. Genomically predicted arabinanases and arabinogalactanases of 2-40

Name[a]	Sde number[c]	Predicted function[d]	Modules[d,e]	amino acids[f]	MW[f]
Arb43A	2809	endo-1,5-a-L-arabinanase	CBM13/GH43	789	85.0
Arb43B	0787	exo-1,5-a-L-arabinofuranosidase	LPB/GH43	362	40.7
Arb43C[b]	0777	exo-1,5-a-L-arabinofuranosidase	GH43	314	35.9
Arb43D	1014	endo-1,5-a-L-arabinanase	CBM13/GH43	472	51.1
Arb43E	0786	endo-1,5-a-L-arabinanase	LPB/GH43	346	38.6
Arb43F	0791	exo-1,5-a-L-arabinofuranosidase	GH43	607	67.2
Arb51A	1767	exo-1,5-a-L-arabinofuranosidase	GH51	534	59.4
Arg53A	0683	arabinogalactanase	GH53	397	44.5
Arg53B	3710	arabinogalactanase	GH53/CBM13	476	52.6
Arg53C	2827	arabinogalactanase	GH53	650	71.1

[a] Acronyms: arb, arabinofuranosidase; arg, arabinogalactanase

[b] Arb43C does not have an N-terminal secretion signal

[c] Gene number having the prefix “Sde”, for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[d] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[e] Module abbreviations: CBM, carbohydrate binding module; GH, glycosyl hydrolase; LPB, lipobox signature sequence;

[f] MW and amino acid count calculated using the protParam tool (http://us.expasy.org/tools/) based on DOE/JGI amino acid sequence translations including any predicted signal peptide.

Table 3-3. Genomically predicted laminarinases of 2-40

Name[a]	Sde number[b]	Predicted function[c]	Modules[c,d]	amino acids[e]	MW[e]
Lam16A	1393	β-1,3-glucanase	GH16/CBM6/CBM6/TSP3/TSP3/TSP3/TSP3/COG3488	1,707	163.3
Lam16B	2927	β-1,3-glucanase	GH16/CBM6/CBM6/EPR(56)/CBM32/CBM32	1,441	158.6
Lam16C	1444	β-1,3-glucanase	GH16/CBM4/CBM32/CBM32	1,184	129.1
Lam16D	3021	β-1,3-glucanase	GH16/CBM32/PSL(48)/TMR	722	77.7
Lam16E	0652	β-1,3-glucanase	CBM6/CBM6/GH16	569	61.4
Lam16F	3121	β-1,3-glucanase	LPB/GH16	742	80.2
Lam16G	2832	catalytic residues missing	LPB/GH16/CBM6/CBM6	877	94.2
Lam81A	2834	β-1,3-glucanase	LPB/CAD/GH81/FN3/FN3	1,238	133.1

[a] Acronyms: lam, laminarinase.

[b] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[c] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[d] Module abbreviations: CAD, cadherin-like domain; CBM, carbohydrate binding module; COG3488, thiol-oxidoreductase like domain; EPR, glutamic acid-proline rich region; FN3, fibronectin type 3 module; GH, glycosyl hydrolase; LPB, lipobox signature sequence; PSL, polyserine linker; TMR, transmembrane region; TSP3, thrombospondin type 3 repeat.

[e] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI gene model amino acid sequence translations including any predicted signal peptide.

Table 3-4. Genomically predicted pectinases and accessory enzymes.

Name[a]	Sde number[b]	Predicted function[c]	Modules[c,d]	amino acids[e]	MW[e]
Pel1A	0943	pectate lyase	PL1/PSL(35)/PSL(29)/FN3/PSL(65)/PL1	1,316	136.4
Pel1B	0937	pectate lyase	PL1	427	46.1
Pel1C	0942	pectate lyase	PL1/PSL(33)/PSL(23)/FN3	769	78.9
Pel1D	3448	pectate lyase	LPB/EPR(24)/PL1	594	63.3
Pel1E	2307	pectate lyase	LPB/PL1	425	45.4
Pel1F	2311	pectate lyase	CBM2/PSL(45)/PSL(31)/PL1	772	81.5
Pel3A	2308	pectate lyase	CBM2/PSL(30)/PSL(20)/PL3	511	52.3
Pel3B	0608	pectate lyase	PSL(24)/FN3/PSL(20)/PL3	452	46.1
Pel3C	3007	pectate lyase	PSL(19)/FN3/PSL(18)/PL3	424	42.9
Pel3D	1703	pectate lyase	CBM13/PL3	392	41.6
Pel9A	2946	pectate lyase	PSL(47)/FN3/FN3/PL9	733	75.6
Pel10A	1051	pectate lyase	CBM2/PSL(54)/PSL(28)/PL10	700	73.3
Pel10B	2947	pectate lyase	LPB/PSL(46)/PL10	574	60.8
Pgl28A	0953	polygalacturonanase	LPB/GH28	463	50.9
Pes8A	3447	pectin methylesterase	LPB/EPR(23)/CE8	1,081	115.1
Pes8B	0944	pectin methylesterase	LPB/CE8	389	42.4
Rgl11A	1650	rhamnogalacturonan lyase	PL11/PSL(33)/CBM2	914	97.6

[a] Acronyms: pel, pectate lyase; pes, pectin methylesterase; pgl, polygalacturonanase; rgl, rhamnogalacturonanase

[b] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[c] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[d] Module abbreviations: CBM, carbohydrate binding module; CE, carbohydrate esterase; EPR, glutamic acid-proline rich region; FN3, fibronectin type 3 module; GH, glycosyl hydrolase; LPB, lipobox signature sequence; PL, polysaccharide lyase; PSL, polyserine linker.

[e] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI gene model amino acid sequence translations including any predicted signal peptide.

Table 3-5. Predicted mannanases of 2-40.

Name[a]	Sde number[b]	Predicted function[c]	Modules[c,d]	amino acids[e]	MW[e]	MS[f]
Man26A	3691	β-mannanase (E.C. 3.2.1.78)	LPB/PSL(52)/CBM10/GH26	506	54.9	
Man5N	0064	β-mannanase	GH5/PSL(32)/CBM10/PSL(44)/CBM2	561	57.6	
Man5O	0656	β-mannanase	GH5/PSL(23)/CBM10/CBM10	507	52.6	
Man5P	0509	β-mannosidase (E.C. 3.2.1.25)	GH5	457	50.9	cm
Man5Q	2541	β-mannanase	LQAC/GH5	850	92.3	ag
Agl27A	1593	α-galactosidase (E.C. 3.2.1.22)	GH27	408	46.5	

[a] Acronyms: agl, α-galactosidase; man, mannanase.

[b] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[c] Predictions of function and module determination by CAZy ModO at AFMB-CNRS

[d] Module abbreviations: CBM, carbohydrate binding module; GH, glycosyl hydrolase; LPB, lipobox signature sequence; LQAC, lipobox-like sequence; PSL, polyserine linker.

[e] MW and amino acid count calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI amino acid sequence translations including any predicted signal peptide.

[f] Protein identified by tandem mass spectrometry in concentrated supernatant from cultures grown in agarose (ag) or carboxymethylcellulose (cm).

Xylan and arabinan metabolism

Tables 3-1 and 3-2, show that *S. degradans* has 26 genes for the degradation of glucuronic acid and/or arabinose substituted xylan (the so called glucuronoarabinoxylans, or GAX), acetylated xylan and xyloglucans. There are seven endoxylanases, which cleave β-1,4-xylan chains, and a full complement of accessory enzymes or modules for the removal of various substituents. Four of the seven xylan depolymerases contain CBMs from families 2, 5 and 10 which bind crystalline substrates (Table 3-1). Interestingly, these same four ORFs also contain one or more extended regions comprised mainly of serine residues. An unusually high number of "polyserine linkers" (PSL) occur between binding and/or catalytic modules in *S. degradans* carbohydrases, and are thought to function as flexible spacers or linkers (56).

Two of the enzymes that annotated as xylan depolymerases are more aptly categorized as multifunctional enzymes, because they also carry accessory modules predicted to aid in degradation of substituted xylans. Of these, the architecture of Xyl/Arb43G-Xyn10D is particularly complex, with three CBMs, two PSLs, a Y94 domain (see Chapter 2), and two catalytic modules. The N-terminal GH43 module annotates as a bifunctional β-xylosidase/α-arabinosidase, while the C-terminal GH10 is a predicted endoxylanase. This catalytic complexity is matched by its predicted binding affinities: the CBM2a annotates as being specific for crystalline cellulose; CBM families 6 and 22 bind single glycan chains such as amorphous cellulose or xylan (29). The other multifunctional xylanase, Xyn11B-Axe4A, features a GH11 endoxylanase module, a CBM10 and a carbohydrate esterase (CE) family 4 acetylxylanesterase module predicted to deacetlyate xylan chains.

There are also a number of accessory proteins, including a β-xylosidase, an acetylxylanesterase, an α-1,2-glucuronidase for removal of the occasional α-linked

Glucuronic acid substituent occurring at the 2 position of xylan backbone residues, and an abundance of xylosidase/arabinosidase proteins containing glycosyl hydorlase family GH43. These GH43 modules are predicted to have a dual function, cleaving β-xylosidase bonds as well as α-arabinosidase bonds.

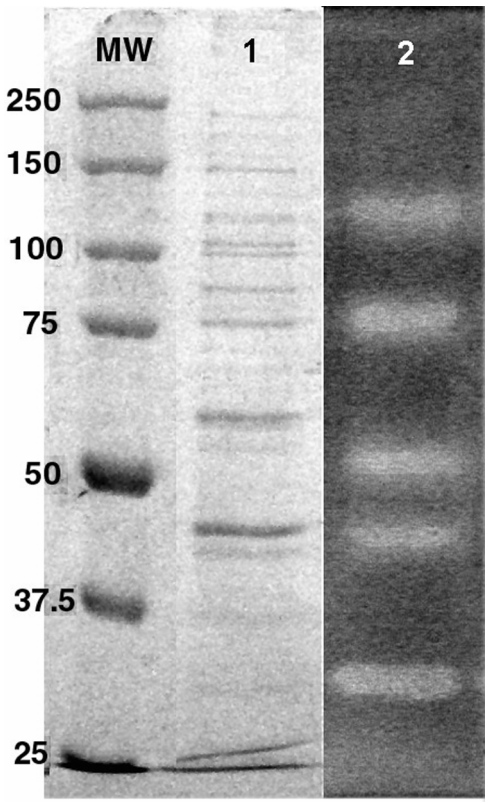
The genome also contains a number of ORFs with predicted specificity for polymers having an arabinose backbone. There are three endo-α-arabinanases, Arb43A and Arb43D and Arb43E, and four exo-α-arabinofuranosidases, Arb43B, Arb43C, Arb43F, and Arb51A. Arb43C does not have an obvious secretion signal, suggesting that it is localized in the cytoplasm. There are also three arabinogalactanases from the GH53 family: Arg53A, Arg53B and Arg53C. These normally degrade substituted arabinans and galactans.

The activity of several predicted xylanases was confirmed by zymogram. Xylanase activity gels reveal five apparent endoxylanases in supernatants of xylan-grown cultures (Figure 3-1). The xylanase activity bands can be correlated to predicted enzymes. For example, there is only one ORF over 100 kDa with predicted endoxylanases activity, Xyl/Arb43-Xyn10D, which at 130 kDa correlates with an activity band of the same weight. Similarly, the ~75 kDa activity band is most likely Xyn10E, with a predicted MW of 75 kDa. The ~55 kDa activity band is not as easily identified, but perhaps may be explained by either the 65 kDa Xyn10B or Xyn10A, at 62 kDa. The activity bands at ~45 kDa and ~30 kDa appear to correlate with Xyn10C and Xyn11A, at 42.3 kDa and 30.4 kDa, respectively.

In order to identify some of the plant-wall carbohydrases, Avicel, CMC, and xylan-grown culture supernatants were analyzed by Mass Spectrometry. One xylan depolymerase, Xyn10E, and an α-xylosidase, Xyl31A, were detected in xylan-grown samples. A β-xylosidase, Xyl3A, was detected in xylan, Avicel and CMC-grown cultures (Table 3-1).

Figure 3-1. Zymogram of xylan-grown *S. degradans* culture supernatant

10 μg xylan grown culture supernatant was electrophoresed into an 8 % SDS-PAGE containing 0.15 % xylan. After renaturation overnight and incubation at 37°C gels were stained in 0.1 % Congo red and destained in 1M NaCl. Lane MW, BioRad (Hercules, CA) precision pre stained molecular weight markers; Lane 1, Coomassie Blue stain of supernatant from stationary phase culture of cells grown in xylan; Lane 2, Zymogram of supernatant from stationary phase culture of cells grown in xylan


MW
1
2
250
150
100
75
50
37.5
25

Laminarin (β-1,3-glucan) metabolism

S. degradans encodes eight endo-β-1,3-glucanases, commonly referred to as laminarinases, seven of which contain GH16 modules while the other has a GH81(Table 3-3). All of the laminarinases have a type II secretion signal sequence. In addition to their diverse CBMs, many enzymes in this group contain other unusual domains, including thrombospondin type 3 (TSP3) and cadherin-like (CADG) calcium-binding motifs, as well as type 3 fibronectin (FN3) modules. An example is found in Lam16A, which has an N-terminal GH16, two CBM6s, three TSP3s, and a C-terminal module with homology to thiol oxidoreductase-like proteins. Lam16B contains an extended glutamate-proline repeat and two CBM32s, while Lam16C also has two CBM32s and an internal CBM4 (Table 3-3). Much like CBM family 6, CBM4s bind individual glycan chains, whereas CBM32 modules have been shown to bind galactose, lactose and sialic acid (19). According to the CAZy ModO analysis, while the predicted lipoprotein Lam16G has the sequence elements of a laminarinase, it appears to lack the catalytic residues which are typical of GH16 laminarinases.

Two laminarinases were cloned and expressed from *E. coli.* Lam16A was expressed as Lam16A:His_6, and Lam16D was expressed as Lam16D:His_6. Western blots using α-hisTag antibodies confirmed the identity of the expressed proteins. Lam16A:His_6 exhibited unusual mobility in SDS-PAGE with an apparent MW of ~240 kDa as compared to its predicted MW of 180 kDa. The reason for this discrepancy is uncertain as it persisted despite modifications to the SDS-PAGE sample buffer or treatment regimen (increasing the SDS and DTT concentrations, increasing boiling time, addition of EDTA to sample buffer). Lam16D:His_6 had an apparent MW of ~90 kDa, which correlates well with its predicted MW

of 82.2 kDa. The cloned laminarinases were subjected to reducing sugar and pNp-conjugate activity assays, the results of which are shown in Table 3-6. Lam16A:His_6 had high specific activity against laminarin, and was not active against any other tested substrate, thus confirming its predicted role as a laminarinase. Lam16D:His_6 had low, but detectable activity against laminarin.

Pectin metabolism

The genome contains 17 enzyme pectinase system consisting of 13 pectate lyases from four different polysaccharide lyase (PL) families, two pectin methylesterases, a GH28 polygalacturonanase, and a PL11 rhamnogalacturonanase, Rgl11A (Table 3-4). Pel1F, Pel3A, Pel10A and Rgl11A have CBM2a modules which have been demonstrated to specifically bind crystalline cellulose (100). As with the xylanases, these CBM2 modules adjoin PSL's. Interestingly, five pectinases contain one or more FN3 modules, including Pel9A with two FN3s, and Pell1A, Pel1C, Pel3B and Pel3C with one each. In addition to the FN3, Pel1A has dual PL1 catalytic modules and three PSLs. Although FN3 modules have been identified in carbohydrases from many different organisms (48), their function remains largely unknown.

Table 3-6. Enzymatic activities of cloned laminarinases.

Substrate[a]	Lam16A:His_6	Lam16D:His_6
Laminarin	20.90 (1.38)[b]	0.08 (0.03)[b]
Barley Glucan	--[c]	--
CMC	--	--
PASC	--	--
Avicel	--	--
p-NP-cellobiose	--	--
p-NP-α-ararabinpyanoside	--	--
p-NP-α-arabinofuranoside	--	--
p-NP-β-xyloside	--	--
p-NP-β-Glucoside	--	--

[a] Recombinant proteins (5 ng) were subjected to reducing sugar assays (RSAs) using 1 % (w/v) solutions of the following substrates: laminarin; barley β-glucan; CMC, carboxymethylcellulose; PASC, phosphoric acid-swollen cellulose. They were also assayed for activity against 5 mM solutions of the indicated para-nitrophenol (p-NP) sugar conjugates. All assays were preformed for 30 minutes at 37°C except for RSAs using Avicel, CMC and PASC as substrates.

[b] Specific activity, expressed as U/mg protein $x10^2$ where 1 U = 1μmol of glucose reducing equivalents released per minute as determined by Nelson-Somogyi reducing sugar assay (44).

[c] --, no detected activity.

Mannan metabolism

The *S. degradans* genome contains five confidently-predicted mannanases and an α-galactosidase accessory, Agl27A, which is predicted to remove alpha-linked galactose side chains from galactomannans (60) (Table3-5). Three mannanases contain CBM10 modules, one of which also has a CBM2a. Certain CBM10 modules have been shown to bind crystalline mannan (55). Man26A has a PSL, a CBM10, and a C-terminal GH26. Man5N has an N-terminal GH5, and internal CBM10 and a C-terminal CBM2a, which are separated from each other by two PSLs, while Man5O has an N-terminal GH5, followed by a PSL and tandem CBM10s.

Two mannanases were detected by Mass spectrometry in culture supernatants. Man5P was detected in CMC-grown supernatant, and Man5Q, was detected in agarose-grown samples; this is the mannose which contains the unusual lipobox-like sequence, LQAC.

Discussion

Genomic analyses, supported by other lines of evidence, including zymograms, mass spectrometry, and assays of cloned proteins, strongly suggest that *S. degradans* contains complete multienzyme systems for the metabolism of the major hemicellulose and pectin components of the plant cell wall. While an exhaustive functional analysis of each system was beyond the scope of this work, some of the relevant points pertaining to each enzyme system merit discussion.

Xylanases and arabinanases

It is, perhaps, reflective of the chemical complexity of glucuronoarabinoxylans (GAX) that so many of the enzymes for their metabolism are predicted to have multiple catalytic activities. In some cases these are accommodated by a single catalytic dual-function module, as exemplified by the four bifunctional GH43-containing proteins. In other cases, such as Xyl/Arb43G-Xyn10D and Xyn11B-Axe4A, separate catalytic modules having different, but related, activities are integrated into proteins containing one or more CBM, These enzymes have complex domain architecture. Furthermore, in Xyl/Arb43G-Xyn10D, the GH43 module is bifunctional. GH43 modules clip xylose and arabinose residues from side chains of arabinoxylans, while the GH10 module cleaves the xylan backbone to xylo-oligomers.

Xyl/Arb43G-Xyn10D has similar complexity. Its CBM2 binds crystalline cellulose, while its CBM6 and CBM22 bind single-chain glycans, likely xylans or arabinanans. Xyl/Arb43G-Xyn10D also contains a Y94 module of unknown function as they are unique to *S. degradans*. The occurrence of Y94 modules in *S. degradans* is limited two xylanases and a protein (which lacks a catalytic site but does have a CBM which may bind xylan), detected in xylan-grown cultures by mass spectrometry. This unusual protein also has two PSLs, believed to impart unusual flexibility. Thus, it is possible that Xyl/Arb43G-Xyn10D can simultaneously bind cellulose and xylan and/or arabinan chains while cleaving xylose and/or arabinose residues from side chains and the backbone of xylan chains. Another multifunctional enzyme, Xyn11B-Axe4A, has both a CE4 acetylxylanesterase module and a GH11 endoxylanase module, presumably imparting the abilities to both deacetylate and cleave the xylan backbone.

Such multifunctional enzymes must confer advantages to bacteria. However, they do not appear to obviate the need for stand-alone accessory proteins, as *S. degradans* has a number of single-function proteins such as Axe2A, which deacetylates xylose residues prior to their cleavage by endoxylanases. Similarly, glucuronic acid residues are removed from side chains by Agu67A, while Xyl3A is predicted to clip xylose monomers from the ends of xylan chains and oligomers.

Genomic analyses provide insight into the cell biology of *S. degradans* glucuronoarabinoxylan (GAX) metabolism. All GAX depolymerases and all but one of the arabino- and xylosidases carry type II secretion signals, indicating that the depolymerization of GAX to arabinose and xylose residues is extracellular. While many of the GAX-degrading enzymes appear to be freely secreted, one xylan depolymerase (Xyn10E) and two arabinan depolymerases (Arb43B and Arb43E) are lipoproteins, implying that at least some of the depolymerization occurs at the cell surface. The downstream degradative processes, namely cleaving GAX oligomers into xylose and arabinose, would also occur on the cell surface since the four GH43 arabinosidase/xylosidases likewise contain lipoboxes. The xylose and arabinose monomers would then be imported to the cell, likely involving active transport and dedicated membrane receptors.

Laminarin metabolism

In general, laminarinases are not as well characterized as cellulases and xylanases (65), limiting the specificity of genomic annotations and functional analyses. Three laminarinases (Lam16F, Lam16G and Lam81A) contain lipobox sequences, and another (Lam16D) has a C-terminal membrane helix, indicating that these four enzymes are anchored to the *S. degradans* outer membrane.

Additionally, Lam16A contains four copies of a thrombospondin type 3-like motif. While the function of this motif is uncertain, it is homologous to eukaryotic calcium binding repeat sequences which are known to participate in protein-protein interactions (68, 93). N Ekborg (38) has presented evidence that the one other TSP3 containing carbohydrase, Agarase E, synthesized by *S. degradans,* is a surface-bound enzyme.

The genomics could not be relied upon to indicate whether these surface-bound laminarinases are involved in depolymerization and/or final processing. Thus, Lam16A:His_6 and Lam16D:His_6 were cloned expressed and assayed for activity. Lam16A:His_6 was highly active on laminarin in reducing sugar assays (Table 3-5), and was not active against pNp-β-glucopyranoside indicating that it is an endo-acting laminarin depolymerase. The specific activity of Lam16A:His_6 is between 40 and 500 times lower than the activities of 8.5, 34.8, and 108 U/mg reported for three laminarinases purified from the cellulolytic fungus *Trichoderma viride* (81). While the explanations for this disparity may include differences between bacterial and fungal enzymes, or a need for synergism between two or more components of the *S. degreadans* laminarinase system, it should be noted that no attempt was made to optimize the reaction

conditions of these enzymes. Therefore the optimal specific activities may be higher than the results obtained in the activity screen performed here. Lam16D:His_6 was only weakly active against laminarin (or active against trace amounts of laminarin oligomers present in the preparation) and was likewise inactive against pNp-β-glucopyranoside. Lam16D contains a single CBM32, a family which is known to bind short oligomers (18). This, together with its apparent substrate preference for laminarin oligomers, suggests that Lam16D processes β-1,3 linked glucose oligosaccharides.

Pectin metabolism

The types of catalytic modules which comprise the pectinase system provide some mechanistic insight into how the bacterium degrades pectin. There are two main classes of lyases for the depolymerization of pectin. Pectin lyases are specific for true pectin, (methylpectate), while pectate lyases act on de-methylated pectin (polypectate) (85). That the pectinase system of *S. degradans* consists entirely of pectate lyases, complemented by two pectin methylesterases, indicates that pectin is first de-methylated to pectate, thus permitting the pectate lyases to cleave the pectate backbone.

As is the case for laminarinases, the relationships between sequence, structure, and function are not well developed for pectinases. Thus, sequence analysis does not provide much insight into which of the pectinases of *S. degradans* may be depolymerases or oligomer processing enzymes. However, there are 15 depolymerases with specificity for pectic substrates, implying that *S. degradans* should fully degrade pectin. In addition to the 13 pectate lyases, *S. degradans* has a polygalacturonan hydrolase, Pgl28A, and a rhamnogalacturonan lyase, Rgl11A, which is predicted to act on rhamnogalacturonans RG-I and RG-II (76). These depolymerases are complemented by the two pectin

methylesterases, giving *S. degradans* a 17 enzyme pectinase system. In comparison, analysis of the genome sequence of the highly pectinolytic plant pathogen *Erwinia carotovora subsp. atroseptica* SCRI1043 reveals a 20 enzyme pectinase system (16).

Apparently much of the polymer degradation by *S. degradans* is carried out on the cell surface and the pectinase system is no different. Seven of the 17 enzymes contain lipoboxes, including three pectate lyases (Pel1D, Pel1E, and Pel10B), the polygalacturonan hydrolase Pgl28A and both pectin methylesterases (Pes8A and Pes8B).

Mannan metabolism

The mannanase system contains five ORFs that annotate as mannan depolymerases (four from GH5 and one from GH6) and an α-galactosidase accessory enzyme (Table 3-5), paralleling the mannanase system of *Cellvibrio japonicus*, which contains three GH5 mannanases and a GH26 mannanase (55). In *C. japonicus* all four depolymerases act on soluble mannans (glucomannan and mannan oligomers), with one of them, Man5C, also active against crystalline mannose (55).

Based on the substrate specificities of the CBMs, Hogg et al (2003) predicted that the *C. japonicus* mannanases that bind to crystalline cellulose act on mannans that are intricately associated with the plant cell wall; the mannanase without a CBM (Man26A in *C. japonicus*) acts on the energy storage polymer, galactomannan, and on mannose oligosaccharides (55). Using the same rationale, in *S. degradans* Man26A, Man5N, and Man5O contain CBMs specific for crystalline substrates, and are thus likely to degrade plant wall mannans, while Man5P and Man5Q lack CBMs, and most likely act in concert with the α-galactosidase to degrade galactomannans. Man5P and/or Man5Q may also act

in mannan oligomer processing. Comparisons of active site homology with Man5A of *Cellvibrio mixtus* indicate that Man5P likely exhibits exo-mannosidase activity (32).

Thus, Man5P appears to be a freely secreted exo-mannosidase while Man5Q, which contains the putative lipobox-like sequence LQAC, appears to be a surface-bound endo-mannanase. That Man26A contains a lipobox sequence as well as a CBM10 indicates that mannan depolymerization also occurs at the cell surface, and may provide a means for *S. degradans* cells to adhere to the plant cell wall.

Concluding thoughts on hemicellulose and pectin metabolism

It is remarkable that *S. degradans* apparently has five complete systems for the metabolism of the hemicellulose and pectin components of the plant cell wall. In many cases, as exemplified by the pectin-degrading system, their complexity is comparable to those found in specialized plant saprophytes or pathogens. This is particularly striking given that *S. degradans* also has complete multienzyme systems for the metabolism of cellulose (chapter 2, this work), agar (101), alginate (21), and chitin (57). Such broad, comprehensive enzymatic ability is thus far unique to *Saccharophagus degradans*. This is illustrated by comparisons to other prolific cellulose-degrading bacteria, such as *Clostridium thermocellum.* Although *C. thermocellum* is a prolific degrader of cellulose and synthesizes endoxylanases and other hemicellulases, it cannot further metabolize these products—suggesting that the function of these enzymes is to increase access to cellulose (36). Similarly, the Gram-negative soil aerobe *Cellvibrio japonicus* (formerly *Pseudomonas cellulosa*) has complete systems for the hydolysis of cellulose, GAX and mannan, however it has only two pectinolytic enzymes and cannot grow on pectic

oligomers or galacturonic acid (40, 49, 76). Another cellulosome-producing anaerobe, *Clostridium cellulovorans* can utilize cellulose, xylan, mannan and pectin (46), however it can not degrade agar, alginate or chitin.

The versatility of *S. degradans* allows it to be saprophytic on algae, invertebrates, and plants. (It is not known whether it invades and feeds off of healthy eukaryotes.) Its plentitude of plant wall degrading enzyme systems, not previously reported in any marine prokaryote, are hypothesized to be an integral part of the global geochemical carbon cycle.

Chapter 4: Expression of *Saccharophagus degradans* Plant Cell Wall Degrading Enzyme Systems on Cognate Substrates

Introduction

Since its isolation from decaying salt-marsh cordgrass Chesapeake Bay (3), *S. degradans* has been identified as a prolific degrader of complex polysaccharides (CP) from algae (21, 101, 114), invertebrates (57), and plants (41), (Chapters 2 and 3, this work). It has also been established as a member of an emerging group of marine bacteria which mediate the turnover of environmental CP (39), and at the time of this writing *S. degradans* is the only marine bacterium shown to metabolize and degrade intact plant matter (Chapter 2, this work).

The plant cell wall is both chemically and structurally complex, containing a structural skeleton of cellulose fibrils surrounded by a matrix of hemicellulose (xylans, arabinans, β-glucans and mannans) and pectin (20, 30, 115). Numerous lines of evidence indicate that *S. degradans* is equipped with multienzyme systems for the degradation and metabolism of the major polysaccharides of the plant cell wall. In addition to the plethora of plant-wall specific carbohydrases identified by genomic analyses, *S. degradans* has been shown to grow using amorphous (CMC) and crystalline (Avicel) forms of cellulose, as well as xylan, barley β-glucan (the so-called "mixed-linkage" β-1,3(4)-glucan; herein referred to as β-glucan or barley glucan) and laminarin as sole carbon sources. *S. degradans* has been shown to grow in pectin (3, 41).

The regulation of complex polysaccharide-degrading enzyme synthesis in organisms with multiple carbohydrase systems has not been well studied. The studies that

have been done so far show that the coordinate regulation of plant cell wall degrading enzyme systems is complex in that a single substrate will often induce the synthesis of multiple enzyme systems. The anaerobe *Clostridium thermocellum*, a cellulositic Gram positve bacterium that synthesizes cellulosomes (11, 13-15), has been shown to constitutively express at least one cellulosome-bound endoxylanase. However, *C. thermocellum* cannot use xylose as a carbon or energy source suggesting that it synthesizes xylanases to gain access to cellulose (36).

