ENDOXYLANASES FROM GLYCOSYL HYDROLASE FAMILIES 5 AND 10: BACTERIAL ENZYMES FOR DEVELOPMENT OF GRAM-POSITIVE BIOCATALYSTS FOR PRODUCTION OF BIO-BASED PRODUCTS

By

FRANZ JOSEF ST. JOHN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006

I dedicate this work to the memory of Josefa Florentine St. John

1940-1990

Wife to one, mother of four, five who miss, evermore

ACKNOWLEDGMENTS

I acknowledge the steadfast support of my family and wife who have vicariously enjoyed graduate school through my continued praise. Without their support this may have been impossible. Further, I thank my advisor, James. F. Preston III, for taking the time to help me grow as a scientist and teaching me that my imagination may be the best scientific tool of all.

I would also like to thank my Ph. D. research committee; Doctors A. Edison, L. O. Ingram, J. Maupin-Furlow and K. T. Shanmugam, for their support and guidance. Their comments and suggestions have helped guide this work to a fruitful conclusion. I thank them.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ENDOXYLANASES FROM GLYCOSYL HYDROLASE FAMILIES 5 AND 10: BACTERIAL ENZYMES FOR DEVELOPMENT OF GRAM-POSITIVE BIOCATALYSTS FOR PRODUCTION OF BIO-BASED PRODUCTS

By

Franz Josef St. John

December 2006

Chair: James F. Preston III
Major Department: Microbiology and Cell Science

Fossil fuels are a nonrenewable resource. Their wide geological distribution and relatively simple acquisition have allowed massive increases in human population and associated energy expenditure over the last century. The current rate of consumption and the expectation of reduced fuel supplies predicate the need to develop new energy sources that may be merged with the current energy infrastructure. As an underutilized renewable resource, lignocellulosic biomass may, through microbial bioconversion, contribute to the environmentally benign production of alternative fuels and chemical feedstocks. A major target for this conversion is methylglucuronoxylan, the predominant structural polysaccharide in the hemicellulose fraction of hardwood and crop residues. The research described herein has focused on xylanolytic bacteria and their secreted endoxylanases that are involved in the depolymerization of the methylglucuronoxylan. In this work, endoxylanases of glycosyl hydrolase families 5 and 10 (GH 5 and GH 10 xylanases) have been characterized with emphasis on the native bacterial host and utilization of the hydrolysis limit products of methylglucuronoxylan. Recombinant constructs of the genes encoding these xylanases have made them available for definitive characterization and for expression in transgenic organisms. $XynA_1$, a multimodular GH 10

xylanase anchored to the cell surface of *Paenibacillus* sp. strain JDR-2, generates aldouronates that are efficiently assimilated and metabolized by this organism. The proximity between $XynA_1$ and the native *Paenibacillus* host, and the proficient utilization of the resulting hydrolysis products, identify a process of vectoral assimilation of methylglucuronoxylan-derived products. A GH 5 xylanase encoded by the *ynfF* gene in *Bacillus subtilis* 168 is directed for catalysis by methylglucuronosyl substitutions on the xylan chain, supporting its application in an accessory role in the overall depolymerization process. Secretion of this GH 5 and a GH 11 endoxylanase by the genetically malleable *B. subtilis* 168, for which the entire genome has been sequenced, recommends it as a target for introduction of genes encoding the GH 10 endoxylanase, $XynA_1$ and aldouronate-utilization enzymes for efficient depolymerization and metabolism of methylglucuronoxylan. These discoveries provide insight needed for the development of second-generation bacterial biocatalysts for the direct conversion of lignocellulosic biomass to alternative fuels and bio-based products.

CHAPTER 1
INTRODUCTION

The political and social stability of the world is dependent on a plentiful supply of energy. For the last century, this supply has been in the form of fossil fuels mined or pumped from the earth. At the current rate of use, it is predicted that approximately half-way through this century, the supply of easily obtainable fossil fuels will be significantly limited. Three recent studies of proven world oil supplies show an average estimate of currently obtainable crude oil reserves of 1188 billion barrels (Energy Information Administration, 2006[a]). In 2003 world oil consumption was estimated at about 29 billion barrels per year (Energy Information Administration, 2006[b]). These values lead to an estimated time of 41 years before the current proven reserves are exhausted. Although there are newly realized additions to the world crude oil reserve identified frequently, consumption is increasing steadily and outpacing discovery (Erickson, 2003). Regardless of the number of years until a severe shortage in crude oil reserves, the inevitable consequence of dependence on this finite resource is troubling. Increasing population and industrialization around the world exacerbate this situation. Furthermore, as crude oil supplies decrease, that which remains becomes more difficult to obtain. These circumstances result in high demand for a diminishing commodity which can stifle worldwide growth and challenge international relations. Fossil fuels are also a primary contributor to "greenhouse" gases (GHG) which could, over the next few decades, have a major effect on global climate. There is no need to wait until this inevitable energy crisis suffocates the world economy. It is not a requirement that we utilize every obtainable drop of crude oil. In fact, it would be extremely irresponsible to do so. Research efforts must focus on increasing energy yield of current alternative energy sources, on developing novel methods for harvesting energy, and decreasing per capita energy consumption. Together, these goals may add up to a sustainable energy future for the world.

Further, investment in this new energy infrastructure may spur equivalent growth as that witnessed during the 20th century in the United States. After all, survival is a strong motive for reform.

Over the last several decades, methods have been developed by which the constituent sugars of lignocellulosics can be converted to clean burning ethanol using microorganisms as biocatalysts (Dien et al., 2003; Ingram et al., 1999; Jeffries, 2005). Large scale application of this technology to renewable resources, such as woody biomass and crop residues, may help to balance the demand for fossil fuels by supplementing the supply with ethanol. In the long term, this could become a major contributor to liquid fuel supplies for transportation needs. Use of ethanol produced from biomass is carbon neutral, meaning that it does not increase the net atmospheric CO_2 concentration. Therefore, it does not contribute to GHG emissions. The process by which this conversion takes place can be broken down into two key steps. The first is a preprocessing step, which is typically required to prepare the complex, recalcitrant lignocellulosic biomass for conversion. The second utilizes engineered microbial biocatalysts to convert the simple sugars released by the pretreatment to ethanol. In some systems, conversion of the sugars in lignocellulose to ethanol has been accomplished to near theoretical yields (Ingram et al., 1998).

Two food crops which are used for large scale ethanol production are corn grain and sugarcane. In the US, corn grain is the primary source for ethanol production representing approximately 2% of liquid fuel use whereas in Brazil, sugarcane has been used to produce enough ethanol to significantly displace gasoline. Another biofuel, biodiesel, used as a diesel fuel replacement, can be prepared from triglyceride rich crops such as soybean, rapeseed and palm. A primary concern frequently raised when considering ethanol or biodiesel production

from food crops is the net energy balance. Recent studies have shown that production of ethanol from corn grain and biodiesel from soybean are energy positive (Dewulf et al., 2005; Farrell et al., 2006). When considering these food crops as substrate biomass, the balance is obtained by considering all inputs, including all costs of farming and biomass preprocessing, and the output costs including the net yield of fuel and all other valuable components. In a recent study, bioconversion of corn grain to ethanol yielded only a 25% energy increase over the energy consumed in the process and biodiesel from soybean yielded a 93% energy increase (Hill et al., 2006). The lower energy yield for corn conversion is primarily attributed to the greater growth requirements of corn and the use of gas-fired boilers for ethanol purification. Although soybean may be a good candidate crop for biodiesel production, analysis of total crop land requirements for use in growing any of theses crops as major sources for fuel is thought to be very ambitious, and would still contribute only a small amount of total liquid fuel requirements (Hill et al., 2006). Large scale food crop cultivation for bioenergy also concerns scientists as it links energy production directly to food production and also increases demand for valuable agricultural land. It is generally acknowledged that the greatest benefit for society will come from production of cellulosic ethanol that results from conversion of lignocellulosics and crop residues (Dewulf et al., 2005; Farrell et al., 2006; Hammerschlag, 2006; Hill et al., 2006; Perlack et al., 2005). Idealized substrates for ethanol production will be lignocellulosics such as crop residues like corn stover, forestry products including pulp and paper mill waste, and renewable energy crops such as switchgrass and poplar. As considered above, bioenergy crops are expected to compete for generally useful agricultural lands. However, they will not directly compete with agricultural crop products. Use of these underutilized renewable biomass resources as substrates for biofuel production greatly reduces the input requirements therefore increasing energy yield. The United

States Department of Agriculture recently published a report entitled "Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply" (Perlack et al., 2005). The analysis presented in this report proposes that it is possible to replace 30% of the US petroleum consumption with biofuels by 2030. Development of bioenergy crops like poplar (hardwoods) and switchgrass (graminaceous plants) along with food crop residues, such as corn stover and sugar cane bagasse, will contribute greatly to ethanol production and accomplishment of the "Billion-Ton Annual Supply" and energy independence.

CHAPTER 2
BIOMASS AND BIOCONVERSION

The Complex Composition of Bioenergy Crops and Agricultural Crop Residues

Energy harvested from the sun through the process of photosynthesis is stored by plants as fixed carbon in the form of several different, tightly associated complex polymers. The combined characteristics of these interacting polymers impart to the plant the strength and resilience that is common in wood products. For this reason, wood has been ubiquitous in the urbanization of society, allowing for relatively affordable, easily constructed homes and buildings. When considering bioconversion processes, these characteristics are referred to as recalcitrance, and are a primary concern for development of processes which liberate the complex sugar components as simple utilizable sugars. The primary components of wood which interact to create this strength and recalcitrance are cellulose, hemicellulose and lignin. Cellulose is the most abundant carbohydrate polymer in the biosphere and is the major structural component of plant life. In bioenergy crops and crop residues, cellulose can range from 36% to 50% total mass (Table 2-1) (Lynd et al., 1999; Scurlock). The second most abundant polymer in biomass is the combined hemicellulose fraction, which in hardwoods is composed primarily of methylglucuronoxylan, and can range from 20% to 35% of total mass (Timell, 1967). This fraction in graminaceous plants such as switchgrass and the crop residue corn stover represents up to 35% of mass and is primarily composed of the polymer methylglucuronoarabinoxylan (Lynd et al., 1999). A secondary hemicellulose from hardwoods, glucomannan, is a minor sugar component representing no more that 4% of mass, and will not be directly considered in this discussion. Lignin is a complex non carbohydrate polymer which is directly associated with recalcitrance of biomass. Renewable biomass sources receiving the most intense scrutiny contain lignin from between 5.5% for early cut harvest of switchgrass growth up to 24% for

common hardwood sources such as poplar (Lynd et al., 1999; Timell, 1967). Together, these interacting polymers impart the characteristics for which wood has been cherished, but they further act to prevent enzymatic accessibility to the individual carbohydrate polymers for direct use of woody biomass as a renewable resource.

Cellulose

Cellulose is a homopolymer composed of repeating β-1,4 linked glucose molecules. When derived from woody biomass it has a degree of polymerization of approximately 10,000 glucose residues, making it the largest naturally occurring polymer (O'sullivan, 1997). Although the word cellulose refers to the β-1,4-linked polymer described above, it also describes the tightly associated crystalline fibers that result from many individual cellulose strands hydrogen bonding together to form the chemically and physically recalcitrant cellulose fiber. In general, cellulose from biomass is composed of many identical molecules which are tightly associated through hydrogen bonding interactions, intermixed with hemicellulose. This tight association between cellulose molecules and fibrils makes this specific polymer recalcitrant to chemical and enzymatic degradation.

Hemicellulose

Hardwood hemicellulose differs from graminaceous sources of hemicellulose primarily by not having arabinose substitutions. As the name suggests methylglucuronoxylan is a β-1,4 linked xylose chain randomly substituted (Jacobs et al., 2001) with α-1,2 linked 4-O-methylglucuronate moieties. Common hardwood sources have methylglucuronate substitutions on one in every ten xylose residues, but there have been reports of hardwoods with significantly higher methylglucuronate content (Hurlbert and Preston, 2001; Preston et al., 2003; Puls, 1997). Detailed ^{13}C-NMR studies of methylglucuronoxylan from the hardwood sweetgum, a candidate bioenergy hardwood, reflect degrees of substitution as high as 1 in every 6 xylose residues.

Methylglucuronoarabinoxylan from graminaceous biomass resources have fewer methylglucuronate substitutions, but can contain arabinose at a ratio of one for every six xylose residues (Puls, 1997; Wilkie, 1979). An additional substitution on both the methylglucuronoxylan and methylglucuronoarabinoxylan polymers is O-linked acetyl groups (Puls, 1997; Timell, 1967). Methylglucuronoxylan has this substitution in either the O2 or O3 position on 70% of all the xylose residues and methylglucuronoarabinoxylan contains approximately half this amount (Puls, 1997; Timell, 1967).

Lignin

Of all the polymers in lignocellulose, lignin is the most complex due to its random amorphous nature. There is little if any in the primary cell wall, but it forms a major component in the secondary wall and the middle lamella. It is composed of characteristic phenyl propane derivatives, randomly linked through carbon-carbon bonds by an enzymatic dehydrogenation process, shedding light on the reason for the complexity. The phenyl propane derivatives differ depending on the lignin source, but in general lignin from most plants is composed of guaiacyl, syringyl and *p*-hydroxyphenol moieties (Fujita and Harada, 2001).

Polymer Interactions Which Create Recalcitrant Tissues

The secondary cell wall in woody tissue is the primary structural component of biomass, which imparts rigidity and strength to the plant and is the primary source of stored carbon composed of cellulose, hemicellulose and lignin. Based on data reviewed by Mellerowicz *et al*. (Mellerowicz et al., 2001), 86% to 88% (v/v) of cells in poplar wood have secondary cell walls consisting of 80% utilizable carbohydrate. This quantifies the value of hardwood as a biomass resource. This wood type is characterized by its high content of methylglucuronoxylan and lower content of lignin when compared to softwoods. These properties make this wood type more amenable to pretreatment processes. Formation of a full secondary cell wall begins with

synthesis of a primary wall. Cellulose microfibrils, which are composed of many individual cellulose molecules, are synthesized on the cell surface by large complexes of cellulose synthase enzymes called rosettes (Ding and Himmel, 2006; Doblin et al., 2002). The layers of cellulose microfibrils that are produced in the primary wall form a net-like cellulose matrix (Fujita and Harada, 2001). Together with this matrix are the primary wall hemicellulose polymers and associated polysaccharides and proteins which lend support to and flexibility for cell expansion (Carpita et al., 2001; Somerville et al., 2004). After the primary wall is synthesized, if the cell is a xylem fiber type, it begins to synthesize a secondary cell wall. The secondary cell wall is much thicker than the primary wall and is typically divided into at least three distinct layers. As can be observed in Figure 2-1, cellulose fiber synthesis in each layer is offset from the next by some degree so that after synthesis of the complete secondary cell wall, there are cellulose fibers bracing the wall from most angles (Fujita and Harada, 2001). It is thought that hemicellulose interacts with and coats the outer layer of cellulose microfibrils to allow for movement, in effect acting as a lubricant and preventing formation of larger cellulose fibers through hydrogen bonding (Ding and Himmel, 2006; Fengel, 1971; Reis et al., 1994). Layers of the secondary cell wall are described as showing a helicoid structure. It has been postulated that it is the intimate interaction with hemicellulose polymers, which are at a greater concentration between secondary cell wall layers that control the formation of this helicoid structure (Reis et al., 1994). Although the mass of cellulose and associated hemicellulose polymers impart strength to the cell wall and the plant, the wall is further reinforced by cross linking with lignin. This randomly formed complex, non-carbohydrate polymer forms ester linkages to various moieties on the hemicellulose chain (Puls, 1997; Reis et al., 1994). The final secondary cell wall structure of mature xylem cells contains three heavy layers of cellulose (Table 2-1) intertwined with

hemicellulose, which is lignified through covalent bonds to lignin. The middle lamella is fully lignified, filling in between individual cells (Fromm et al., 2003). The complex matrix formed by these three associated materials can be compared to reinforced concrete and the outermost layer of lignin filling the middle lamella acts to glue it all together.

Softwood conifers are composed of the same three major components discussed above. However, the hemicellulose fraction is not composed of methylglucuronoxylan as in hardwood, but rather consists of two different polymers. The minor polymer is similar to methylglucuronoarabinoxylan of the graminaceous plants, but has a greater amount of methyglucuronate substitution and no acetyl groups. The predominant hemicellulose polymer in softwood is galactoglucomannan. This polymer is an O-acetylated polymer composed of terminally branched galactose, glucose and mannose in the ratio 1 : 1 : 3. Generally, softwood is more heavily lignified than hardwoods and is considered to be less efficient in bioconversion processes. Further, the pulp and paper and construction industries are major consumers of softwood supplies. It is possible that future bioconversion endeavors may combine all wood sources for bioconversion, but current objectives primarily include hardwoods and switchgrass as biofuel crops and waste agricultural residues, e.g., corn stover and sugar cane bagasse.