Another cellulosome-producing bacterium*, Clostridium cellulovorans,* has more diverse metabolic capabilities than *C. thermocellum*, being able to grow on cellulose, xylan, mannan and pectin (35, 46). Two of its cellulases, ExgS and EngE, are expressed constitutively; however its other cellulases are subject to substrate induction (45). For example, growth on cellulose results in the coordinate induction of cellulases EngH and EngG, xylanases, arabinofuranosidases, and mannanases. Similarly, xylan induces cellulases and mannanases. Interestingly, its pectinase PelA is only induced during growth on pectin, and growth in pectin induces the synthesis of cellulases but not xylanases (46, 47). Although *C. cellulovorans* xylanases are induced during growth on cellulose, they are induced at the highest levels during growth on xylan (47). All of the plant wall degrading carbohydrases in this organism are repressed by glucose (46).

In *Thermobifida fusca*, a cellulolytic actinomycete, growth in xylan does not strongly induce cellulase production (70). Thus, the regulation of CP degrading enzyme systems does not appear to fit one general model, even between *C. thermocellum* and *C. cellulovorans* which are closely related.

To gain further insight into the regulation of the plant cell wall degrading enzyme systems of *S. degradans* during growth in plant-wall derived carbon sources, cell and supernatant fractions grown in sole carbon source Avicel, xylan, β-glucan, laminarin, glucose and *Spartina alterniflora,* were assayed for activity against cellulose, xylan, β-glucan and laminarin. Patterns of enzyme system regulation emerged and were analyzed.

Materials and methods

All of the procedures and assays in this study were performed in triplicate, with the results averaged and subjected to statistical analyses.

Growth of S. degradans in sole carbon sources

S. degradans strain 2-40 (ATCC43961^T) has been maintained in the laboratory since its isolation in 1988 on Difco® Marine Agar 2216 (Difco, Detroit, MI). Cultures were prepared in two-liter Erlenmeyer flasks containing 1 L of minimal medium (MM; per L: 23 g Instant Ocean Sea salts, 1 g Yeast extract, 50 mmol Tris buffer pH 7.4 and 0.5 g NH_4Cl) amended with 0.2 % of one of the following carbon sources: glucose, Avicel, xylan, barley β-glucan, laminarin, or *Spartina* leaves cut into ca. 2 cm sections. Before inclusion in growth medium laboratory-grown *Spartina alterniflora* was harvested by cutting the entire plant 1cm above the soil line. The harvested *Spartina* was stored in a desiccator at room temperature until use. To prepare MM containing *Spartina* the dried plants were cut into ~2 cm sections and added to the medium prior to autoclaving. Flasks of *Spartina* MM were autoclaved for one hour to ensure sterility. Cultures were inoculated with 2.5 ml of a *S. degradans* culture which was grown to late log/stationary-

phase in glucose MM. The cells were washed twice and resuspended in 1.25 ml of MM without an added carbon source. Growth was enumerated by viable plate counts (VPC) on ½ MA and the generation time (T_{gen}) during log-phase growth was calculated. Additionally, cultures grown in glucose, xylan, β-glucan and laminarin were examined spectrophotometrically (A_{600}). Culture purity was assessed microscopically and on LB agar. At each timepoint a 15ml aliquot was removed, with the cell fraction in a 10ml portion collected by centrifugation and washed twice and resuspended in 10 ml Tris/IO (50 mM Tris, pH 7.4, 2.3 % (w/v) Instant Ocean). The supernatant fraction was passed through a 0.2 μm filter. Resuspended pellet and filtered supernatant fractions were stored at -80°C for enzyme activity and protein quantification assays. Five ml were used for VPC and turbidity measurements. Flasks containing MM without an added carbon source (no C-source controls) did not support growth of *S. degradans*.

The protein concentrations in filtered supernatant and resuspended pellet fractions were measured using the Pierce bicinchoninic acid assay (BCA; Peirce, Rockford, IL) according to the manufacturer's instructions.

Enzyme Activity Assays

Cell and supernatant fractions were analyzed by the microplate adaptation of the Nelson-Somogyi reducing-sugar assay (44). Samples were assayed for activity against Avicel, PASC, CMC, xylan, β-glucan and laminarin. Specific activities were calculated as U/mg protein, where 1 U = 1μmol reducing sugar equivalent released/minute.

Results and Discussion

Growth parameters

S. degradans grew in all tested carbon sources. The generation times and maximum viable cell densities are summarized in Table 4-1. As expected, cultures grown in glucose had the shortest generation time (T_{gen}) at 51 minutes, and cultures grown in Spartina had the longest T_{gen} at 280 minutes. The observed T_{gen} in glucose is in good agreement with previous findings (112), confirming that *S. degradans* grows fastest on the monosaccharide, glucose (21, 101, 112).

The generation times (Table 4-1) increase with the complexity of the carbon source. Glucose, a monosaccharide, is expected to be directly imported into the cell by the phosphotransferase system and metabolized (2). Laminarin (from *Laminaria digitata*; Sigma, St. Louis MO) contains β-1,3-linked glucose with an average degree of polymerization (DP) of 25-30 residues, about 5 % of which contain a β-1,6-branch linked glucose residue (91). Barley β-glucan (medium viscosity; Megazyme, Ireland) is a linear polymer, having an average DP of 1,200 residues. The ratio of β-1,3 linkages to β-1,4 linkages is about 1:2 (http://www.megazyme.com). Increasing in compositional complexity is xylan. The β-1,4-linked backbone of the xylan substrate (Birchwood; Sigma, St. Louis MO) is periodically substituted with *O*-acetyl and 4-*O*-methylglucuronic acid residues (89). Avicel, as a homopolymer, is less compositionally complex than xylan; however it presents recalcitrant crystalline regions (discussed in chapter 2). *S. degradans* growing in Avicel has a T_{gen} of 155 minutes. Finally, the most complex of these substrates is *Spartina*, in which the T_{gen} is nearly twice that attained during growth in Avicel.

Surprisingly, β-glucan supported a significantly higher peak population than did glucose ($P<0.05$). Substrate recalcitrance clearly affects culture population dynamics. Cultures grown in *Spartina* and Avicel exhibited a conventional phase of logarithmic growth and an extended stationary phase (Figures 4-1 and 4-2), suggesting that they grew rapidly on the easily accessible substrate and more slowly on the recalcitrant regions. As anticipated, glucose, xylan, β-glucan and laminarin were more readily utilized as shown by a characteristic growth curve whereby logarithmic growth was followed by a brief stationary phase and sharp decline phase (Figures 4-3, 4-4, 4-5, and 4-6).

Table 4-1. Growth of *S. degradans* in polysaccharide carbon sources and glucose.

Sole carbon source[a]	T_{gen} (min)[b]	CFU (max)[c]	Std Dev[d]
Spartina	280	2.2×10^7	5.7×10^6
Avicel	155	7.0×10^8	4.0×10^7
Xylan	90	4.5×10^8	6.1×10^7
β-glucan	71	3.6×10^9	2.9×10^8
Laminarin	60	8.4×10^8	1.2×10^8
Glucose	51	1.4×10^9	2.1×10^8

[a] Cultures of *S. degradans* were grown in medium containing 0.2 % of each of the specified substrate as sole carbon source.

[b] Generation time was calculated during log-phase growth as determined by viable plate count (VPC).

[c] Average cell density at peak culture population as determined by VPC.

[d] Calculated standard deviation of maximum cell density ($n = 3$).

Figure 4-1. Growth and specific carbohydrase activities of *S. degradans* grown in sole carbon source Avicel minimal medium. Cultures of *S. degradans* were grown in sole carbon source Avicel minimal medium and assayed for activity against Avicel, PASC, CMC, xylan, β-glucan, and laminarin using the Nelson-Somogyi reducing sugar assay. Activity detected at each time point is expressed as specific activity in U/mg protein where 1 U = 1 μmol reducing sugar equivalent released per minute. Activity against phosphoric acid swollen cellulose (PASC) is not shown to improve figure clarity.

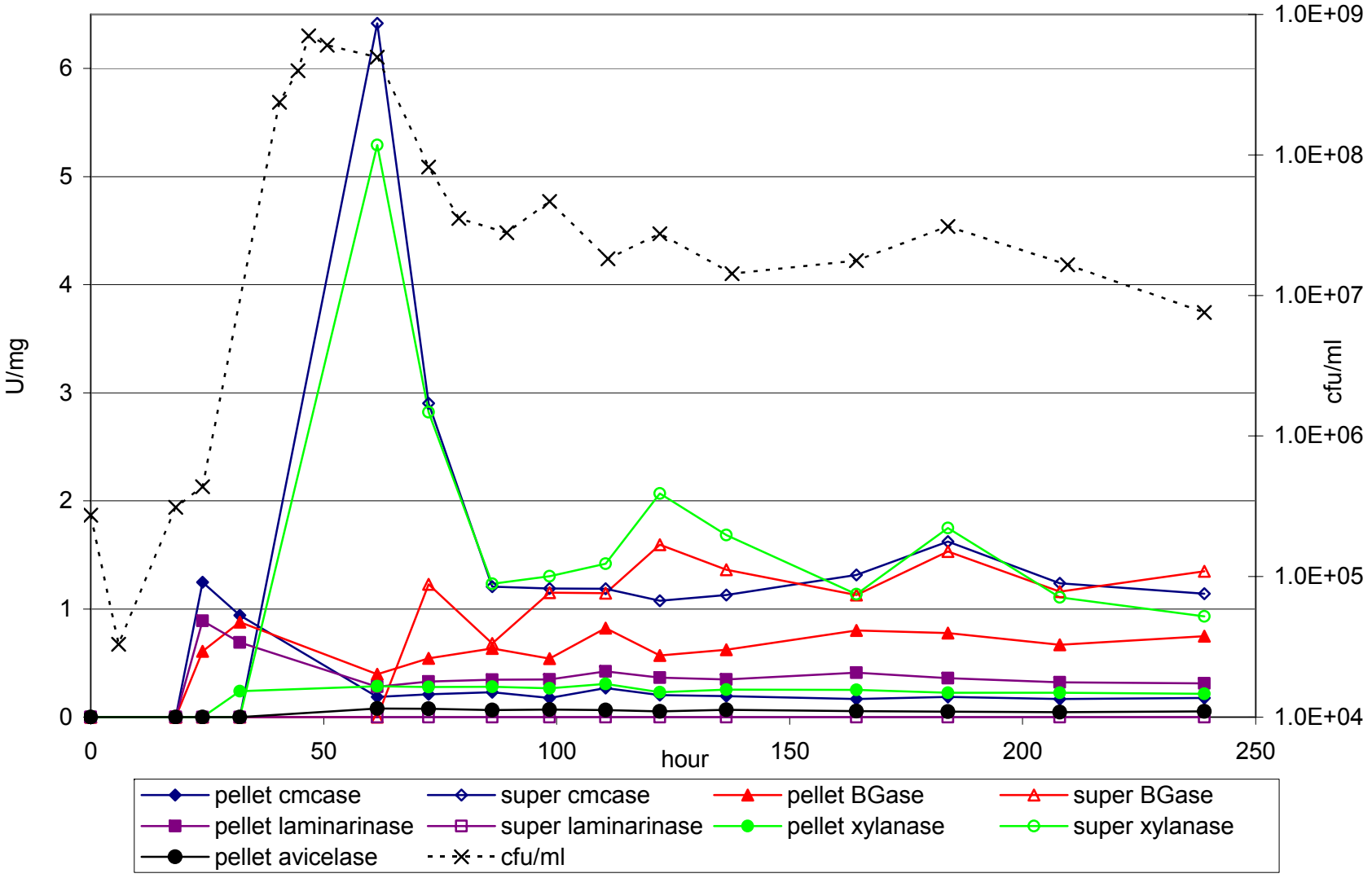

U/mg
cfu/ml
hour
0
1
2
3
4
5
6
0
50
100
150
200
250
1.0E+04
1.0E+05
1.0E+06
1.0E+07
1.0E+08
1.0E+09
pellet cmcase
super cmcase
pellet BGase
super BGase
pellet laminarinase
super laminarinase
pellet xylanase
super xylanase
pellet avicelase
cfu/ml

Figure 4-2. Growth and specific carbohydrase activities of *S. degradans* grown in sole carbon source *Spartina* minimal medium. Cultures of *S. degradans* were grown in sole carbon source *Spartina* minimal medium and assayed for activity against Avicel, PASC, CMC, xylan, β-glucan, and laminarin using the Nelson-Somogyi reducing sugar assay. Activity detected at each time point is expressed as specific activity in U/mg protein where 1 U = 1 μmol reducing sugar equivalent released per minute. Activity against phosphoric acid swollen cellulose (PASC) is not shown to improve figure clarity.

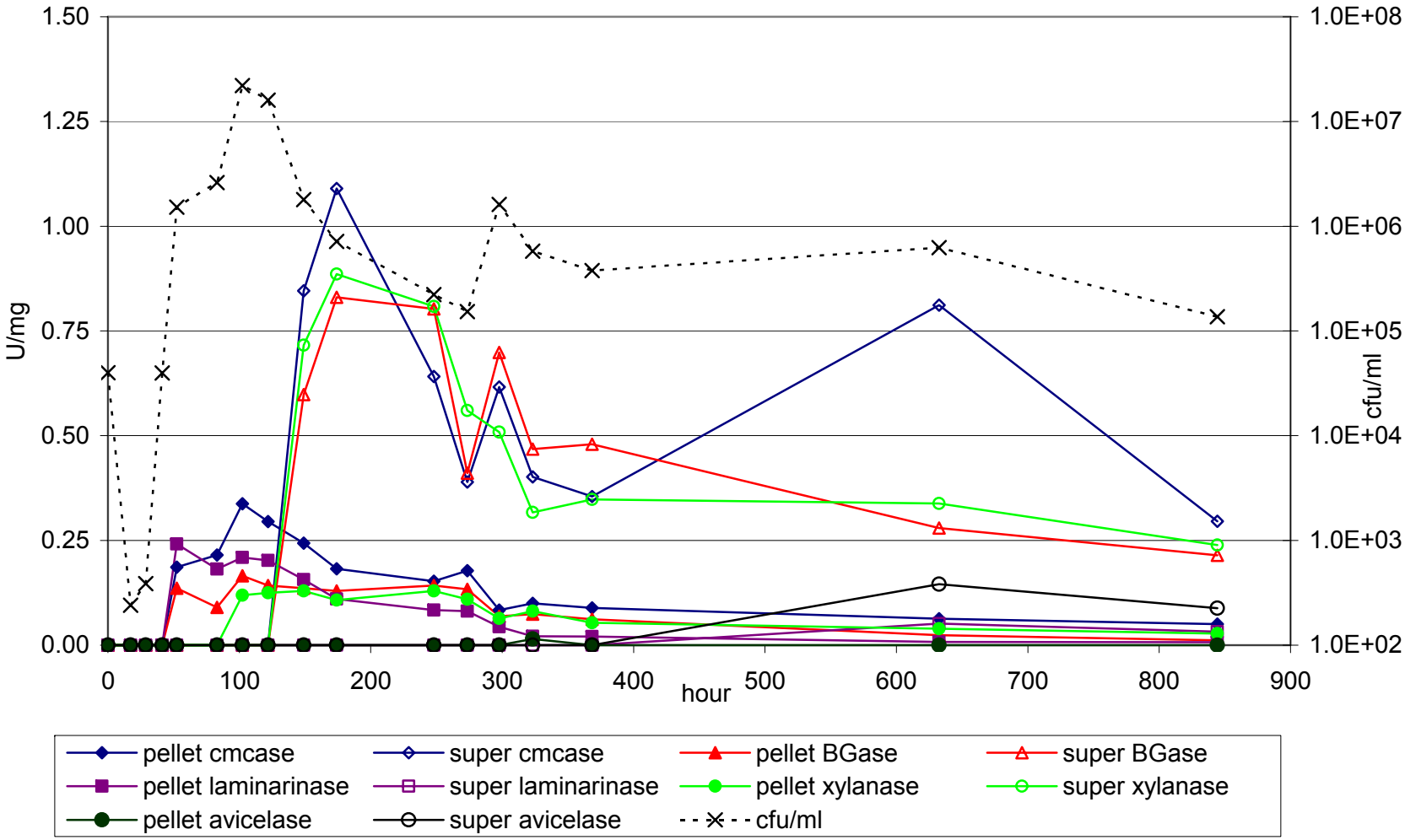

1.50
1.25
1.00
0.75
0.50
0.25
0.00
U/mg
1.0E+08
1.0E+07
1.0E+06
1.0E+05
1.0E+04
1.0E+03
1.0E+02
cfu/ml
0
100
200
300
400
hour
500
600
700
800
900
pellet cmcase
super cmcase
pellet BGase
super BGase
pellet laminarinase
super laminarinase
pellet xylanase
super xylanase
pellet avicelase
super avicelase
cfu/ml

Figure 4-3. Growth and specific carbohydrase activities of *S. degradans* grown in sole carbon source glucose minimal medium. Cultures of *S. degradans* were grown in sole carbon source glucose minimal medium and assayed for activity against Avicel, PASC, CMC, xylan, β-glucan, and laminarin using the Nelson-Somogyi reducing sugar assay. Activity detected at each time point is expressed as specific activity in U/mg protein where 1 U = 1 μmol reducing sugar equivalent released per minute. Activities not shown were not detected in reducing sugar assays.

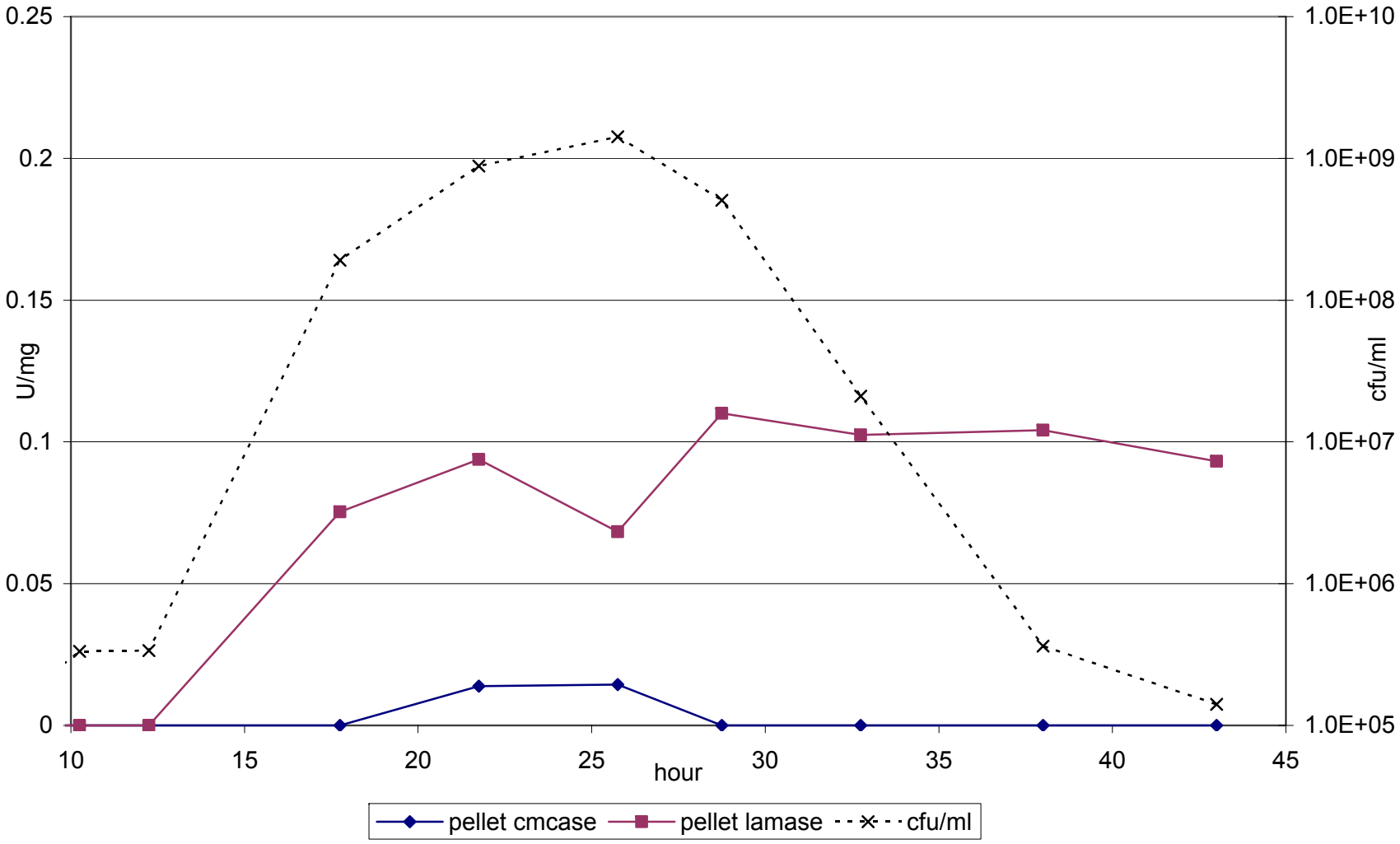

0.25
0.2
0.15
0.1
0.05
0
U/mg
1.0E+10
1.0E+09
1.0E+08
1.0E+07
1.0E+06
1.0E+05
cfu/ml
10
15
20
25
30
35
40
45
hour
pellet cmcase
pellet lamase
cfu/ml

Figure 4-4. Growth and specific carbohydrase activities of *S. degradans* grown in sole carbon source xylan minimal medium. Cultures of *S. degradans* were grown in sole carbon source xylan minimal medium and assayed for activity against Avicel, PASC, CMC, xylan, β-glucan, and laminarin using the Nelson-Somogyi reducing sugar assay. Activity detected at each time point is expressed as specific activity in U/mg protein where 1 U = 1 μmol reducing sugar equivalent released per minute. Activities not shown were not detected in reducing sugar assays.

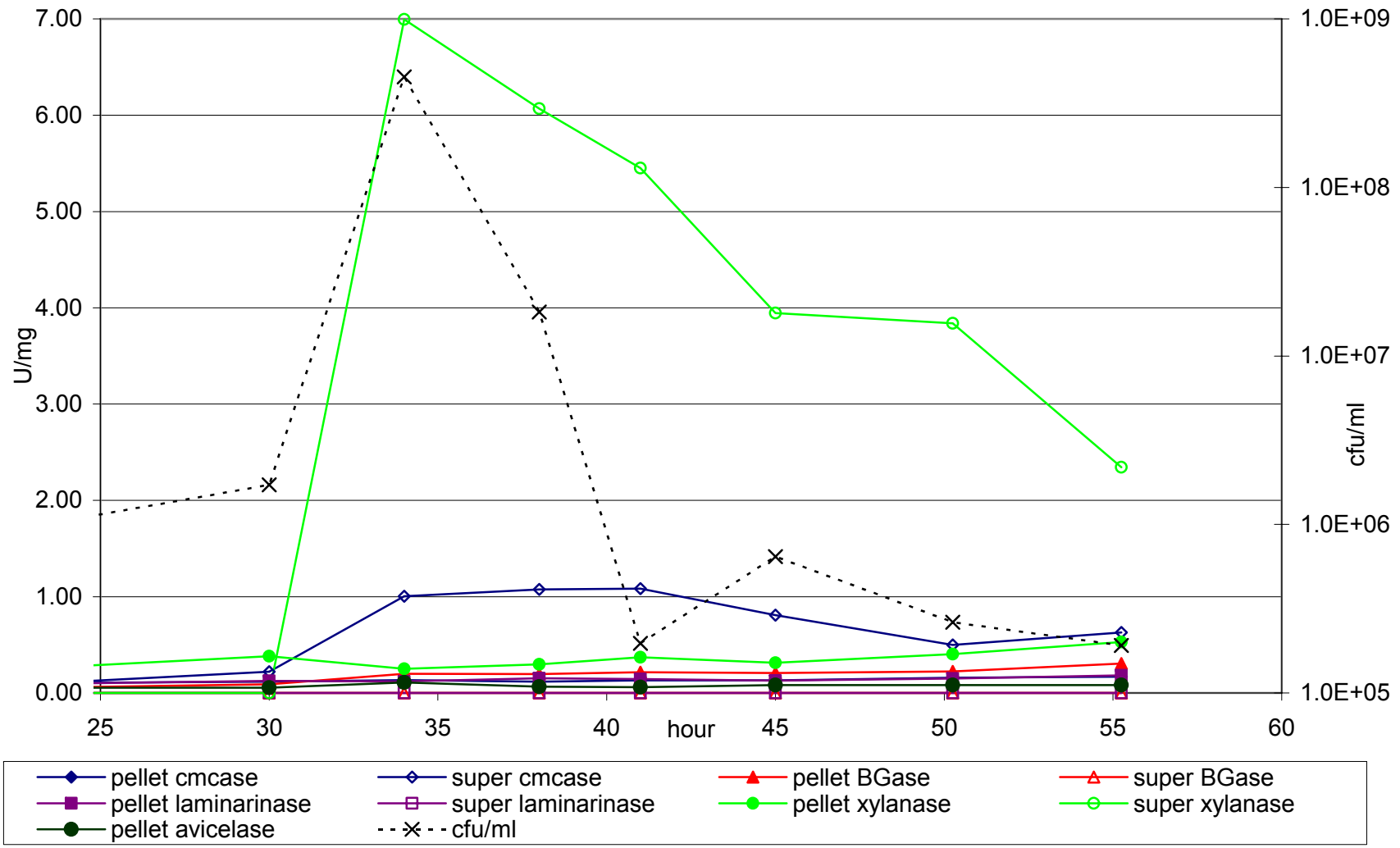

7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
U/mg
1.0E+09
1.0E+08
1.0E+07
1.0E+06
1.0E+05
cfu/ml
25
30
35
40
hour
45
50
55
60
pellet cmcase
super cmcase
pellet BGase
super BGase
pellet laminarinase
super laminarinase
pellet xylanase
super xylanase
pellet avicelase
cfu/ml

Figure 4-5. Growth and specific carbohydrase activities of *S. degradans* grown in sole carbon source β-glucan minimal medium. Cultures of *S. degradans* were grown in sole carbon source β-glucan minimal medium and assayed for activity against Avicel, PASC, CMC, xylan, β-glucan, and laminarin using the Nelson-Somogyi reducing sugar assay. Activity detected at each time point is expressed as specific activity in U/mg protein where 1 U = 1 μmol reducing sugar equivalent released per minute. Activities not shown were not detected in reducing sugar assays, with the exception of activity against PASC, which was omitted to improve figure clarity.