Pretreatment of Lignocellulose

Due to the complex nature of lignocellulosics, utilization of the embedded carbohydrates requires preprocessing, which usually includes a physical and chemical pretreatment. Following this, all pretreatment methods in development require supplementation with fungal cellulolytic hydrolase mixtures (e.g., Spezyme CP). The most established pretreatment methods have recently been reviewed. In this particularly constructive effort, several laboratories worked together to obtain equivalent, directly comparable data (Kim and Holtzapple, 2005; Kim and Lee, 2005; Liu and Wyman, 2005; Lloyd and Wyman, 2005; Mosier et al., 2005a; Teymouri et

al., 2005; Wyman et al., 2005b). The seven preprocessing methods studied used NREL standard methods for data collection and result analysis. Results were reported as total sugars released after the pretreatment step and also after the subsequent enzyme hydrolysis. Besides the review of the individual preprocessing methods and general results of each (Wyman et al., 2005a), further articles published simultaneously (same volume) combined complete data sets of all seven pretreatments for comparative analysis (Eggeman and Elander, 2005). They also provided an in-depth economic analysis of the most promising approaches (Mosier et al., 2005b) and considered the effect of preprocessing on the biomass structure (Lloyd and Wyman, 2005). Here I briefly consider these processing methods to better understand the requirements for subsequent bioconversion to ethanol.

Description of Biomass Pretreatment Methods

The pretreatment methods compared include dilute sulfuric acid, flow-through, pH controlled water, ammonia fiber explosion (AFEX), ammonia recycle percolation (ARP) and lime pretreatment. Each pretreatment method is reviewed to better understand how each affects biomass. The first applies mild acid conditions (0.5-3.0% H_2SO_4) and temperatures and pressures from 130°C to 200°C and 3atm to15atm, respectively. In this method, the acid, temperature and pressure function to liberate the hemicellulose in an almost completely utilizable form. In addition, dilute acid treatment disrupts the normally recalcitrant cellulose for effective subsequent enzymatic hydrolysis (Liu and Wyman, 2005). The flow-through pretreatment method was simultaneously compared to an improved method termed partial flow-through pretreatment (PFP), both of which are similar in principle to the dilute acid pretreatment. Both of these methods apply temperatures of 190-200°C and pressure of 20-24 atm. Partial flowthrough pretreatment was superior to water flow-through pretreatment because the final solubilized xylan was more concentrated by the reduced water elution volumes. This

significantly affects downstream costs associated with product recovery for fermentation. Furthermore, this method can remove significant lignin content prior to enzyme hydrolysis (Mosier et al., 2005a). The pH controlled water pretreatment method uses high temperature (160-190°C) and high pressure (6-14 atm) for times between 10 and 30 minutes. The application of high temperature and neutral pH water has several advantages over the acid hydrolysis methods. Hemicellulose is solubilized and primarily remains as small oligomers, reducing the conditional formation of side products such as furfural, which can inhibit fermentations. Furthermore, this method significantly reduces the lignin content to increase enzymatic hydrolysis (Teymouri et al., 2005). Ammonia fiber explosion (AFEX) is a promising pretreatment method which applies ammonia at elevated temperatures (70-90°C) and pressures (15-20 atm) for only short periods of time (<5 minute). The inherent solvent properties and volatility of ammonia are the characteristics which allow this unique approach to disrupt the biomass. The explosion occurs with the sudden release of pressure resulting in rapid expansion of the volatile heated ammonia. The method does not remove any component of the biomass, but rather disrupts it sufficiently to allow near complete enzyme hydrolysis (Kim et al., 2006). ARP uses a 10-15% (w/w) ammonia solution at 150-170°C and 10-20 atm of pressure for 10-20 minutes. This method is somewhat similar to the PFP described above, but uses an aqueous ammonia solvent rather than acidic water. As with PFP, this method has the advantage of obtaining significant lignin removal, but requires downstream separation of solubilized carbohydrates for maximum realized sugar yield (Kim and Holtzapple, 2005; Kim and Holtzapple, 2006a). Use of lime ($CaOH_2$) in pretreatment applies opposite chemistry as compared to acid based pretreatments. Combination of lime and oxygen with lignocellulose degrades hemicellulose and cellulose by the peeling reaction (endwise reducing end β-

elimination), releasing glucoisosaccharinate and xylosaccharinate. In general, this pretreatment method produces a complex mixture of degradation products, and allows removal of a large percentage of lignin (Wyman et al., 2005a). This treatment, as with all the discussed pretreatment methods, significantly improves enzymatic hydrolysis of the resulting pretreated biomass.

Analysis of the Methods to Determine Which Pretreatment Protocols are Most Effective

From the pretreatment studies of corn stover described above, comparisons were performed by analysis of total sugar yields (Eggeman and Elander, 2005). With a standard enzyme amount of 15 filter paper units Spezyme CP per gram glucan applied post pretreatment, all methods yielded no greater than 86% of the total carbohydrate, indicating that they are all candidates for further refinement and optimization. The lowest yield (86.8%) was obtained with an optimized lime treatment and three of the seven methods resulted in yields over 90%. These included dilute acid treatment yielding 92.4%, AFEX yielding 94.4% and flow-through yielding 96.6%. These three pretreatment methods differ greatly in their resulting carbohydrate product mixtures. The dilute acid treatment converted approximately 83% of the total xylose content to free xylose and slightly more than 2% to xylooligomers. This achieved near complete removal of xylan from the cellulose and lignin matrix and provides a likely explanation for the almost complete hydrolysis of cellulose (~92%) observed with the subsequent cellulase treatment. AFEX is by far the most interesting case and resulted in the second best sugar yield (94.4%). The pretreatment did not release any sugars, but resulted in a much altered biomass structure, which facilitated near complete enzymatic hydrolysis. The best results were obtained with the flow-through pretreatment method. With this approach, almost complete xylose conversion occurred, but it was primarily in the form of xylooligomers (~92%). Only a small amount of free xylose was detected (~4.5%).

Economic analysis of these preprocessing methods showed that no single method had a clear advantage (Eggeman and Elander, 2005). The dilute acid, AFEX, ARP and lime pretreatments were each estimated to require a similar fixed capital investment. As an example of this, in the dilute acid method, the primary cost was associated with equipment requirements needed to handle the corrosive conditions, and a minor cost was associated with chemical supply requirements. The opposite case was observed for AFEX pretreatment, which requires costly pure ammonia, which is less corrosive then acid. Use of pure ammonia is a substantial cost although AFEX plants are designed to recover most by condensation. Dilute acid, AFEX and lime pretreatments resulted in the lowest total fixed capital per gallon of annual ethanol production capacity making these the best current methods for large scale pretreatment. Both dilute acid and AFEX were at $3.72/gallon and lime was significantly lower at $3.35/gallon.

Except for AFEX, it is clear that all these pretreatment methods act primarily by removing xylose or lignin, and in some cases, a significant amount of both. Since the hemicellulose and lignin fractions are thoroughly embedded within the cellulose matrix, it seems likely that methods which remove either will significantly alter cellulose accessibility and render it susceptible to enzymatic hydrolysis. In the case of AFEX, no degradation of the treated biomass is apparent, but based on the subsequent ability for almost complete enzyme hydrolysis it seems likely that this process readily alters the lignin and xylan association with cellulose (Kim and Holtzapple, 2006b; Teymouri et al., 2005). Although it is thought that residual lignin in biomass inhibits enzymes added for hydrolysis, (Berlin et al., 2006) these studies show no indication of this. Dilute acid preprocessing does not remove lignin, but enzyme hydrolysis resulted in the third highest yield of carbohydrate. AFEX, as mentioned, retained the majority of its lignin content and resulted in the second highest yield of carbohydrate. This suggests that 15 FPU

Spezyme CP/g glucose may be a wasteful amount of hydrolytic activity for complete hydrolysis following some pretreatments.

Enzyme Systems for Utilization of Glucuronoxylan

Although further research into pretreatment methods is required, it seems likely that the methods reviewed above approach their optimal performance. Limitations with the current methods for ethanol production from biomass include the high cost of pretreatment and the cost of commercial enzyme preparations required to obtain maximum yield. Advancements which lower the cost of bioconversion of lignocellulosics to ethanol are likely to come from the development of less expensive, more robust enzyme systems with a greater range of enzymatic activities and development of robust microbial biocatalysts. The latter research direction includes biocatalyst advances such as: decreasing growth requirements and increasing the substrate range, development of hydrolytic enzyme secretion systems to reduce commercial enzyme use, and in general, optimizing a specific biocatalyst for use with a specific preprocessing and bioconversion method. The ultimate biocatalyst would secrete most, if not all of the required hydrolytic activity and efficiently transport and ferment hydrolysis limit products to fuel ethanol. Research in this direction will facilitate advances for low-cost, high yield bioconversion processes.

Dilute acid pretreatment is currently being developed as a leading pretreatment method. Other than the energy and chemical requirements discussed above, limitations specific to this process include formation of acid hydrolysis side products such as furfural and α-1,2-glucuronoxylose. Furfural forms from the acid and heat catalyzed dehydration of xylose. Formation of this side product reduces process efficiency in two ways. First, it reduces the net convertible xylose concentration and secondly, furfural is known to inhibit microbial growth and fermentation (Zaldivar et al., 1999). The aldouronate, α-1,2-glucuronoxylose, results from acid

hydrolysis of methylglucuronoxylan and methyglucuronoarabinoxylan due to the stability of the α-1,2 glycosyl linkage, which is thought to form an internal lactone between the carboxylate moiety on the glucuronic acid and a hydroxyl on the substituted xylose while under acidic conditions (M. E. Rodriquez, A. Martinez, S. W. York, K. Zuobi-Hasona, L. O. Ingram, K. T. Shanmugam, J. F. Preston, Abstr. 101st ASM General Meeting, abstr.O-21, 2001) (Jones et al., 1961). Whereas the arabinose and acetyl linkages are considered acid labile, the stability of the α-1,2 glucuronosyl linkage allows for the buildup of singly substituted aldouronates, which are unable to be utilized by any current biocatalyst. Considering that the frequency of substitution of methylglucuronate is at least 1 for every 10 xylose residues, this suggests at best, a bioconversion process can only recover 90% of the total xylose fraction. This does not take into account the significant potential contribution made by free glucuronic acid to the net ethanol yield.

In the short term, feasible goals can be met by developing enzyme systems which function efficiently to allow reduced pretreatment of biomass, in effect, lowering the total cost of preprocessing. Limited pretreatment with dilute acid would allow for reduced energy and/or acid consumption in the pretreatment process and would also lower the formation of furfural and other fermentation inhibiting compounds. The resulting pretreated biomass would still have a significant content of polymerized xylose and may require enzymatic treatment to fully release fermentable carbohydrate. For this reason, enzymes which degrade hemicellulose are primary research targets to facilitate utilization of methylglucuronoxylan and methyglucuronarabinoxylan by biocatalysts.

As detailed above, these two polymers make up the second most abundant carbohydrate in bioenergy crops and agricultural crop residues and unlike the chemically simple, but physically

recalcitrant cellulose polymer, methylglucuronoxylan and methylglucuronoarabinoxylan are chemically complex and require a battery of enzymes with a wide range of activities to fully degrade them to simple sugars. As shown in Figure 2-2 these activities include several different xylanases, an α-glucuronidase, acetyl esterases, arabinofuranosidases, and lignin esterases (not shown). Xylanases have a primary role in the degradation of xylan as they reduce the large linear polymer to small xylooligomers and small substituted xylooligomers. Accessory enzymes such as the α-1,2-glucuronidase are known to have activity on small substituted hydrolysis products resulting from xylanase digestion, but not on the intact polymer (Nagy et al., 2002; Nurizzo et al., 2002). This research will address how xylanases of glycosyl hydrolase (GH) family 5 and 10 function to hydrolyze polymeric methylglucuronoxylan.

Throughout this dissertation, methylglucuronoxylan is considered an idealized substrate consisting of a β-1,4-xylan substituted with α-1,2-linked 4-O-methylglucuronate moieties. Based on the acid labile character of the less significant substitutions on the methylglucuronxylan and methylglucuronoarabinoxylan polymers, pretreatment utilizing limited dilute acid conditions may result in methylglucuronoxylan being the primary retained polymer. Processing of methylglucuronoxylan by the three major families of xylanase enzymes is depicted in Figure 2-3. Xylanases of glycosyl hydrolases family 5 (GH 5) are the newest xylanases to be characterized. This work (Chapter 5) details the current understanding of this novel xylanase family. Although all indications are that it is specific for the hydrolysis of methylglucuronoxylan, resulting products are thought to be too large for direct utilization by biocatalysts. The abilities of this enzyme may well complement the activities of the other two primary xylanase families.

Xylanases from families GH 10 and GH 11 are relatively well characterized. Both of these xylanase families are known to produce primarily xylobiose and xylotriose as primary neutral

limit products of methylglucuronoxylan. However, while GH 10 xylanases yield the smallest substituted aldouronate, aldotetrauronate ($MeGAX_3$) (Fig 2-3) which is substituted directly on the nonreducing terminal xylose with α-1,2 glucuronate, GH 11 xylanases yield aldopentauronate ($MeGAX_4$) which is substituted penultimate to the nonreducing terminal xylose (Biely et al., 1997). This slight difference has significance in that substrates for the xylanolytic α-1,2-glucuronidase accessory enzyme can only hydrolyze methylglucuronate from xylooligomers when it is substituted directly on the nonreducing terminal xylose (Nagy et al., 2002; Nagy et al., 2003; Nurizzo et al., 2002). Further, most bacterial α-1,2-glucuronidase enzymes are intracellular and supporting research has indicated that $MeGAX_3$ is the largest aldouronate which is readily transported for catabolism (G. Nong, V. Chow, J. D. Rice, F. St. John, J. F. Preston, Abstr. 105th ASM General Meeting, abstr.O-055, 2005) (Shulami et al., 1999; St. John et al., 2006). Figure 2-3 depicts the processing of methylglucuronoxylan as an idealized substrate for utilization by bacterial biocatalysts.

Although GH 10 and GH 11 xylanases share identical hydrolytic mechanisms (as with GH 5) these two families differ in primary protein fold. Catalysis of the β-1,4 xylan chain proceeds by a double displacement mechanism with retention of anomeric configuration. Figure 2-4 identifies the structural differences and presents the common mechanism by which these xylanases function. The different limit aldouronates resulting from GH 10 and GH 11 xylanases result from steric interactions between the substituted xylan polymer and the binding cleft of these two xylanases with different protein folds.

For consistency, throughout the following chapters and just below, 4-O-methylglucuronoxylan will be abbreviated $MeGAX_n$ and corresponding substituted xylooligomers as $MeGAX_x$, where x equals the number of xylose residues. In sections

considering arabinose substitutions the name methylglucuronoarabinoxylan will be used and arabinose substituted xylooligomers will be denoted as AX_x, where x equals the number of xylose residues.

The following chapters contain the analysis of xylanases from glycosyl hydrolase families 5 and 10 and explore their mode of action and their hydrolysis products on the substrate $MeGAX_n$. By developing a strong understanding of how these enzymes act to hydrolyze $MeGAX_n$, how they function to benefit the native bacterial host in $MeGAX_n$ utilization and how they may facilitate enzyme systems for complete hydrolysis and utilization of $MeGAX_n$, we may better employ these enzymes in development of bacterial bioconversion processes and next-generation biocatalysts.

Table 2-1. Composition of potential biofuel crops and other biomass sources

Biomass Resource		Carbohydrate Composition	Non-carbohydrate Composition
Hardwood	*Populus tremuloides* (Poplar)[a]	48% cellulose 24% glucuronoxylan	21% lignin
	Betula papyrifera (Paper Birch)[a]	42% cellulose 35% glucuronoxylan	19% lignin
Herbaceous	Switchgrass (early cut)[b]	40.7% cellulose 35.1% hemicellulose	5.5% lignin
	Switchgrass (late cut)[b]	44.9% cellulose 31.4% hemicellulose	12% lignin
Crop residue	Corn stover[b]	36.4% glucan 18.0% xylan	16.6% lignin
	Wheat straw[b]	38.2% glucan 21.2% xylan	23.4% lignin
	Sugarcane Bagasse[c]	32-48% cellulose 19-24% xylan	23-32% lignin

Adapted From:

[b] **Timell, T. E.** (1967). Recent progress in the chemistry of wood hemicelluloses. Wood Sci Technol *1*, 45-70.

[a] **Lynd, L. R., C. E. Wyman, and T. U. Gerngross** (1999). Biocommodity Engineering. Biotechnol Prog *15*, 777-793.

[c] **Scurlock, J.** (http://bioenergy.ornl.gov/papers/misc/biochar_factsheet.html). Bioenergy feedstock characteristics (Oak Ridge: Oak Ridge National Laboratory, Department of Energy).

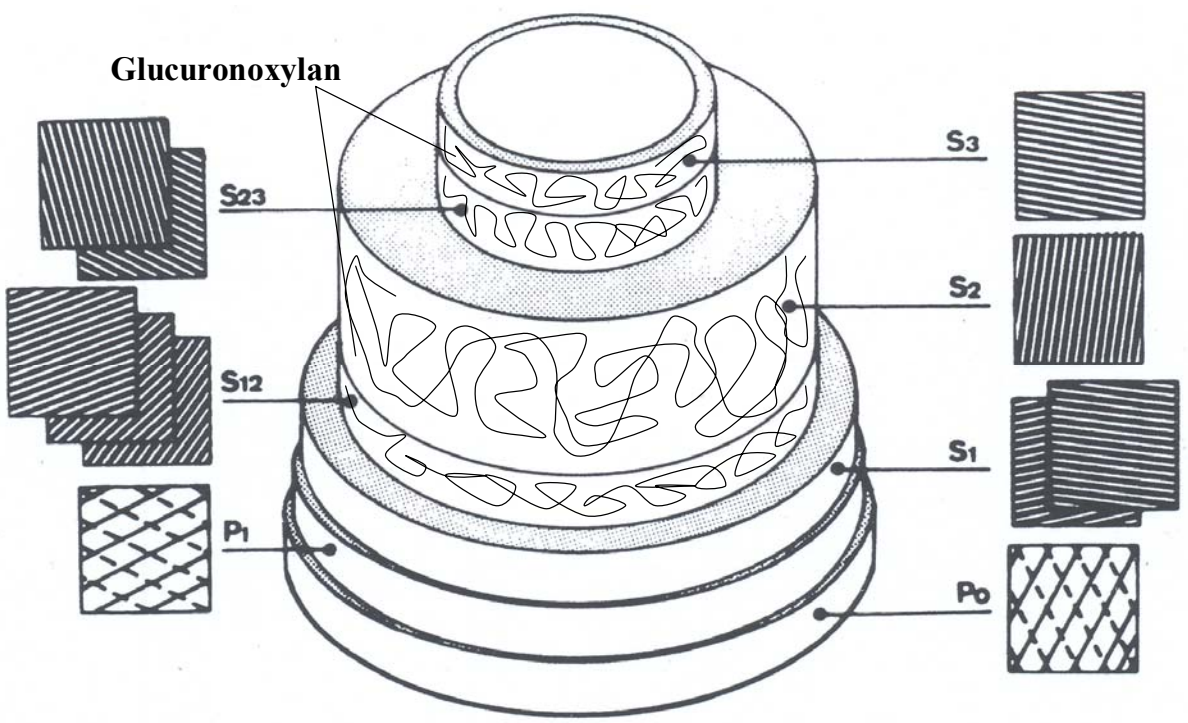

Figure 2-1. Pattern of cellulose fiber deposit in different layers of the primary and secondary cell wall. A "P" designation refers to layers of the primary cell wall while an "S" designation refers to layers of the secondary cell wall. Glucuronoxylan is thought to be more concentrated at the interface between secondary cell wall layers. Figure adapted from, **Fujita, M., and H. Harada** (2001). Ultrastructure and formation of wood cell wall, In Wood and Cellulosic Chemistry, D. N.-S. Hon, and N. Shiraishi, eds. (New York: Marcel Dekker), pp. 1-49.

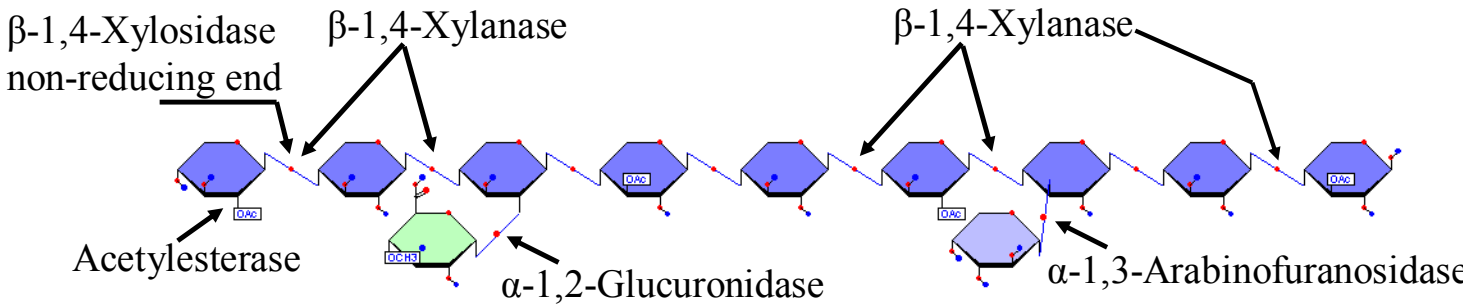

Figure 2-2. Common structural elements and sites of enzymatic hydrolysis which degrade methylglucuronoxylan and methylglucuronoarabinoxylan.

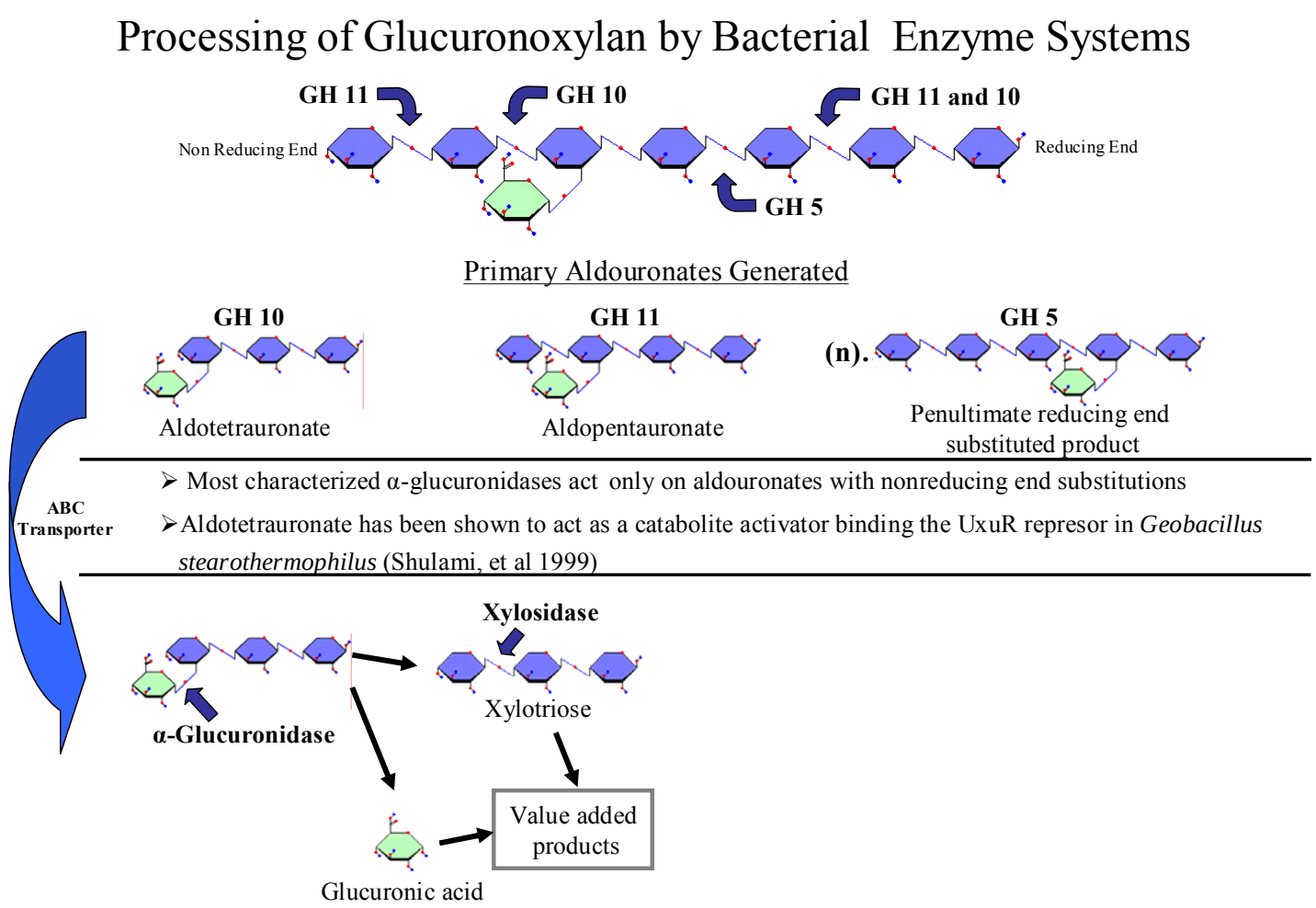

Figure 2-3. Generation of hydrolysis products by different families of xylanases highlighting the intricate role of GH 10 xylanases in complete pentosan utilization with respect to the other xylanases.

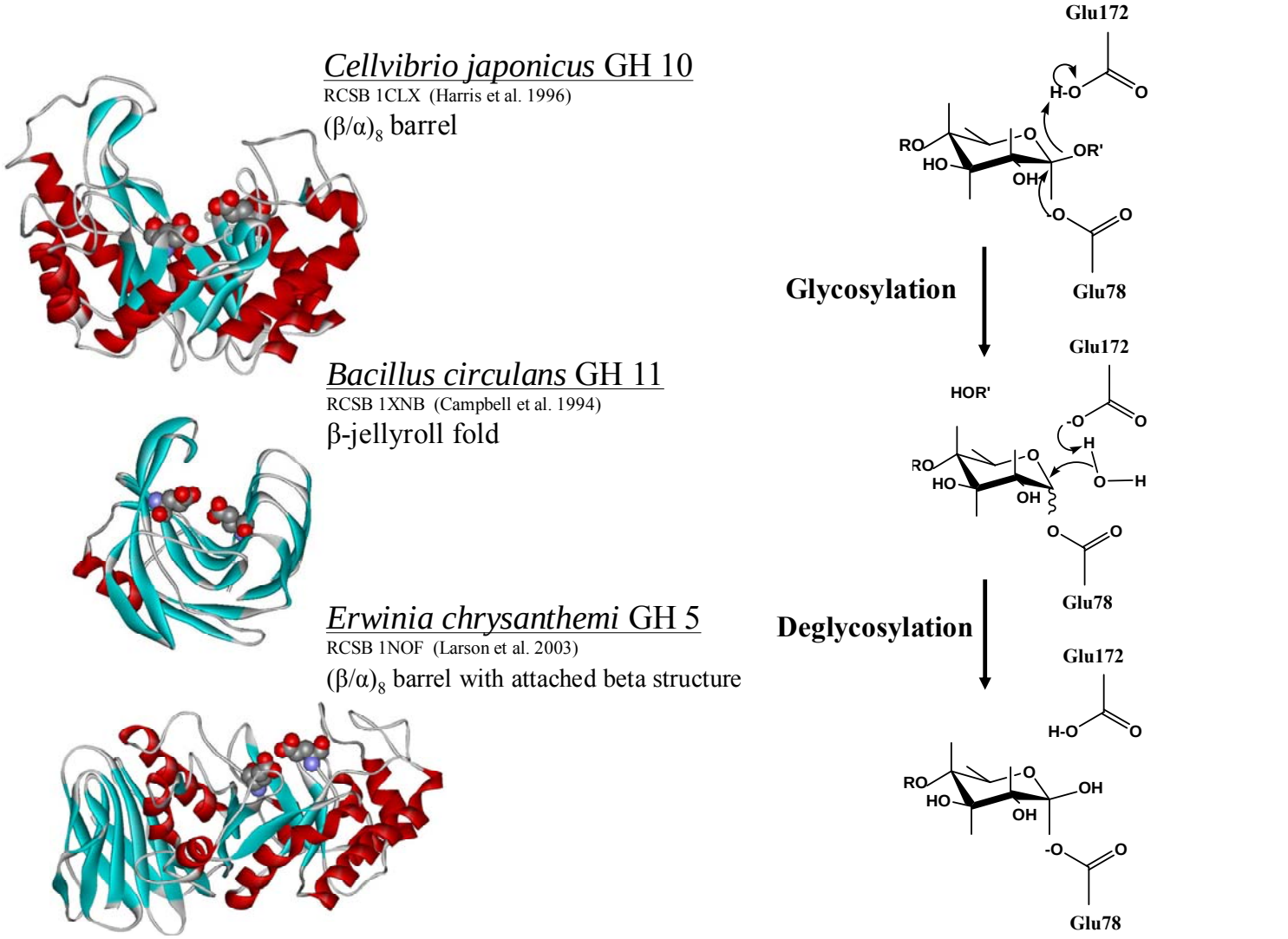

Figure 2-4. Xylanase structure and function. Diverse xylanase structures catalyze identical reactions by the same mechanism.

CHAPTER 3
FAMILY 10 GLYCOSYL HYDROLASES: STRUCTURE, FUNCTION AND PHYLOGENETIC RELATIONASHIPS

A *Paenibacillus* sp. (strain JDR-2) has been isolated that is capable of efficient utilization of $MeGAX_n$. Studies of this organism have attributed this ability to the production of a large multimodular GH 10 xylanase. This 154 kDa secreted protein has eight separate modules which contribute functions for efficient hydrolysis and utilization of polymeric xylan. Two different modules are involved with substrate association while another module is involved in cell surface localization. The proximity between the cell, substrate and hydrolysis products which results from the combined function of these appended modules are thought to facilitate vectoral or directional utilization of xylan hydrolysis products (St. John et al., 2006). Analysis of *Paenibacillus* sp. strain JDR-2 and characterization of $XynA_1$ is presented in Chapter 4. This chapter reviews GH 10 xylanases and endeavors to establish functional themes through analysis of associated modules and application of phylogenetics.

Xylanases of Glycosyl Hydrolase Family 10

Glycosyl hydrolase family 10 (GH 10) xylanases are arguably the best studied and understood family of xylanolytic enzymes. Their substrate is the ubiquitous β-1,4-linked xylose backbone of the major xylans of hardwood and crop residues, the primary sources of biomass for bioconversion to ethanol. It is expected that GH 10 xylanases have a leading role in the degradation of $MeGAX_n$, allowing for subsequent turnover of this major biomass component. To date, they have been found in all three domains of life. The Carbohydrate Active Enzymes database (CAZy) (www.cazy.org/CAZY/) has 175 bacterial GH 10 entries and 110 eukaryote entries (Davies and Henrissat, 1995; Henrissat, 1991; Henrissat and Bairoch, 1993). As with any sequence database in this era of genomics, most of the sequences have been deposited in conjunction with genome sequencing projects. Hence, only a few have been studied for kinetic

properties and fewer have received detailed molecular analysis to understand mechanistic enzyme substrate interactions.

Enzyme Structure and Mechanism

The primary unit of GH 10 xylanases is the catalytic domain (CD). This module typically ranges in size from 30 kDa to 40 kDa. Several examples of GH 10 xylanases have been crystallized and x-ray diffraction for structural analysis revealed a common α/β_8 protein fold. As with many endo acting glycosyl hydrolases, GH 10 xylanases have a substrate binding cleft that appears from crystal structures to run the breadth of the enzyme (Figure 2-3). The binding site of GH 10 xylanases (as with most glycosyl hydrolases) is composed of a series of subsites that position and bind individual xylose residues. The nomenclature for describing the organization of subsites has been reviewed (Davies et al., 1997). It is based on the convention for the naming of polymeric carbohydrates and the point of hydrolysis within the enzyme. Subsites are numbered increasing from the point of hydrolysis with negative designation toward the nonreducing terminus (glycone) (left) and a positive designation in the direction of the reducing terminus (aglycone) (right). Hydrolysis of xylan occurs through a double displacement mechanism with retention of anomeric configuration (Davies and Henrissat, 1995; Gebler et al., 1992; Henrissat et al., 1995) (Fig. 2-4). Two glutamate residues have been identified which catalyze this hydrolysis, one acting as the primary nucleophile and the other as the proton donor.

Modular Characteristics of GH 10 Xylanases

Often GH 10 xylanases are associated within their translated protein product, with additional separately folding domains. Although a variety of different functional domains have been identified, the majority represent carbohydrate binding modules (CBM). These separately folding modules are β-sheet structures that bind target carbohydrates, not necessarily xylan. Further, it is common for there to be modular repeats such that there are multiple modules of the

same family. The largest GH 10 xylanase in the CAZy database has eight separate modules and is 194 kDa in mass. Six of the modules represent three different CBMs and a seventh module, an additional enzymatic activity.

No direct interaction has been identified between a GH 10 CD and its associated CBM. The modules are generally connected through linker regions that in some cases have characteristic protein sequences. It is generally thought that linker regions lack structure, having the singular task of connecting two functional domains. In only two cases has a CBM been crystallized together with its cognate CD (Fujimoto et al., 2000; Pell et al., 2004a). In these studies the linker did not yield a precise electron density map for structure analysis. The tethering action of the linker sequence between a CD and CBM identifies a simple spatial relationship required for enhancement in CD function. This contrasts with the concept of a coordinated interaction in which accessory modules may directly interact with catalytic modules for enhanced functionality (Akin et al., 2006; Irwin et al., 1998; Sakon et al., 1997). Boraston et al. have recently reviewed the structure function and classification of CBM modules (Boraston et al., 2004). Conventional wisdom suggests that these modules help target the CD to the expected substrate, thereby increasing the localized substrate concentration. In contrast to idealized kinetic systems with a soluble substrate, the recalcitrant, composite character of lignocellulosic biomass requires enzymes to be targeted to specific regions of the substrate for effective hydrolysis. The frequent occurrence of GH 10 xylanases that do not contain an appended carbohydrate binding module suggest that in many cases CBMs are not necessary for desired function.