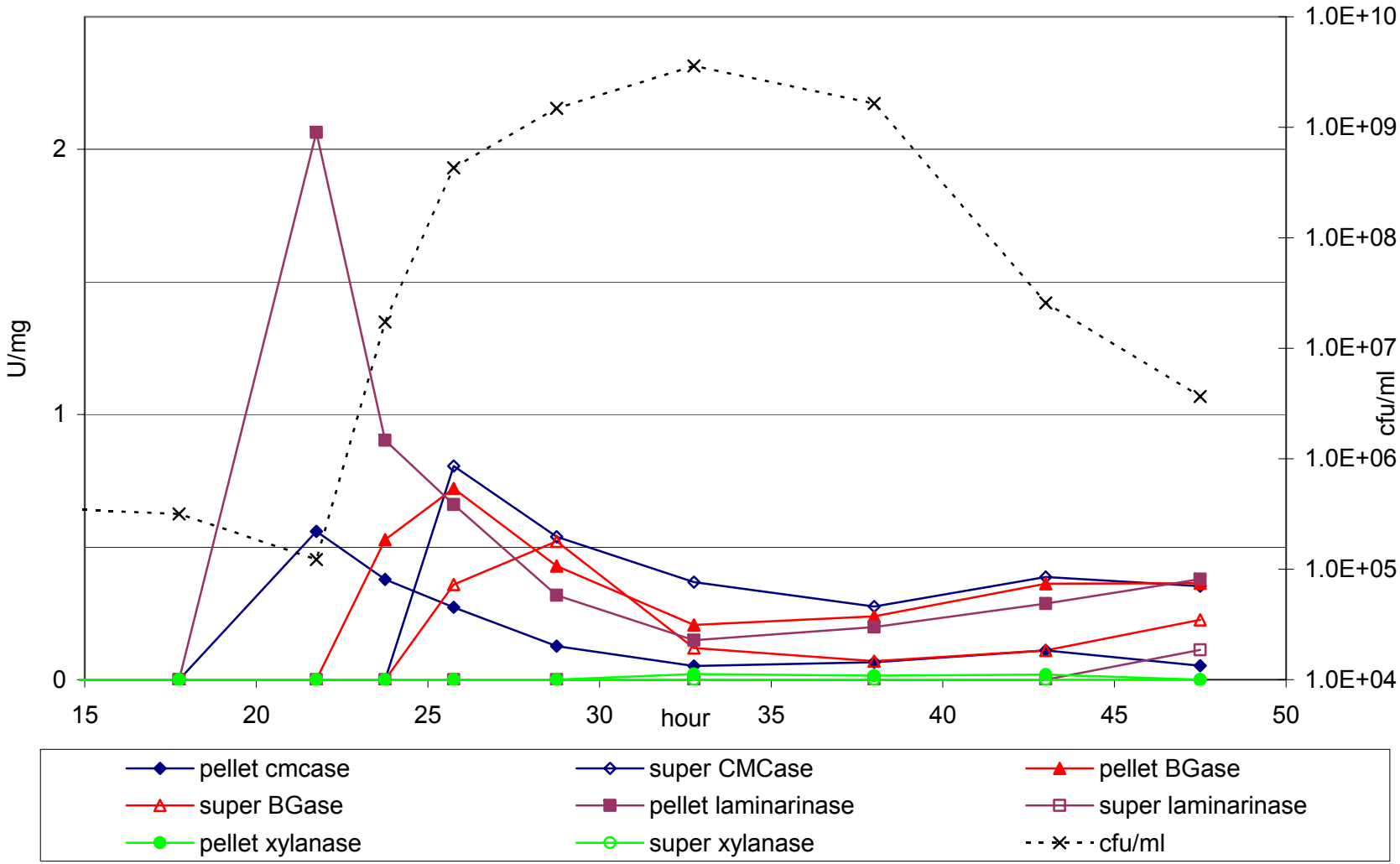
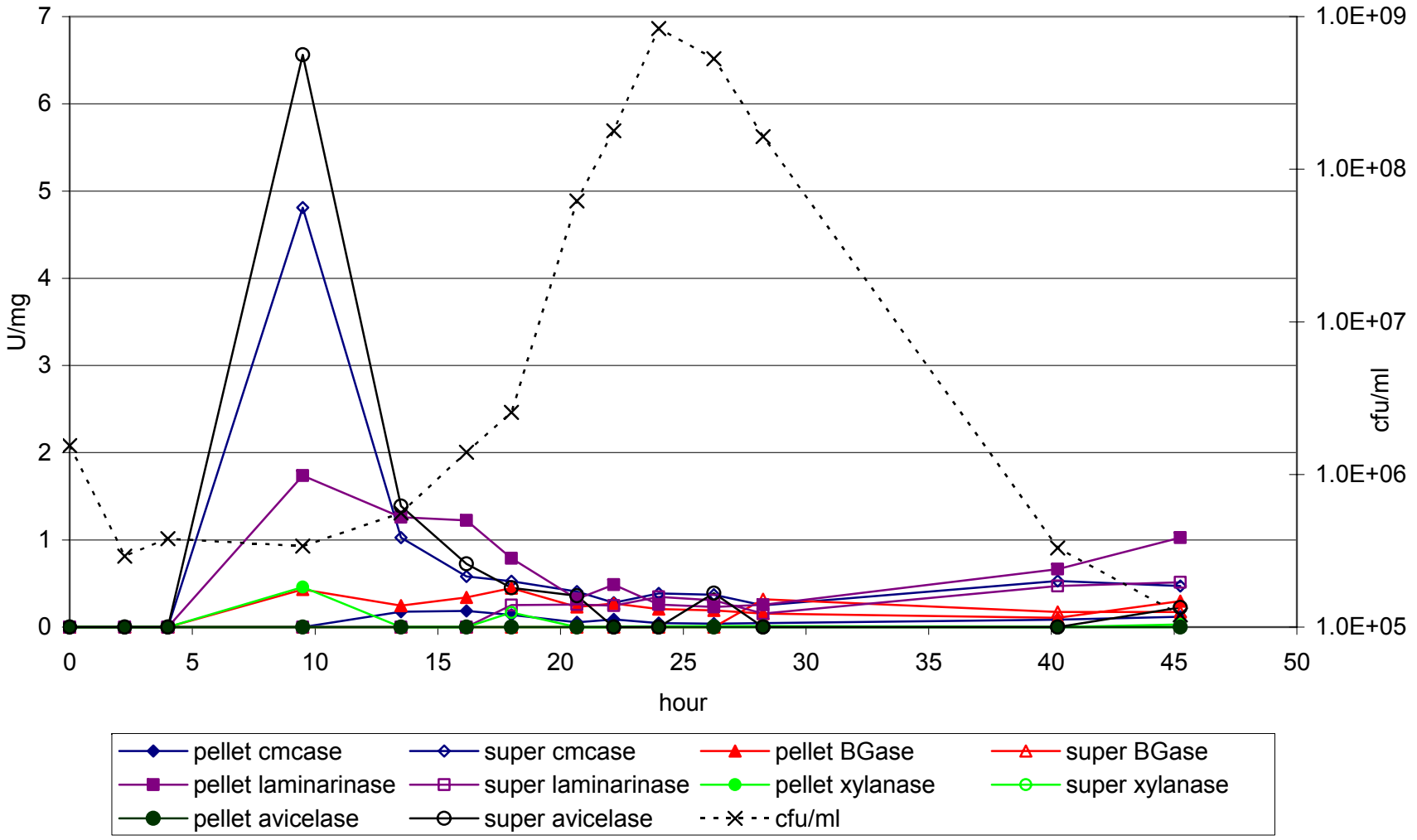

U/mg
cfu/ml
hour
0
1
2
15
20
25
30
35
40
45
50
1.0E+04
1.0E+05
1.0E+06
1.0E+07
1.0E+08
1.0E+09
1.0E+10
pellet cmcase
super CMCase
pellet BGase
super BGase
pellet laminarinase
super laminarinase
pellet xylanase
super xylanase
cfu/ml

Figure 4-6. Growth and specific carbohydrase activities of *S. degradans* grown in sole carbon source laminarin minimal medium. Cultures of *S. degradans* were grown in sole carbon source laminarin minimal medium and assayed for activity against Avicel, PASC, CMC, xylan, β-glucan, and laminarin using the Nelson-Somogyi reducing sugar assay. Activity detected at each time point is expressed as specific activity in U/mg protein where 1 U = 1 μmol reducing sugar equivalent released per minute. Activities not shown were not detected in reducing sugar assays, with the exception of activity against PASC, which was omitted to improve figure clarity.


7
6
5
4
3
2
1
0
U/mg
1.0E+09
1.0E+08
1.0E+07
1.0E+06
1.0E+05
cfu/ml
0
5
10
15
20
25
30
35
40
45
50
hour
pellet cmcase
super cmcase
pellet BGase
super BGase
pellet laminarinase
super laminarinase
pellet xylanase
super xylanase
pellet avicelase
super avicelase
cfu/ml

Enzyme activities

As with the agarase, alginase and chitinase systems (41, 101, 112), cultures grown in glucose did not express complex carbohydrase activity, except for trace levels of cell-associated activity vs. CMC and laminarin (Table 4-2). Furthermore, glucose strongly suppressed protein secretion (Table 4-3).

The cellular fractions of Avicel and xylan-grown cultures had similar activity against all tested substrates (Table 4-2). Growth in Avicel induced the highest specific activities against all substrates except xylan, with homologous induction by xylan predictably exceeding it. The supernatant fractions of Avicel-grown cultures had activity against CMC, PASC, xylan, and β-glucan whereas supernatants from cultures grown in xylan were active against CMC and xylan only (Table 4-3).

Despite the fact that it was not detected in these assays, it is likely that there is activity against Avicel in the supernatant of Avicel-grown cultures. However such activity was probably obscured by CBM carrying enzymes that act upon crystalline cellulose, binding to it and partitioning with residual Avicel in the cell fraction.

Cultures grown in medium containing *Spartina*, the plant from which *S. degradans* was originally isolated (3), had cellulase, xylanase, β-glucanase and laminarinase activity. Like the case with Avicel, *Spartina* grown cultures secreted cellulases, xylanases and β-glucanases (Table 4-3). *Spartina* grown cultures also had higher levels of cellulase activity in the cell fraction than cultures grown in xylan.

Table 4-2. Maximum specific activities of cell fractions during growth on tested carbon sources

Carbon source	Growth phase[a]	Specific activity vs.						Protein (net)[c]
		Avicel	CMC[b]	PASC[b]	Xylan	β-glucan	Laminarin	
Spartina	L/S	0.02	0.34	0.31	0.13	0.17	0.24	120
Avicel	L	0.10	0.94	1.16	0.25	0.88	0.69	105
Xylan	L/S	0.11	0.13	0.14	0.25	0.20	0.12	845
BG	L	0.00	0.27	0.00	0.00	0.72	0.66	249
Lam	L	0.00	0.12	0.06	0.00	0.45	0.79	354
Glucose	L/S	0.00	0.01	0.00	0.00	0.00	0.07	717

[a] Growth phase of maximum specific activity: L, logarithmic growth; S, stationary phase. Transitions between phases are indicated with slash marks, i.e., L/S, log/stationary phase transition point.

[b] CMC, carboxymethylcellulose; PASC, phosphoric-acid swollen cellulose

[c] Net protein present in sample in μg/ml.

Table 4-3. Maximum specific activities of supernatant fractions during growth on tested carbon sources

Carbon source	Growth phase[a]	Specific activity vs.						Protein (net)[c]
		Avicel	CMC[b]	PASC[b]	Xylan	β-glucan	Laminarin	
Spartina	S	0.15	1.09	0.87	0.89	0.83	0.00	16
Avicel	S	N.D.[d]	2.91	3.23	2.07	1.59	0.00	41
Xylan	L/S	0.00	1.00	0.00	6.99	0.00	0.00	8
BG	L/S	0.00	0.54	0.00	0.00	0.52	0.00	46
Lam	L/S	0.00	0.39	0.37	0.00	0.00	0.35	44
Glucose	L/S	0.00	0.00	0.00	0.00	0.00	0.00	0

[a] Growth phase of maximum specific activity: L, logarithmic growth; S, stationary phase. Transitions between phases are indicated with slash marks, i.e., L/S, log/stationary phase transition point.

[b] CMC, carboxymethylcellulose; PASC, phosphoric-acid swollen cellulose

[c] Net protein present in sample in μg/ml.

[d] N.D., not determinable under assay conditions. In Avicel-grown cultures, secreted enzymes with activity against Avicel would remain bound to Avicel present in the pellet fraction.

The activity assay results showed that cellulose and xylan broadly induced the synthesis of *S. degradans* plant-wall active carbohydrases. However, cellulose induced the highest levels of enzyme secretion. In cultures grown in Avicel, xylan and *Spartina,* most of the enzyme activity was in the cell fraction during early growth phases, but during later growth phases it was in the supernatant fraction. This metabolic "switch" from cell-bound to secreted activity is best shown in Figures 4-1 and 4-2, and is similar to results obtained during studies of the agarase and chitinase systems (41, 112, 114). In contrast, growth on β-glucan and laminarin induced laminarinase and β-glucanase activity that remained with the cell fraction during all growth phases.

S. degradans, grown in laminarin as a sole carbon source, had activity against CMC, indicating that β-1,4-glucanases are induced by laminarin. Similarly, barley glucan grown cultures had laminarinase activity, showing that β-1,3-glucanases are induced by barley glucan (Tables 4-2 and 4-3). This cross induction is consistent with the cleavage of barley β-1,3(4)-glucan backbones by β-1,3- and β-1,4 glucanases (25). Yet, predictably, the highest levels of laminarinase or β-glucanase activity were induced by the homologous substrate. Furthermore, β-glucan and laminarin induced the highest activity in the cell fraction during all growth phases suggesting that the degradation of both substrates are primarily cell-bound processes.

The results of these assays support the assertion that the regulation of plant wall carbohydrase degrading systems in *S. degradans* is quite complex. In the cell fraction, crystalline cellulose and xylan induced the full spectrum of activities required to metabolize plant material, consistent with the fact that cellulose and xylan are the most abundant polymers in the plant wall. Xylan induced lower, but significant, cell-bound and

secreted cellulase activity than did Avicel, suggesting that *S. degradans* encounters cellulose and xylan together in nature.

Summary and conclusions

The regulation of the plant cell wall degrading enzyme systems of the marine bacterium, *Saccharophagus degradans* turns out to share features with the anaerobic soil bacterium, *Clostridium cellulovorans* (45). In both organisms, sole carbon source Avicel induces xylanases and other hemicellulases. Furthermore, while growth in sole carbon source xylan induces synthesis of cellulases and non-xylanase hemicellulases, the highest levels of xylanase activity are induced during growth on xylan. Cellulose and xylan induce activity against β-glucan and laminarin, while β-glucan and laminarin induce cellulases, but not xylanases. Also, in both bacteria, glucose strongly represses carbohydrase synthesis and expression.

In *C. cellulovorans* pectinase PelA was induced by pectin but not by cellulose or hemicellulose (45, 46) On the other hand, pectin did induce cellulases but not xylanases. Thus, for both *C. cellulovorans* and *S. degradans*, there appears to be a hierarchy in which the major plant cell wall polymers, cellulose and xylan, broadly induce cellulases and hemicellulases, while other polymers, like pectin, induce a narrower spectrum of enzyme systems such as pectinases and cellulases. That cellulose and xylan did not induce the pectinases, and pectin did not induce xylanases, suggests that *C. cellulovorans* and *S. degradans* can down regulate (or not induce) certain enzyme systems in the absence of the target substrate. This entire, complicated, regulatory motif would allow *S.*

degradans to conserve metabolic resources, while remaining able to utilize the ubiquitous cell-wall polymer, cellulose.

Chapter 5: Large Calcium-binding Proteins Which May Function in Enzyme Attachment and/or Cell Adhesion

Introduction

Analyses of the recently-completed genome sequence of *Saccharophagus degradans* (http://genome.ornl.gov/microbial/mdeg/), a complex marine carbohydrate degrading prokaryote, reveals the presence of seven ORFs coding for proteins with molecular weights ranging from 274,000 to 1,568,000 Daltons. Each of these megaproteins contains calcium-binding motifs that are homologous to domains which participate in protein-protein /adhesion interactions in eukaryotes. Among the 7 proteins, there are six distinct calcium-binding motifs, some of which are multicopy, viz 35 thrombospondin type 3 (TSP3) repeats.

TSP3 domains have been best studied in Human thrombospondin which is a 420 kDa adhesive glycoprotein, secreted by platelets. Thrombospondin associates with the platelet membrane, contributing to clot formation adhering to fibrinogen, fibronectin and collagen *in vitro* (68). Thrombospondin contains multiple copies of three distinct types of repeat sequences, the third of which (TSP3) is believed to be responsible for cellular adhesion (68). According to the Protein Families database of alignments (PFAM, http://www.sanger.ac.uk/Software/Pfam/) 159 known proteins contain TSP3 domains. Of these, 102 are eubacterial proteins. Eighty of the bacterial proteins containing TSP3 repeats also contain an outer-membrane protein domain OmpA, fourteen are proteins which are predicted to function in either cell to surface or cell to host adhesion, and four are predicted carbohydrases.

Other calcium-binding motifs within the megaproteins of *S. degradans* include three different cadherin-like sequences (CADG, (33); CADH, (102); Calx-β, (96, 97)), and sequences homologus to known calcium-binding sequences (EF Hand, (63)) or integrin-like motifs involved in calcium-mediated protein-protein binding or signaling (FG-GAP, (58)). While a definitive role for these domains in prokaryotes has not been demonstrated, they occur in proteins which are hypothesized to function in cell to cell or cell to surface adhesion (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00028).

Each of these calcium-binding domains contains a conserved sequence motif, designated DxDxDG for its core of conserved aspartic acid and glycine residues (93). In Eukaryotes, they serve different functions including adhesion, assembly of multiprotein structural complexes, or cell-to cell signaling, (93). The functional genomics of these motifs has not been reported in prokaryotes and, consequently the automated annotation system of DOE/ONRL did not classify the proteins containing them.

While the genomic analysis discovered these Ca^{2+} binding megaproteins, it did not suggest their function. BLASTP searches failed to identify such proteins in *E. coli*, but similar proteins are found in distantly related marine bacteria such as *Shewanella oneidensis* (50). Two possible functions in *S. degradans* could be in cell /cell interactions or in the assembly of multienzyme complexes on the cell surface. In fact, *S. degradans* is believed to be quite genetically promiscuous (74, 109) with substantial cell/cell contact and, also does form large surface protuberances (Appendix II) that localize carbohydrases to the surface (21, 112). The protuberances resemble the *Clostridium thermocellum* cellulosome, in which enzymes are assembled via calcium-binding, interacting dockerin and cohesin domains (9, 11, 13). Since the annotation did not identify any conserved

cohesin or dockerin sequences, this could be an example of converging evolution as the dockerin sequence, much like the Ca^{2+} binding motifs within the megaproteins of *S. degradans*, contains the DxDxDG motif (93).

This note reports the genomic analysis of these unusual proteins and focus on the largest prokaryotic protein to date, showing that it is expressed, exported as a megaprotein and that it resides on the surface of this gram negative bacterium.

Materials and Methods

Genomic analyses

The genome of *S. degradans* was analyzed by BLASTP (http://www.ncbi.nlm.nih.gov/blast/) to assess overall homology to known calcium-binding or adhesion-related proteins. Genes with intriguing homologies were analyzed by the Simple Modular Architecture Research Tool (SMART; http://smart.embl-heidelberg.de/) and the Protein Family database (PFAM; http://www.sanger.ac.uk/Software/Pfam/search.shtml) to identify conserved domains. Genes containing calcium-binding motifs were further analyzed by the database of lipoproteins (DOLOP; http://www.mrc-lmb.cam.ac.uk/genomes/dolop/), and the LipoP 1.0 server (lipoP; http://www.cbs.dtu.dk/services/LipoP/) to identify predicted lipoprotein anchoring sequences, as well as the signalP (http://www.cbs.dtu.dk/services/SignalP/) and TmPred webservers (http://www.ch.embnet.org/software/TMPRED_form.html) to identify N-terminal secretion signal sequences and transmembrane anchoring regions, respectively.

Mass Spectrometry and proteomic analyses of S. degradans culture supernatants

Supernatants from agarose, alginate, Avicel, CMC, and xylan-grown cultures were concentrated to ~25X by centrifugal ultrafiltration using Centricon™ or Microcon™ devices (Millipore). Protein concentrations were determined using the BCA protein assay (Pierce, Rockland IL). Samples were exchanged into 100 mM Tris buffer, pH 8.5, containing 8 M urea and 10 mM DTT and incubated 2 hours at 37°C to denature and reduce the proteins. After reduction, cysteine residues were alkylated by the addition of 1 M iodoacetate to a final concentration of 50 mM and incubated at 25°C for 30 minutes. The samples were exchanged into 50 mM Tris, 1 mM $CaCl_2$, pH 8.5 using Microcon™ devices. The denatured, reduced, and alkylated samples were digested overnight at 37°C using proteomics grade trypsin (Promega) at a 1:50 enzyme to substrate ratio. Digestions were stopped by the addition of 99 % formic acid to a final concentration of 1 % and analyzed by RPHPLC-MS/MS at the UMCP College of Life Sciences CORE Mass Spectrometry facility using a Waters 2960 HPLC linked to a Finnagin LCQ tandem Mass Spectrometer. All peptide fragment masses were analyzed by the peptide analysis packages SEQUEST and MASCOT (37, 86), and compared to amino acid sequence translations of all gene models in the 2-40 draft genome and to the non-redundant Mass Spectrometry Database (ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam). Peptide identity matches were evaluated using the accepted thresholds of statistical significance specific to each program.

Cloning and expression of a sub region of a large Ca^{2+} binding protein

A 1203 base-pair region (corresponding to bases 33,936 to 35,040) of gene Sde2049 of the finished genome, designated cabA, was subcloned using the primers CabA-PstI-F: AAAACTGCAGCGTTCGTTATTTCGGTGCTA and CabA-HindIII-R: CCCAAGCTTTACACGACCGTAAAGGCG which were purchased from Invitrogen (Frederick, MD). PCR reactions (50 µl) used standard parameters and conditions for tailed primers and Proof Pro® *Pfu* Polymerase (Continental Lab Products, San Diego, CA) and included 0.5 µl of *S. degradans* genomic DNA as the template. PCR products were cloned into pETBlue2, and the resulting plasmids were transformed into *E coli* DH5α by electroporation and blue/white screened on LB/amp/X-gal. Plasmids were recovered, singly digested and visualized by agarose electrophoresis for size confirmation.

The plasmids were transformed into the expression strain by heat shock. *E coli* BL-21(DE3)pLysS, transformants were selected on LB agar containing ampicillin and chloramphenicol and incubated overnight at 37°C. Production of an appropriate-sized His-tagged protein was confirmed by comparing pre-induced and induced (1 mM IPTG) cell lysates in western blots using 1/5000 anti-HisTag® monoclonal primary antibody (Novagen, Madison, WI) and 1/7500 goat anti-mouse HRP conjugated secondary antibody (BioRad, Hercules, CA). Blots were developed colorimetrically with the OPTI-4CN kit (BioRad).

Production and purification of recombinant protein

Expression cultures were grown to A_{600} of 0.6 to 0.8 in 500 ml or 1 liter broths of LB containing ampicillin and chloramphenicol. The cultures were then induced by 1 mM

IPTG incubated for four hours at 37°C or 16 hours at 20°C. Culture pellets were harvested by centrifugation (5000 x *g*, 20 min) and frozen overnight at -20°C. Cells were thawed on ice and suspended in 4ml urea lysis buffer (8 M Urea, 100 mM NaH_2PO_4, 25 mM Tris, pH 8.0) per gram wet pellet weight. The samples were clarified by centrifugation at 15,000 *g*. The resulting supernatant was mixed with Nickel-NTA resin (QIAGEN, Valencia, CA) per the manufacturer's instructions. After 1 hour at 25°C, the slurries were washed twice with urea lysis buffer, pH 7.0.

The recombinant protein, designated mini-CabA:His_6, was eluted in 1 M urea, 25 mM Tris pH 7.4, 500 mM NaCl, 20% glycerol containing 250 mM imidazole. Void, wash and elution fractions were surveyed for HisTag® production in western blots as described above. Elution fractions containing the recombinant proteins were pooled and exchanged into Storage Buffer (20 mM Tris pH 7.4, 10 mM NaCl, 10 % glycerol) using Centricon™ centrifugal ultrafiltration devices (Millipore, Bedford, MA) and frozen at -80°C until further analysis.

Production of miniCabA:His_6 antiserum

A standard mini-format 8 % SDS-PAGE gel (67) was prepared with one lane containing BioRad Precision Plus® molecular weight markers (BioRad, Hercules, CA) and the others containing ~15 µg mini-CabA:His_6. After staining with Coomassie Blue (66), gel bands corresponding to miniCabA:His_6 were excised, washed in Millipore water and sent to Sigma-Genosys company (The Woodlands, TX) to be used as antigens in the production of antiserum using two New Zealand white rabbits as the host animals. Preimmune serum was collected, as were three production bleeds collected at two-week

intervals starting one month after the initial injection. Immune and preimmune sera were analyzed by western blot.

Dialysis of xylan-grown S. degradans culture supernatant

S. degradans was inoculated into 500 ml each of minimal medium (MM; per L: 23 g Instant Ocean Sea salts, 1 g Yeast extract, 50 mM Tris buffer pH 7.4 and 0.5 g NH_4Cl) amended with 0.2 % (w/v) xylan. The cultures were incubated at 27°C and shaken at 200 rpm. Cultures were grown to early stationary phase and centrifuged to pellet the cells. The supernatant was decanted and retained. One 250 ml portion of the culture supernatant was passed through a 0.2 μm filter, and a 5 ml aliquot of each sample was frozen at -80°C. The remaining filtered and unfiltered supernatants were concentrated to approximately $1/30^{th}$ original volume by centrifugal ultrafiltration using Amicon Ultra-15® devices (Millipore, Bedford, MA) with a 100,000 Dalton molecular weight cutoff (MWCO). The concentrated samples were dialyzed for 24 hours at 12°C against three one liter portions of 1 % (w/v) SDS, 25 mM Tris pH 7.4, 15 mM EDTA in polyvinylidene fluoride (PVDF) dialysis tubing with a 1,000,000 Dalton MWCO (Spectrum laboratories, Rancho Dominguez, CA). The resulting retentate was exchanged into 25 mM Tris pH 7.4 using Amicon Ultra-15 devices (Millipore) with 10,000 Dalton MWCO to remove excess SDS.

The protein concentrations of the concentrated dialysates, and the corresponding pre-dialysis samples, were estimated using the BCA protein assay (Pierce, Rockford, IL) according to the manufacturer's instructions.

SDS-PAGE and Western Blots of Dialysates

In an attempt to visualize CabA as synthesized by *S. degradans* the xylan-grown culture supernatant dialysates were analyzed by SDS-PAGE western blots using α-mini-CabA serum. From the unfiltered and 0.2 μm filtered xylan-grown culture supernatants, the following samples were prepared: unconcentrated culture supernatant, concentrated culture supernatant (pre-dialysis), and the dialyzed concentrated culture supernatant. To increase the likelihood of detectable CabA fragments entering the gel two additional sets of samples were digested using trypsin as follows. Samples were denatured and reduced by incubation in 8 M urea, 100 mM Tris pH 8.5, 10 mM DTT for 1 hr at 37°C, after which they were exchanged into 50 mM Tris pH 8.5, 1 mM $CaCl_2$ using Amicon Microcon® centrifugal filtration devices with a 5,000 Dalton MWCO and amended to 25 μg/ml and 50 μg/ml with sequencing-grade trypsin (Promega, Madison, WI). Digestion was performed for 30 minutes at 37°C and stopped by addition of SDS-PAGE sample buffer and freezing at -80°C until further analysis. Fifteen micrograms of each sample were loaded into duplicate 4 % SDS-PAGE gels which had a 4 % stacking gel (67). To detect any protein of interest which did not enter the resolving gel, the stacking gels were processed identically to their corresponding resolving gels. One gel was stained with SYPRO Ruby Orange® protein stain (Molecular Probes, Eugene, OR), and the other was transferred to nitrocellulose membrane and processed as a western blot using a 1/250 dilution of α-mini-CabA antiserum in PBST (0.01 M NaH_2PO_4, pH 7.4, 0.137 M NaCl, 0.1 % Tween-20) containing 3 % (w/v) dry nonfat milk. The secondary antibody was horseradish peroxidase conjugated Goat anti-rabbit IgG (BioRad) diluted 1/5000 in PBST. The blots were developed colorimetrically using the Opti-4CN kit (BioRad) and

photographed with a Canon Powershot 70 digital camera. SYPRO® stained gels were visualized by UV transillumination and imaged using a GelDoc 2000 (BioRad).

Immunofluorescence microscopy

Saccharophagus degradans cultures were grown to late log/early stationary-phase in minimal medium containing 0.2 % glucose (MM+G) and 0.2 % xylan (MM+X). Cells were harvested by centrifugation and washed twice in phosphate buffered saline (PBS). The washed cells were suspended in ½ volume PBS and 20 µl was smeared onto microscope slides which were coated with polylysine to improve cell adhesion (Sigma, St. Louis, MO). Cell suspensions were allowed to dry at room temperature, following which the slides were incubated at 60°C for 10 minutes to heat-fix the cells to the slides.

For immunofluorescence microscopy, 15 µl of a 1/16 dilution of α-mini-cabA antiserum in PBS was piptetted onto the cell smear and incubated at room temperature for 20 minutes in a moist chamber. The slides were then washed by two 5 minute immersions in 150 ml PBS, after which 15 µl of secondary antibody (Goat anti-rabbit IgG FITC-cojugated; Sigma) diluted 1/32 in PBS was added. Secondary antibody was incubated for 20 minutes, after which the slides were washed twice, as before in PBS. Specimens were viewed on a Zeiss Axiophot® epifluorescence microscope (Carl Zeiss, Germany) under phase-contrast and UV illumination. Samples were photographed using the on-microscope camera with 800 ASA 35 mm film (Kodak, Rochester, NY). Pre-immune serum and samples prepared using PBS in place of primary antibody were used as experimental controls.

Results and Discussion

S. degradans encodes seven megaproteins containing at least one Ca^{2+} binding motif (Table 5-1). These all have a MW >250 kDa, a low isoelectric point and comparatively weak identity to even their most closely related proteins. Cab A has 35 TSP3 repeats, and four copies of an IPT/TIG domain, characterized by an immunoglobulin-like fold (23). RhsA and RhsB contain an uncommon YD motif which binds heparin (78) and is found in Rhs-proteins of *E. coli* (110) and teneurin neural protein (78). CabE is a putative lipoprotein, and all but CabB and CabD carry N-terminal secretion signals (Table 5-1).

Table 5-1. Megaproteins containing calcium-binding motifs

Name[a]	Sde number[b]	amino acids[c]	MW[c]	pI[c]	Features[d,e]	Best BLASTP hit[f]	I, S (length)[g]	MS/MS[h]
cabA	2049	14,609	1,568	4.18	TSP3 (x35), EF hand (x7), IPT/TIG	OmpA-like protein SO4025 of *Shewanella oneidensis* MR-1	35, 50 (442)	av ag alg cm xy
cabB[i]	3233	4465	465.9	3.45	CADG (x30), PEPQ[j]	COG2931: RTX-like Ca2+-binding proteins *Magnetococcus* sp. MC-1	36, 53 (4231)	
cabC	3323	3477	348.5	3.72	calx β, TSP3 (x2), CADH (x3)	FN3-like protein *Shewanella oneidensis* MR-1	29, 44 (2096)	xy
cabD[i]	0798	3474	359.6	3.78	FG-GAP (x2), CADH (x2), CADG (x8),VCBS (x5)	COG2931: RTX-like Ca binding *Magnetococcus* sp MC-1	26, 41 (2318)	
cabE	3470	2836	302.2	3.97	LP-box, PEPQ[j], TSP3 (x2)	EP3-3 procyclin precursor *Trypanosoma brucei brucei*	69, 73 (63)	av ag alg cm xy
rhsA	3020	2762	303.5	4.75	FG-GAP(x8) YD	Rhs family protein *Vibrio vulnificus*	26, 41 (978)	av alg xy
rhsB	3249	2513	274.1	4.91	FG-GAP(x8) YD (x2)	Rhs family protein *Vibrio vulnificus*	23, 37 (2103)	ag alg av cm xy

[a] Acronyms: cab, calcium-binding; rhs, originally found in *E. coli* and named for conserved Arginine Histadine and Serine residues.

[b] Gene number having the prefix "Sde", for *Saccharophagus degradans* as assigned in Jun 15, 2005 genome assembly (http://genome.ornl.gov/microbial/mdeg/15jun05/mdeg.html).

[c] MW in kDa, amino acid count, and isoelectric point (pI) calculated using the protParam tool at http://us.expasy.org/tools/ based on DOE/JGI amino acid sequence translations including any predicted signal peptide.

[d] Sequence features identified by BLASTP (http://www.ncbi.nlm.nih.gov/BLAST/), DOLOP (http://www.mrc-lmb.cam.ac.uk/genomes/dolop/), SMART analysis (http://smart.embl-heidelberg.de/), signalP (http://www.cbs.dtu.dk/services/SignalP/) or TmPred (http://www.ch.embnet.org/software/TMPRED_form.html)

[e] Feature abbreviations: CADG, Distroglycan type cadherin-like domain (33); CADH, cadherin-like domain (102); Calx β, Calx-beta calcium-binding motif similar to that of sodium/calcium exchange proteins (96, 97); EF hand, conserved motif commonly found in calcium-binding proteins (63); FG-GAP, integrin-like calcium binding sequence likely involved in ligand binding (58); LPB, lipobox signature sequence; TSP3, thrombospondin type 3 calcium-binding repeat sequence of ~15 amino acids (68); VCBS repeat sequence is present in high copy numbers (~35) in large proteins from genera *Vibrio, Colwellia, Bradyrhizobium,* and *Shewanella*. While the function is unknown a role in adhesion has been proposed (http://www.tigr.org/tigr-scripts/CMR2/hmm_report.spl?user-access&password-access&acc=TIGR01965); YD, repeat sequence found in Rhs proteins of *E. coli* and in teneurin (a neural protein in eukaryotes). In teneurin the YD repeat has been shown to bind the carbohydrate heparin (78).

[f] Protein with the highest homology as identified by BLASTP

[g] Percentage of amino acids which are identical (I) or have similar properties (S) to those in the best BLASTP hit. The number in parentheses indicates the length of sequence homology in amino acid residues.

[h] Protein identified by tandem mass spectrometry in concentrated supernatant from cultures grown in agarose (ag) or carboxymethylcellulose (cm).

[i] CabB and CabF do not have a predicted signal sequence

[j] CabB contains the EP rich sequence: EPEPQPEPQPEPEPETPEE; CabD contains:

PEPQPEPQPQPEPEPEPEPEPEPEPEPEPEPEPE PEPEPEPQ

[k] MVGC sequence is a possible lipobox sequence as identified by LipoP server (http://www.cbs.dtu.dk/services/LipoP/)

The properties of the calcium-binding motifs within these megaproteins open the possibility that one or more of them are involved in the formation of multiprotein complexes and/or adhesion. Although sequence analysis does not predict which of these motifs may interact in such a scenario, it is intriguing to note the patterns apparent in their distribution among these proteins. For example, the distroglycan-type cadherin-like domain CADG is only found in the putative cytoplasmic proteins CabB and CabD, while the TSP3-like domains occur on CabA, CabC, and CabE—all of which are secreted. CabD, contains another repeat sequence which has been identified in genomically-sequenced marine bacteria, the so-called VBCS repeat. VBCS repeats occurs in high number (~35 copies) among large proteins identified in the bacterial genera *Vibrio, Bradyrhizobium, Colwellia,* and *Shewanella* , and is believed to function in surface adhesion (http://www.tigr.org/tigr-scripts/CMR2/hmm_report.spl?user-access&password-access&acc=TIGR01965). The integrin-like motifs designated FG-GAP occur on RhsA, RhsB, and CabD (Table 5-1). RhsA and RhsB contain both FG-GAP motifs and YD repeats, normally involved in carbohydrate binding, suggesting a possible dual role in substrate-binding and protein interaction. One or more of these Ca^{2+} binding domains may interact though calcium bridges in lock and key fashion.

These megaproteins are expressed in *S. degradans*. Five of the seven (Table 5-1) were detected by mass spectroscopy in the supernatant fractions of stationary-phase *S. degradans* cultures. Furthermore, CabA, CabE and RhsB were detected in cultures grown in sole carbon source Avicel, agarose, alginate, carboxymethylcellulose and xylan. In the Avicel grown sample the peptide fragments

of CabA that were detected amounted to 17% coverage of the total protein sequence and multiple peptides were detected from the N-terminal, middle and C-terminal regions of the protein, strongly suggesting that CabA is synthesized intact (Figure 5-1).

CabA is particularly noteworthy because of its thirty five TSP3 repeat sequences and remarkable size—a predicted 1.5 million Daltons at 14,609 amino acids. Its domain architecture is shown in Figure 5-2. BLASTP analysis revealed that CabA does not have any known end-to-end homologue; the best identified homology occurs over a 442 amino acid region and is to SO4025, a predicted outer membrane protein from *Shewanella oneidensis* strain MR-1 which contains three TSP3 repeats. Parsed sequence, however, blasted against multiple databases (NIH, Entrez, BlastP) revealed shorter regions of around 125 amino acids, that have some identity with *Clostridium thermocellum* S-layer protein (e-19) and, elsewhere, serine-threonine kinases and protein kinase receptors (e-17)

Proteins as large as CabA are rare—even among Eukaryotes. Examples include the vertebrate muscle protein titin, which has an estimated molecular weight of 3 MegaDalton (MDa), and contains over 38,000 amino acids. Titin contains between 240 and 300 domains with homology to immunoglobulin and fibronectin, which are believed to specifically interact with ligands on other muscle proteins such as myosin (107).

Figure 5-1. Sequence map of peptide fragment coverage of CabA detected in Avicel grown culture supernatant. . Avicel grown culture supernatant was digested with trypsin and subjected to peptide sequencing by tandem mass spectroscopy as described in the materials and methods. The figure shows the MASCOT analysis software output with confidently detected peptides indicated in red type.

Nominal mass (M_r): **1567073**; Calculated pI value: **4.18**
NCBI BLAST search of 2_2149 against nr
Unformatted sequence string for pasting into other applications

Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **17%**

Matched peptides shown in **Bold Red**

```
   1 MKRFFGLSAS LVAFLFSATV SAAVNLDFRA TTSLYETEEE FHYTLAISDC
  51 RAVTSIKIGA AGALEAYDLS DATPKADGSG GCSINVSLAL AEGNAYNPAA
 101 EVNLLNEPTQ FYSEDFSGDS AVPTIQFESV QITSDISTII GVNGSQNTQY
 151 LVATARVADD VDVSHVVFSV TGLRASNLRA NGGLVHKAKE DAFLQVEGRR
 201 VFPATYDQQL LHFVQALPNS LSYEEITRDV LVLVEAAVAD SRGRQANFSE
 251 LTTLGDNLNE GILSWRLSQQ RLQFTNVLDS AQLIPLVNYE FRGEVPLVGA
 301 GAGVSYSSSD DSLLHVTPSG VVYPLAESSE ELFITLSYAD LPDISIPVSL
 351 DFSKQLIGLE LEGLSGNQPL VLPSLNSFHP LPAISAVFDD GSKSALNRSV
 401 KPVLSMAESA SALLDLDSSV PQIRAKSAID STSPVSVSLG LDLFSDISVN
 451 FMVAAEDGLP EIELVSSNGS LPRAVAVNSV FEMVANARDD VGISSVEFWV
 501 DDGLVARKET RPYSISLPIA ETMVGKKLSV YAVAIDTAGK KTTTPIKVVE
 551 VTAEQKVTFP DFEFSLPYDG QRIIEGTPFN ARVAINLGKA VPPEKSAITY
 601 VEFFADGRKL GQTYYAAIEQ DKTSEEFYEL WQYQITAPEI STNETSLALE
 651 ARVFVGNQHK DAGGKIIRIK NNDSPKAKII QPLTGAIATA GQKVRVSLQG
 701 ADDTLGLGTR YQLLVNDEVV DTKKYIKPSA DFSAFTVQEY GYDSHVEVFE
 751 FAITEAMIGQ TLKIVGRIED FHQEISESEP LLLPIKGDQA PTIALSHPIE
 801 GQHIVSGLPV ELRANASDDL GIKRVDFFVN SKLVGSDAAA PFSFVFTTPE
 851 NIQTEQPLAI YAVVTDSAGK VATSNTVNAT LGKDETPPVV NISSPAVTAK
 901 HAGVDVSEVV EGSKFVLKVT GYDNVEVQSL TLYGVRKEAG MPYELTGNFD
 951 DVLTGEQFAP QPVPGNVNAF SALKLLEAPQ FKNDVGVPFD SYPIKVEARD
1001 LAGNVSTVEV VVGVFADDKP VAVRAKADKA GYYVNDVAVI ETLATDDLSV
1051 HAVRLALYLG ADATPFYEKT VRAADGITPA DILVHAFTVP LSEFGLANEN
1101 QVVRIELQVE DGNGTLSDVL LSDGTDLRPH LRILKDAQKP ALALASPVEG
1151 SPLYTGLSQS FEYGLSDNAG LQSIVVRNGA TVVHSKTFSA NEKVYSGSFS
1201 TAIGSGDSLD LSVEVTDIHQ NTTQDTWTFK VRNDAPPSIS IRTPAAGARL
1251 YEGEQFTANV LVDDDREVRS IVFFVRNATT GESLEEFSVN VNSAKGASEY
1301 YSKSLRVPHR NDGEVMELGV RATDSASQIT EALFQIDIID DDEAPYVSMV
1351 EPSTAITVLP GESFKVAGAA NDNIYIDSIQ AVLTDANGVE TPLEWLAFGR
1401 KDQVENITIP NPGTFGGIVA STRMKTDFQG RITLPDSFTA RAGETFKLSV
1451 RALDRGINVG NSLSVDFTVA LDEEPPVIDI SSPAKNLYVL QDVIFKANVK
1501 DNVSLASVKV YVEASGKERE LLRSYEGLQG NNNIAVPKSG DTEISIDISE
1551 YTSFANGNVP LSFIVEATDT AGNVSTKSQL SSILPDSRPS VSVQALTDLY
1601 SGAYTKITLD IADDYVTPNN PVRWVGMLSS LKGEDGTSPR AFTSAVYAFA
1651 DDNLIDYPLS TSLADLYANA IEPTILQNNT HTLQVVSAAY PDYSGSSFQL
1701 TAGSFALKAA ESKLNLLSKA GPLDYIVNFK HEDVAEFLFE ITVYSQDECG
1751 LGVTTQQVQS AGELDLTNYA SDKTISEISI KVFALNGVAL PVKQLDVKLA
1801 RLKGISTEFI YTLHILDRNS ASGLAVVQPL QPSVVSSVST SKTQGLWVPR
1851 NEWLESVALY SLGTDRYELF RAQDQNLAPV KSWVLKQDEL VPDLSLRSPE
1901 QGSEFVAGEA IPFKLYLSGS EYLDAIKIKQ NYNLPLVEYF PGWKSSQFEF
1951 PIRIPNNYTF DEYTLTVLGV DKSGNETEFN VTYPVGINEK PEFKFVAFES
```

```
2001 YKVNGKPLKS IREAARLNYG EFWVRQGETF DVEGLFSDDA SLSHIELSRL
2051 NYDGSVAETI FEESMGNSCP EYPVKSVQRK IAVEFDRAEP TEYQFKLIDS
2101 LGQITTRTFL VHPIANVAPA LRFVSPAQNQ EVASGSFNLQ VGIVLADDRF
2151 LSGSNIQMYA NGLLLNAPRP IEATDLDVDD YDLSVEQKFA EIYDEFERNY
2201 GVEIAKEYGR PSSPYALSRN YAVDLPAGLS SNVTEIELEA YVTDSDGAVA
2251 FSRVTLNVLP DDIAPIVSEL ALPDGAIEGT DFTFTYRGAD NVLVNQLELY
2301 RGYGVLLNSG EYIKSTPDLL RTIDDIPRID DRPITTNNID TPQYSQLVNT
2351 PKLRDIAATL SVNYADVARF DTLITVVVRD TNNEVSLTNS VTTRPDERPV
2401 LDIVEPLNNA TVVENTPLYV NVNAFDDVAI AYVNLVATYG NGQPAYQMRL
2451 KAPPYNYSVP VPAYDANNAA NNLITISVEA VDSYGVNYQD VDAHRAEETV
2501 QVQIKPDQGP LVAIGVPENN SEVTEGRYLL VQVNATDDVG IQQVVLNIDG
2551 LKAGTKTVTD LVYPYDFVVH IPYGEAGSNL KFTAYATEIR KNGVARTAET
2601 PTPINVKVMK DATAPVIKVF NPVADSAITE GRPFTYNLDV TDNVEVRSVR
2651 LELLAHITDE NNDVSKKSVG SQLLLGAPYY GVMSVGKIED YVPAGLPVPD
2701 QLLMTFKITA LDGAGNNSVV ERAMVLKRNQ KPSVNSMKTL DARGYNLGDI
2751 TEVTEGRGIV ISVLASDPEA GVDSVTLYQS LSVGGVEGAY HVVGEDVALP
2801 FQFHLTTPTG RVGEVLRFKA EAVDVDGYAS GISPVLKSLT LTADKPPTAK
2851 IVKPDNKNTV VINGENIEVW VEAFDDLGAD GIDRVVFYVN DKPVATVYDS
2901 MGERNGSVAQ NHIYTTSLET PPGVDGYYIY AIAYDKLGQV GQSQTIMVGS
2951 IKDTVEPELN VLYPFDNEII TTAETLRTAV SVRDLGLSIE HVKQYWVRQY
3001 QVESGAWVAL EESEIELFRD DTQSAGAGLP SSDPENYYYV YWADFTNGNI
3051 LKRTDARNER VKVVTEVKTP DHVVTEETVY EVGLPIAERR YIQPNSSDIN
3101 AGKDIYYTAV DQYHSLERTG GMVVAWASQT PLMFEQGLNL KEFDKQGANI
3151 PRTGLYLMDA ADEAFSEGNG NVHVYSSLLN PAAEVFAGTI SSVKAEQNII
3201 VAGKMGFEPA EFIPDDELAK SLLESNGLTD DEIEAVLEQV FGEAANGGAF
3251 VSKLETEVKA DWLGGDTGAE GSGELYYSNL GGELLVFNTV NADNNFGLPY
3301 ILAGRIDLPY SDVYGIDTKD NIAFVANGNG GVQVIDISDL GAPYHVGYIK
3351 PNGFTRDVKI AGRYAYIAAS HEGVVVADVA DPSMPIIAKI DTLGIANRIE
3401 IVGNRLFVTD MAGEGFTSAV NVVDISDPFQ PKMQSVVELQ PARKDLVSDG
3451 VYDVKVLGNL IYSSVLYSDQ EDIPAQSVVE IVDLNKLSQR AVDPSVPVMV
3501 NRAVVGNDIA PMDIALARGA IQIASGKQGV NRLELPELTV LEHSPYYEQF
3551 DVSTATHNII IELSAVLGQS IDLAQYIQVF EGNPALGGVD ITDQFSLGFE
3601 QRFNSTANAY EDHRRIIKLE HFDNNQFYAN QRYYVRVKAG LPALTGYPLG
3651 TDYEFSFVTA ANDGAAPQIT SICTLVSIQN GESGCRAAGS INGGTEIAVR
3701 GLNFSDKPTL EIGGQALVVQ SYTPATIDDP YGTIFAKTAP NYAGPAAVKV
3751 TNKQNLSDTV IGGYTYVDIL EISFIDPAVV RVVQAGEGDE VSVVGFGFHP
3801 GVKLTAYKSG VPSSAIVSKV NGDSLSLYSA EKMTWVVPDF SDGAGQPYRG
3851 FVDVQISDDL GRSFLLQNAL FYGRLVINRS IETADLITPK KVDEYLEPPS
3901 SNVKSLDRLP PGEIVDLAVD TDLNWVYVLG RGVQYKKGLT DKGMWLDKVS
3951 SKDAFYTYYA PSWISLVKYD RTDLTEAAPR HGFGYYDLPQ DLVATDIALD
4001 KEYLYVTATG YQLDRINTPY EDTNVLLVYD RELRDEVDIP EEGKTRDILY
4051 SLPLPIEHTP TKAIIKDKLV FIAAGSDGVV VVNIADPVKP SVVTRLTQGT
4101 VAGKAKALFV WDIELVGKQL HVVAGLTASP STKYRFVFDI TKPSLPQLGA
4151 SNFVNISAPL VNEKAFTTGG NLQLVDSTYP QYLRADGSYN GFGFSIPGTV
4201 TSANSLVSSS VFLSREPLKG DNNKSPNIGY SLGGEKVSFF NSGSGALSRF
4251 YLGAYDVSDR KNINLLDALV VDITRSAADA GVVEDGRETV VMTDDGLALF
4301 AVNRWMDGSA QSFRSSLNFV DLLVLDLAST SPAKGASGVH LDKPIVLNFT
4351 QPIGYSEDEL LSYINLHKDN GTADLESVAF TLSFEGDDRR TVVVTPEDSL
4401 ESNQRYIVNL TGVASSRRTQ GLFDYALDFV TGSATGPELE FISLEPSVSD
4451 VDGGLVNVVL NNGGNAPSFM VSGQAASVIS STPNDDGSET YSVQVPGNSP
4501 GPASLRVVRN DGDTVEVVGA LYYVEPLLLN SISPSAGSLE GGTIVTIKGQ
4551 GFRSGSNEMS VRFGPYFAAT DSIKVIDSET LQVTVPGGRI GSVDVTVEVI
4601 GKQQGVLSNA FEYLQPMQSL IRAGEKSKST YYDAVRDPSG TFLVVAAGTD
4651 GVVIVNTDAS TYTANEENPL NPEDLLELID QDGDEQDDRL VANIRLPGGY
4701 SALSVATYFE RGFDRVFVLG ARVMGEQAAD PKLFTLSFDA IDISQSTIIE
4751 ELVLPVNAAR HIEIENNRVL VGGDKGGLVI VDSYLQDKQY TFDAYPMPQQ
4801 GAVLDVAAFN NESIKDSLYA LVGGAFHYGD NKLISEETVG AGGFYIATNG
4851 PENGFQVLAS LDVPASKVVV DGNVAYLAAG SSGLVVVDIS NPAKPAILSR
4901 VANIGFVYDI GLSGGFIYAA LGENGVKAID VTNPYAPAVR EGFDNLETGF
4951 ISTVVAGAYA AYGLGYDEFS SIITVLPDAV LKIHHIDPVN GLLDIGLDGF
5001 AKFIVRFNKA IDLHSANLGL FSVSDSAGRD LSISTQIINN DAIIKLAQSE
```

5051 VDQLTVGDLL DVNIQAGVAA VKPGNGNDYL VLYELKKPVT KQFTYRGERA
5101 EMPEIDSVVP RRVPVNTAKT VTVSVRNASE NLTDIRFFLG ALELNLTGIQ
5151 VSGDDGEVTI AEVVVPAIAN AGLYDLRVEV KHNSLWVSDT LRGAVAVDAP
5201 IEIQSISPLW GPISGGSRIT ITGNGFEPGT SISETIKLRL GSLPLVGLDV
5251 YSTTEMTALT PPGSPGKHTL VATDRHGRTS VVDNEETFGY GLKQLAAISP
5301 SRVNINELLE ADSGAVIATT GRLQTTLDPT GKDLVVFNGV AAVPDNYRVA
5351 SFDVENPLQP LMVGGAYALD SYSVTGSNTI GDYAFGLDSL DIYQSRELEE
5401 NVWRNRLYVA SGNGGIARLN LDEEAGMQFI SEQPAGNGEE SYDAPSKHIV
5451 AIEKAGYSVL AGSVSPNGKT PSQPDPICVH AIGGEYSGNI ELHNYIAPFD
5501 PIYVGELPSG QSPIFIHKYD DWVFSGGARK SVGWVPGACG TTPPFTAGPL
5551 AVSNNTVNVI NLVDKNVTYD VVFSAGVMDV AVYGNHLIAA LGAEGIEIVP
5601 LHNLEQHLKI NFDGLQAVAA TATKLDVFGS TLFVSSGRGV VVIDIADIDA
5651 PRILSAGNSE KVESILVKKG TLVAGASDGL KTFEFTHSFV SDFGVREGGL
5701 VEAGDNLSLT IAFNELTTID SVIADGNVLV EVVDTVTGEA SSVSVAVAPV
5751 DEAEGLATEY TVTFDRVPNA QMRLAVNDAR NGRGGSIWAK SVRTFNVAQE
5801 GDIDVQIHRV DGGLFAVNNQ PLIAITGAGF DGVTEVYVNN YPVDSAQIIS
5851 DSLIEIPAKA FEMAGIVLEP GQHNIAVVSG SFRATWHGAL LVGDSLDGLS
5901 EADFKLSADS ALTTGGETIE IDSTRQVILP GTIAVLKKHR DPDYLIHTGR
5951 SEFDEVFGEY TIDLRDDVKT LQKFSFNLPA VIDPELYDLY LRIPNGSAHS
6001 SADSDTYHKI FVGSISYTQA TGIEIDLPNY PPHVIGAGQL VDDILFIGLK
6051 EGATTRNAYN RFLLEYGLEI FDINIWERPV RLSQLALDAP VHGIAVVGNS
6101 LLLANDSAGV QLVDVTNLNQ PLLLTGMSIP GYRAMDVDFD FVRGLGVAAA
6151 AHPFDTGILR FFELEQGELV PAQGFGSIEF SEEGDDPLRL FGTPADVQWL
6201 SDGSLYVAFV RDQQVFVAIF NDLPNAAQYE VIAIDRLKVS KPSDVSMHVQ
6251 NGQIILKSES AANAPEDLDG SAQDVVVYTR TASGYELSYW QLLQNGHSEI
6301 IENSGEVFTS VADGMQVVNG GAFAVSSVEP FANADLGQGE TVTIGFNDLF
6351 NTDSVHVTEN IEILGADGQP IPADQYELVA ENTLQGGRLL VKFSNQLTYN
6401 GVITLSVLES LTNLDGRQLN EAFTASYTML AGERPFIDSV VREENGELLN
6451 HYFHANGTEV AIVRGARFGT ESDQLEVYIG TELINASDVT LVDGETLKLN
6501 VPDLNVDGVS ASMPLRIVRK DSGLAYVLYG ALSILPQIEI DDINPQTGPP
6551 QGGNYVDIYG RGFSHYSNVL FGGTLAGDLK VYSSNHIRVR APAGSFGEVE
6601 VTTTSDLFTS EVAVSPTAYF YTGQGTGSVN LANDQSSPVA AIAMQNQVLY
6651 AVTGGAFEVV NTSGDITANL SSTTARLVLA DISDPVHPII LEKEFAGSSF
6701 PYHFDVQGGV SPKGFIDLYV NDSLLIVLGG KQLFLFDITL PTDPLRLQTV
6751 NLTHEANSIA VEGSTVVVSH ANGISMYQLV DEALRSAGSL SLQQLGSVPG
6801 KVKIYSDSLW VALPTAKQVI EVELASGTYG IKRRVDTVDN VDAGIANKLA
6851 ASDLVVVNNL LLVATGSSAS VQAYVLNSET SATPVASINL AYLVSQGTLS
6901 ANNLAVWGQT LYVVAEQGDV QIFDISNWLL GEFTAAVELQ NYFSVVGNAS
6951 SMAFSPSAIY VGSSFAYIGD EPAENPIDED VTRMGGGLTT VVNTDLLIHA
7001 QWPAVSGYLP YQDAIRVSFN RVLNNQQLQN NAQGLITLTE VGGGVITGFV
7051 SQQIDNDGAN LIFVPSQPLV AGNLYEFNID QSIQTLHGQT LPNDYTFRFT
7101 AVEGAQPVIA DVFPIKGSWR GGEIITITGN GFDSNSVIEV GGFEVPSVDV
7151 LSVSENTLEF ILPGLNSAPQ DNRLVAISVL NGPLQATDLS AFTYVTDPTL
7201 SAIGSYNRTS QTLDKNDKKF YFNAGEVIGM QGKGLSADTH IRVNGKDVPN
7251 VVEERYQELS FVVPNNTVGV LNVEISNDGF AEDTIADTTL SVVLDGNKQL
7301 SDVSKVKQVG NNLFTFSNSE VRIYSIQDST VPAFISKFDA GGVIRDIAVD
7351 GYDIAVLLAD GATVHYYSLA NIYSAQFVNA FTNQSGLNAS EIIIRNGETY
7401 LLADSSVLVG SYNGYELLAL DVSFSIVELT IAEDALYLIG ATQIEKRLLS
7451 DLSQAVGTVY GDFSGLLQAS VNNQYLTIVQ PYGVQLFDTY QFDENALLDS
7501 LSFNSRIYSA SLVGDLLTLH TASNVAVYDV DYDQTYGLQI SWKADVDNEG
7551 SRLTATKHSW INSGVLHWLD STSYYSVELP FANLELVDPY LIGSESDVIS
7601 LTVQRQPLDW SVSELSVKRS SDGVAVVGGT VLQGNALSFS AITGSYEPGE
7651 AYDLQLTTLP NQSLDGGHIN LDMPRRIVAS NAFDQAEIAI KRFAPATAVA
7701 GQTVEFTVYG QGLNTVERLV LGSLELNSSQ FNASADGSRI TFNAAVNSVG
7751 LVSVAVSQGT TTDGILAALN VVPALSIDSV ESSSNLGTTS VSNLGGDQIT
7801 VNGQGFTHDL TVHWLRAGVQ MVPDASNKVA FHLSGGEITL NAPAAQPGQA
7851 YEVAIKRVST GELVFSGEQN QLLAVDNSKP KVTVVSAPSY QNALILEGDE
7901 ALVLASLNGF SVTKTFKDYS DLATTSIDVS NRFVLSQLSD TRIKLGLREG
7951 FALEHNALYT FVISGITDTA QNLPFDGEGM IAGVYTTSFV ANDTLPAKLD
8001 SIRLALNGSY VESTTSLKRG AAYTFIPQAE DNYQTSAQLR YQYRVSFDGG
8051 LNFVNNWTQS EEFTLQVPSY ASNLVLKVRV TDASANAIEK NFNVVTVPAS

```
 8101 VQLSSNPQFP AFFTNPEWVE ELNPSEIGFH LVGDLELIKT AEIKVFDDGV
 8151 WQPALVDSST GKTSIQFNQP RLSDLNGQTT IPVRLRVAYG LQSETTIANF
 8201 AESYELNPDS TAPTLAFVAP GNGGFVPRGE QVDVILRSFD QYGIDRVELC
 8251 KDSASANPFD DAQACSILSN PNSFKFTVDP NAVNAIVLHA KAIDTNGLES
 8301 AVATLTLNPY DGREGAPELR VISPLAGDTF HAGETINVQL YLHKLVEAEL
 8351 YLDINGDPSA NGSLGPISVA RTTASPYTVN TTVTLPADIT RSGVLLLRAE
 8401 ADSEGRALKA QTIVNLIKDD GIEAEPVLEL TPSDRILSGA SLFAKATVDQ
 8451 AMDDMASQSR LTIEENAVAE NFAYEENIVY TTANAVSALT ATAHLQDRSG
 8501 HSKSLQSSLT MTEYFVDDAT LLLEGDPLWV TADAVVFPSV NGGTLIAQNH
 8551 YEQGYRLVNA AGEVLFSAQY GHLKQLSYQG GYLHALVDSV QGKLVQVFAV
 8601 ANTINGVVIS EAGAKPIQGE YIGAIGNQHF IRYGAVIGVT EFDSNSQEWV
 8651 DLLGTAIDVS AVKVQIDNSA IYVLGNEKLF VYELTNLNTL TLANEYEHGV
 8701 TSPVGFTIQN HALYIWNDDQ IAHYQQAGNE ITVLNTLSYT GKVVNAKWDA
 8751 NITWLLIENS LTGNYWQALE SGEYVGQWMN AATKLFFGAS DVLYIENNDV
 8801 FARPRTNPSV ASVTVLTEAT ATGFDLIISP VIPRAQYQVW DANGEPVNYT
 8851 IAINDDSSLR IKLSKRNVNS GETVSFTTTY SGTLQYATLP DANGTVLLFV
 8901 PTNNAVLAQG SKLPVYVSSG DDVVTIQLNG EAAPASYLNG SMVWLQNGDQ
 8951 ASDVSVSATT AQSYDSVNYS VVETTPVQTT INIGGFENNA IFTEGDVLNF
 9001 NYTAQSNNGD TLQYVDAQLR DFNNNILWNA TLTSAEGSVA LVLPNVNVQE
 9051 NYRLVVNAYY NDSYTKASGN RLLRLSPSYA LPEPRLDGLP ARVFSGGTLS
 9101 LQVANNLDGV SAFITVRDGS GAILLTGGEQ LVFTVPEQVT ELNVDVRLEN
 9151 GLGNVSSSRF NAQVDVGFDV SLSTLTHAFN VIRPEVGGYW YANGNQLTHS
 9201 GGATKVFEAE VTAIERVGAR LLVALKGFGV AFVDVATLEI ISLQQFTNSI
 9251 TALAHNNGVL AVNEAGQLAF YAIRANALEL IVKHNNLQPH TIAVSNQGFA
 9301 VSAANTLRVF GINGNELYSY TAIAPIYDVL VQANAIFAAI DNSIVRFDTS
 9351 LNIQATIADI HADKLVAHGG DILAISENEE SLTVIDARHK DVLSKLAQHA
 9401 VAMGVGVNNA VWTGGKLLLG DSRNTVVEIT ESLDYVAPLY RASRYGHQQK
 9451 GSVTDVAINQ GAVIAAANNV GAYIVRKSDD EKWTQSTYPA TEFTQAADQV
 9501 EITANAYYVL QKNLGRVVRI DRKTSAVKTL INNANLSLMA VTPSGLFVIE
 9551 EETVHLLNLA NGDVIRSFTL PAGTSAIGIA PLELGAVIVT DTGLLLSANT
 9601 SLKTLSQGLY AVKNIATNGK LIAVTTDATL SFPVSANAMA VTQREFESAI
 9651 TAVQLSEGVA VVATANNRVH VVSVFDINTD LVTFNSASKV SSIDIDGDMV
 9701 AFGLGVNGVQ VEQLGALKYY PYPVVNSPLS KTEYDLNQSL ELTINSNSGA
 9751 VRTSVSVNDA TIVDVNAVAE TIKTPIPKSI ANGQNFSTLV AVESLAGEQV
 9801 SADALTLRVQ SQNEIYNGFS VSVDVPQDSW WPEPLAIKAV ISDSDYGIAL
 9851 VEYYLSSNEV GPYQLIAKQY GPNYQIVRDF DTSYDGYYIK AVAIDEFGNS
 9901 TESVPVQFSR FQDQVQPTAS VTAAGNALTG VYAVIGVPYN ISVEAQDDAS
 9951 GIASIKIFEN ETLVAANFNS AQLTYSANPQ TIGTVRYRAE VTDNAGNVFN
10001 SNKDLSLVAD QYPEITGVST PAQIREASAF NVTVSAADDA GLGRLEVDFG
10051 STTYSHNFNG AKAGSKTFAI TDMRASRLTS SVDVPVLVRV FDTKEQLTQN
10101 STNFVTVVQD AAPVATGVSL STNQLAFYGS KVTVTINGLQ NTDDGTALNV
10151 QLLDATGDAA QPLYTATINA SNTSSRSKEI AIPTTSLPNN ELLLQVKLRD
10201 ELGQEAFSSI VPVVVHQKPN LIRFTPTLTG LNENTFAVGS SANFRVEVLD
10251 AFATPVQGIG VKWSLKNLQT SQISSKGIVA TNSNGLAEVA ISTVLQQGQY
10301 EVIAEIPTYP AIADTRHVMS ITTGPTAEVR FAYIDAVMAS ESFVLNLSAH
10351 DAAGNFVATE STRTVQLTMP TGFTLPPQNS VQSNSITLND GVASVVVNSA
10401 TQAGSYAIPV VSSGAVIGFD HDNKANTSAQ TVSELPVTVV HNVPAQALLV
10451 DDNGNVQPDG DVAVYGTKHT AQLSVMLVDE FANLIDTGEY AFTVSATGSA
10501 TINASNSSGV ANVVNGYGGF SVQNTEFETV TVSIVGVPEG LETINSTSTF
10551 RLTFNQQPPE LISAAFTAEA NSNVLVAKFS YDEALAMSAG GSSYPSLRQN
10601 GVAVDGVWSL VESELIFTPT QLIRLNTCYE YDTSASDLEG AARNDAIVIE
10651 QRELCSPRYR LAGQPYDAVY VGQELDVEFE TYDGTSRPRY LSIAMGGTVL
10701 TSGYYSLPNT GVFNSATGTY FTNKLVTVPN FFSNELFVEG STVSLALEVT
10751 NNSYAGFAVN DQQYRVFSEE GDFDGDGISN LAEIESGTLL PYSTDSNNDG
10801 VLDINEDSDG DGLTNGFEIA QGTSINLRDT DGDLLWDNLE LQYGTQPLNP
10851 DSDGDGITDG VEVGSQPFTN PNSPDSDGDG ILDGVEISLG MNPNNGDDAT
10901 QDRDGDGLDN ITEIDLGTGV NDPDSDDDGL SDGVEVNDRL SNPLSADTDG
10951 DGIPDGQEQG DDILIADETA PVVNVVSPVR ELPLLRGQTV NFTAQVTDAG
11001 IVQSTSFRVN DQQIATDSTA PYQTTWTVPE TFANQLLLEV IAHDTALNQG
11051 TSDIIIYDVI DDPLTTVTGF VTDASGNPLE GVDVSAAGVS SVTVADGSFN
11101 LVNVPMALGD IQVLAVGNIA GESVSALSIP VAPNCCDAGS ITEAGTIILE
```