There are 46 families of carbohydrate binding modules in the CAZy database assigned on the basis of sequence similarity and hydrophobic cluster analysis (Henrissat and Bairoch, 1993;

Tomme et al., 1995). Currently, eleven are found associated with GH 10 xylanases. These include members of families 1, 2, 3, 4, 5, 6, 9, 10, 13, 15 and 22. The occurrence of these modules within GH 10 xylanases varies greatly. For instance, of all the different 106 families of glycosyl hydrolases in the CAZy database, CBM 22 modules are primarily found associated with bacterial and plant GH 10 xylanases (Table 3-1). CBM 9 modules are only associated with GH 10 xylanase that already have a CBM 22 module, but for these modules, only bacterial associations are known. CBM's of families 2, 3, 5, 6 and 10 are primarily found in bacterial enzymes, with only a few from fungal glycosyl hydrolases. Many of these modules are associated with a variety of enzymatic activities. To exemplify an extreme distribution, CBM family 5 has only a single module associated with a GH 10 xylanase, but has about 200 entries in the CAZy database associated mainly with chitinase and cellulase enzymes, while family 15 has only two entries in the database and both are associated *Cellvibrio* GH 10 xylanases. Of all the families, CBM family 13 is probably the most diverse, with representatives in bacterial, fungal, plant and mammalian proteins. These modules are only common in GH 10 xylanases from *Streptomyces* (Table 3-1). CBM 1 modules are common in fungal GH 10 xylanases, but are found more often in other families of glycosyl hydrolase enzymes of fungal origin.

CBM classification by target substrate

A relatively new classification system for carbohydrate binding modules identifies their target substrate rather than their protein fold. Type A modules bind crystalline substrates such as crystalline cellulose, which is not necessarily the target of the associated catalytic module. Type B modules bind soluble substrates, which are usually the intended substrate for the catalytic module and Type C modules bind small soluble sugars such as cellobiose. The CBM modules listed above which are found appended to GH 10 catalytic modules have representatives in each of these types. Modules 1, 2a (see below), 3, 5, 10 classify as Type A, modules 2b (see below),

4, 6, 15 and 22 classify as Type B and modules 9 and 13 classify as Type C (Boraston et al., 2004). The following descriptions clarify their carbohydrate binding preferences detailing the differences between the three types.

CBM modules common in bacterial GH 10 xylanases and their general architectural arrangement

Modules of CBM family 22 are associated with GH 10 xylanases. There are examples of this module associated with GH 10 xylanases of bacterial and plant origin. Even with this diversity, there is only one example of a characterized CBM 22 not from bacterial origin. Early studies assigned a thermostabilizing function to these domains as removal of the module resulted in decreased thermal stability of the cognate xylanase CD (Fontes et al., 1995). It was soon realized that these domains have a primary role in binding carbohydrate polymers. Most CBM 22 modules are located N-terminal to the CD and are often observed as a duplicate or triplicate set (Ali et al., 2001b; St. John et al., 2006). Xyn10B of *Clostridium thermocellum* has a unique CBM 22 configuration and also has been well characterized (Charnock et al., 2000; Dias et al., 2004; Xie et al., 2001). In this case the CD is flanked on both sides by a single CBM 22 module. Substrate binding studies showed that while the module on the C-terminal side of the CD has affinity for xylan, the N-terminal localized CBM 22 has no detectable affinity for tested substrates. Crystal structure analysis of the functional module revealed a β-sandwich structure with a small cleft for binding substrate sugars. The tandem N-terminal CBM 22 modules of Xyn10A of *Clostridium josui* expressed together showed similar carbohydrate binding properties as the C-terminal located CBM 22 of Xyn10B from *Clostridium thermocellum* described above (Ali et al., 2005a).

Xyn10C of *Clostridium thermocellum* has a single CBM 22 in the N-terminal region. In a recent report, absorption assays with the recombinantly expressed CBM showed that it bound

best to acid-swollen cellulose and ball milled cellulose but native affinity polyacrylamide gel electrophoresis (NAPAGE) analysis showed the greatest gel retardation with birch wood xylan. Although in these studies substrate affinity for this CBM 22 is not definitive, results showed a 4-fold activity increase between the Xyn10C CD expression product and the native Xyn10C protein product, confirming the generally accepted role of CBM modules (Ali et al., 2005b). The nonbacterial contribution to CBM 22 characterizations comes from the ruminal protozoan *Polyplastron mutivesiculatum*. Xyn10B of this protozoan has a single N-terminal CBM 22 as described above for Xyn10C of *Clostridium thermocellum*. While it was shown to bind xylan, it did not function to enhance catalytic activity (Devillard et al., 2003).

In two cases, CBM 22 modules have been shown to bind mixed linkage β-1,3-1,4 glucan chains. Work with Xyn10B of *Clostridium stercorarium* did not differentiate between an N-terminal duplicate of CBM 22 modules, but showed that they only slightly increased activity on oat spelt xylan with respect to the separate catalytic domain. Unexpectedly, activity on β-1,3-1,4 glucan, which was very low with the separate catalytic domain, was higher than activity on oat spelt xylan for the native non truncated enzyme (Araki et al., 2004). Previous work with this enzyme showed that the CBM modules facilitated binding to cellulose (Ali et al., 2001b). XynA of *Thermotoga maritima* has an identical CBM 22 arrangement. Very detailed studies of these modules identified major differences between the first (CBM 22-1) and second (CBM 22-2) (left to right) modules. Meissner and colleagues showed by NAPAGE that CBM 22-2 bound β-1,3-1,4 glucan, β-1,3-1,4 xylan, and β-1,4 xylans while CBM 22-1 failed to show separate affinity for these potential substrates (Meissner et al., 2000).

Their wide-spread diversity and apparent variety of specificity make CBM 22 modules interesting platforms to study carbohydrate epitope recognition and binding. Further structural

work may lead to sugar binding cleft engineering for development of tools in biotechnology. Thorough studies of CBM 22 modules clearly show that sequence based determination of the presence of these modules cannot confidently be correlated to a specific function, it is clear that these modules are involved with binding carbohydrate polymers.

CBM 9 modules are frequently associated with CBM 22 modules. These modules are usually positioned just C-terminal to the CD and their presence is most common in modular GH 10 xylanases which already have a CBM 22 module to the N-terminal side of the CD. In several cases, modular GH 10 xylanases from thermopiles have tandem CBM 9 modules. Defining research, characterizing the second CBM 9 module of *T. maritima* Xyn10A (CBM 9-2) showed that this module had high affinity for small soluble oligosaccharides, including glucose, xylose, cellobiose, xylobiose. This attribute classifies these modules as Type C. Binding affinities for oligomers over two residues did not increase, indicating that the binding epitope recognized no more that two sugar residues. This type of CBM also displayed affinity toward xylans and cellulose of all types (Ali et al., 2001a; Clarke et al., 1996; Notenboom et al., 2001). The Mechanism of binding was characterized using sodium borohydride reduced polymers. Replacement of the hemiacetyl reducing terminal sugar with sugar alcohols prevented binding of CBM 9-2. Subsequent crystal structure analysis supported these findings revealing that every hydroxyl group of the reducing terminal sugar in a cyclic conformation interacted with the protein via multiple hydrogen bonding interactions (Boraston et al., 2001; Notenboom et al., 2001).

Analysis of CBM 9-1 of *T. maritima* Xyn10A failed to identify a functional role for this very similar module. Modeling of CBM 9-1 using CBM 9-2 as template and a sequence alignment of many CBM 9 sequences showed that CBM 9-1 as well as all other CBM 9 modules

in the same modular position in a tandem arrangement lacked the structurally characterized conserved sugar binding amino acids identified in CBM 9-2. Based on the differences between CBM 9-1 and CBM 9-2, it has been proposed that two subfamilies be designated (CBM 9a and 9b). In all of the CBM 9 tandem arrangements, the first CBM 9 classifies as a CBM 9a and the second CBM 9 classifies as a CBM 9b (Notenboom et al., 2001).

Many other GH 10 xylanases, including three in the alignment discussed above, have single CBM 9 modules. The three included in the alignment and the single CBM 9 of *Paenibacillus* sp. strain JDR-2, a mesophilic, aggressive xylan utilizing organism, show near complete conservation of the key residues attributed to sugar binding in CBM 9-2 of *T. maritima*. This suggests that GH 10 xylanases, in which there is a single CBM 9 module, it may function similar to that of CBM 9-2 of *T. maritima*.

The concise studies performed with Xyn10A CBM 9b of *T. maritima* identified a role for these modules in binding of reducing terminal sugars. Although this module showed affinity for the reducing ends of xylan and cellulose, it had a much higher association constant with cellobiose that with xylobiose, possibly indicating a preference for binding of cellulose.

Modules of CBM family 2 can bind crystalline cellulose and xylan. While there are eleven sequences within the GH 10 family that contain this domain, there are about 200 in the database from glycosyl hydrolase families of chitinases and various cellulases. This family has been grouped into two subfamilies designated CBM 2a (Type A) and CBM 2b (Type B). While CBM 2a binds to crystalline cellulose, CBM 2b has been shown to bind soluble xylan. The difference in structure that changes substrate specificity is attributed to a single amino acid switch which reorients a tryptophan for binding to xylan (Simpson et al., 2000). Based on this analysis, out of the eleven CBM 2 modules found in GH 10 xylanases, only one is classified as a

CBM 2b. The other ten classify as CBM 2a, presumably having specificity for crystalline cellulose. More examples of CBM 2b modules are associated with GH 11 xylanases. CBM 2a modules bind crystalline cellulose irreversibly (Creagh et al., 1996) but are thought to be mobile, allowing for movement on the surface of cellulose crystalline fibers (Jervis et al., 1997). An example of a CBM 2a module from a GH 6 cellobiohydrolase has been shown to disrupt crystalline cellulose (Din et al., 1994), revealing a synergism between the CBM 2 module and the associated CD. Thermodynamic and structural analysis of these modules conclude that binding of crystalline cellulose occurs through an entropic driven process, probably due to displacement of water molecules between the cellulose surface and the near planar face of the carbohydrate binding module (Creagh et al., 1996; McLean et al., 2000).

Family 3 CBM modules bind crystalline cellulose. This family is of notable interest for cellulases of GH family 9. It can be found in five GH 10 xylanases, three of which have two separated modules. These modules have been divided into three subfamilies. Although both CBM 3a and 3b bind to the surface of crystalline cellulose, CBM 3a differs from CBM 3b primarily in a loop structure which contributes to substrate binding (Jindou et al., 2006). Further, CBM 3a modules are associated with scafoldin components of the cellulosome (Shimon et al., 2000) where CBM 3b modules are enzyme localized (Gilad et al., 2003). The last subtype, CBM 3c, is a glycosyl hydrolase family 9 CD fusion domain which is thought to feed or guide the cellulose chain into the GH 9 catalytic domain. This type of association is attributed to processive degradation of cellulose (Irwin et al., 1998; Sakon et al., 1997).

CBM modules of family 6 bind soluble polymeric sugar substrates. There are seven GH 10 xylanases containing this Type B CBM module. They differ from other group B CBM domains in that the substrate binding location is a ridge rather than the typical cleft of the β-

sandwich structures. They have been shown to bind a variety of soluble substrate sugars with similar affinities. A CBM 6 module from *Cellvibrio mixtus* endoglucanase 5A has revealed two binding sites, each with unique substrate specificities. Binding of xylan was specific for cleft A which could also bind cellooligosaccharides, while cleft B also bound cellooligosaccharides, but was specific for β-1,3-1,4-glucans (Boraston et al., 2003; Henshaw et al., 2004; Pires et al., 2004).

Xylanases from *Streptomyces* spp. have CBM 13 modules. Of the 10 CBM 13 modules associated with GH 10 xylanases, 7 are in *Streptomyces* sequences (Table 3-1). These modules have similarity to the lectin like B-chain of ricin toxin which has specificity for galactose. Each CBM is composed of a triplicate repeat of approximately 40 amino acids and each repeat is a separate site for carbohydrate binding. CBM 13 modules are selective for pyranose sugars with generally low association constants at each site. Upon binding of polymeric xylan there is a "cooperative and additive" effect (Notenboom et al., 2002), increasing the affinity for this substrate more than by a simple additive result of the three contributing sugar binding sites (Boraston et al., 2000; Fujimoto et al., 2002; Notenboom et al., 2002). Studies have indicated that the three binding sites (α, β, γ) accommodate three different xylooligomers (Scharpf et al., 2002).

CBM modules of families 4, 5, 10 and 15 are rare in GH 10 xylanases. CBM modules of family 4 have been identified in about 30 sequences from the CAZy database. Only one of these sequences is a GH 10 xylanase which has an N-terminal tandem set. All the others which have been identified are associated with various β-1,4 and β-1,3 glucanases. Structural studies have characterized this module family as having a β-sandwich jelly roll fold (Johnson et al., 1996b). Binding of soluble carbohydrates occurs within a binding cleft. The bottom of the cleft

is lined with hydrophobic residues and the walls have hydrophilic residues for hydrogen bonding interactions with the carbohydrate polymers. Several subfamilies of CBM 4 modules have been identified. In general, the CBM 4 modules bind substrate for the associated catalytic module (Simpson et al., 2002; Zverlov et al., 2001). The first studies of this module family were performed with the N-terminal tandem CBM 4 modules from *Cellulomonas fimi* CenC. These modules were specific for soluble β-1,4 glucan and did not associate with xylan (Brun et al., 2000; Johnson et al., 1996a; Johnson et al., 1996b; Tomme et al., 1996). Xyn10A of *Rhodothermus marinus* was found to have a related tandem N-terminal set of modules and substrate binding studies showed that although they had a low affinity for soluble cellulose they showed specificity for xylans (Abou Hachem et al., 2000). Structural analysis of the second module of this system allowed the researchers to postulate differences which dictate substrate specificity between those from *C. fimi* CenC that bound soluble cellulose and those from *R. marinus* Xyn10A that bind xylan (Simpson et al., 2002). The CBM 4 modules from *T. neapolitana* Lam16A, a laminarinase (β-1,3-glucan), do not bind soluble cellulose but are specific for various β-1,3 linked glucan polymers (Zverlov et al., 2001). The diversity of carbohydrate binding in this family is similar to that found in family 22 modules. Both are classified together as Type B CBMs in a larger superfamily (Sunna et al., 2001).

There is only one example of a CBM 5 module in GH 10 xylanases. These modules are thought to bind cellulose but the primary associated enzymatic activity is a chitinase. The CBM 5 of *Erwinia chrysanthemi* endoglucanase Z has been structurally characterized and the authors correlated it with CBM 5 modules associated with chitinase enzymes (Brun et al., 1995; Brun et al., 1997). At present, family 15 CBMs have only been found in two enzymes. Both are GH 10 xylanases of *Cellvibrio* (Table 3-1). With these modules, association constants increase up to

xylohexaose, indicating there are 6 subsites in the binding cleft. Although no natural substituted polymer such as $MeGAX_n$ achieved as high an association constant as observed with xylohexaose, affinity in the worst case decreased by only one-half, not a significant decrease in the measure of association. These modules are thought to efficiently bind decorated xylan because the O2 and O3 hydroxyls (substituted positions in native xylan) of most xylan binding subsites are solvent exposed (Szabo et al., 2001). Only a single CBM 10 module has been found associated with GH 10 xylanases. These 45 amino acid modules have a hydrophobic side involved with cellulose binding. The mechanism of association with crystalline cellulose is similar to CBM 2 modules with coplanar aromatic amino acid residues stacking on the cellulose surface (Millward-Sadler et al., 1995; Ponyi et al., 2000; Raghothama et al., 2000).

Fungal modules

Of all the domains associated with GH 10 xylanases, CBM 1 modules are strictly found in sequences from fungal enzymes. However it is not restricted to xylanases, being primarily found in fungal cellobiohydrolases and cellulases. These small 36 amino acid structures have four highly conserved cysteine residues involved in the formation of disulfide bridging (Kraulis et al., 1989). This module has been shown to facilitate association of cellobiohydrolases and cellulases with cellulose (Carrard et al., 2000; Gilkes et al., 1991). In GH 10 xylanases these modules are usually located to the far N or C-terminal region, some distance from the CD. One report showed that the CBM 1 of a GH 7 reducing terminal cellobiohydrolase (Cbh1) from *Penicillium janthinellum* had a disruptive effect on cellulose that enhanced activity (Boraston et al., 2004). No research has determined possible differences between the CBM 1 modules in cellulose active fungal enzymes and those in fungal GH 10 xylanases. It may be that similarities among these small modules are significantly high to discourage such endeavor. If the primary purpose of this

module is to associate the fungal GH 10 CD with cellulose, it serves a similar role as several bacterial CBM modules.

Other modules and sequences from GH 10 xylanase

Surface Layer Homology (SLH) domains anchor proteins to the cell surface. SLH domains have several roles in bacterial physiology. With respect to GH 10 xylanase and glycosyl hydrolase function, these domains are often arranged as C-terminal sets (up to three separate domains) and are involved in anchoring the associated enzyme to the cell surface. They are also involved as a primary surface anchoring mechanism for the multicomponent lignolytic cellulosome complex produced by several *Clostridium* spp. Bacterial surface binding studies of several SLH module sets have identified two mechanisms of binding. Several studies have identified binding to secondary cell wall polysaccharide (SCWP). These binding sites consist of carbohydrates associated with the peptidoglycan cell wall. Genetic verification for this binding mechanism was obtained from a *csaB* gene knockout in *Bacillus anthracis* (Mesnage et al., 2000). Other results indicate that SLH domains bind directly to the cell wall peptidoglycan layer. Recent work with the SLH C-terminal triplicate of the scafoldin dockerin binding protein (SdbA) from *C. thermocellum* found that it bound to the peptidoglycan layer of *Escherichia coli*. This was in contrast to the SLH doublets from the xylanases, Xyn10A and Xyn10B of *C. josui* and *C. stercorarium*, respectfully. In this case, these SLH domains displayed specificity for the *Clostridia* SCWP extract and reduced affinity for hydrofluoric acid extracted secondary cell wall polymer (Zhao et al., 2006a). This preference for binding native peptidoglycan suggests that the SLH modules from the SdbA protein may be used in biotechnology applications. Recently, similar binding selectivity was observed between the two surface layer proteins (Slp1 and Slp2) and the cellulosome anchoring protein (Anc1) of *C. thermocellum* (Zhao et al., 2006b). Studies of $XynA_1$ and Xyn5, both GH 10 xylanases from different *Paenibacillus* spp. (Ito et al., 2003; St.