11151 SPVIRVGYYS VQTRATLSSD FSVVTLSELT PETLLGIDVL AMDFVTLSEP
11201 QQAIVDEYVA NGGALVVHRD PINLEAEIPN LPGDPARGSG GLSNYYAYRV
11251 DAYSPIMNGK YGNYSSLTES SSVRGTVDVR TFADGVVPVL SLETNYRSNN
11301 AVSFYYPWGE GVVYFGAIHF GTTEATDEHE TVYWPNLLTW AYRQALPDSD
11351 EDGLPDIEEL YNGTNYLNAD TDGDLLLDGF EVQFGFDPLV ADDASLDSDG
11401 DGLSTLQEQE FNSHPNDGDT DDDSLNDYLE AMSSYISNPR SSDTDWDGLT
11451 DYEEFNSGTN PNVQDTDSDG VLDYDEINTY YTDPTRMDSD GDGLIDSFEV
11501 GHDYFDPNNA TDAFADEDGD GIINREEILT HGTDFLLADS DQDGLSDGEE
11551 LLAGLPPLNP DYDGDGLLDG EDYNPLLADS TAPQVMLAEP TTATLTRGET
11601 VRFVPSFIEE GRIVSVDYYI DNTYYQTSTN ANFDLEYTLG YGDSVSVYMV
11651 ATDTVDLQGQ SESVQFSLQD DAGARLYGRV IDMFGNTVSG ATVTFEDELV
11701 YTDANGNYEF TNVSSLALHA TVAAEVYYGN EQLIGSSVMF DPNTNGEHSI
11751 PDIVVDSVAQ LFLPFAENES NRIVTEMHNF NVSSTFNLYG NGYQWDIYSS
11801 GYVGDGADNT FDSSEILLIN GSSFPGVSYA EIVNDGRQIV LTSKAMSGLD
11851 VHREIYVPSN GYFARWLDVL ENPTANDITV TLGWDSNYGI SSNRYLAYSS
11901 NDDGLFTVDD YAIAIGETTL NDDLAGHVYG SALNAVRPSS VSYNNSDDLT
11951 RYYQVTVPAG GRRILMQFGV QHSELEPLLE EIQKIEEGDP AILVGLTREQ
12001 QRDIINFNFR KDSDLDGISD VDEVNEGTDP FNSDSDFDGL HDGFERKYGF
12051 SPSFDANVDE SGLDSDSDGL TNLEEQTNGT NPLESDTDGD QMPDGWEVKN
12101 NLNPVKDDAS LDNDRDSINN LQEYMEGTNP VFNEVYESDL DSDGIADGWE
12151 VRYGLDPTVD DSGLDFDSDG LINIEEFIRF TNPTIADTDQ DGLPDGYEVD
12201 IGTSPVNARY QVEAGSYSES SCVLDDNGVA CWGNTSYGIN NVPTLINPVE
12251 IAVGSYGACA IDDLGVSCWG SSTYYPAVPK NFINPRKLAT SFADANTCVL
12301 DDEGVYCWGE RDGYGIADAP ELLRPEDISV SSSHACALDA VGVVCWGNNN
12351 YGQTNVPPLS NPTDVEVGDN YSCAVNDSGV VCWGSSSYSL NQATVTDGRV
12401 SLISAGPSNR CELVDNELSC LGYTNYGVQQ PPAISGVVDV SVGSHYACAI
12451 GAEGVKCWGN TSDRYGMNPA FNFAADTDDD GMPDLWERAI GSNPLVDDSN
12501 EDINADGITN IVEYQLSLHG EDVDNDGMSN GWETNFGLNP LVADGDNDTD
12551 GDGLTNLEEF VHHSNPLLAD TDSDGSYDGV EVANGGSPIL PRYQVDPGRL
12601 HTCALDEQGV QCWGYSSSNI TNVPALFNPK QVASGYYFSC ALDDNGVTCW
12651 GQNNYAVTTV PSLVNPKKII TGSYHACALD DDGVTCWGNS NYGLLDVPTL
12701 SNPRDIAAGE NFTCALDDSG VVCWGLSSNG NLTPPKLSNV QTLSAGANSA
12751 CANTADGVVC WGADQSVAQP MSFTAATQID ADGTSYCSLY EDNIECWGGS
12801 NGYGQLTIPD FANAYYVGTG YEHNCALDDN GLHCWGRNDY TQTSPIAVFV
12851 FDLDRDLDGI PNLWETKHGL NPDDSSDATA EPNQDGVTNL MIYQASMSDG
12901 DQDNDGIPNE YEFAHGLNAF NGIIGDADRD GITDLEEYTL GLSASDSSDA
12951 LGDLDFDGLS NVMELRLLGT DPLESNGSNT TGGLIDKLEP NDSIEQAQSV
13001 DGLFSLNYSA DIGDENGNNT SETIPHVTID APDHTNSYDY YSFTIVESNQ
13051 VGYFDVDYGQ SSGVAFDPYL TLFDSSGLQI ATSDDYSTSA GAGGSVHSHD
13101 SFLKHTFTDA GVYVIRVGRC CTGVVNGSYQ LQVSIDGASV TVADADNDGV
13151 AVTYEEFFGF SDSNPADALS DGDADGLSNY EEFLNSTNPT NIDTDGDGLT
13201 DGREVFGGFG PHSFVDSDND GFGDDWEMFF FANLLTDGSG DSDQDGLTDA
13251 EEWALYTDPT HSDTDGDGLE DYEEINAYST SPIRVDTDGD GLADGWELDN
13301 GFDPLASGEE VLDGDFDNLA NLQEFQLGTN PSLADTDGDN VNDDIDSHPL
13351 DTNEQSDNDA DGIGDNADTD DDNDGFSDDD EINVFGSDPL VANTNVDGDE
13401 YPDGYDSDND NDGVDDTNDA FPFDPSESLD TDLDGIGNNA DEDDDNDGLI
13451 DTIETQLGSN ALLADTDQDG LSDLYEFENG LDLLVDDAAN DPDADGLSNV
13501 AEYEAGTSPQ LADTDGDTYL DGWELANGYS PLIPRYQVAL AENHACMLED
13551 GGEITCWGNE ENGQTVVPSI ANATSIAAAD YSYSCAANDT GIECWGRSYS
13601 NSLNPPNMQN VTALASGYYH TCAANAEEVS CWGYDVYGGT YAPDISGVKK
13651 LALGYYQSCA LTESGVTCWG GGSEEAESQN LINPVDLSAG RNHACAIDAG
13701 GVTCWGNNSY NQLDVPDLIN PTRVASGFEH NCALDSNGVV CWGSNESAQL
13751 DVPSLSNPYN VFAAGNYSCA LDDTGLVCWG ARYNSNFDYY VPPDTDGDGV
13801 IDEDEIAAGF DINNPNDVGL DVDGDGLTAY EEYLLGTDNN LADTDGDLIE
13851 DAIDQYPLNA SRFVQIGTGS VLLVSSHSGN LEGYEGALKS IGYDVSTLIA
13901 DGEGGISFEQ LDGFNKVYWR DDGIISGGES VMRNYLDSGR CIVVDSEDFG
13951 YYWGGALPNE YFGLEEMLYA GSFGTINTAG SLEDEIGELE PENSEDSISP
14001 VPSSSQYYSN FAWQPNENAE AILNGVASET IYPLALHNYS SEVGYHAYIL
14051 GVGLAILGET NFKKVAGAIF NQCTSEEEAL NKLDSDSDSL PDVFEQAYGT
14101 DINLADTDGD GLSDYQELAN RLNPLVADDI EALGDSDGDG LSTYDELNLG
14151 LDPFNSDSDG DGLTDLVEYN SGIYDPLNPD TNGNQFSDGF EHSYGSDLLL

```
14201 DGDEDNDGLT NLEESNAGSS PFSTDVDGDQ LSDVYEIQRG LLPGKFDSDG
14251 DGWSDSFAIR YGIDFDPTGD EDGDGLTNSE ENIAGSSPFE SDTDRDGLSD
14301 YYEAKMSFTL SYASDSDGDG VSDIRELFIT DTNPLDADSA IALQNNISTI
14351 AVYVNDLSAE IYSQDGSLNL GWQGSYLLSS TSSKLSVNGV SLSSNNAIQR
14401 CAFYQANYLS CSYNDAFSHV ESEETWSLVP HVIDQILVSR EITIPNLDYG
14451 LVRYVDSFQN IGTEEKTISV DITNTYAADN NNVRIHTSNG QDYLTASDQY
14501 SVFSDYSNTY SAIAHVYSDG STVAPTGIGH SGRGDWQLGY SAVTLQPGQT
14551 KRLVHAYYVH NSYEELLQAA NDAANNHGSN ILQGISSEEK SELINFSVCA
14601 INDTEGPCQ
```

In the human pathogen, *Staphylococcus aureus*, a 1.1 MDa protein designated Ebh, has been identified through genomic sequencing and a subcloned fragment of Ebh has been shown to specifically bind to human fibronectin, suggesting a role as a host colonization virulence factor (22). Similarly, CabA is a megadalton protein with protein interactive regions. It is known that *S. degradans* was isolated from decaying plant matter, but it is not known if it is a plant pathogen and so the function of Cab A in plant cell colonization or attachment remains speculative.

S. degradans was probed with anti-mini-CabA antiserum and FITC-conjugated secondary antibody to determine whether CabA resided on the cell surface. Under epifluorescence microscopy, *S. degradans* grown on sole carbon source glucose or xylan labeled brightly (Figure 5-3). Those treated with preimmune serum or PBS instead of primary antibody were not labeled (Figure 5-3, inset B).

To isolate CabA, *S. degradans* cultures, grown in xylan to stationary-phase, were harvested and their supernatants concentrated and dialyzed in 1 million Dalton MWCO dialysis tubing under denaturing conditions (1 % SDS) with 15 mM EDTA to inhibit metalloproteases. SYPRO® staining of SDS-PAGE gels revealed the presence of proteins which were unable to fully enter the 4 % stacking gel whereas the largest available markers, 250 kDa (precision plus MW markers; BioRad, Hercules, CA) ran at the dye front (Figure 5-4, panel A). In western blots using anti-mini-CabA these proteins were detected by the antibody, confirming that the protein was CabA (Figure 5-4, panel B). Among the many controls, lanes containing trypsin-digested samples did not contain the large α-mini-CabA-reactive protein.

Figure 5-2. Domain architecture of CabA. The solid black vertical bar represents the N-terminal secretion signal, blue-filled ovals represent IPT/TIG immunoglobulin-like domains (23), and the red-filled bars represent 15 amino acid tsp3-like repeats (68). Scale bar represents 200 amino acids and the dashed box encloses the region of CabA which was subcloned and expressed as mini-CabA.

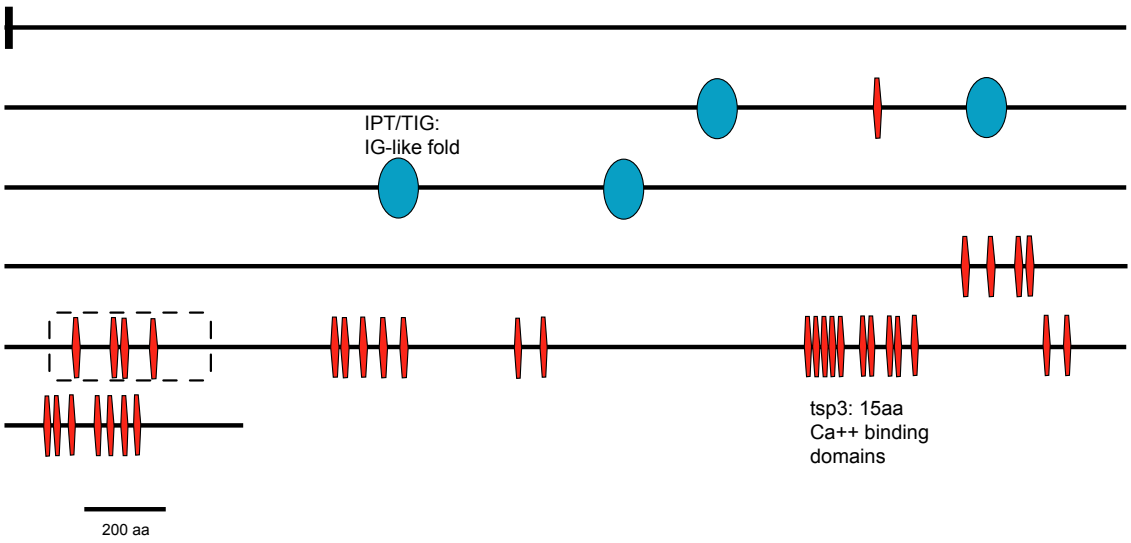

IPT/TIG:
IG-like fold
tsp3: 15aa
Ca++ binding
domains
200 aa

Figure 5-3. Labeling of cells by anti-CabA. Epifluorescence micrograph of *S. degradans* cells grown in xylan to late log phase and labeled with antiserum raised against mini-CabA. The majority (~ 80 %) of the cells visible in phase contrast fluoresced brightly under UV illumination. Cells grown in glucose and Avicel similarly fluoresced. Phase contrast image of cells shown in inset A. Controls included secondary antibody alone and preimmune serum which did not fluoresce. Secondary antibody alone is shown in inset panel B.

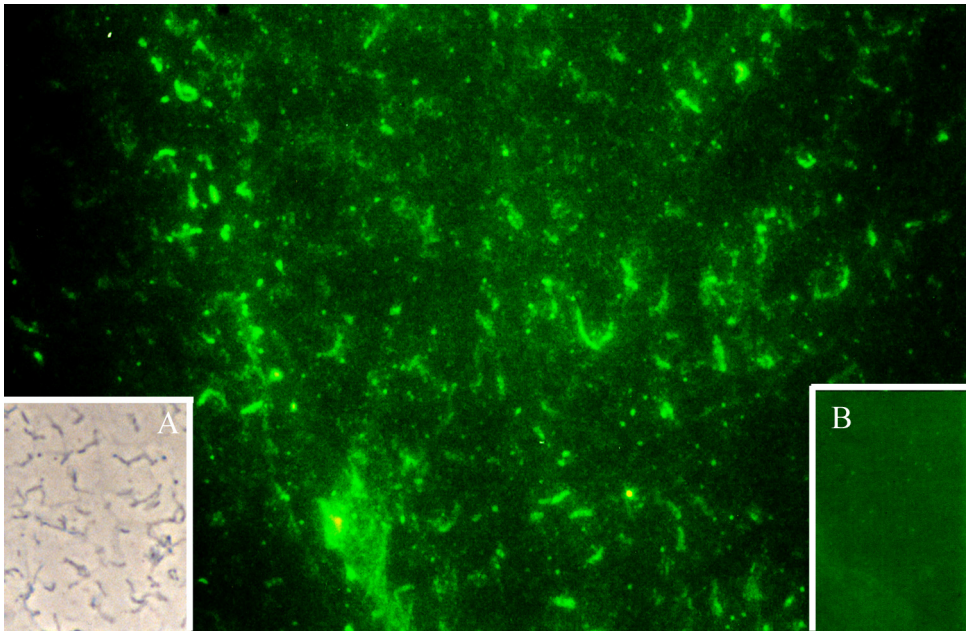
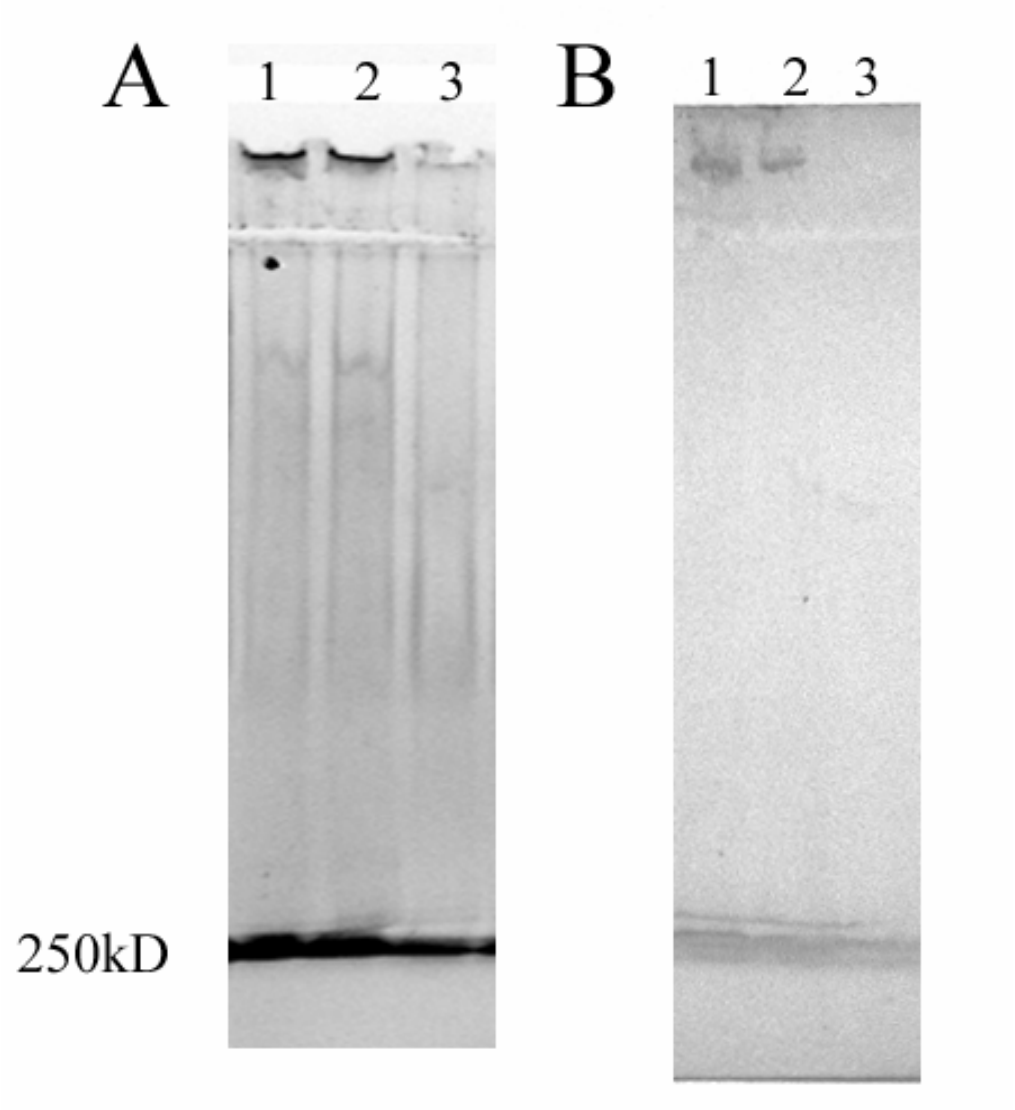

A
B

Figure 5-4. SDS-PAGE and western blot analysis of concentrated pre and post dialysis samples of xylan-grown *S. degradans* culture supernatants. Early stationary-phase cultures were centrifuged, with the resulting supernatant concentrated in Amicon Ultracon-15 ultrafiltration device with 100,000 Dalton molecular weight cut-off (MWCO). Concentrated supernatants were dialyzed in PVDF dialysis tubing with a 1,000,000 Dalton MWCO, after which they were further concentrated and analyzed by SDS-PAGE and western blots. Panel A, 4% SDS-PAGE gel stained for total protein using SYPRO ruby orange®. Panel B, Western blot of identically prepared 4 % SDS-PAGE developed with α-mini-CabA antiserum and HRP-conjugated goat anti-rabbit secondary antibody. Lane 1, concentrated, unfiltered, dialysate; Lane 2, concentrated, 0.2 μm filtered dialysate; Lane 3 concentrated, unfiltered, dialysate digested with 25 ug/ml trypsin. Note the heavily-stained bands in lanes 1 and 2 of panel A which correspond to the same lanes in panel B labeled by α-mini-CabA, and reduced concentration of visible and antibody-labeled (lane 3). The position of the 250 kDa molecular weight marker (which ran at the dye front of these 4 % gels) is shown in panel A.


A
1 2 3
B
1 2 3
250kD

Because of its detection by mass spectrometry culture supernatants grown in CP carbon sources, it is concluded that during growth on complex carbon CabA is constitutively expressed by *S. degradans* and secreted intact (14,587 residues, less the N-terminal secretion signal sequence). Furthermore, immunofluorescence microscopy reveals that it is associated with the cell surface. As the culture supernatants analyzed by mass spectrometry came from stationary-phase cultures, it may be that CabA is 'shed' in later growth stages.

That such an energetically expensive protein as CabA is apparently expressed under various growth conditions, suggests that CabA has a function which is either essential or very advantageous for *S. degradans* survival. The function(s) of CabA remain to be determined, especially the notion that it is a multipurpose molecule; it may be an S-layer protein and/or participate in the formation and organization of multienzyme complexes. Alternatively, CabA may function in cell to cell or cell to surface adhesion (Appendix II), and/or in signal transduction.

Chapter 6: Concluding Remarks and Future Directions

Saccharophagus degradans is the first marine bacterium shown to utilize crystalline cellulose as a sole carbon and energy source. The cellulase system of *S. degradans* is unlike that of other sequenced cellulolytic organisms, containing 12 endoglucanases and a single non-reducing end cellobiohydrolase. This is a departure from the accepted paradigm of *Trichoderma reesei,* in which the synergistic activities of reducing-end and non reducing-end cellobiohydrolases, as well as endoglucanases, is required for efficient degradation of crystalline cellulose (6, 84, 106). *S. degradans* is capable of growth on Avicel and on filter paper which demonstrates that the genomically identified cellulase system is both complete and functional, suggesting that synergistic activities of the abundant endoglucanases compensate for the reducing-end cellobiohydrolase.

Functional predictions of the specificity of the catalytic and substrate binding modules of the cellulase system were incorporated with the identifiable mechanisms of surface attachment and polyserine linkers to create the schematic of cellulose degradation shown in Figure 2-7. This model provides the framework for further study of the *S. degradans* cellulase system, allowing hypothesis-based testing of the proposed location, binding specificity and reaction specificity of each cellulase system enzyme.

Cloning, and expression, followed by activity and binding assays of the CBM proteins and the glycanases of uncertain function are needed to authoritatively determine their roles in the degradation of cellulose and other complex

polysaccharides. The Cbm proteins Cbm2B and Cbm2C were shown to be expressed by *S. degradans* during growth in Avicel and xylan, respectively, suggesting that they do participate in CP degradation somehow. Although neither protein was active against plant wall polymers or against the relevant para-Nitrophenol conjugates, they were both shown to specifically bind crystalline cellulose. Additional analyses are needed to resolve the functions of these, and the other Cbm proteins. Possible functions include substrate binding, enhancement of cellulose degradation through disruption of cellulose fibrils, or enzymatic activities which remain undetected. As such, subcloning regions of unknown function and assaying them against a wide range of oligomeric substrates may be needed to determine their activity. Such proteins or subcloned portions thereof may need to be tested in combination with other enzymes to determine if they interact synergistically. As there are no reports of an organism with as many Cbm proteins as *S. degradans,* the organism provides an unprecedented opportunity to study the role of non-catalytic carbohydrate binding proteins in cellulose degradation. Other modules warranting further analyses include the putative novel CBMs: the FCL module of Cbm32D, the tandem modules within Gly43M, designated LGL which have homology to the lectin Concanavalin A, as well as the Y94 and Y95 modules discussed in chapter two. Characterization of these modules will most likely result in the establishment of new CBM families, with the potential to extend our understanding of protein-carbohydrate interactions. Alternatively, one or more these novel modules may function in protein-protein interactions. In either case, their study will make significant contributions to the study of bacterial carbohydrase structure and function.

Similarly, characterization of the glycanases of unknown function is likely to extend the understanding of the relationships between carbohydrase sequence, structure and function. The active sites of these enzymes differ from the studied examples within their respective GH families, and the enzymes did not catalyze the same reactions as the known examples. As the carbohydrases of *S. degradans* evolved to function in marine environments, their active sites may contain adaptations allowing them to function under ionic conditions and salt concentrations which terrestrial enzymes do not encounter. Thus, further analyses of the carbohydrases of *S. degradans* should include comparisons of active site sequences and reaction chemistry with terrestrial enzymes with similar functions. Such studies should provide insights into the adaptations required for carbohydrase function under the physiochemical conditions which occur in marine environments.

In addition to its efficient cellulolytic capabilities, *S. degradans* has all the enzyme systems needed to degrade every class of polymers commonly present in the cell walls of plants, including xylan, β-(1,3(4))-glucan, β-1,3-glucan, arabinan, pectin and mannan. That *S. degradans* can grow on each of these polymers as a sole carbon source (this work; also ref (41) implies that each of these systems are complete and functional as well, which is confirmed y the fact that *S. degradans* can grow on and degrade plant material in monoculture, a feat which often requires consortia of microorganisms. The differential induction of the plant cell wall degrading enzyme systems described in chapter four indicates that *S. degradans* can coordinately regulate these systems, and furthermore that the organism has a highly sophisticated means of sensing and differentiating between mixtures of complex polysaccharide

substrates. These findings suggest that *S. degradans* has excellent potential as a model organism for studying the global regulation and coordination of complex multienzyme systems.