John et al., 2006) have shown that the C-terminal triplicate SLH module anchors the GH 10 xylanase to the cell surface. These triplicate SLH domains as well as the linker region to their N-terminal have homology to the same region of the SdbA protein discussed above. Based on this homology, these two xylanases may bind with specificity similar to those examples above which bind native peptidoglycan with no requirement for SCWP. Currently, there may be enough characterized examples available to allow for sequence based determination of amino acid functionality and define the differences between these two similar modules that bind different polysaccharides on bacterial cell surfaces.

Characteristic linker regions connect modules in some enzymes. Although the ascribed function of linker regions in glycosyl hydrolases is that they connect together functional domains, the identification of linkers with unique amino acid sequences has made them an interesting topic of research. These unique sequences are characterized as having very high content of specific amino acids. These include the serine rich linker (Sr) (*Cellvibrio*, *Pseudomonas*, and *Saccharophagus*) (Hall et al., 1989), the asparagine rich linker (Nr) (*Ruminococcus*), the proline and threonine rich linker (PTr) (*Caldibacillus*, *Caldicellulosiruptor*, and *Cellulomonas*), the proline and glutamate rich linker (PEr) (*Colwellia*, *Pseudomonas*, and *Saccharophagus*), and the proline and glycine rich linker (PGr) (*Thermobifida*). The serine rich linker regions of XynA and XynC (an arabinofuranosidase) of *Cellvibrio japonicus* have been characterized (Table 3-1) (Black et al., 1997; Black et al., 1996; Ferreira et al., 1990). Initial studies with XynA determined that the linker sequence was not required for activity and substrate binding functions. Removal of this intervening sequence resulted in lower activities. Although this could be attributed to other functions, it was concluded that it resulted from reduced flexibility of the CD with respect to the CBM.

A completely novel linker sequence has been identified in XynB of *Neocallimastix patriciarum*. It is composed of 12 tandem repeats of the core amino acid sequence TLPG followed by 45 tandem repeats of the octapeptide XSKTLPGG (X=S, K or N). This linker region connects a C-terminal family 1 CBM. Research to elucidate this modular system failed to obtain good expression for functional analysis but showed that the CD sequence coded for a functional GH 10 xylanase (Black et al., 1994).

As discussed above, CBM 22 modules were originally thought to confer thermostabilizing properties to GH 10 xylanases. New research shows it is possible that these conclusions resulted from the presence of the linker sequence. The 18 amino acids connecting the CBM 22 module with its cognate GH 10 CD has recently been shown to attribute thermo stabilization and resistance to proteolysis (Dias et al., 2004).

Glycosyl Hydrolase Accessory Module Discussion

The descriptions above regarding the modules appended to GH 10 CDs exemplify the functional diversity common in glycosyl hydrolase families for lignocellulose degradation. Whether associated with a xylan or crystalline cellulose binding domains the assumed goal of these modules is to facilitate interaction with the substrate. Modules like family CBM 2, which target crystalline cellulose, may have roles in xylan hydrolysis by GH 10 xylanases that cannot easily be determined. Endeavors to distinguish functionality of these modules with respect to the GH 10 catalytic core may facilitate development of applications for efficient enzymatic hydrolysis of lignocellulosics.

Bacterial GH 10 domain architecture. As can be observed in Figure 3-1, common domain arrangements are evident. Significant modular arrangements include: CBM 22 modules are localized to the N-terminal region of the GH 10 CD (except for one GH 10 of *C. thermocellum*), all CBM 9 modules are localized to the C-terminal side of the CD and all but one

is associated with CDs that also have a CBM 22 module. All sequences which have SLH modules for possible cell surface anchoring also have both CBM 22 and 9 modules. CBM 3 modules are always immediately flanked by proline and threonine-rich linker regions and are only found in *Caldibacillus* and *Caldicellulosiruptor*. In several cases there are two of these in the same xylanase. CBM 2 modules are in GH 10 xylanases from *Cellvibrio*, *Sacharophagus*, *Cellulomonas*, *Streptomyces* and *Thermobifida* (Table 3-1). These modules are also often flanked by a proline and threonine-rich or a serine-rich linker sequence. Predictions of GH 10 xylanase function can be proposed based on common architectural module associations and an understanding of the function of these carbohydrate binding modules.

Information concerning the method by which the CBM facilitates CD activity can also be deduced from positional relationships. While many of the CBM modules of bacterial GH 10 xylanases are usually in a specific position with respect to the CD, some modules are not. The CBM 1 module is only found associated with fungal CDs. Of the fourteen which have this domain, seven have it toward the N-terminal and seven have it toward the C-terminal. From this, it seems that the only purpose of this module is to ensure localization to the lignocellulose substrate.

From Figure 3-1 and the brief description of associated modules above, we can imagine a mode of action for these GH 10 xylanases based on their respective module assemblages. The *Thermoanaerobacterium saccharolyticum* (P36917) and *Paenibacillus* species JDR-2 (62990090) xylanase would be expected to associate with soluble polymers such as xylan or β-1,3-1,4 glucan with their N-terminal CBM 22 domains. The CBM 9 module is expected to bind the reducing terminus of a cellulose chain fixing the catalytic module in place and the C-terminal SLH modules should anchor this enzyme system to the cell surface. The combined properties of

these appended modules favor simultaneous substrate and cell surface localization, perhaps increasing hydrolysis product recovery by the cell through a process of vectoral transport. The triplicate family 22 CBM in the N-terminal region of the *Arabidopsis thaliana* GH 10 xylanase (Q9SM08) is expected to facilitate substrate localization for this CD. How these enzymes function in *A. thaliana* is difficult to determine but it can be imagined that they may function in expansion of the cell wall. The *Caldibacillus cellulovorans* xylanase (7385020) has a C-terminal localized double CBM 3 set. These modules would be expected to bind the crystalline surface of cellulose and the N-terminal CBM 22 would promote association with soluble substrate. A similar mode of action can be imagined for the xylanase from *Cellulomonas fimi* (73427793). The irreversible binding and mobile character of the C-terminal CBM 2 module would allow an associated CD to translate the surface of cellulose crystals, in search for substrate. The *Streptomyces coelicolor* (Q8CJQ1) modular xylanase is the simplest of all the examples. The CBM 13 module in the C-terminal region is expected to associate with soluble xylan and increase localized substrate concentration to enhance enzymatic efficiency.

These examples offer a glimpse into the possible mode of action for several modular GH 10 xylanases. Although these descriptions are not absolute, they provide a framework for development of methods which utilize these enzymes for complex biomass degradation.

Hydrolysis of Substituted Xylans by GH 10 Xylanases

Xylan hydrolysis by GH 10 xylanases primarily result in the limit products xylose, xylobiose, xylotriose and small substituted xylooligomers. Early studies using GH 10 xylanases from *Cryptococcus albidus* and *Streptomyces lividans* to digest methylglucuronoxylan resulted in the characterization of aldotetrauronate ($MeGAX_3$) as the smallest substituted xylooligosaccharide (Fig. 3-2) (Biely et al., 1997). Similar work digesting insoluble wheat arabinoxylan with the GH 10 xylanase, XylA from *Thermoascus aurantiacus* resulted in two

small substituted limit products. Arabinofuranose-xylobiose (AX_2) with the substitution in the O3 position of the nonreducing xylose of xylobiose and arabinofuranose-xylotriose (AX_3) with the same substitution on the middle xylose of xylotriose resulted as 50% and 30% respectively of the total arabinofuranose (Ara*f*) substituted products (Fig. 3-2) (Vardakou et al., 2003). These biochemical methods have recently been validated with detailed structural studies of two GH 10 xylanases together with these limit products (Fujimoto et al., 2004; Pell et al., 2004b).

Xylan has been reported to have a three fold helical symmetry. Binding subsites of GH 10 xylanases and CBM modules specific for xylan accommodate this characteristic. Native xylan is usually substituted at the O2 or O3 hydroxyl positions (Chapter 2). Substitutions in these positions along the xylan chain can either be accommodated into the protein structure or exposed to the solvent so as not to interfere with subsite xylan interaction. Specific interactions can be understood from the positioning of the O2 and O3 hydroxyl in the subsite bound xylose residue relative to the protein structure. If a subsite orients the bound xylose moiety such that these positions of the xylose are sterically confined by protein structure, no substitution can be accommodated in that position. For a subsite to bind a substituted xylose moiety in the xylan chain, there can either be a pocket into which the substitution can fit in the protein tertiary structure, or the O2 and O3 hydroxyl positions can be solvent exposed away from the protein surface. As will be seen below, substituted hydrolysis products can also result from subsite flexibility. Resulting substituted hydrolysis limit products reflect subsite accommodation by GH 10 xylanases.

Hydrolysis of Methylglucuronoxylan

Crystal structure analysis of GH 10 xylanases from *Streptomyces olivaceoviridis* (Xyn10A) and *Cellvibrio mixtus* (Xyn10B) have provided molecular level determination of subsite interactions of the methylglucuronosyl moiety on the xylan chain (Fujimoto et al., 2004; Pell et

al., 2004b). The limit product, $MeGAX_3$, was cocrystallized with an active site mutant and structure analysis revealed binding of this hydrolysis product in the -3 through -1 subsites and +1 through +3 subsites. Binding of $MeGAX_3$ reflected enzyme substrate interactions, indicating the -3 and +1 subsites accommodate methylglucuronosyl substitutions. For *C. mixtus* Xyn10B, the -3 subsite methylglucuronosyl could not be modeled as electron density was diffuse, but for the same position in *S. olivaceoviridis* Xyn10A electron density was clear. In this position the O2 hydroxyl is solvent exposed and the substituted methylglucuronate is extended up into solvent. No interactions between the methylglucuronate and protein were identified to explain the clear electron density observed for Xyn10A. In the +1 subsite, the O2 position points into the protein. A pocket in this position accommodates O2 substituted methylglucuronosyl moieties. For *S. olivaceoviridis* Xyn10A, diffuse electron density did not allow modeling, indicating that the protein has minimal interactions with this carbohydrate residue but is structured to allow unrestricted access in this position. In the case of *C. mixtus* Xyn10B, clear electron density was observed for the methylglucuronosyl in this position. In Xyn10B, the +1 subsite has more xylose-binding interactions then other GH 10 xylanases, and while in the methylglucuronosyl pocket the glucuronate moiety is hydrogen bound to two separate amino acid residues. The additional stability in this position is used to explain the clear electron density for the methylglucuronate and significantly increased activity with respect to other xylanases on the polymeric substrate $MeGAX_n$ (Fujimoto et al., 2004; Pell et al., 2004b). Identification of the methylglucuronosyl pocket within the aglycone +1 subsite suggests that GH 10 xylanases may have evolved to address this specific O2 hydroxyl substitution.

Hydrolysis of Methylglucuronoarabinoxylan

GH 10 xylanase crystal structure analysis of Ara*f* substituted hydrolysis products, AX_2 and AX_3, did not identify conserved Ara*f* protein interactions. Results for Xyn10B of *C. mixtus* and

Xyn10A of *S. olivaceoviridis* were comparable. Xyn10B binding of AX_2 and AX_3 in the glycone subsites resulted in clear xylose modeling in subsites -1 through -2 and -1 through -3, respectively. In both cases the Ara*f* substitution yielded clear electron density. The Ara*f* of AX_2 had two alternative conformations. In one, Ara*f* hydrogen bonds with the protein and in the other, similar to the positioning of Ara*f* in AX_3, the O3 hydroxyl hydrogen bonds to the O5 endocyclic oxygen of the xylose in subsite -3 having no direct interaction with the protein. Xyn10A is similar to this, but electron density is not clear for Ara*f* of AX_2 in the -1 through -2 subsites. AX_3 however resulted in clear modeling of the Ara*f* moiety. In this case the O3 hydroxyl of Ara*f* hydrogen bonded with two separate positions within Xyn10B.

Interactions of AX_2 and AX_3 in the aglycone subsite region identified possible xylose subsite binding flexibility. Xyn10B of *C. mixtus* did not have clear electron density data for AX_2, but the xylotriose backbone of AX_3 modeled into subsite +1 through +3 as expected. For both enzymes, no Ara*f* moiety could be clearly modeled in the aglycone sites. For Xyn10C of *S. olivaceoviridis*, the oligomers only allowed modeling of xylobiose in subsite +1 through +2 with the third xylose of AX_3 not clear in subsite +3. Based on the modeling for xylose residues in the +1 and +2 subsites for both oligomers, it is assumed Ara*f* is positioned in these subsites for AX_2 and AX_3, respectively. In the case of AX_3, the +2 subsite xylose was slightly displaced from the binding subsite suggesting that Ara*f* was wedging into an awkward position. It is a good indication that this flexibility in arabinose accommodation is normal as Xyn10C was used to generate AX_3 as the major Ara*f* substituted hydrolysis product of wheat arabinoxylan. AX_2 was produced by hydrolysis of the same with Xyn10B. Subsite binding of this oligomer into the expected aglycone subsites did not allow modeling. Further the authors identified restrictions of

the O3 hydroxyl of xylose in the +2 subsite of Xyn10B, suggesting that accommodation of AX_3 would be more difficult then in Xyn10C.

It is apparent that glycone subsites of both enzymes can accommodate O2 linked glucuronosyl in the -3 and an O3 linked Ara*f* in the -2 subsites. These substitutions occur as the O2 and O3 in these positions are solvent exposed. For aglycone subsites, O2 glucuronosyl substitutions in the +1 subsite are easily accommodated within a pocket. Ara*f* accommodation in this area of the catalytic cleft seems to be variable among xylanases. The differences between these two enzymes can be highlighted by the fact that Xyn10B of *C. mixtus* was used to produce AX_2 and Xyn10C of *S. olivaceoviridis* was used to produce AX_3. The latter, as positioned in the aglycone binding region of Xyn10C, revealed a flexibility of xylose binding in the +2 subsite which helps explain how the Ara*f* in this position is accommodated. Xyn10B was suggested not to have this flexibility for O3 linked Ara*f* in the +2 subsite, but based on hydrolysis product analysis must accommodate it in the +1 subsite.

Hydrolysis of Rhodymenan by GH 10 xylanases

Only one reported study has considered the hydrolytic products of GH 10 xylanases on substrates other than β-1,4-linked xylans. Rhodymenan, a β-1,3-1,4-linked xylan digested with the two GH 10 xylanases of *Cryptococcus albidus* and *Streptomyces lividans* discussed above, resulted in the hydrolysis limit product xylosyl-β-1,3-xylosyl-β-1,4-xylose (Biely et al., 1997).

GH 10 Xylanase Substrate Binding Cleft Studies

An important expectation has recently been addressed, which changes the way we must consider synergy of methylglucuronoxylan hydrolysis between different families of xylanases. This expectation was that the smallest methyglucuronate substituted hydrolysis product released by a GH 11 xylanase, aldopentauronate ($MeGAX_4$), would be further hydrolyzed by a GH 10 xylanase with release of xylose and generation of aldotetrauronate. $MeGAX_4$ is substituted

penultimate to the nonreducing terminal xylose of xylotetraose and this methylglucuronosyl substitution would be expected to guide the substrate into the +1 subsite where the methylglucuronosyl can be accommodated. The additional xylose would then be expected to lie across the active site residues and hydrolysis would release xylose. In this interesting study, four different GH 10 xylanases did not have activity on this substrate (Kolenova et al., 2006). However, hydrolysis of the substrate aldohexauronate, in which the methylglucuronosyl moiety is substituted on the middle xylose of xylopentaose, resulted in release of xylobiose. This study may have identified a substrate requirement for GH 10 xylanases. The inability of these GH 10 xylanases to use $MeGAX_4$ as substrate but use $MeGAX_5$, indicates that binding of xylose to the -1 subsite does not occur with a single xylose (Kolenova et al., 2006). This reflects strong binding of xylose at the -2 subsite compared to binding at the -1 subsite.

In a similar study, the GH 10 xylanase of *T. aurantiacus* (Xyn10) was shown to use an O3 Ara*f* substitution in the -2 subsite as a substrate specificity determinant (Vardakou et al., 2005). It was determined that multiple interactions between the Ara*f* moiety and amino acids in the enzyme stabilized the interaction with this substituted substrate more than with unsubstituted substrate. The purpose of this interaction was validated with comparison of activity on xylotriose and AX_3 in which the arabinose was substituted O3 on the nonreducing terminal xylose. The results showed a four-fold higher activity on the Ara*f* substituted substrate.

An interesting comparison between the above result and the previous discussion, considering hydrolysis of $MeGAX_4$, is that a single xylose extending across the active site from the glycone region was hydrolyzed, suggesting that the +1 subsite binds xylose without additional interactions into the +2 subsite. In the work described for GH 10-catalyzed hydrolysis of $MeGAX_4$, a single xylose residue extending across the active site into the glycone region was

not hydrolyzed, but the larger xylobiose was hydrolyzed. It seems possible that the methylglucuronosyl substitution in $MeGAX_4$ may limit the binding of xylose in the -1 subsite.

These findings suggest that GH 10 and GH 11 xylanases may not function synergistically. Rather, GH 11 xylanases may hinder the full potential of a GH 10 xylanase. Identifying how decorated substrates interact with the catalytic cleft of GH 10 xylanases is important. This knowledge can be used to develop enzyme mixtures for efficient hydrolysis of target biomass substrates. Studies to determine the functional properties of GH 10 xylan binding subsites will also help to develop synthetic xylanases with engineered characteristics.

Phylogenetic Relationships of Glycosyl Hydrolase Family 10 Xylanases

Phylogenetic analysis of 241 GH 10 CD sequences is presented in Figure 3-3. A complete list of compared sequences and their attributes is presented in Table 3-2. The tree identifies three major branches of divergence (A, B and C). The first major branch, A, diverges to plants (A_1) and a bacterial clade (A_2). The bacteria in this clade closely associate with the B branch bacterial clade which contains most members of the phytopathogenic genus *Xanthomonas*. This large bacterial clade is made larger, grouping close to the C_1 bacterial subgroup which diverges from the major C branch. The C_1 clade contains interesting bacterial genera such as *Rhizobium*, *Agrobacterium*, *Synechococcus*, *Anabaena* and *Nostoc*. The divergent C_{2a} branch leads to most fungal sequences. It diverges into two major fungal clades one of which splits into a *Streptomyces* clade. This association is of notable interest as filamentous prokaryotic *Streptomyces* spp. have similar cell structure and morphological stages as some fungi. The third fungal group is composed entirely of *Fusarium* which branches separately from other fungi. Sequences 1-59 are comprised of almost entirely bacterial species. Many of these sequences cluster by bacterial genus or enzyme characteristics, such as the associated modular architecture.

For instance, sequences 22 through 44 are all highly modular enzymes consisting of similar modular type and architecture.

Plant and Related Bacterial GH 10 Xylanase.

The phylogenetic tree highlights associations between GH 10 xylanases of plants and bacteria which can only be attributed to close evolutionary origin or interaction. Bacterial sequences from 189 through 206 which originate partially from branches B and all of A_2 are closest to plant sequences but represent such a diverse assemblage of genera that it is difficult to draw conclusions. However, the remaining sequences in branch B (177 through 188) and those in the closely associated C_1 clade contain bacteria with clear similarities or associations to plants. These represent sequences from the phytopathogenic genera *Xanthomonas* and *Argobacterium* and the well studied plant pathogen *Pseudomonas syringae*. Also included are three different genera of cyanobacteria, the nonsulfur purple photosynthetic bacterium *Rhodopseudomonas palustric* and two rhizosphere nitrogen-fixing plant endosymbionts.

Similarities between these sequences may arise from common ancestry or common evolutionary ascendancy defined by prolonged plant-bacterial interaction. It would be expected that plant GH 10 xylanases have a role in expansion of the cell wall and may have the inherent capability for hydrolysis of highly substituted xylans. Plant pathogens probably would benefit from these same properties found in plant GH 10 xylanases as they are expected to perform a similar task in a similar environment.

Due to these possibilities, GH 10 xylanases from plant pathogens may have interesting characteristics when compared to the same from saprophytic microorganisms. Although the goal of each of these enzymes is considered to be the same, the substrate for saprophytes is not unaltered plant tissue but is rather decaying biomass, occurring through the function of saprophytic microbial consortia. The combined activities of many hydrolytic enzymes within

this environment may present a significantly altered substrate. These GH 10 xylanases may have evolved high turnovers on simplified substrate vs. others, having lower rates on complex substrates.

Fungal and *Streptomyces* Association

GH 10 xylanases from fungal origin are intriguing in that they seem significantly less complex then the modularly diverse bacterial xylanases. Of the two major fungal clades, the one containing sequences 127 through 162 have seven sequences with an appended CBM 1 module, six having this module in the N-terminal domain. The other clade, containing fungal sequences 97 through 113 (17 sequences) contains seven which have a CBM 1 module in the C-terminal. From this, it seems that the difference between the two clades including the positioning of the CBM 1 module is reflected within the sequence of the CD. A clade for the genus *Streptomyces* intervenes between the two fungal groupings. Every sequence in this clade has a C-terminal CBM module. Most have CBM 13 modules (β-trefoil), but there are four CBM 2 modules, two found with *Cellulomonas fimi* sequences in this clade. These are the only two which are not *Streptomyces* spp.

Bacterial GH 10 Xylanases: Tools to Work With

Sequences 1 through 79 and also 89 through 96 are primarily bacterial. Approximately 52% of these sequences contain accessory modules. Several small clades come directly off the C_2 branch but most branch from the C_{2b} (Fig. 3-3). In this subgroup, sequences 1 through 9, 14 through 21 and 45 through 59 consist of CDs only. Of these, sequences 1 through 9 do not contain detectable secretion signal sequences and are therefore considered intracellular. Sequences 10 through 13 and 22 through 44 are all modular, many showing a common modular architecture consisting of a CBM 22 and CBM 9 appended to the CD. Several of these that

group closely, including *Paenibacillus* sp. strain JDR-2, also have SLH modules involved with cell surface anchoring.

Conclusion

This review emphasizes the wide distribution and significant diversity of glycosyl hydrolases of family 10 xylanases. The array of accessory modules often found associated with GH 10 xylanases highlights possible functional variability and suggests that directed effort to develop xylanases to facilitate preprocessing may benefit from inclusion of these modules. Substrate binding studies of GH 10 xylanases have revealed the details describing the interaction of the GH 10 catalytic cleft with substitutions on the methylglucuronoxylan chain. For subsites which bind xylose orienting the O2 or O3 hydroxyls into the protein, substitutions can only be accommodated by open secondary structure such as the existence of a pocket as in the case of O2 substituted methylglucuronate in the +1 subsite, or by subsite flexibility as suggested for binding of AX_3 in the +2 subsite. Aglycone substrate binding accepts methyglucuronate in the -3 subsite and O3 substituted Ara*f* in the -2 subsite. These positions are solvent exposed and generally display little to no interactions between the protein and the substrate appendage. The product variability found for hydrolysis of methyglucuronoarabinoxylan suggest that minor amino acid changes within the xylan binding cleft may contribute to to large differences in hydrolysis product profiles. Even though the catalytic cleft is well conserved, differences in the ability of GH 10 xylanases can occur as a result of appended accessory modules and variations in the catalytic cleft.

Phylogenetic tree analysis identified interesting associations between the GH 10 xylanases from different organisms. Although the role of GH 10 xylanases has not been determined in plants, the number of available sequences from *Arabidopsis* and other plant genera suggests that they may have some function in cell wall alteration. Proximity in the phylogenetic tree identifies

a close similarity of GH 10 xylanases from plants and several plant pathogens and photosynthetic organisms. Other than the distribution of the CBM 22 modules which are significant in plant sequences, and the CBM 1 module that is strictly fungal, all other modules and the CBM 22 module are found only in bacterial GH 10 xylanases. This again, highlights the significant diversity available to accentuate GH 10 catalytic abilities and identifies bacterial GH 10 xylanases as a significant biotechnology resource for bioengineering and development of next-generation bacterial biocatalysts.

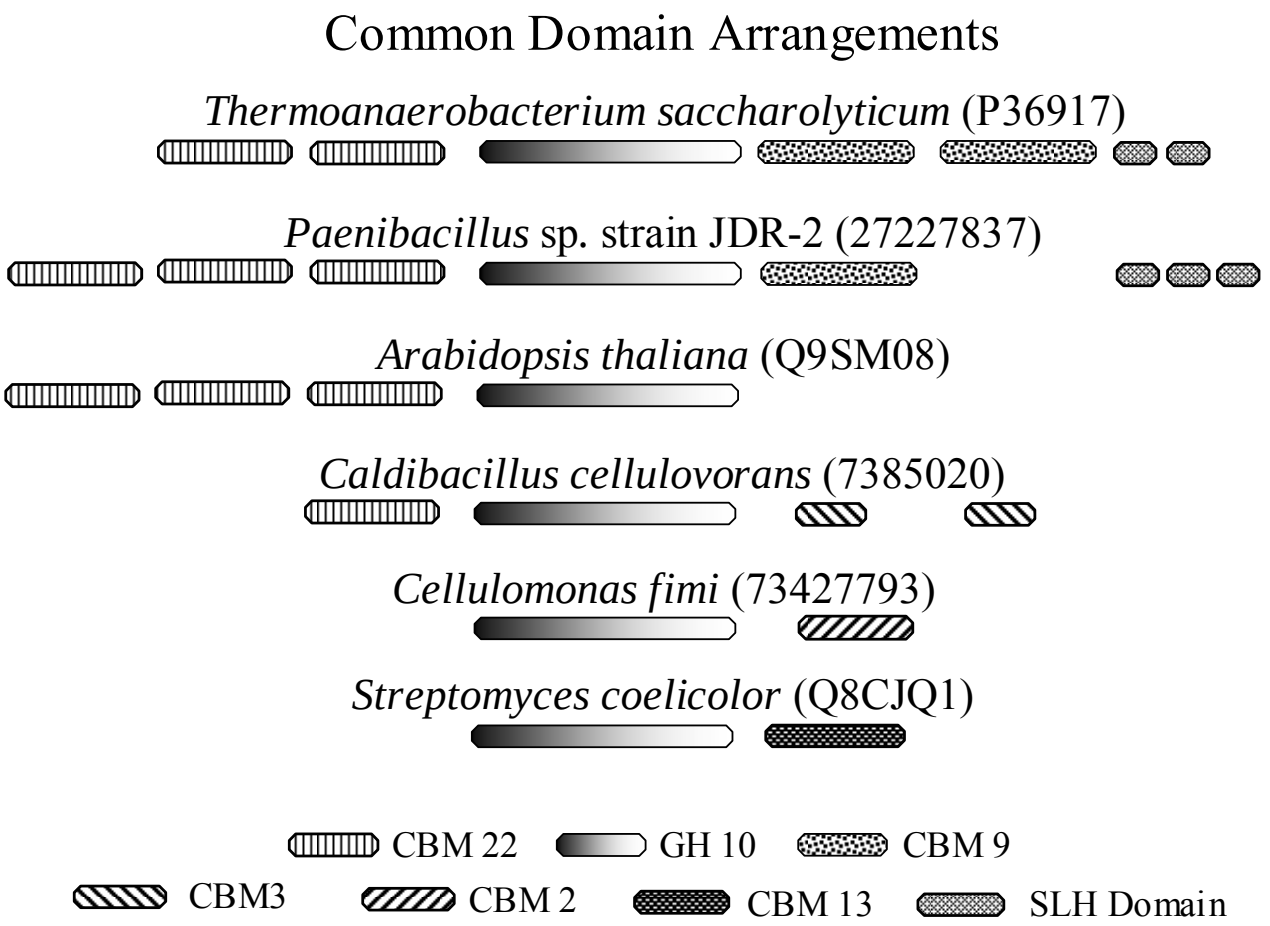

Figure 3-1. Common domain arrangements found in GH 10 xylanases.

Table 3-1. Distribution by bacterial genus of carbohydrate binding modules and other functional domains associated with GH 10 xylanases.

Genus	Family of Carbohydrate Binding Module										Linking Sequence					SLH	Other Domain
	2	3	4	5	6	9	10	13	15	22	Nr	Sr	PEr	PTr	PGr		
Aeromonas						●				●							
Anaerocellum										●							
Bacillus						●				●							
Caldibacillus		●								●				●			
Caldicellulosiruptor		●			●					●				●			●[a]
Cellulomonas	●			●		●				●				●			●[b]
Cellvibrio	●				●		●		●			●					
Clostridium					●	●				●						●	●[c]
Colwellia													●				●[d]
Cytophaga						●				●							
Eubacterium						●				●							
Fibrobacter					●												
Nonomuraea								●									
Paenibacillus						●				●						●	
Prevotella										●							
Pseudomonas												●	●				
Rhodothermus			●														
Ruminococcus											●						●[e]
Saccharophagus	●				●					●		●	●				●[f]
Streptomyces	●							●									●[g]
Thermoanaerobacterium						●				●						●	
Thermobifida	●														●		
Thermotoga						●				●							

[a] GH 5 cellulase module and a truncated GH 43 module. [b] Chitin binding module and a deacetylation module. [c] Esterase module. [d] Cadherin repeat module and *Salmonella* repeat of unknown function. [e] GH 11 module. [f] Chitin binding module and a truncated GH 43 module. [g] GH 62 module and a chitinase module.

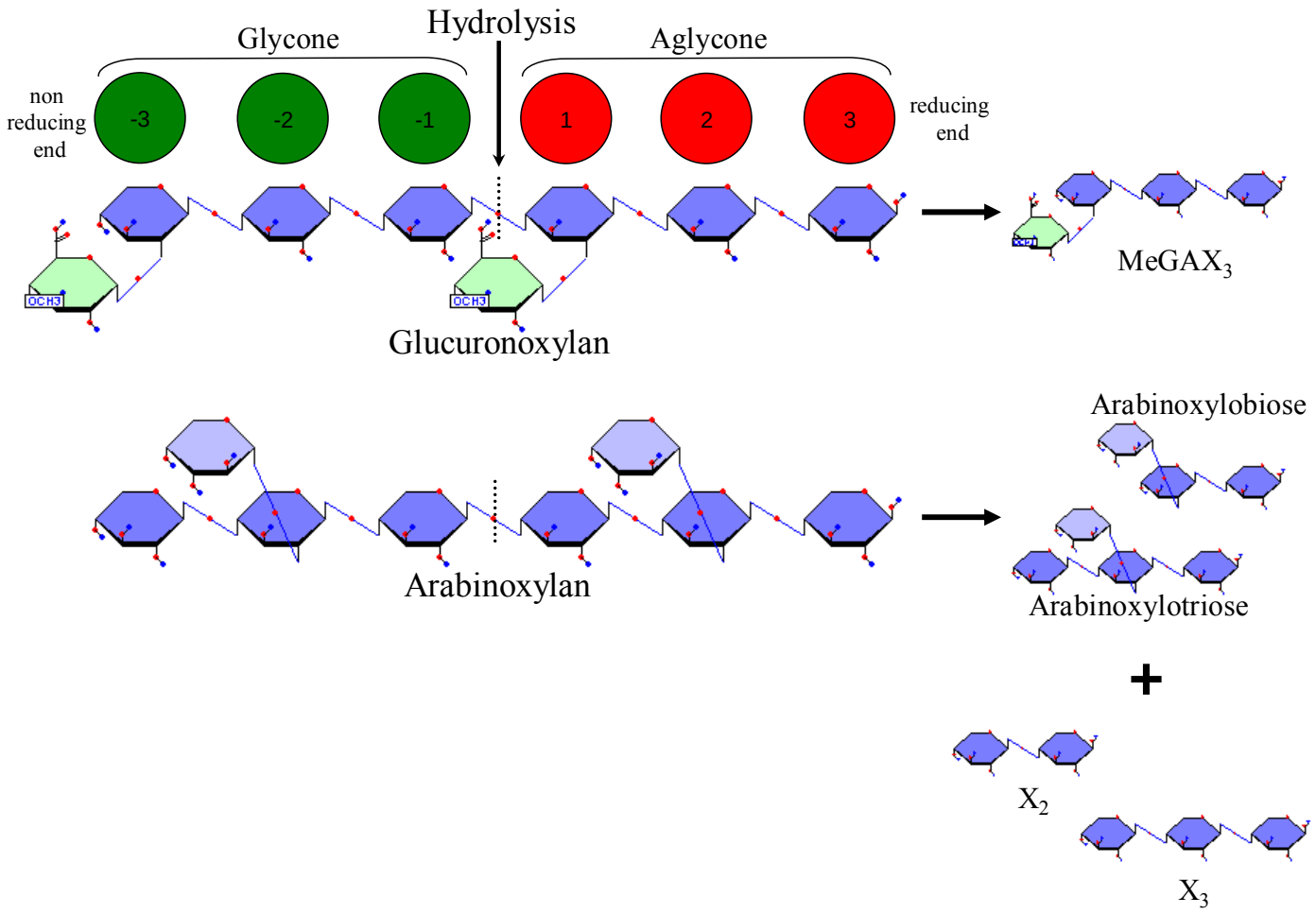

Figure 3-2. Products formed by the hydrolysis of methylglucuronoxylan and methyglucuronoarabinoxylan by a glycosyl hydrolase family 10 xylanase. Substituted hydrolysis limit products are determined by the interaction between the substitutions and binding subsites.