Comparisons of cultures of *S. degradans* grown in Avicel and *Spartina* reveal that relatively little crystalline cellulase activity (as measured by activity against Avicel) is actually needed to degrade plant matter. In cultures grown on *Spartina* detectable activity against Avicel did not appear in the supernatant fraction until around 600 hours of growth, by which time the 2 cm segments of *Spartina* leaves were no longer intact and largely degraded. This suggests that plant material can be substantially degraded by the activity of amorphous cellulases acting cooperatively with hemicellulases such as xylanases and glucanases. The appearance of activity against Avicel in late culture suggests that *S. degradans* has already metabolized the more easily-degraded plant cell wall components and is adjusting enzyme synthesis to begin utilizing the most-recalcitrant "skeletons" of crystalline cellulose. These findings indicate that one must be careful when viewing the world of cellulose degradation through the prism of one extensively-studied system, such as *T. reesei.*

It also appears that many surprising revelations will come from future studies of the calcium-binding megaproteins (Table 5-1). Resolving the function of proteins such as CabA, with its unprecedented size and 35 TSP3 domains will doubtless provide insight into the roles of calcium-binding domains, whether they participate in intracellular or extracellular signaling, or in protein-protein, cell-to-cell and/or cell-to-surface adhesion. The diversity of calcium-binding motifs within these seven proteins suggests there is the potential for important discoveries on all of these fronts.

Lastly, to determine the significance of *S. degradans* in the marine and estuarine carbon cycle, it is essential to determine its geographic distribution. It is clear that *S. degradans* has unprecedented metabolic capability; because of its tolerance of wide variations in temperature and salinity, the degradative abilities of *S. degradans* suggest that the bacterium might play a significant role in the turnover of carbon in estuarine and coastal marine environments. Comprehensive studies to determine the extent of its geographic distribution will determine if *S. degradans* is indeed a primary contributor to the marine carbon cycle.

Appendix I: Sequence analysis of calcium-binding repeats potentially involved in surface enzyme display

Introduction

Sequence analysis of the genome of *Saccharophagus degradans* has revealed the presence of type 3 thrombospondin-like repeats (TSP3) which may be involved in calcium-mediated protein interactions. The presence of these, and similar acidic repeat sequences in some of the mega proteins which are the focus of chapter 5 suggests a possible role in surface enzyme attachment or cell to cell or cell to surface adhesion. Interestingly, in *S. degradans* there are two carbohydrases which each contain four copies of a sequence which contains a TSP3-like motif. One of these TSP3-containing carbohydrases is the laminarinase Lam16A (chapter 3, this work) and the other is the agarase AgaE (38). Amino acid sequence alignment reveals similarities and differences between the TSP3-like motifs found within CabA, Lam16A, and AgaE of *S. degradans* and the dockerin domains of the scaffoldin proteins CipA and CipB of *Clostridium thermocellum*. The data of this section are supplemental analyses which complement chapter 5.

Materials and Methods

The thirty-five TSP3 15 amino-acid repeat sequences identified by Pfam analysis (http://www.sanger.ac.uk/Software/Pfam/) in CabA (gene Sde2049, Table 5-1), Lam16A (Sde1393, Table 3-3), and AgaE (Sde2655; 38) of *S. degradans* were analyzed by the ClustalW 1.8 multiple alignment tool at the Baylor College of Medicine web site (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html).

Results and Discussion

In CabA, the 35 identified acidic fifteen amino acid repeat sequences with homology to thrombospondin type 3 (tsp3) domains align over a 13aa region which contains nine conserved amino acids. The CabA consensus TSP3 sequence is dsdgdglxdxxev (Figure A1-1). That the TSP3 sequences contain the conserved calcium-binding motif DxDxDG suggests a function involving calcium-binding. These repeats are somewhat similar to dockerins number one and two from CipA of *Clostridium thermocellum*, containing 5 and 7 conserved residues, respectively (Figure A1-2). The observed similarities and differences between the sequences of the TSP3s of CabA and the dockerin sequences raise the possibility that the TSP3s of *S. degradans* may have a function similar to dockerin. That the sequences are not more closely related suggests that if the TSP3s do have a dockerin-like function, they may be an example of convergent evolution.

A closer look at the sequence of CabA reveals that the TSP3 repeats identified by Pfam analysis lie within regions rich in acidic amino acid residues and, therefore the identified TSP3 may actually be part of a larger domain which is not well enough conserved with known sequences to identify by pattern searches such as Pfam. Eight of these sequences were identified by manual review of the sequence of CabA and aligned by ClustalW (Figure A1-3). The first 13 amino acids of this alternate repeat contains the conserved TSP3 sequence (dsdgdglxdxxev) followed by a well-conserved Gfdpxn sequence and another DxDgDg motif. That this alternative repeat

contains two copies of the DxDxDG calcium-binding motif suggests a domain which may interact with another protein through two calcium bridges.

Similarly, the regions of Lam16A and AgaE which contain TSP3 motifs as identified by Pfam occur within regions of acidic amino acid sequence. These regions align over a 29aa stretch which begins with a 12aa sequence very similar to the CabA TSP3 consensus. This sequence is: DsDnDGvxdsxDqCpntPagtxvxavGCpvtq (Figure A1-4, A). The motifs identified in Lam16A and AgaE are quite similar to each other, but also share the DxDxDG motif with the alternative repeats of CabA (Figure A1-4, B). However, there are also conserved differences between the alternative motifs of CabA and those of the carbohydrases, particularly in the middle of the sequences (Figure A1-4, B). While these sequence analyses do not allow predictions of interactive specificities, it is intriguing to note that the mega protein CabA and two carbohydrases, Lam16A and AgaE, contain conserved motifs which contain one or more copies of the calcium-binding motif DxDxDG. The similarities and differences between the CabA motifs and the carbohydrase motifs suggest that these may be protein interactive domains which function analogously to the dockerins and cohesins of *C. thermocellum*. Although it may be that one of the CabA motifs may interact directly with one or more carbohydrase motif, it is also possible that there is another, unidentified intermediary which bridges the two motifs. Clearly many significant discoveries are likely to result from further study of the calcium-binding proteins of *S. degradans*.

Figure A1-1. ClustalW alignment of the thrombospondin type3-like repeats of CabA.

The 35 thrombospondin-like repeats identified in CabA are shown aligned by ClustalW. Note the prevalence of conserved glutamic acid and aspartic acid residues and that the consensus sequence, dsdgdglxdxxev contains the characteristic calcium-binding signature sequence DxDxDG (93).

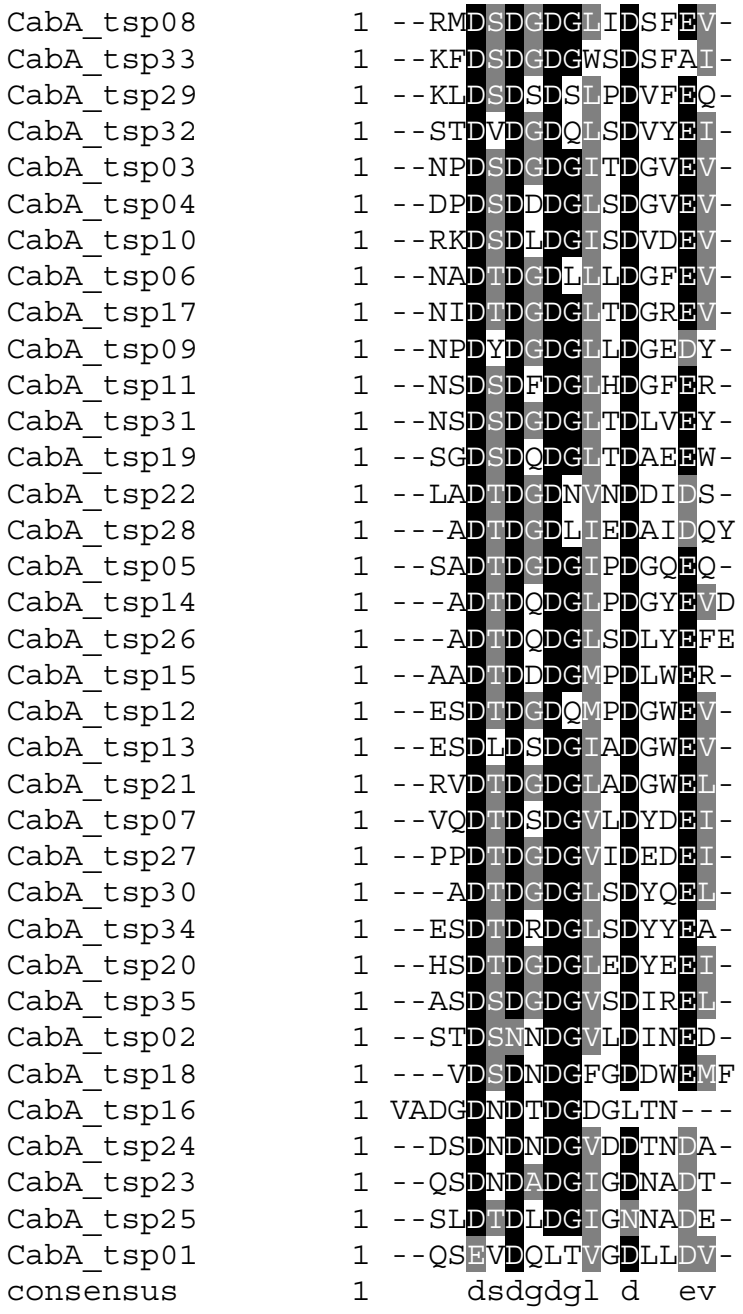

CabA_tsp08 1 --RMDSDGDGLIDSFEV-
CabA_tsp33 1 --KFDSDGDGWSDSFAI-
CabA_tsp29 1 --KLDSDSDSLPDVFEQ-
CabA_tsp32 1 --STDVDGDQLSDVYEI-
CabA_tsp03 1 --NPDSDGDGITDGVEV-
CabA_tsp04 1 --DPDSDDDGLSDGVEV-
CabA_tsp10 1 --RKDSDLDGISDVDEV-
CabA_tsp06 1 --NADTDGDLLLDGFEV-
CabA_tsp17 1 --NIDTDGDGLTDGREV-
CabA_tsp09 1 --NPDYDGDGLLDGEDY-
CabA_tsp11 1 --NSDSDFDGLHDGFER-
CabA_tsp31 1 --NSDSDGDGLTDLVEY-
CabA_tsp19 1 --SGDSDQDGLTDAEEW-
CabA_tsp22 1 --LADTDGDNVNDDIDS-
CabA_tsp28 1 ---ADTDGDLIEDAIDQY
CabA_tsp05 1 --SADTDGDGIPDGQEQ-
CabA_tsp14 1 ---ADTDQDGLPDGYEVD
CabA_tsp26 1 ---ADTDQDGLSDLYEFE
CabA_tsp15 1 --AADTDDDGMPDLWER-
CabA_tsp12 1 --ESDTDGDQMPDGWEV-
CabA_tsp13 1 --ESDLDSDGIADGWEV-
CabA_tsp21 1 --RVDTDGDGLADGWEL-
CabA_tsp07 1 --VQDTDSDGVLDYDEI-
CabA_tsp27 1 --PPDTDGDGVIDEDEI-
CabA_tsp30 1 ---ADTDGDGLSDYQEL-
CabA_tsp34 1 --ESDTDRDGLSDYYEA-
CabA_tsp20 1 --HSDTDGDGLEDYEEI-
CabA_tsp35 1 --ASDSDGDGVSDIREL-
CabA_tsp02 1 --STDSNNDGVLDINED-
CabA_tsp18 1 ---VDSDNDGFGDDWEMF
CabA_tsp16 1 VADGDNDTDGDGLTN---
CabA_tsp24 1 --DSDNDNDGVDDTNDA-
CabA_tsp23 1 --QSDNDADGIGDNADT-
CabA_tsp25 1 --SLDTDLDGIGNNADE-
CabA_tsp01 1 --QSEVDQLTVGDLLDV-
consensus 1 dsdgdgl d ev

Figure A1-2. ClustalW alignment of the TSP3 repeats of CabA with dockerins of *Clostridium thermocellum*. Dockerins number one and two from the scaffoldin proteins CipA are shown aligned with the TSP3 repeats of CabA from *S. degradans*. Note the conservation between the TSP3 repeats of CabA and residues 1, 5, 9, 12, and 13 of CipA and residues 1, 3, 5, 6, 9, 12, and 13 of CipB.

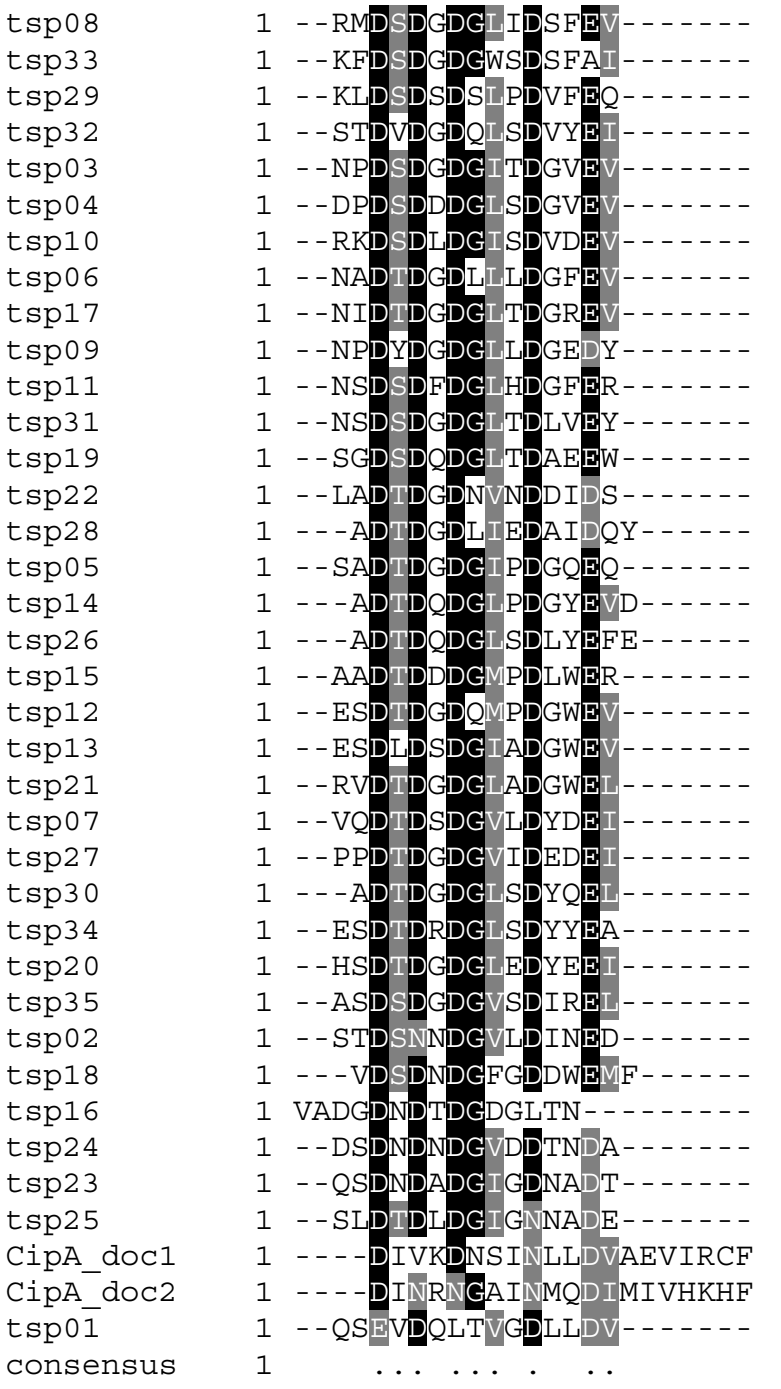

tsp08 1 --RMDSDGDGLIDSFEV-------
tsp33 1 --KFDSDGDGWSDSFAI-------
tsp29 1 --KLDSDSDSLPDVFEQ-------
tsp32 1 --STDVDGDQLSDVYEI-------
tsp03 1 --NPDSDGDGITDGVEV-------
tsp04 1 --DPDSDDDGLSDGVEV-------
tsp10 1 --RKDSDLDGISDVDEV-------
tsp06 1 --NADTDGDLLLDGFEV-------
tsp17 1 --NIDTDGDGLTDGREV-------
tsp09 1 --NPDYDGDGLLDGEDY-------
tsp11 1 --NSDSDFDGLHDGFER-------
tsp31 1 --NSDSDGDGLTDLVEY-------
tsp19 1 --SGDSDQDGLTDAEEW-------
tsp22 1 --LADTDGDNVNDDIDS-------
tsp28 1 ---ADTDGDLIEDAIDQY------
tsp05 1 --SADTDGDGIPDGQEQ-------
tsp14 1 ---ADTDQDGLPDGYEVD------
tsp26 1 ---ADTDQDGLSDLYEFE------
tsp15 1 --AADTDDDGMPDLWER-------
tsp12 1 --ESDTDGDQMPDGWEV-------
tsp13 1 --ESDLDSDGIADGWEV-------
tsp21 1 --RVDTDGDGLADGWEL-------
tsp07 1 --VQDTDSDGVLDYDEI-------
tsp27 1 --PPDTDGDGVIDEDEI-------
tsp30 1 ---ADTDGDGLSDYQEL-------
tsp34 1 --ESDTDRDGLSDYYEA-------
tsp20 1 --HSDTDGDGLEDYEEI-------
tsp35 1 --ASDSDGDGVSDIREL-------
tsp02 1 --STDSNNDGVLDINED-------
tsp18 1 ---VDSDNDGFGDDWEMF------
tsp16 1 VADGDNDTDGDGLTN---------
tsp24 1 --DSDNDNDGVDDTNDA-------
tsp23 1 --QSDNDADGIGDNADT-------
tsp25 1 --SLDTDLDGIGNNADE-------
CipA_doc1 1 ----DIVKDNSINLLDVAEVIRCF
CipA_doc2 1 ----DINRNGAINMQDIMIVHKHF
tsp01 1 --QSEVDQLTVGDLLDV-------
consensus 1 ... ...

Figure A1-3. Alignment of alternative repeat sequence identified within CabA. At least eight copies of a larger acidic repeat sequence which contain a TSP3 repeat were visually identified in the protein sequence of CabA. The consensus sequence of this so-called "CabA alt repeat" is DtDgDgvidgxEixxGfdpxnxxdaxlDxdgDgltxxxey. Note the similarity of the first 13 amino acids of this "alt" repeat to the consensus TSP3 sequence, dsdgdglxdxxev.

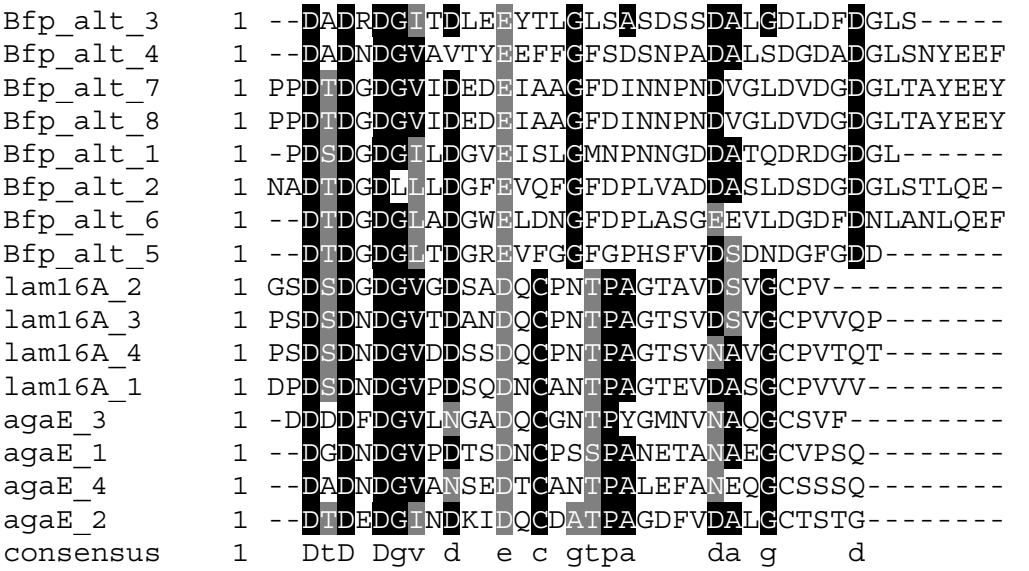

CabA_alt_7 1 PPDTDGDGVIDEDEIAAGFDINNPNDVGLDVDGDGLTAYEEY
CabA_alt_8 1 PPDTDGDGVIDEDEIAAGFDINNPNDVGLDVDGDGLTAYEEY
CabA_alt_1 1 -PDSDGDGILDGVEISLGMNPNNGDDATQDRDGDGL------
CabA_alt_5 1 --DTDGDGLTDGREVFGGFGPHSFVDSDNDGFGDD-------
CabA_alt_6 1 --DTDGDGLADGWELDNGFDPLASGEEVLDGDFDNLANLQEF
CabA_alt_2 1 NADTDGDLLLDGFEVQFGFDPLVADDASLDSDGDGLSTLQE-
CabA_alt_3 1 --DADRDGITDLEEYTLGLSASDSSDALGDLDFDGLS-----
CabA_alt_4 1 --DADNDGVAVTYEEFFGFSDSNPADALSDGDADGLSNYEEF
consensus 1 DtDgDgvidg Ei Gfdp n da lD dgDglt ey

Figure A1-4. Alignment of the regions of Lam16A and AgaE which contain TSP3-like repeats. A, Note the strong conservation of 25 of the 29 aligned residues in the TSP3-containing motifs of the laminarinase Lam16A and the agarase AgaE. The TSP3-like region is found within the first twelve residues. B, Alignment of the TSP3-containing motif within Lam16A and AgaE with the alternate repeat from CabA. Note similarities including the conserved DxDxDG motif within the first 15 residues and the conserved differences in the middle of the sequences.

A


CabA_alt_7 1 PPDTDGDGVIDEDEIAAGFDINNPNDVGLDVDGDGLTAYEEY
CabA_alt_8 1 PPDTDGDGVIDEDEIAAGFDINNPNDVGLDVDGDGLTAYEEY
CabA_alt_1 1 -PDSDGDGILDGVEISLGMNPNNGDDATQDRDGDGL------
CabA_alt_5 1 --DTDGDGLTDGREVFGGFGPHSFVDSDNDGFGDD-------
CabA_alt_6 1 --DTDGDGLADGWELDNGFDPLASGEEVLDGDFDNLANLQEF
CabA_alt_2 1 NADTDGDLLLDGFEVQFGFDPLVADDASLDSDGDGLSTLQE-
CabA_alt_3 1 --DADRDGITDLEEYTLGLSASDSSDALGDLDFDGLS-----
CabA_alt_4 1 --DADNDGVAVTYEEFFGFSDSNPADALSDGDADGLSNYEEF
consensus 1 DtDgDgvidg Ei Gfdp n da lD dgDglt ey

B


Bfp_alt_3 1 --DADRDGITDLEEYTLGLSASDSSDALGDLDFDGLS-----
Bfp_alt_4 1 --DADNDGVAVTYEEFFGFSDSNPADALSDGDADGLSNYEEF
Bfp_alt_7 1 PPDTDGDGVIDEDEIAAGFDINNPNDVGLDVDGDGLTAYEEY
Bfp_alt_8 1 PPDTDGDGVIDEDEIAAGFDINNPNDVGLDVDGDGLTAYEEY
Bfp_alt_1 1 -PDSDGDGILDGVEISLGMNPNNGDDATQDRDGDGL------
Bfp_alt_2 1 NADTDGDLLLDGFEVQFGFDPLVADDASLDSDGDGLSTLQE-
Bfp_alt_6 1 --DTDGDGLADGWELDNGFDPLASGEEVLDGDFDNLANLQEF
Bfp_alt_5 1 --DTDGDGLTDGREVFGGFGPHSFVDSDNDGFGDD-------
lam16A_2 1 GSDSDGDGVGDSADQCPNTPAGTAVDSVGCPV----------
lam16A_3 1 PSDSDNDGVTDANDQCPNTPAGTSVDSVGCPVVQP-------
lam16A_4 1 PSDSDNDGVDDSSDQCPNTPAGTSVNAVGCPVTQT-------
lam16A_1 1 DPDSDNDGVPDSQDNCANTPAGTEVDASGCPVVV--------
agaE_3 1 -DDDDFDGVLNGADQCGNTPYGMNVNAQGCSVF---------
agaE_1 1 --DGDNDGVPDTSDNCPSSPANETANAEGCVPSQ--------
agaE_4 1 --DADNDGVANSEDTCANTPALEFANEQGCSSSQ--------
agaE_2 1 --DTDEDGINDKIDQCDATPAGDFVDALGCTSTG--------
consensus 1 DtD Dgv d e c gtpa da g d

Appendix II: Morphology and Cell-surface Features of *Saccharophagus degradans* during growth on polysaccharides

Introduction

Shortly after its isolation from Chesapeake Bay, it was noted that *S. degradans* exhibits unusual morphology during growth on complex polysaccharides (CP) (4). While cells grown in glucose have a smooth surface and relatively uniform appearance drastic alterations in surface topology are seen during growth in CP carbon sources (112, 113). Along with the works already described in this volume, ancillary studies investigating the morphology of *S. degradans* during growth on CP were performed using Scanning and Transmission electron microscopy (SEM and TEM, respectively). The following section is a gallery of the more unusual cell features and morphologies which were consistently observed during CP metabolism by *S. degradans*.

Materials and Methods

Cells were harvested at the desired growth stage by centrifugation, washed twice in 20mM PIPES buffer, pH 6.8, resuspended in ½ volume PIPES and fixed by the addition of 50% glutaraldehyde to 1% final concentration. Additional cell samples may be prepared by fixing before washing in PIPES to determine if this enhances preservation of surface features and ES. For transmission electron microscopy (TEM) a 10µl drop of fixed cell suspension was adsorbed onto 200 mesh nickel grids (Electron Microscopy Sciences, Fort Washington, PA) for 30 seconds and stained for 15 seconds with 1% (w/v) uranyl acetate (Electron Microscopy Sciences). Specimens

were viewed on a Zeiss EM/10CA transmission electron microscope (Carl Zeiss inc., Germany). Samples for scanning electron microscopy (SEM) were prepared by passing 1ml of fixed cell suspension through a 0.2µm pore size Nucleopore 13mm polycarbonate filters (Whatman, Middlesex, UK). Filters and adherent cells were washed 3 x 10 minutes in Millonig's buffer (0.12M Phosphate buffer pH 7.4, 0.25% NaCl) and post-fixed 30 minutes in 2% (v/v) osmium tetraoxide (OsO_4). The samples were washed 4 x 10 minutes in dH_2O and dehydrated in ethanol series (10 minutes each in 75%, 95%, 100%, 100%, and 100% ethanol). The samples were then critical point dried in CO_2 using a Denton DCP-1 critical point dryer (Denton Vacuum, Moorestown, NJ). Dried filters were mounted on SEM specimen stubs and coated with ~10nm gold/palladium alloy using a Denton DV-502A vacuum evaporator (Denton Vacuum). Specimens were viewed on a Hitachi S-4700 ultra high resolution scanning electron microscope (UHR-SEM).

Cell to cell contact

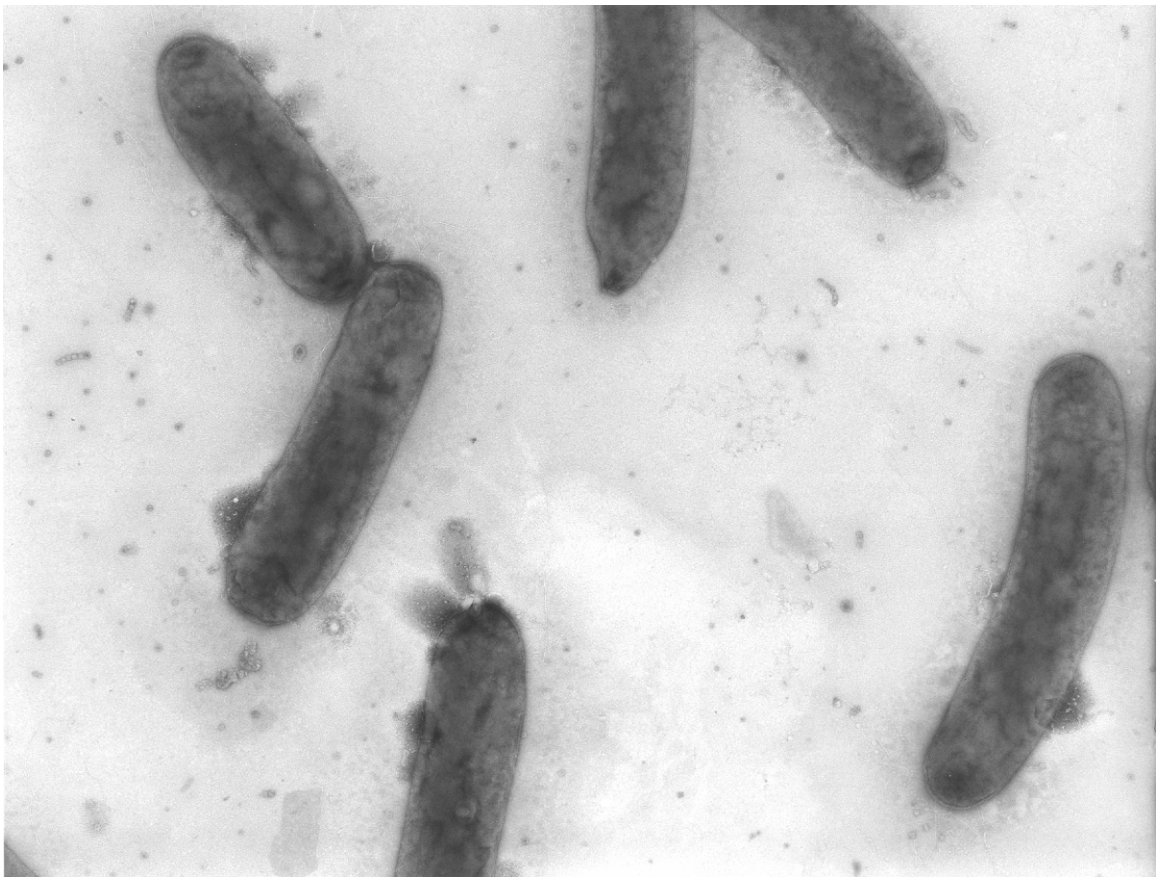


Figure A2-1. TEM of stationary-phase glucose grown culture of *S. degradans*.

Note generally smooth cell surface and polar contact/adhesion

Cell to cell contact

Figure A2-2. TEM of log phase agarose grown cells. Cellular contact appears to involve densely-staining surface features

Cell to cell contact

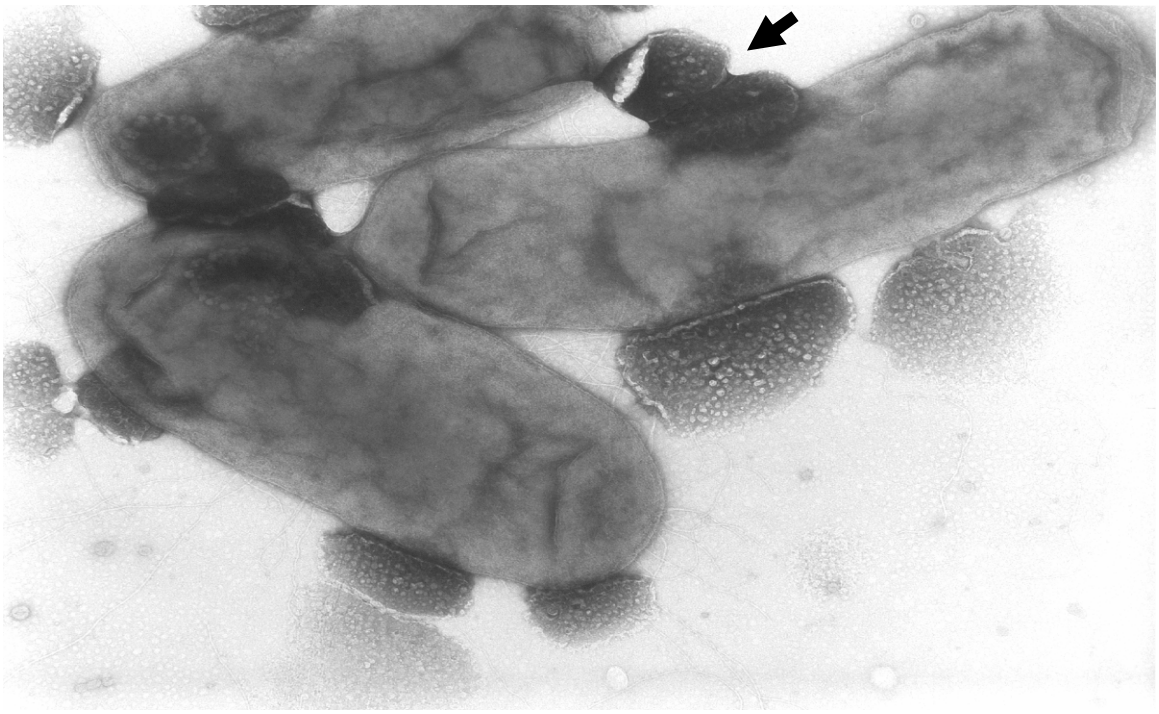


Figure A2-3. TEM of log phase agarose grown cells. Note membrane to membrane contact and large surface features with granular staining pattern. Also apparent contact between the two large extracellular structures (ES) indicated by the arrow.

Cell to cell contact

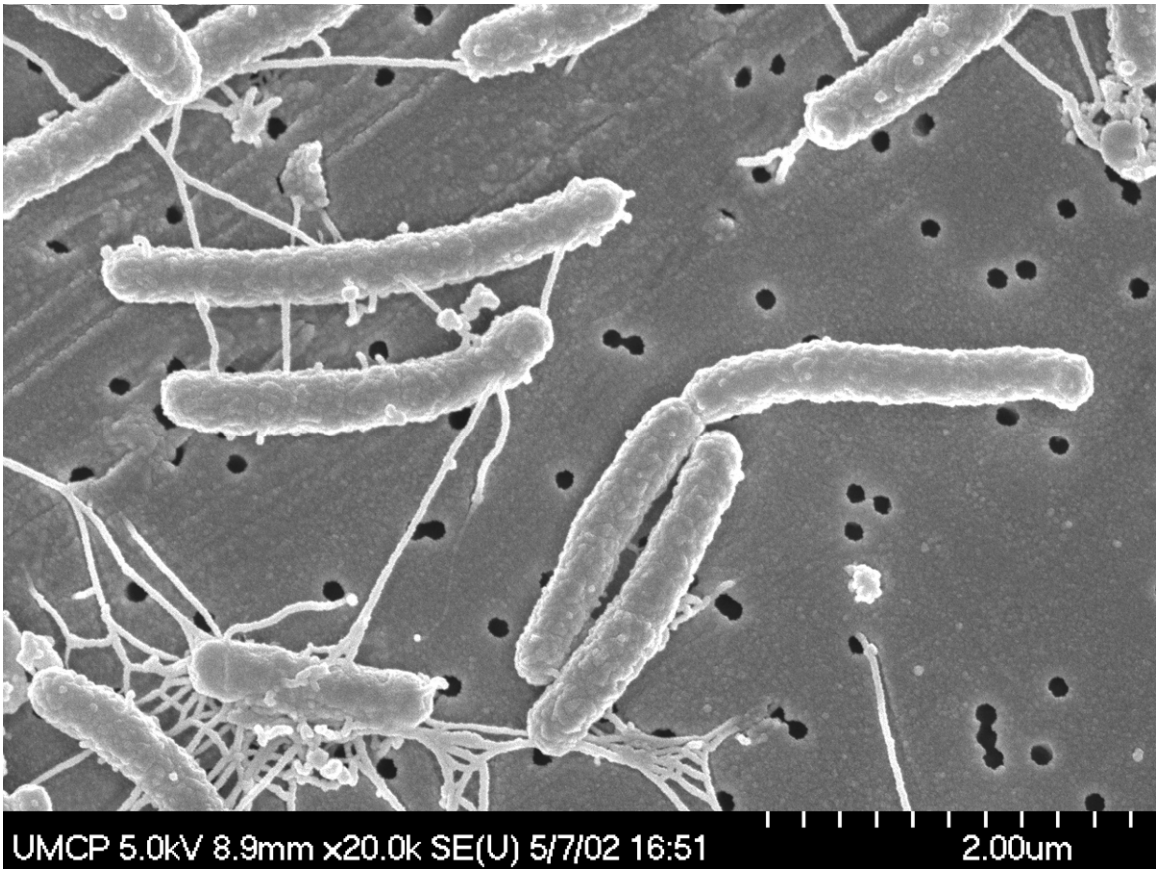


Figure A2-4. SEM of log phase agarose grown cells. Note cell to cell contact similar to that seen in Figure A2-2 and irregular surface with tubule-like appendages.

Cell to cell contact

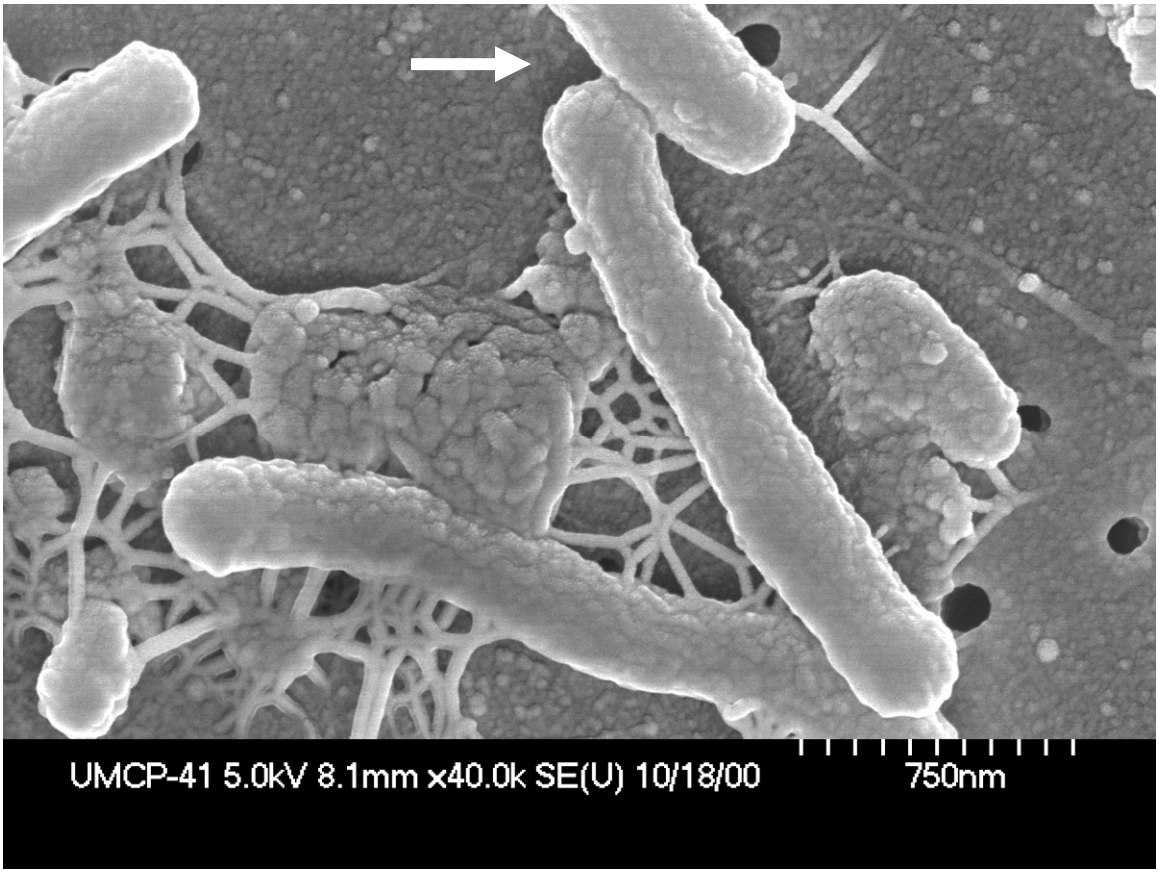


Figure A2-5. SEM of log phase agarose grown cells. Arrow indicates apparent cell to cell contact.

Cell to cell contact

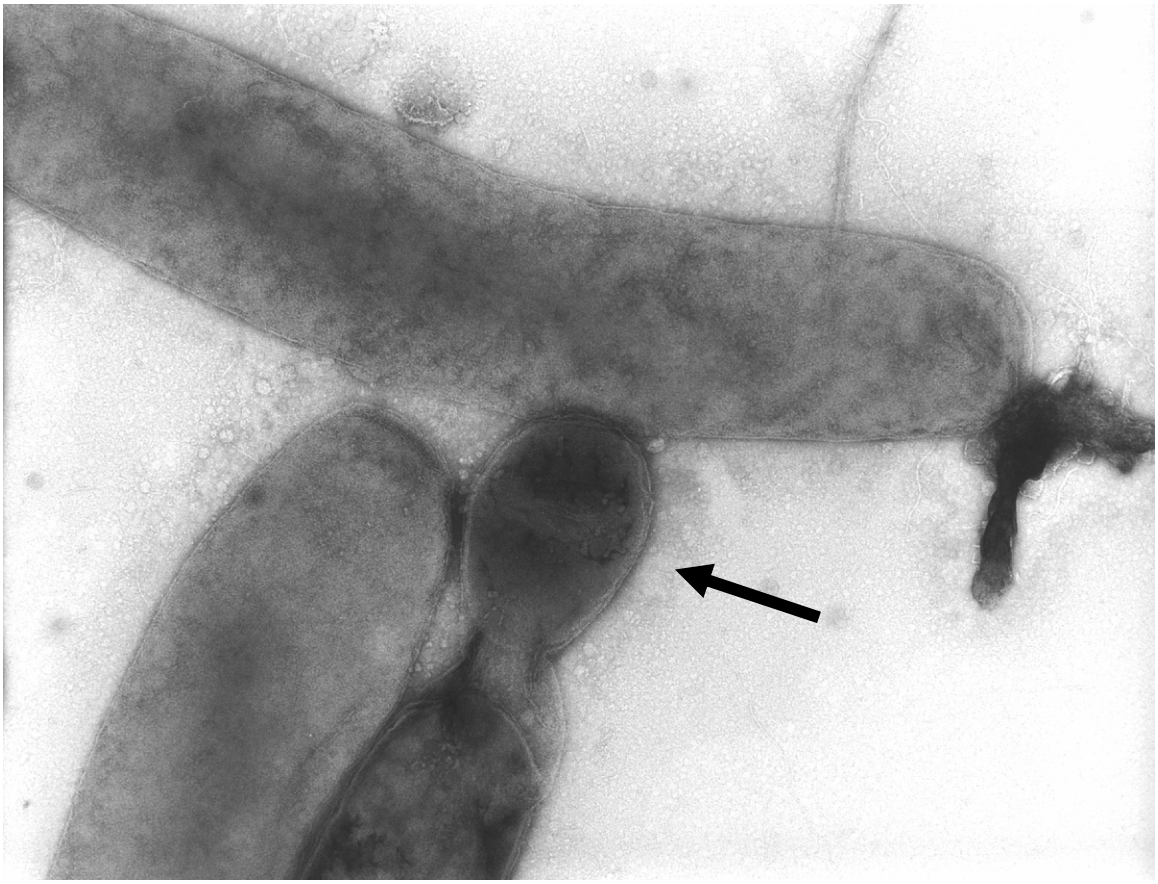


Figure A2-6. TEM of stationary phase glucose grown cells. Note cell to cell contact mediated by an apparent membrane-bound ES or coccoid cell (arrow), as well as tubule-like appendage and densely-stained structure on cell pole.

Large extracellular structures (ES)

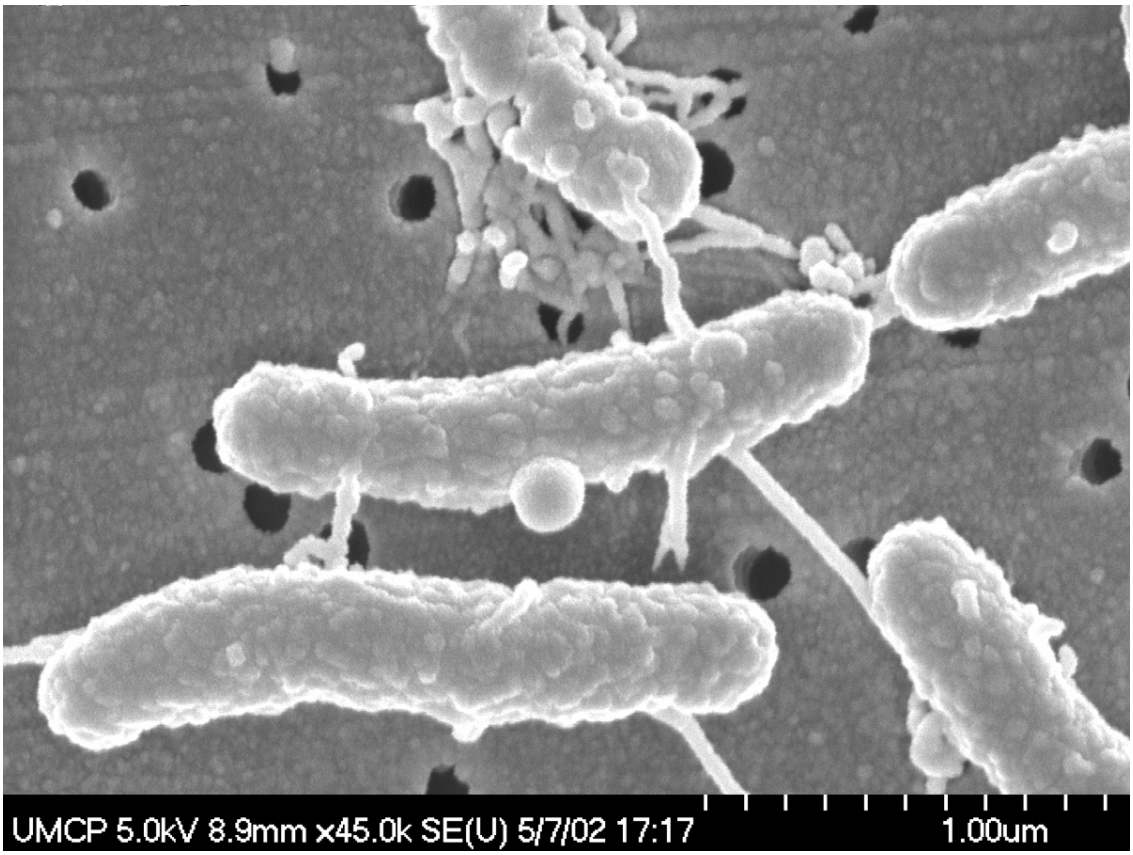


Figure A2-7. SEM of log phase agarose grown cells. Note large surface protuberance on cell in center of the image as well as smaller protuberances and tubule-like appendages.

Large extracellular structures (ES)

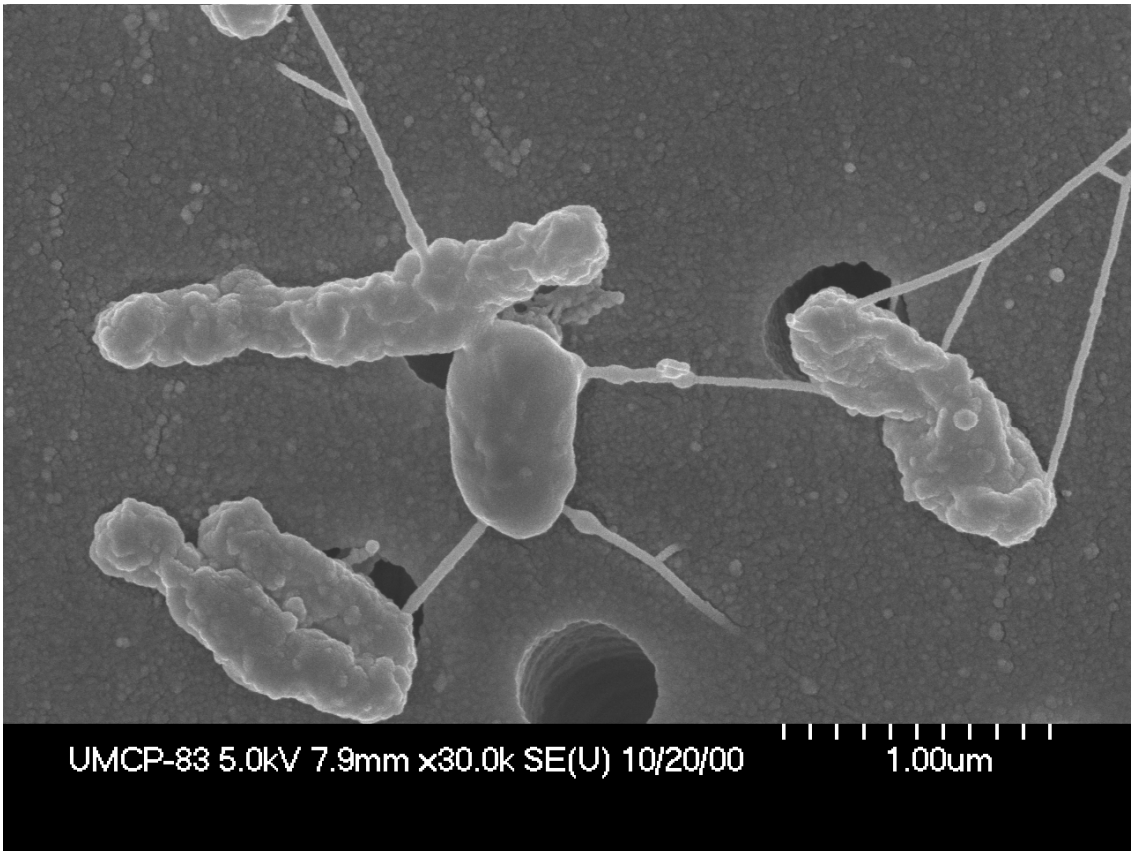


Figure A2-8. SEM of stationary phase agarose grown cells. Cell surfaces have pronounced topology and tubule-like appendages.

Large extracellular structures (ES)

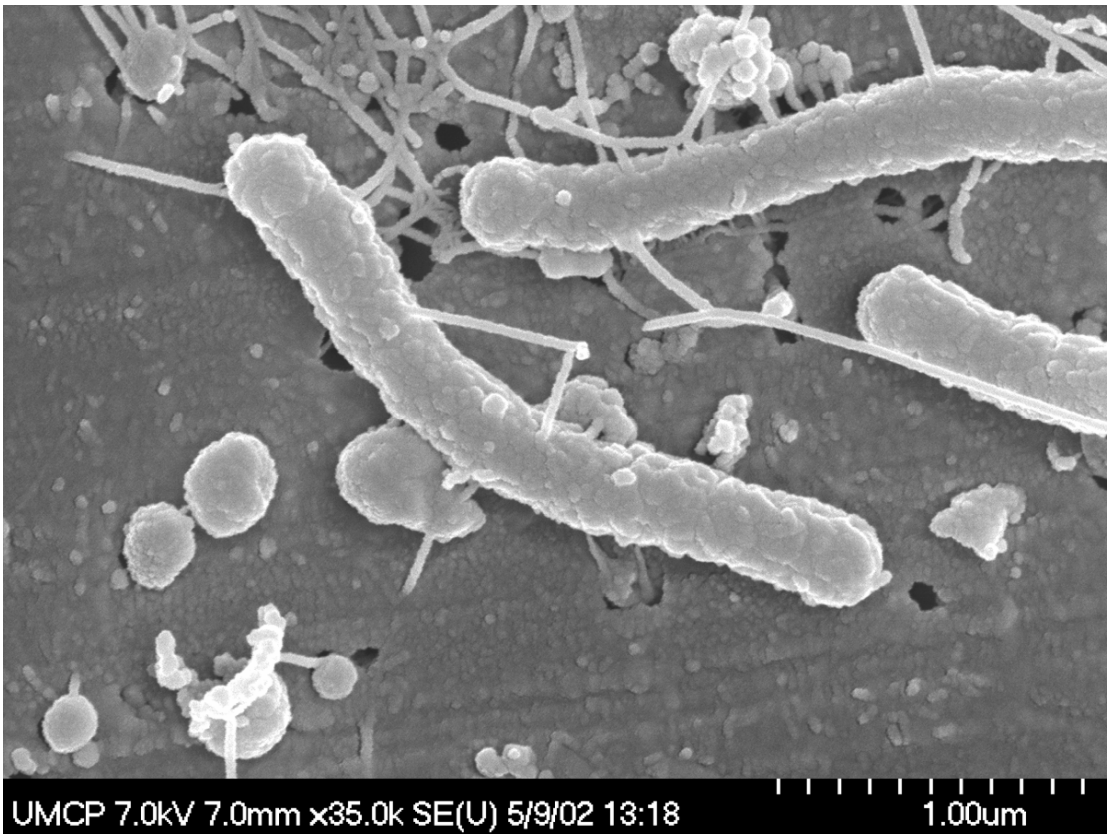


Figure A2-9. SEM of log phase agarose grown cells. Abundance of small protuberances and some larger Bleb-like ES

Large extracellular structures (ES)

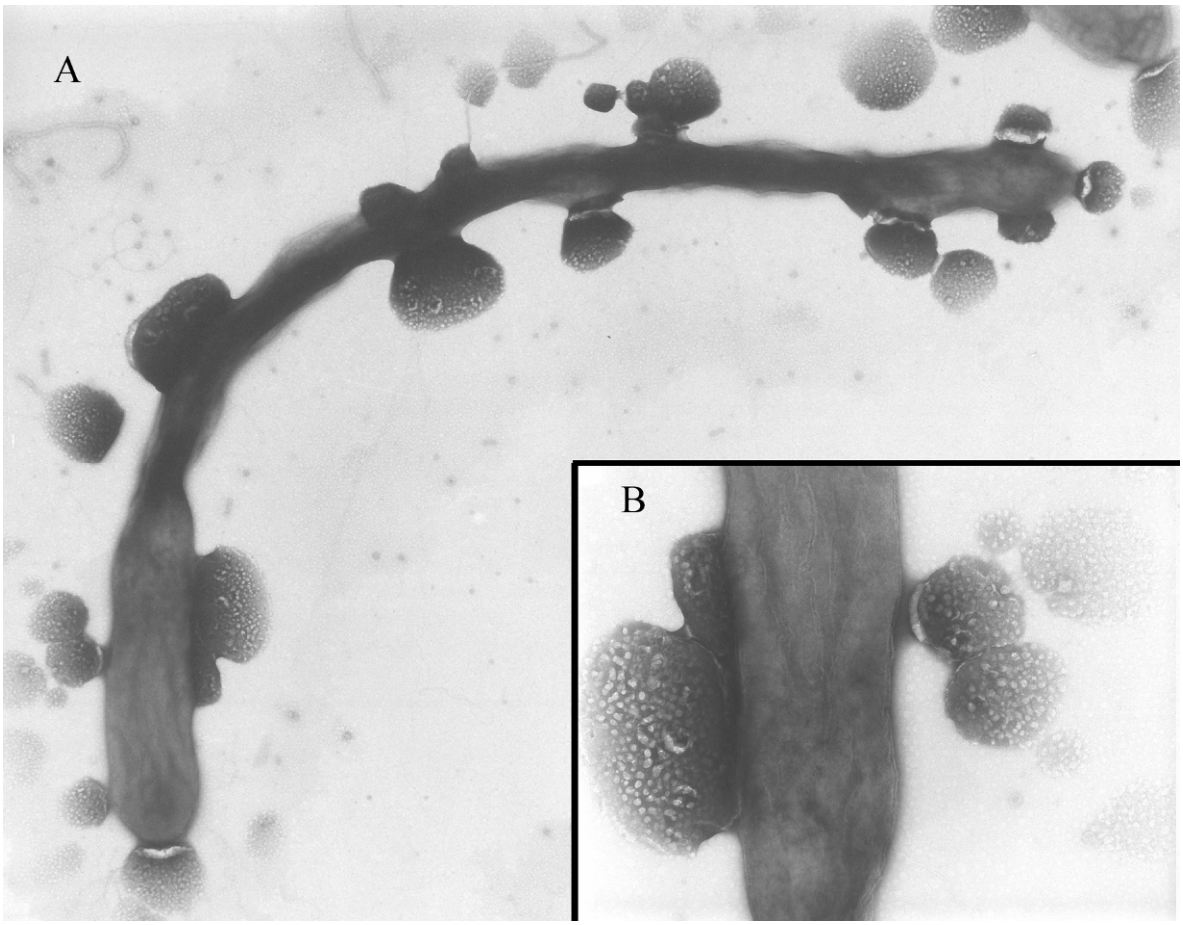


Figure A2-10. TEM of late log phase agarose grown cell. A, cell with elongated morphology and abundance of large ES which stain with a characteristic, granular appearance. B, close-up of large ES which appear to be separating from the cell surface. Note similarity to ES shown in Figure A2-3.

Tubule/Fibril-like cellular appendages

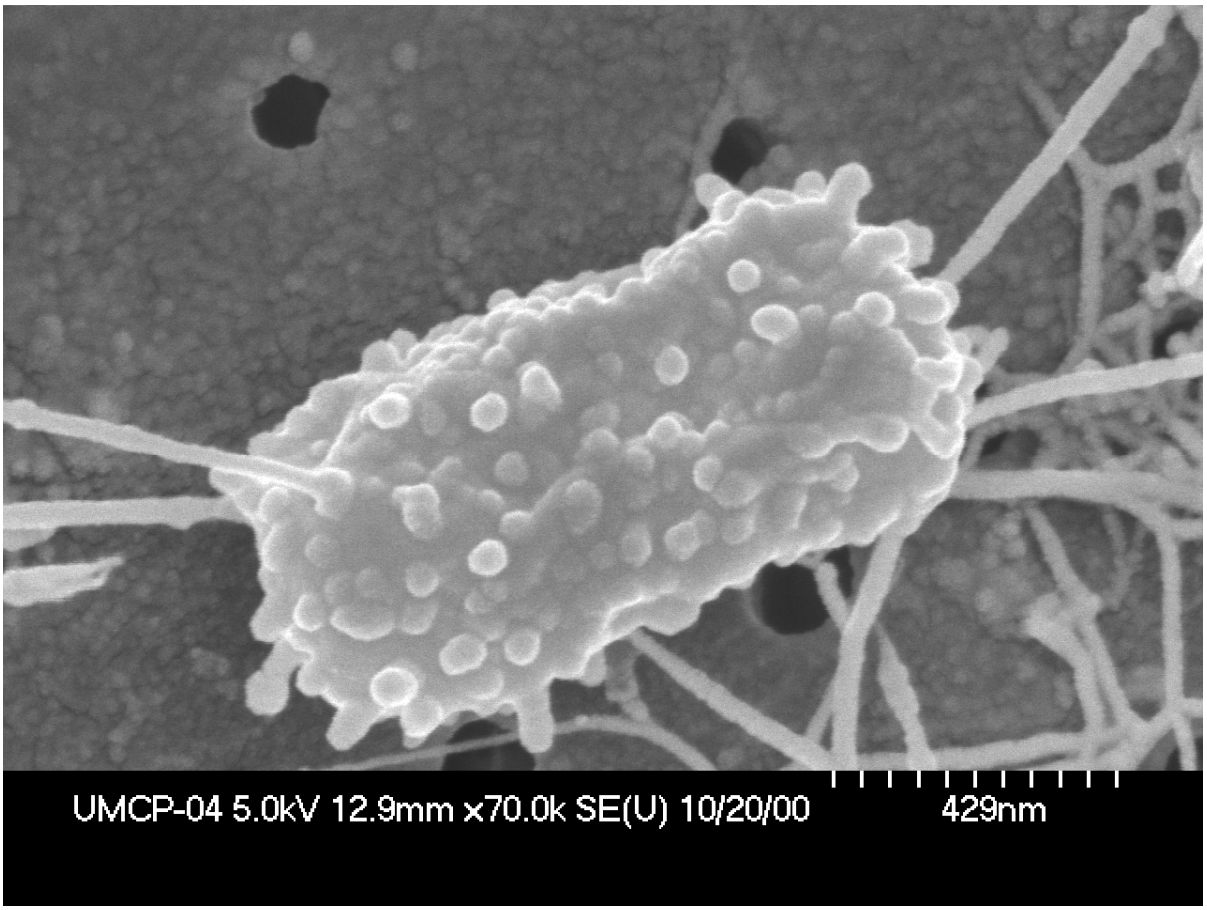


Figure A2-11. SEM of stationary-phase agarose grown cell. Cell exhibiting abundance of appendages similar to those in Figures A2-8 and A2-9. Some of the small protuberances appear to be elongated, suggesting that they may be appendages in the process of formation.