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties

	Organism	Accession Number	Module Architecture	Secreted
	Caldicellulosiruptor saccharolyticus	144299	CD	No
	Caulobacter crescentus CB15	16127272	CD	No
	thermophilic anaerobe NA10	O24820	CD/PTr/CBM3/PTr/Cel5	No
	Ampullaria crossean	66474472	CD	No
	Eucalyptus globulus subsp. *globulus*	88659658	CD	No
	Hordeum vulgare subsp. *vulgare*	Q8GZB5	CD	No
1	*Aeromonas punctata*	61287936	CD	No
2	*Bacillus* sp. BP-23	3201483	CD	No
3	uncultured bacterium	Q7X3W7	CD	No
4	*Geobacillus stearothermophilus*	499714	CD	No
5	*Geobacillus stearothermophilus*	73332107	CD	No
6	*Bacillus alcalophilus*	37694736	CD	No
7	*Bacillus* sp.	662884	CD	No
8	*Thermobacillus xylanilyticus*	O69261	CD	No
9	*Butyrivibrio fibrisolvens*	48963	CD/tAes	No
10	*Thermotoga* sp. strain FjSS3-B.1	Q9WWJ9	CBM22(2)/CD/CBM9(2)	Yes
11	*Thermotoga maritima*	Q60037	CBM22(2)/CD/CBM9(2)	Yes
12	*Thermotoga neapolitana*	Q60042	CBM22(2)/CD/CBM9(2)	Yes
13	*Thermotoga* sp. strain FjSS3-B.1	Q9R6T4	CBM22(2)/CD/CBM9(2)	Yes
14	*Clostridium stercorarium*	216419	CD	Yes
15	*Clostridium stercorarium*	23304849	CD	Yes
16	*Geobacillus stearothermophilus*	P40943	CD	Yes
17	*Bacillus* sp. NG-27	2429332	CD	Yes
18	*Bacillus firmus*	34978678	CD	Yes
19	*Bacillus halodurans*	56567273	CD	Yes
20	*Bacillus* sp.	216371	CD	Yes

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
21	*Bacillus halodurans*	22597186	CD	Yes
22	unidentified	39749821	tCBM22/CD/CBM9t	No
23	*Caldicellulosiruptor saccharolyticus*	2645425	CBM22(2)/CD	Yes
24	*Caldicellulosiruptor* sp. Rt8B.4	P40944	CBM22(2)/CD	Yes
25	*Anaerocellum thermophilum*	1208895	CBM22(2)/CD	Yes
26	*Caldicellulosiruptor saccharolyticus*	2645417	CBM22(2)/CD	Yes
27	*Caldicellulosiruptor saccharolyticus*	40646	CD/PTr/CBM3/PTr/Cel5	Yes
28	*Caldicellulosiruptor* sp. Tok7B.1	4836168	CD/PTr/CBM3/PTr/CBM3(2)/PTr/Cel5	Yes
29	unidentified	39749823	CD	No
30	*Caldicellulosiruptor* sp. Tok7B.1	4836167	CBM22(2)/CD/PTr/CBM3(2)/PTr/CBM3/PTr/GH43t/CBM6	Yes
31	*Aeromonas punctata*	3810965	CBM22/CD/CBM9(2)	Yes
32	*Bacillus* sp. BP-23	3201481	CBM22(2)/CD/CBM9(2)	Yes
33	*Cellulomonas fimi*	1103639	Deac/CBM22/CD/CBM9	No
34	*Cellulomonas pachnodae*	5880612	CBM22(2)/CD/CBM9/CBM5/ChtBD3	ND
35	*Clostridium thermocellum*	144776	CBM22/CD/CBM9(2)/SLH(3)	Yes
36	*Thermoanaerobacterium thermosulfurigenes*	Q60046	CBM22(2)/CD/CBM9(2)/SLH(3)	Yes
37	*Thermoanaerobacterium saccharolyticum*	P36917	CBM22(2)/CD/CBM9(2)/SLH(2)	Yes
38	*Thermoanaerobacterium* sp.strain JW/SL-YS 485	Q60043	CBM22(2)/CD/CBM9(2)/SLH(3)	Yes
39	*Clostridium stercorarium*	5360744	CBM22/CD/CBM9	Yes
40	*Clostridium stercorarium*	23304851	CBM22/CD/CBM9	Yes
41	*Clostridium josui*	12225048	CBM22/CD/CBM9/tSLH	Yes
42	*Paenibacillus* sp. JDR-2	62990090	CBM22(3)/CD/CBM9/SLH(3)	Yes
43	*Paenibacillus* sp. W-61	27227837	CBM22(2)/CD/CBM9/SLH(2)	Yes
44	*Caldibacillus cellulovorans*	7385020	CBM22/CD/PTr/CBM3/PTr/CBM3	Yes
45	*Pseudoalteromonas atlantica* T6c	109701250	CD	Yes
46	*Prevotella ruminicola*	P48789	CD	Yes
47	*Xanthomonas axonopodis* pv. *citri* str. 306	21110686	CD	Yes

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
48	*Xanthomonas campestris* pv. *vesicatoria* str. 85-10	78038341	CD	No
49	*Xanthomonas campestris* pv. *campestris* str. 8004	66575835	CD	Yes
50	*Xanthomonas campestris* pv. *campestris* str.	21115393	CD	Yes
51	*Bacteroides ovatus*	450852	CD	Yes
52	uncultured bacterium	56709936	CD	Yes
53	*Flavobacterium* sp. MSY2	68525474	CD	ND
54	*Saccharophagus degradans* 2-40	89951878	CD	Yes
55	*Cellvibrio japonicus*	5690438	CD	Yes
56	*Cellvibrio mixtus*	37962277	CD	TAT
57	*Caldicellulosiruptor saccharolyticus*	2645419	CD	No
58	*Dictyoglomus thermophilum*	973983	CD	ND
59	uncultured bacterium	Q8VPE4	CD	Yes
60	*Clostridium cellulovorans*	47716661	CBM22/CD	Yes
61	*Clostridium thermocellum*	4850306	CBM22/CD	Yes
62	*Polyplastron multivesiculatum*	Q9U0G1	CBM22/CD	Yes
63	*Clostridium thermocellum*	P51584	CBM22/CD/CBM22/Est	Yes
64	*Ruminococcus flavefaciens*	P29126	GH11/Nr/CD	Yes
65	*Epidinium caudatum*	28569972	CD/CBM13	No
66	*Butyrivibrio fibrisolvens*	P23551	CD	Yes
67	*Eubacterium ruminantium*	974180	CBM22/CD/CBM9	No
68	*Cellvibrio japonicus*	38323070	Sr/CBM15/CD	ND
69	*Cellvibrio mixtus*	757809	Sr/CBM15/CD	ND
70	*Cellvibrio japonicus*	45520	CBM2/Sr/CBM10/Sr/CD	Yes
71	*Saccharophagus degradans* 2-40	89952176	CBM2/Sr(2)/CD	Yes
72	*Neocallimastix patriciarum*	Q02290	CD/XSKTLPGG(45)/CBM1	Yes
73	*Clostridium thermocellum*	P10478	Est/CBM6/CD	Yes
74	*Rhodopirellula baltica* SH 1	32446690	CD	No

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
75	*Thermotoga* sp.	Q60044	CD	No
76	*Thermotoga neapolitana*	Q60041	CD	No
77	*Thermotoga maritima*	Q9WXS5	CD	No
78	*Clavibacter michiganensis* subsp. *michiganensis*	31559721	CD	Yes
79	*Streptomyces turgidiscabies*	57338460	CD/ChNT	Yes
80	*Aspergillus nidulans* FGSC A4	40745311	CD	Yes
81	*Fusarium oxysporum*	Q8TGC2	CD	Yes
82	*Fusarium oxysporum*	Q8TGC3	CD	Yes
83	*Fusarium oxysporum* f. sp. *lycopersici*	O93976	CD	Yes
84	*Fusarium oxysporum*	Q8TGC4	CD	Yes
85	*Fusarium oxysporum*	19912845	CD	Yes
86	*Fusarium oxysporum*	19912843	CD	Yes
87	*Fusarium oxysporum*	21699819	CD	Yes
88	*Fusarium oxysporum*	19912853	CD	Yes
89	*Prevotella ruminicola*	P72234	CBM22/tCD	Yes
90	*Streptomyces avermitilis* MA-4680	29605742	CD	Yes
91	*Thermobifida fusca* YX	71916922	CD	Yes
92	*Streptomyces avermitilis* MA-4680	29608643	CD/CBM2	Yes
93	*Thermobifida alba*	P74912	CD/PGr/CBM2	Yes
94	*Thermobifida fusca* YX	71917054	CD/PGr/CBM2	Yes
95	*Colwellia psychrerythraea* 34H	71145740	CD	Yes
96	*Cryptococcus adeliensis*	O13436	CD	No
97	Patent 5693518	3015123	CD/CBM1	Yes
98	*Aspergillus nidulans* FGSC A4	40742582	CD/CBM1	Yes
99	*Aspergillus oryzae*	83775646	CD	Yes
100	*Penicillium funiculosum*	53747929	CD/CBM1	Yes
101	*Talaromyces emersonii*	21437253	CD/CBM1	Yes

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
102	*Magnaporthe grisea* 70-15	39973147	Bpht/CD	Yes
103	*Neurospora crassa* OR74A	32416834	CD	Yes
104	*Neurospora crassa* OR74A	32413873	CD	Yes
105	*Magnaporthe grisea*	24496243	CD	Yes
106	*Gibberella zeae*	50844266	CD	Yes
107	*Neurospora crassa* OR74A	32407695	CD/CBM1	Yes
108	*Humicola grisea*	P79046	CD/CBM1	Yes
109	*Magnaporthe grisea* 70-15	39963865	CD/CBM1	Yes
110	*Aureobasidium pullulans* var. *melanigenum*	84469404	CD	Yes
111	*Gibberella zeae*	56555501	CD	No
112	*Agaricus bisporus*	O60206	CD	Yes
113	*Phanerochaete chrysosporium*	Q9HEZ1	CBM1/CD	Yes
114	*Streptomyces coelicolor*	Q8CJQ1	CD/CBM13	Yes
115	*Streptomyces lividans*	P26514	CD/CBM13	Yes
116	*Streptomyces olivaceoviridis*	Q7SI98	CD/CBM13	No
117	*Streptomyces thermocyaneoviolaceus*	Q9RMM5	CD/CBM13	Yes
118	*Streptomyces thermoviolaceus*	38524461	CD/CBM13	Yes
119	*Streptomyces avermitilis*	Q9X584	CD/CBM13	Yes
120	Patent 6300114	34606109	CD/CBM13t	Yes
121	*Nonomuraea flexuosa*	Q8GMV6	CD/CBM13	Yes
122	*Cellulomonas fimi*	73427793	CD/PTr/CBM2	Yes
123	*Cellulomonas fimi*	144425	CD/PTr/CBM2	Yes
124	*Streptomyces chattanoogensis*	Q9X583	CD/CBM13/GH62	ND
125	*Streptomyces coelicolor*	Q9RJ91	CD/CBM2	Yes
126	*Streptomyces halstedii*	Q59922	CD/CBM2	Yes
127	*Alternaria alternata*	Q9UVP5	CBM1/CD	Yes
128	*Cochliobolus carbonum*	49066418	CD	Yes

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
129	*Claviceps purpurea*	O74717	CD	Yes
130	*Fusarium oxysporum*	P46239	CBM1/CD	Yes
131	*Fusarium oxysporum* f. sp. *lycopersici*	O59937	CBM1/CD	Yes
132	*Gibberella zeae*	50844270	CD	Yes
133	*Magnaporthe grisea*	22415585	CBM1/CD	Yes
134	*Coniothyrium minitans*	11876710	CBM1/CD	Yes
135	*Fusarium oxysporum* f. sp. *lycopersici*	O59938	CD	Yes
136	*Gibberella zeae*	50844272	CD	Yes
137	*Cryptovalsa* sp. BCC 7197	53636303	CD	Yes
138	*Neurospora crassa* OR74A	32410597	CD	Yes
139	*Hypocrea jecorina*	6705997	CD	Yes
140	*Magnaporthe grisea* 70-15	39951799	CD	Yes
141	*Magnaporthe grisea*	Q01176	CD	Yes
142	*Agaricus bisporus*	Q9HGX1	CD	No
143	*Volvariella volvacea*	Q7Z948	CBM1/CD	Yes
144	*Emericella nidulans*	95025700	CD	Yes
145	*Emericella nidulans*	Q00177	CD	Yes
146	*Aspergillus oryzae*	83772405	CD	Yes
147	*Thermoascus aurantiacus*	P23360	CD	Yes
148	*Aspergillus oryzae*	15823785	CD	Yes
149	*Aspergillus oryzae*	83766611	CD	Yes
150	*Aspergillus sojae*	Q9P955	CD	Yes
151	*Aspergillus terreus*	68161138	CD	Yes
152	*Aspergillus oryzae*	O94163	CD	Yes
153	*Aspergillus oryzae*	83775732	CD	Yes
154	*Penicillium canescens*	55792811	CD	Yes
155	*Penicillium simplicissimum*	P56588	CD	No

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
156	*Penicillium purpurogenum*	Q9P8J1	CD	Yes
157	*Aspergillus aculeatus*	O59859	CD	Yes
158	Patent 6197564	14480380	CD	Yes
159	*Aspergillus kawachii*	P33559	CD	Yes
160	*Penicillium chrysogenum*	46406032	CD	Yes
161	*Penicillium chrysogenum*	83416731	CD	Yes
162	*Penicillium chrysogenum*	P29417	CD	Yes
163	*Rhizobium etli* CFN 42	86282913	CD	Yes
164	*Rhizobium leguminosarum* bv. *trifolii*	88657052	CD	Yes
165	*Anabaena variabilis* ATCC 29413	75701321	CD	No
166	*Nostoc* sp. PCC 7120	Q8YNW3	CD	No
167	*Pseudomonas syringae* pv. *phaseolicola* 1448A	71555629	CD	Yes
168	*Pseudomonas syringae* pv. *syringae* B728a	63258442	CD	Yes
169	*Caulobacter crescentus* CB15	16127035	CD	TAT
170	*Synechococcus elongatus* PCC 7942	81169090	CD	Yes
171	*Synechococcus elongatus* PCC 6301	56685123	CD	Yes
172	*Acidobacterium capsulatum*	13591553	CD	No
173	*Agrobacterium tumefaciens* str. C58	17740854	CD	Yes
174	*Agrobacterium tumefaciens* str. C58	15157542	CD	Yes
175	*Bradyrhizobium japonicum* USDA 110	27377352	CD	TAT
176	*Rhodopseudomonas palustris* BisB18	90104203	CD	Yes
177	*Xanthomonas oryzae* pv. *oryzae* KACC10331	58428646	CD	No
178	*Xanthomonas oryzae* pv. *oryzae* MAFF 311018	84369769	CD	No
179	*Xanthomonas oryzae* pv. *oryzae*	Q9AM29	CD	No
180	*Xanthomonas campestris* pv. *vesicatoria* str. 85-10	78038346	CD	Yes
181	*Xanthomonas axonopodis* pv. *citri* str. 306	21110692	CD	Yes
182	*Xanthomonas campestris* pv. *campestris* str. 8004	66575838	CD	Yes

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
183	*Xanthomonas campestris* pv. *campestris* str.	21115397	CD	Yes
184	*Xanthomonas campestris* pv. *vesicatoria* str. 85-10	78038344	CD	Yes
185	*Xanthomonas oryzae* pv. *oryzae* KACC10331	58428645	CD	Yes
186	*Xanthomonas oryzae* pv. *oryzae*	12658424	CD	Yes
187	*Xanthomonas axonopodis* pv. *citri* str. 306	21110690	CD	No
188	*Xanthomonas oryzae* pv. *oryzae* MAFF 311018	84369768	CD	Yes
189	*Rhodothermus marinus*	P96988	CBM4(2)CD	Yes
190	*Fibrobacter succinogenes* S85	11526752	CD/CBM6	Yes
191	*Fibrobacter succinogenes* S85	9965987	CD/CBM6t	Yes
192	*Fibrobacter succinogenes* S85	9965986	CD/CBM6	Yes
193	*Cellvibrio japonicus*	45524	CBM2/Sr/CBM6/Sr/CD	Yes
194	*Saccharophagus degradans* 2-40	89952852	GH43t/CBM6/Sr/CBM2/Sr/CBM22/CD	Yes
195	*Pseudomonas* sp. ND137	57999823	Sr/CD	Yes
196	*Saccharophagus degradans* 2-40	89949430	CD/Sr(2)/ChtBD3	Yes
197	*Clostridium acetobutylicum* ATCC 824	15004819	CD	Yes
198	*Clostridium acetobutylicum* ATCC 824	15004757	CD	Yes
199	*Cytophaga hutchinsonii* ATCC 33406	110280325	CD/CBM22	Yes
200	*Cytophaga hutchinsonii* ATCC 33406	110281120	CD/CBM9	Yes
201	*Colwellia psychrerythraea* 34H	71145380	PEr/CD/Cad/DUF823	ND
202	*Colwellia psychrerythraea* 34H	71143508	CD	Yes
203	*Pseudomonas* sp. PE2	25137524	PEr/CD	No
204	*Saccharophagus degradans* 2-40	89949572	PEr/CD	ND
205	*Cytophaga hutchinsonii* ATCC 33406	110281182	CBM22/CD	Yes
206	*Rhodopirellula baltica* SH 1	32446276	CD	Yes
207	*Arabidopsis thaliana*	O81754	CBM22/CD	Yes
208	*Arabidopsis thaliana*	O81751	CBM22/CD	No
209	*Arabidopsis thaliana*	O81752	tCD/CBM22/CD	No
210	*Medicago truncatula*	92868656	CBM22/CD	Yes

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
211	*Triticum aestivum*	40363757	CD	Yes
212	*Oryza sativa*	19920133	CBM22/CD	No
213	*Oryza sativa* (japonica cultivar-group)	Q7XFF8	CD	Yes
214	*Zea mays*	Q9ZTB8	CD	No
215	*Carica papaya*	Q8GTJ2	CBM22/CD	Yes
216	*Arabidopsis thaliana*	Q9ZVK8	CD	No
217	*Arabidopsis thaliana*	O81897	CD	No
218	*Arabidopsis thaliana*	O82111	CD	No
219	*Arabidopsis thaliana*	Q9SZP3	CD	No
220	*Thermosynechococcus elongatus* BP-1	22295628	CD	No
221	*Bacillus pumilus*	20386142	CD	No
222	*Clostridium thermocellum*	37651955	CBM22/CD	Yes
223	*Ampullaria crossean*	Q7Z1V6	CD	No
224	*Oryza sativa* (japonica cultivar-group)	28411931	CBM22(4)/CD	No
225	*Nicotiana tabacum*	73624749	CBM22(3)/CD	No
226	*Nicotiana tabacum*	73624751	CBM22(3)/CD	No
227	*Populus tremula* x *Populus tremuloides*	60656567	CBM22(3)/CD	No
228	*Arabidopsis thaliana*	Q9SM08	CBM22(3)/CD	No
229	*Arabidopsis thaliana*	Q9SYE3	tCBM22/CD	No
230	*Arabidopsis thaliana*	O80596	CBM22(4)/CD	No
231	*Oryza sativa* (japonica cultivar-group)	29788834	CBM22/tCBM22/CD	No
232	*Oryza sativa* (japonica cultivar-group)	15528604	CBM22/CD	No
233	*Oryza sativa* (japonica cultivar-group)	55168219	CD	No
234	*Oryza sativa* (japonica cultivar-group)	15528602	CD	No
235	*Hordeum vulgare*	P93186	CD	No
236	*Hordeum vulgare* subsp. *vulgare*	71142590	CBM22/CD	No

Table 3-2. Glycosyl hydrolase family 10 sequences included in phylogenetic studies and some of their properties. Continued.