Tubule/Fibril-like cellular appendages

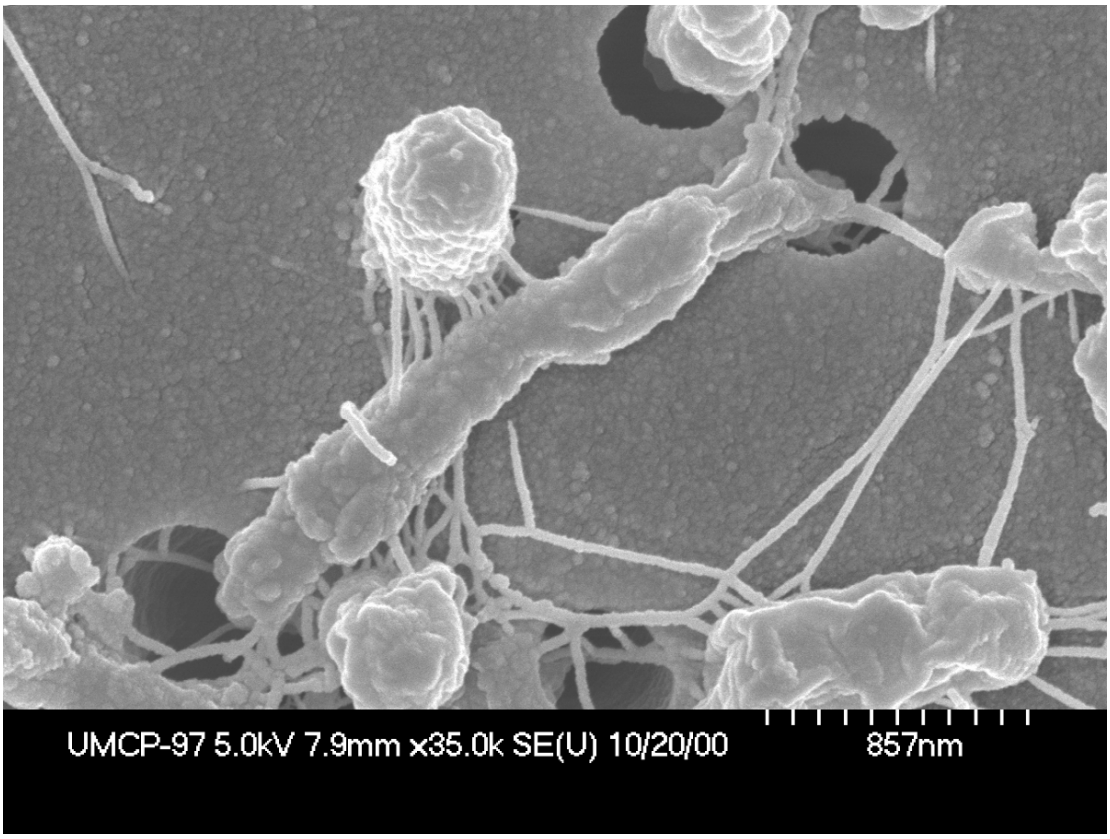


Figure A2-12. SEM of stationary phase agarose-grown cells. Note appendages connecting cell and large extracellular structure.

Tubule/Fibril-like cellular appendages

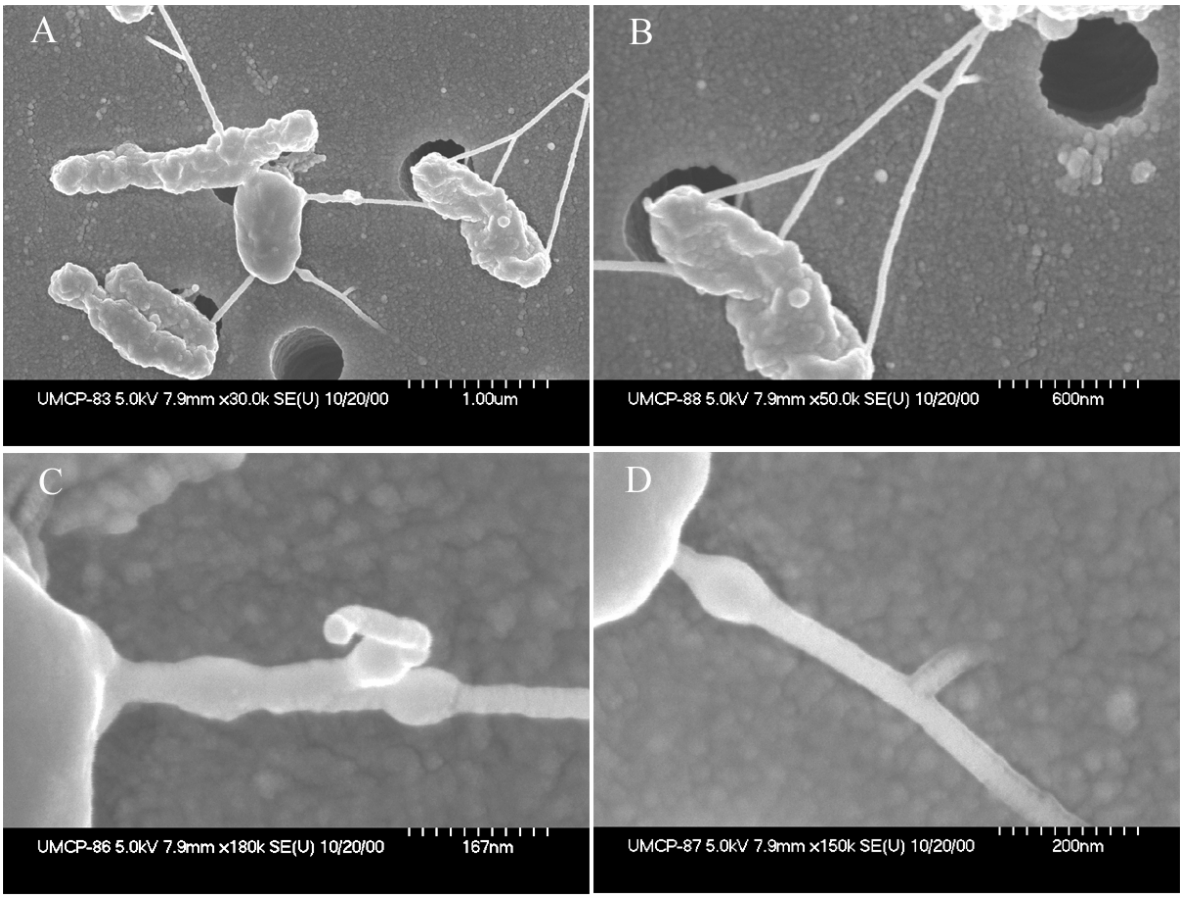


Figure A2-13. Nodules and branching of tubule/fibril-like appendages. A, group of stationary phase agarose grown cells with appendages. B, apparent convergence and cross-linking between appendages. C, nodules and apparent branching of tubule from center cell in panel A. D, close up of other branched appendage showing nodule-like swelling.

References

1. Alldredge, A.L. and M.W. Silver, Characteristics, dynamics and significance of marine snow. *Progress In Oceanography*, 1988. **20**(1): 41-82.
2. Andrews, K.J. and E.C. Lin, Selective advantages of various bacterial carbohydrate transport mechanisms. *Fed Proc*, 1976. **35**(10): 2185-9.
3. Andrykovitch, G. and I. Marx, Isolation of a new polysaccharide-digesting bacterium from a salt marsh. *Appl Environ Microbiol*, 1988. **54**: 3-4.
4. Andrykovitch, G., et al., Chitin induced expression of cellulosome-like structures in a marine bacterium. *Abstracts of the American Society for Microbiology*, 1994.
5. Azam, F., OCEANOGRAPHY: Microbial Control of Oceanic Carbon Flux: The Plot Thickens. *Science*, 1998. **280**(5364): 694-696.
6. Barr, B.K., et al., Identification of two functionally different classes of exocellulases. *Biochemistry*, 1996. **35**(2): 586-92.
7. Bassler, B.L., et al., Chitin utilization by marine bacteria. Degradation and catabolism of chitin oligosaccharides by *Vibrio furnissii*. *J Biol Chem*, 1991. **266**(36): 24276-86.
8. Bayer, E.A., et al., Cellulose, cellulases and cellulosomes. *Curr Opin Struct Biol*, 1998. **8**(5): 548-57.
9. Bayer, E.A. and R. Lamed, Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J Bacteriol*, 1986. **167**(3): 828-36.
10. Bayer, E.A. and R. Lamed, The cellulose paradox: pollutant par excellence and/or a reclaimable natural resource? *Biodegradation*, 1992. **3**(2-3): 171-88.
11. Bayer, E.A., E. Setter, and R. Lamed, Organization and distribution of the cellulosome in *Clostridium thermocellum*. *J Bacteriol*, 1985. **163**(2): 552-9.
12. Beguin, P. and J.P. Aubert, The biological degradation of cellulose. *FEMS Microbiol Rev*, 1994. **13**(1): 25-58.
13. Beguin, P. and M. Lemaire, The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Crit Rev Biochem Mol Biol*, 1996. **31**(3): 201-36.
14. Beguin, P., J. Millet, and J.P. Aubert, Cellulose degradation by *Clostridium thermocellum*: from manure to molecular biology. *FEMS Microbiol Lett*, 1992. **79**(1-3): 523-8.
15. Belaich, J.P., et al., The cellulolytic system of *Clostridium cellulolyticum*. *J Biotechnol*, 1997. **57**(1-3): 3-14.

16. Bell, K.S., et al., Genome sequence of the enterobacterial phytopathogen Erwinia carotovora subsp. atroseptica and characterization of virulence factors. *PNAS*, 2004. **101**(30): 11105-11110.

17. Berkenheger, I. and U. Fischer, Competition for polymers among heterotrophic bacteria, isolated from particles of the Equatorial Atlantic. *Int Microbiol*, 2004. **7**(1): 13-8.

18. Boraston, A.B., et al., Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J*, 2004. **382**(Pt 3): 769-81.

19. Boraston, A.B., et al., Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and "galactose-binding" domains. *J Mol Biol*, 2003. **327**(3): 659-69.

20. Carpita, N.C., Structure and Biogenesis of the Cell Walls of Grasses. *Annu Rev Plant Physiol Plant Mol Biol*, 1996. **47**: 445-476.

21. Chakravorty, D., *Cell Biology of Alginic Acid degradation by Marine Bacterium 2-40*. 1998, University of Maryland: College Park.

22. Clarke, S.R., et al., Analysis of Ebh, a 1.1-Megadalton Cell Wall-Associated Fibronectin-Binding Protein of Staphylococcus aureus. *Infect. Immun.*, 2002. **70**(12): 6680-6687.

23. Collesi, C., et al., A splicing variant of the RON transcript induces constitutive tyrosine kinase activity and an invasive phenotype. *Mol Cell Biol*, 1996. **16**(10): 5518-26.

24. Cosgrove, D.J., et al., The Growing World of Expansins. *Plant Cell Physiol.*, 2002. **43**(12): 1436-1444.

25. Cote, F., S. el Ouakfaoui, and A. Asselin, Detection of beta-glucanase activity on various beta-1,3 and beta-1,4-glucans after native and denaturing polyacrylamide gel electrophoresis. *Electrophoresis*, 1991. **12**(1): 69-74.

26. Coutinho, P.M. and B. Henrissat, *Carbohydrate-active enzyme server*. 1999.

27. Coutinho, P.M. and B. Henrissat, *The modular structure of cellulases and other carbohydrate-active enzymes: an integrated database approach*, in *Genetics, biochemistry and ecology of cellulose degradation*, K. Ohmiya, et al., Editors. 1999, Uni Publishers Co: Tokyo. p. 15-23.

28. Coutinho, P.M., et al., Why are there so many carbohydrate-active enzyme-related genes in plants? *Trends Plant Sci*, 2003. **8**(12): 563-5.

29. Czjzek, M., et al., The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. *J Biol Chem*, 2001. **276**(51): 48580-7.

30. Darvill, A.G., et al., Structure and function of plant cell wall polysaccharides. *J Cell Sci Suppl*, 1985. **2**: 203-17.

31. d'Enfert, C., A. Ryter, and A.P. Pugsley, Cloning and expression in Escherichia coli of the Klebsiella pneumoniae genes for production, surface

localization and secretion of the lipoprotein pullulanase. *Embo J*, 1987. **6**(11): 3531-8.

32. Dias, F.M., et al., Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 revealed by the crystal structure and kinetics of Cellvibrio mixtus mannosidase 5A. *J Biol Chem*, 2004. **279**(24): 25517-26.

33. Dickens, N.J., S. Beatson, and C.P. Ponting, Cadherin-like domains in alpha-dystroglycan, alpha/epsilon-sarcoglycan and yeast and bacterial proteins. *Curr Biol*, 2002. **12**(6): R197-9.

34. Distel, D.L., et al., Teredinibacter turnerae gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae). *Int J Syst Evol Microbiol*, 2002. **52**(6): 2261-2269.

35. Doi, R.H., et al., Cellulosome and noncellulosomal cellulases of Clostridium cellulovorans. *Extremophiles*, 1998. **2**(2): 53-60.

36. Dror, T.W., et al., Regulation of Major Cellulosomal Endoglucanases of Clostridium thermocellum Differs from That of a Prominent Cellulosomal Xylanase. *J. Bacteriol.*, 2005. **187**(7): 2261-2266.

37. Ducret, A., et al., High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci*, 1998. **7**(3): 706-719.

38. Ekborg, N.A., *The Agarolytic System of Saccharophagus degradans strain 2-40: Enzyme Activity and Protein Localization*, in *Cell Biology and Molecular Genetics*. 2005, University of Maryland: College Park.

39. Ekborg, N.A., et al., Saccharophagus degradans gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides. *Int J Syst Evol Microbiol*, 2005. **55**(4): 1545-1549.

40. Emami, K., et al., Evidence for temporal regulation of the two Pseudomonas cellulosa xylanases belonging to glycoside hydrolase family 11. *J Bacteriol*, 2002. **184**(15): 4124-33.

41. Ensor, L., S.K. Stotz, and R.M. Weiner, Expression of multiple insoluble complex polysaccharide degrading enzyme systems by a marine bacterium. *J Ind Microbiol Biotechnol*, 1999. **23**: 123-126.

42. Foreman, P.K., et al., Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus trichoderma reesei. *J. Biol. Chem.*, 2003. **278**(34): 31988-31997.

43. Gonzalez, J. and R.M. Weiner, Phylogenetic characterization of marine bacterium strain 2-40, a degrader of complex polysaccharides. *Int J Syst Evol Microbiol*, 2000. **50**: 831-834.

44. Green, F., 3rd, C.A. Clausen, and T.L. Highley, Adaptation of the Nelson-Somogyi reducing-sugar assay to a microassay using microtiter plates. *Anal Biochem*, 1989. **182**(2): 197-9.

45. Han, S.O., et al., Regulation of Expression of Cellulosomes and Noncellulosomal (Hemi)Cellulolytic Enzymes in Clostridium cellulovorans during Growth on Different Carbon Sources. *J. Bacteriol.*, 2004. **186**(13): 4218-4227.

46. Han, S.O., et al., Regulation of Expression of Cellulosomal Cellulase and Hemicellulase Genes in Clostridium cellulovorans. *J. Bacteriol.*, 2003. **185**(20): 6067-6075.

47. Han, S.O., et al., Transcription of Clostridium cellulovorans Cellulosomal Cellulase and Hemicellulase Genes. *J. Bacteriol.*, 2003. **185**(8): 2520-2527.

48. Hansen, C.K., Fibronectin type III-like sequences and a new domain type in prokaryotic depolymerases with insoluble substrates. *FEBS Lett*, 1992. **305**(2): 91-6.

49. Hazlewood, G.P. and H.J. Gilbert, Structure and function analysis of Pseudomonas plant cell wall hydrolases. *Prog Nucleic Acid Res Mol Biol*, 1998. **61**: 211-41.

50. Heidelberg, J.F., et al., Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis. *Nat Biotechnol*, 2002. **20**(11): 1118-23.

51. Henrissat, B., A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, 1991. **280 (Pt 2)**: 309-16.

52. Henrissat, B. and A. Bairoch, New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, 1993. **293 (Pt 3)**: 781-8.

53. Henrissat, B., et al., in *Enzyme functionality: Design, Engineering and Screening*, A. Svendsen, Editor. 2003, Marcel Dekker: New York. p. 15-34.

54. Henrissat, B., T.T. Teeri, and R.A. Warren, A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Lett*, 1998. **425**(2): 352-4.

55. Hogg, D., et al., The modular architecture of Cellvibrio japonicus mannanases in glycoside hydrolase families 5 and 26 points to differences in their role in mannan degradation. *Biochem J*, 2003. **371**(Pt 3): 1027-43.

56. Howard, M.B., et al., Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium Microbulbifer degradans. *Protein Sci*, 2004. **13**(5): 1422-1425.

57. Howard, M.B., et al., Genomic analysis and initial characterization of the chitinolytic system of Microbulbifer degradans strain 2-40. *J Bacteriol*, 2003. **185**(11): 3352-60.

58. Hynes, R.O., Integrins: bidirectional, allosteric signaling machines. *Cell*, 2002. **110**(6): 673-87.

59. Irwin, D.C., S. Zhang, and D.B. Wilson, Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from Thermobifida fusca. *Eur J Biochem*, 2000. **267**(16): 4988-97.

60. Jindou, S., et al., alpha-Galactosidase Aga27A, an enzymatic component of the Clostridium josui cellulosome. *J Bacteriol*, 2002. **184**(2): 600-4.

61. Juncker, A.S., et al., Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci*, 2003. **12**(8): 1652-62.

62. Jung, H.C., J.M. Lebeault, and J.G. Pan, Surface display of Zymomonas mobilis levansucrase by using the ice-nucleation protein of Pseudomonas syringae. *Nat Biotechnol*, 1998. **16**(6): 576-80.

63. Kawasaki, H. and R.H. Kretsinger, Calcium-binding proteins 1: EF-hands. *Protein Profile*, 1995. **2**(4): 297-490.

64. Kosugi, A., K. Murashima, and R.H. Doi, Characterization of two noncellulosomal subunits, ArfA and BgaA, from Clostridium cellulovorans that cooperate with the cellulosome in plant cell wall degradation. *J Bacteriol*, 2002. **184**(24): 6859-65.

65. Krah, M., et al., The laminarinase from thermophilic eubacterium Rhodothermus marinus . Conformation, stability, and identification of active site carboxylic residues by site-directed mutagenesis. *European Journal of Biochemistry*, 1998. **257**(1): 101-111.

66. Kurien, B.T. and R.H. Scofield, Heat mediated quick Coomassie blue protein staining and destaining of SDS-PAGE gels. *Indian J Biochem Biophys*, 1998. **35**(6): 385-9.

67. Laemmli, U.K., Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature*, 1970. **277**: 680-685.

68. Lawler, J. and R. Hynes, The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J. Cell Biol.*, 1986. **103**(5): 1635-1648.

69. LeNours, J., et al., The Structure and Characterization of a Modular Endo-β-1,4-mannanase from *Cellulomonas fimi*. *Biochemistry*, 2005. **44**(38): 12700-12708.

70. Lin, E. and D.B. Wilson, Regulation of β-1,4-Endoglucanase Synthesis in *Thermomonospora fusca*. *Appl Environ Microbiol*, 1987. **53**(6): 1352-1357.

71. Ljungdahl, L.G. and K.E. Eriksson, *Ecology of Microbial Cellulose Degradation*, in *Adv Microb Ecol*. 1985, Plenum Press: New York. p. 237-299.

72. Lou, J., K. Dawson, and H. Strobel, Role of phosphorolytic cleavage in cellobiose and cellodextrin metabolism by the ruminal bacterium Prevotella ruminicola. *Appl. Environ. Microbiol.*, 1996. **62**(5): 1770-1773.

73. Lynd, L.R., et al., Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev*, 2002. **66**(3): 506-77, table of contents.

74. Mazel, D., Integrons and the Origin of Antibiotic Resistance Gene Cassettes. *ASM News*, 2004. **70**(11): 520-525.

75. McBride, M.J. and S.A. Baker, Development of techniques to genetically manipulate members of the genera Cytophaga, Flavobacterium, Flexibacter, and Sporocytophaga. *Appl Environ Microbiol*, 1996. **62**(8): 3017-22.

76. McKie, V.A., et al., A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. *Biochem J*, 2001. **355**(Pt 1): 167-77.

77. McNeil, M., et al., Structure and function of the primary cell walls of plants. *Annu Rev Biochem*, 1984. **53**: 625-63.

78. Minet, A.D., et al., Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J Cell Sci*, 1999. **112 (Pt 12)**: 2019-32.

79. Moran, M.A. and R.E. Hodson, Dissolved Humic Substances of Vascular Plant Origin in a Coastal Marine Environment. *Limnol Oceanogr*, 1994. **39**(4): 762-771.

80. Narita, S., S. Matsuyama, and H. Tokuda, Lipoprotein trafficking in Escherichia coli. *Arch Microbiol*, 2004. **182**(1): 1-6.

81. Nobe, R., et al., Purification and characterization of laminaran hydrolases from Trichoderma viride. *Biosci Biotechnol Biochem*, 2003. **67**(6): 1349-57.

82. Notenboom, V., et al., High-resolution crystal structures of the lectin-like xylan binding domain from Streptomyces lividans xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry*, 2002. **41**(13): 4246-54.

83. Nurizzo, D., et al., Cellvibrio japonicus alpha-L-arabinanase 43A has a novel five-blade beta-propeller fold. *Nat Struct Biol*, 2002. **9**(9): 665-8.

84. Palonen, H., M. Tenkanen, and M. Linder, Dynamic Interaction of Trichoderma reesei Cellobiohydrolases Cel6A and Cel7A and Cellulose at Equilibrium and during Hydrolysis. *Appl. Environ. Microbiol.*, 1999. **65**(12): 5229-5233.

85. Papi, R. and D. Kyriakidis, Overexpression of the pectin lyase gene of Pseudomonas marginalis in Escherichia coli and purification of the active enzyme. *Biotechnol Appl Biochem*, 2003. **37**(Pt 2): 187-94.

86. Perkins, D.N., et al., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 1999. **20**(18): 3551-67.

87. Poretz, R.D. and I.J. Goldstein, An examination of the topography of the saccharide binding sites of concanavalin A and of the forces involved in complexation. *Biochemistry*, 1970. **9**(14): 2890-6.

88. Pugsley, A.P., C. Chapon, and M. Schwartz, Extracellular pullulanase of Klebsiella pneumoniae is a lipoprotein. *J Bacteriol*, 1986. **166**(3): 1083-8.

89. Puls, J., et al., *Xylobiose and Xylooligomers*, in *Biomass Part A Cellulose and Hemicellulose*, W. WA and K. ST, Editors. 1988, Academic Press: San Diego, CA. p. 528-536.

90. Qiu, X., et al., Isolation and analysis of two cellulase cDNAs from Orpinomyces joyonii. *Gene*, 2000. **245**(1): 119-26.

91. Read, S.M., G. Currie, and A. Bacic, Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydr Res*, 1996. **281**(2): 187-201.

92. Ricklefs, R.E., *Ecology*. third ed. 1990, New York: W. H. Freeman and Co. 896.

93. Rigden, D.J. and M.Y. Galperin, The DxDxDG motif for calcium binding: multiple structural contexts and implications for evolution. *J Mol Biol*, 2004. **343**(4): 971-84.

94. Saloheimo, M., et al., Swollenin, a Trichoderma reesei protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *Eur J Biochem*, 2002. **269**(17): 4202-4211.

95. Salyers, A.A., A. Reeves, and J. D'Ella, Solving the problem of how to eat something as big as yourself: diverse bacterial strategies for degrading polysaccharides. *J Ind Microbiol*, 1996. **17**: 470-476.

96. Schwarz, E. and S. Benzer, The recently reported NIbeta domain is already known as the Calx-beta motif. *Trends Biochem Sci*, 1999. **24**(7): 260.

97. Schwarz, E.M. and S. Benzer, Calx, a Na-Ca exchanger gene of Drosophila melanogaster. *Proc Natl Acad Sci U S A*, 1997. **94**(19): 10249-54.

98. Schwarz, W.H., The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol*, 2001. **56**(5-6): 634-49.

99. Seydel, A., P. Gounon, and A.P. Pugsley, Testing the '+2 rule' for lipoprotein sorting in the Escherichia coli cell envelope with a new genetic selection. *Mol Microbiol*, 1999. **34**(4): 810-821.

100. Simpson, P.J., et al., The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J Biol Chem*, 2000. **275**(52): 41137-42.

101. Stotz, S.K., *An agarase system from a periphytic prokaryote*. 1994, University of Maryland: College Park.

102. Takeichi, M., Cadherins: a molecular family important in selective cell-cell adhesion. *Annu Rev Biochem*, 1990. **59**: 237-52.

103. Tamaru, Y., et al., Formation of protoplasts from cultured tobacco cells and Arabidopsis thaliana by the action of cellulosomes and pectate lyase from Clostridium cellulovorans. *Appl Environ Microbiol*, 2002. **68**(5): 2614-8.

104. Tamm, L.K., A. Arora, and J.H. Kleinschmidt, Structure and assembly of beta-barrel membrane proteins. *J Biol Chem*, 2001. **276**(35): 32399-402.

105. Tans-Kersten, J., Y. Guan, and C. Allen, Ralstonia solanacearum Pectin Methylesterase Is Required for Growth on Methylated Pectin but Not for Bacterial Wilt Virulence. *Appl. Environ. Microbiol.*, 1998. **64**(12): 4918-4923.

106. Tomme, P., R.A. Warren, and N.R. Gilkes, Cellulose hydrolysis by bacteria and fungi. *Adv Microb Physiol*, 1995. **37**: 1-81.

107. Tskhovrebova, L. and J. Trinick, Properties of Titin Immunoglobulin and Fibronectin-3 Domains. *J. Biol. Chem.*, 2004. **279**(45): 46351-46354.

108. Vaaje-Kolstad, G., et al., Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21. *J Biol Chem*, 2005. **280**(12): 11313-9.

109. Vaisvila, R., et al., Discovery and distribution of super-integrons among Pseudomonads. *Molecular Microbiology*, 2001. **42**(3): 587-601.

110. Wang, Y.D., S. Zhao, and C.W. Hill, Rhs elements comprise three subfamilies which diverged prior to acquisition by Escherichia coli. *J Bacteriol*, 1998. **180**(16): 4102-10.

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

112. Whitehead, L., *Complex Polysaccharide Degrading Enzyme Arrays Synthesized By a Marine Bacterium*. 1997, University of Maryland: College Park.

113. Whitehead, L., et al., *Membrane structures in a marine bacterium are synthesized coincidentally with insoluble biopolymer degradation*, in *Abstracts of the American Society for Microbiology 95th General Meeting*. 1995. p. 572.

114. Whitehead, L.A., S.K. Stosz, and R.M. Weiner, Characterization of the agarase system of a multiple carbohydrate degrading marine bacterium. *Cytobios*, 2001. **106 Suppl 1**: 99-117.

115. Zablackis, E., et al., Characterization of the Cell-Wall Polysaccharides of Arabidopsis thaliana Leaves. *Plant Physiol.*, 1995. **107**(4): 1129-1138.

116. Zhang, S., D.C. Irwin, and D.B. Wilson, Site-directed mutation of noncatalytic residues of Thermobifida fusca exocellulase Cel6B. *Eur J Biochem*, 2000. **267**(11): 3101-3115.

117. Zhang, Y.-H.P. and L.R. Lynd, Cellulose utilization by Clostridium thermocellum: Bioenergetics and hydrolysis product assimilation. *PNAS*, 2005. **102**(20): 7321-7325.

118. Zobell, C.E. and S.C. Rittenburg, The occurrence and characteristics of chitinoclastic bacteria in the sea. *Journal of Bacteriology*, 1937. **35**: 275-287.