	Organism	Accession Number	Module Architecture	Secreted
237	*Hordeum vulgare*	71142588	CBM22/CD	No
238	*Triticum aestivum* (bread wheat)	Q9XGT8	CD	Yes
239	*Hordeum vulgare*	P93185	CD	No
240	*Hordeum vulgare*	14861199	CBM22/CD	No
241	*Hordeum vulgare* subsp. *vulgare*	71142586	CBM22/CD	No

CD refers to a GH 10 catalytic module, Aes refers to esterase/lipase module, Cel5 refers to a GH 5 cellulase module, Deac refers to deacetylase domain, ChtBD3 refers to chitin-binding domain, Est refers to esterase, Cad refers to Cadherin repeat domain, DUF823 refers to *Salmonella* repeat of unknown function, Nr refers to asparagine rich domain, Sr refers to serine rich domain, PEr refers to proline glutamate rich domain, PTr refers to proline threonine rich domain, PGr refers to proline glycine rich domain, XSKTLPGG refers to unique *Neocallimastix* linker, GH 62 refers to arabinofuranosidase domain, ChNT refers to chitinase N-terminal domain, Bph refers to bacterial phosphatase, t refers to a predicted truncation and is positioned to the side of the truncation.

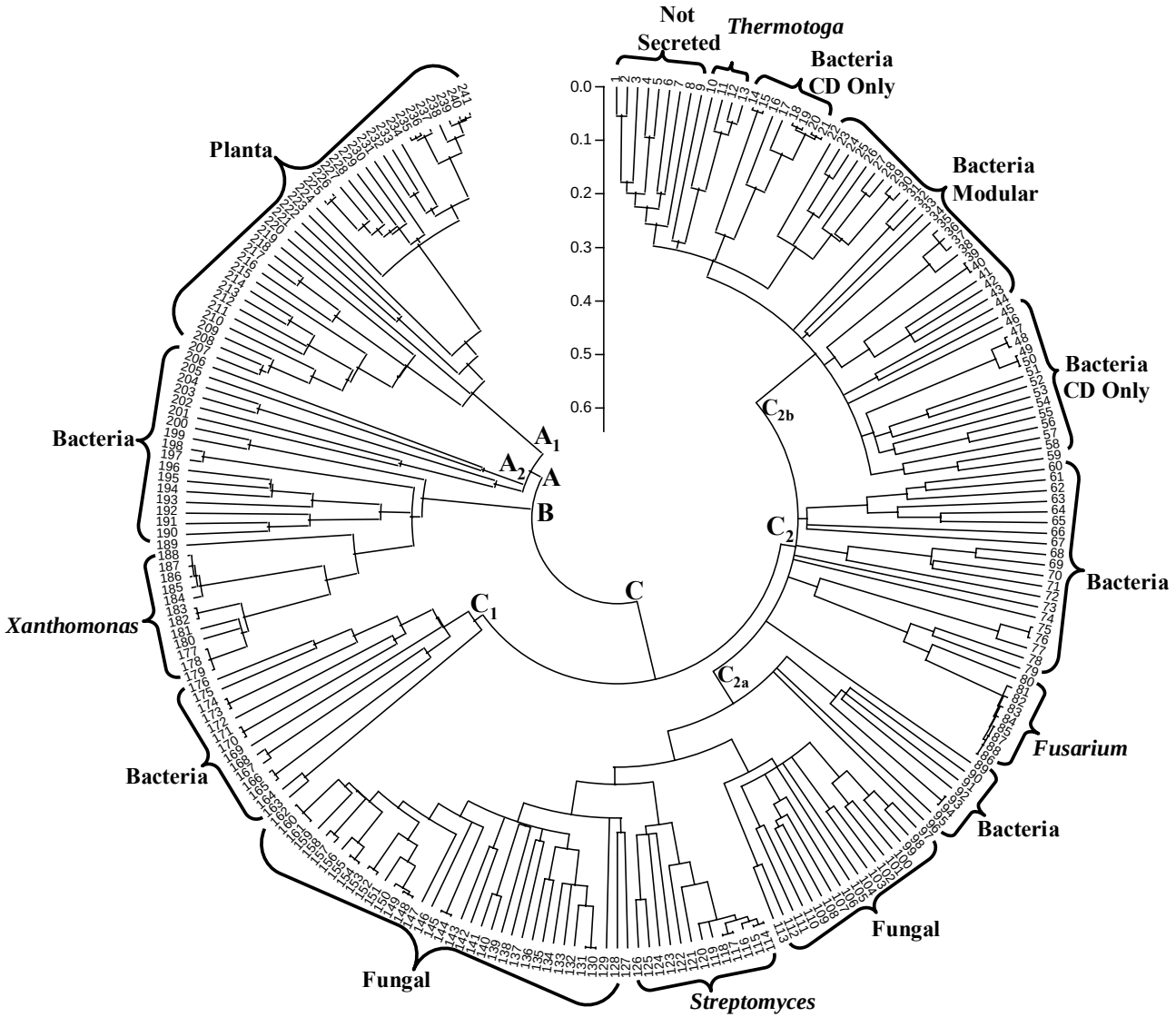

Figure 3-3. Phylogenetic distribution of catalytic domains of glycosyl hydrolase family 10 xylanases. Numbering corresponds to Table 3-2. Full xylanase sequences were aligned in MEGA 3.1 and the highly conserved catalytic domain was trimmed out. These sequences were realigned and used to generate a Neighbor-joining Bootstrap phylogenetic tree.

CHAPTER 4
Paenibacillus SPECIES STRAIN JDR-2 AND $XynA_1$: A NOVEL SYSTEM FOR GLUCURONOXYLAN UTILIZATION

Introduction

A version of this chapter has previously been published as a peer reviewed manuscript in the Febuary 2006 issue of Applied and Environmental Microbiology.

Increasing cost and demand of fossil fuels highlights the need to develop efficient methods to utilize renewable resources for conversion to alternative energy sources such as fuel ethanol (Aldhous, 2005; Sun and Cheng, 2002). Supplementing the energy infrastructure with ethanol may help to shift economic dependence from petroleum-based energy. Microbial biocatalysts, both yeast and bacteria, have been developed for the conversion of glucose derived from cellulose and pentoses from hemicellulose to ethanol (Dien et al., 2003; Ingram et al., 1999; Jeffries and Jin, 2004; Jin et al., 2004), and similar approaches with bacteria have been successfully applied to the formation of value-added products such as optically pure lactic acid (Dien et al., 2002; Zhou et al., 2003a; Zhou et al., 2003b). Current research efforts are directed at improving the pretreatment processes to maximize the release of fermentable pentoses as well as glucose, and also further development of the biocatalysts for specific fermentations of these sugars. The adoption of microbial strategies for the efficient depolymerization and assimilation of hemicellulose derived carbohydrates offer promise for maximizing the conversion of the hemicellulose fraction of lignocellulosic biomass to alternative fuels and biobased products (Preston et al., 2003).

Hemicellulose represents 20% to 30% of lignocellulosic biomass and methylglucuronoxylan ($MeGAX_n$) is the predominant form of hemicellulose found in hardwood and crop residues (Preston et al., 2003; Singh et al., 2003). This polymer consists of β-1,4 linked xylan in which 10% to 20% of the xylose residues are periodically substituted with α-1,2-4-O-

methylglucuronic acid (MeGA) moieties. Complete enzymatic hydrolysis of $MeGAX_n$ requires the combined action of several families of glycosyl hydrolases, including β-1,4-endoxylanase, α-1,2-4-O-methylglucuronidase and β-1,4-xylosidase (Preston et al., 2003). Secreted microbial xylanases that catalyze the depolymerization of $MeGAX_n$ are primarily represented by two families of glycosyl hydrolase, GH 10 and GH 11, based on sequence similarity and hydrophobic cluster analysis (http://afmb.cnrs-mrs.fr/CAZY), (Gilkes et al., 1991; Henrissat and Bairoch, 1993; Henrissat and Davies, 1997).

In bacteria capable of utilizing $MeGAX_n$, the metabolism of the aldouronates generated by enzyme-catalyzed depolymerization is dependent on their assimilation and cleavage of the MeGA substitution. Most substrate and structural studies of α-glucuronidases, the enzymes required to initiate complete degradation of MeGA substituted xylooligosaccharides, have clearly established that only aldouronates in which MeGA is linked to the nonreducing terminal xylose are suitable substrates (Nagy et al., 2002; Nurizzo et al., 2002). This distinguishes the role of GH 10 xylanases from GH 11 xylanases in generating products for direct assimilation and metabolism. This argument is further supported by evidence that aldotetrauronate acts as a catabolic signaling molecule for its further metabolism (Shulami et al., 1999). Studies of the glucuronic acid utilization gene cluster of *Geobacillus stearothermophilus* have identified a putative $MeGAX_3$ transporter in an operon composed of genes involved with the degradative and catabolic processing of glucuronoxylan. The *uxuR* gene product, a DNA binding protein, was found to be a self-regulating element of this operon that acts to repress transcription. Binding of $MeGAX_3$ by UxuR alleviates repression. From this, it appears that GH 10 xylanases play a prominent role, both directly and indirectly, in processing of $MeGAX_n$ for its complete catabolism. There is no evidence to support a similar role for the GH 11 xylanases. It is possible

that GH 11 xylanases act to hydrolyze polymeric xylan primarily into shorter fragments that can then be further acted upon by GH 10 xylanases and β-xylosidases. (Pell et al., 2004a).

Another factor affecting the efficiency of metabolism is the localization of the xylanase relative to the cell. The cellulosome, found primarily in anaerobic *Clostridium* spp. and some ruminant microorganisms, and the xylanosome from aerobic soil bacteria, often have associated GH 10 xylanases (Bayer et al., 2004; Doi et al., 2003; Jiang et al., 2004). These extracellular surface anchored complexes often display a variety of enzymes from several glycosyl hydrolase families with diverse functions. *Clostridium thermocellum* has a well described cellulosome with twenty-six glycosyl hydrolases (62% cellulases, 23% xylanases, 15% other) and several associated esterase activities that contribute to hydrolysis of the lignocellulose complex (Doi and Kosugi, 2004). Localization of this complex to the surface of the organism presumably allows efficient utilization of hydrolysis products, which may provide a competitive advantage in an anaerobic niche. Extracellular GH 10 xylanases may also occur as large multi-modular surface-anchored enzymes separate from other glycosyl hydrolases. Representative GH 10 xylanases from *Clostridium*, *Thermoanaerobacterium*, *Caldicellulosirupter, Thermotoga, Promicromonospora, Paenibacillus* and several other genera have been shown to have similar modular architectures.

Here we describe the properties of an extra-cellular multidomain endoxylanase from an aggressively xylanolytic *Paenibacillus* sp. (strain JDR-2). The association of $XynA_1$ with cell wall preparations indicates an anchoring role for the SLH domains near the C-terminus of the 155 kDa enzyme. The marked preference of this organism for polymeric $MeGAX_n$ as a growth substrate compared to xylose or the aldouronates generated by the action of the GH 10

endoxylanase, supports a role for this enzyme in the vectoral processing of $MeGAX_n$ for subsequent transport and metabolism.

Materials and Methods

Isolation and identification of *Paenibacillus* sp. strain JDR-2. *Paenibacillus* sp. strain JDR-2 was isolated from fresh cut discs (5 cm diameter by 2-4 mm thick) of sweetgum stem wood (*Liquidamber styraciflua*) incubated about one inch below the soil surface in a sweetgum stand for approximately three weeks. Discs were suspended in 50 ml sterile deionized water and sonicated in a 125 Watt Branson Ultrasonic Cleaner water bath for 10 min. The sonicate was inoculated into 0.2% (w/v) sweetgum (SG) $MeGAX_n$ containing the mineral salts of Zucker and Hankin (Zucker and Hankin, 1970) at pH 7.4 and incubated at 30 °C. The SG $MeGAX_n$ was prepared and characterized by ^{13}C-NMR as described previously (Hurlbert and Preston, 2001; Jones et al., 1961; Kardosova et al., 1998). Isolated colonies were passed several times in $MeGAX_1$ broths and agars until pure. A culture growing on 0.2% $MeGAX_n$ Z-H medium was cryostored by mixing 0.5 ml of exponentially growing culture with 0.5 ml 50% (v/v) sterile glycerol and freezing at -70 °C. The purified isolate was submitted to MIDI Labs (http://www.midilabs.com) for partial 16s rRNA sequencing. The organism was identified as *Paenibacillus* sp. with 96% identity to *Paenibacillus granivorans* by blastn submission of 530 nucleotides of sequenced 16s rRNA. The organism has been deposited with the Bacillus Genetic Stock Center (BGSC), (http://www.bgsc.org).

Growth studies. A common protocol was applied in the maintenance and analysis of *Paenibacillus* sp. strain JDR-2 cultures. Each time a culture was prepared for study, a sample from the cryostored stock culture was transferred into 4 ml of 0.5% SG $MeGAX_n$ Z-H medium in 16 x 100 mm test tubes. After 36 to 48 hours of growth the culture was plated on agar medium containing 1.0% yeast extract (YE) and 0.5% oat spelt xylan in Z-H and grown for 36 to

48 hours until appropriately sized colonies were observed. In a slight deviation, inoculum from the cryostored culture was plated directly onto the agar medium and grown for 48 to 72 hrs before picking an isolated colony. All colonies regularly displayed the expected phenotype, i.e. a clearing zone on the opaque oat spelt xylan background with the expected colony morphology. For the various growth studies described below a single colony was inoculated into medium specified for the particular experiment. All growth was performed at 30°C.

Growth optimization studies were performed aerobically in 16 x 100 mm test tubes containing 4 ml volumes of medium and optical densities of cultures were measured at 600 nm with a Beckman DU500 series spectrophotometer with a 16 x 100 mm test tube holder. Individual 4 ml cultures for study were inoculated with 200 µl (5% volume) of an exponentially growing culture (4 ml medium of 1.0% YE in Z-H). For these test tube cultures, agitation was achieved by setting a test tube rack in a large flask holder on a New Brunswick G-2 gyrotory shaker at an angle of approximately 45°. Under these conditions, rotation at 200 rpm yielded the best agitation when compared to simple rotation.

Studies comparing *Paenibacillus* sp. strain JDR-2 utilization of $MeGAX_n$, with or without xylose or glucose as co-substrates, were performed in 125 ml baffle flasks with shaking at 150 rpm on a G-2 gyrotory shaker. Cultures were initiated by the addition of 4 ml (8% volume) of Z-H mineral salts washed cells from an overnight culture (25 ml) of 1.0% YE Z-H medium. Growth was monitored using an HP Diode Array spectrophotometer at 600 nm in a 1.00 cm cuvette. For these cultures, sample dilutions were performed to obtain OD 600 nm readings between 0.2 and 0.8 absorbance units and the resulting value was corrected by the dilution factor. Culture aliquots were centrifuged, supernatants filtered and carbohydrate utilization was measured by HPLC using a complete modular Waters chromatography system comprised of a

600 controller, 610 solvent delivery unit, 2410 RI detector and a 710B WISP automated injector. Carbohydrate separation was achieved with a Bio-Rad HPX-87H column running in 0.01 N H_2SO_4 with a flow rate of 0.8 ml/min at 65 ° C. Data analysis was performed using Waters Millennium Software.

The differential utilization of $MeGAX_n$ and $XynA_1$ CD generated products from $MeGAX_n$ as growth substrates by the organism was evaluated by the initiation of 50 ml cultures with 4 ml (8% volume) of Z-H washed cells from 25 ml overnight cultures in 0.5% SG $MeGAX_n$ Z-H medium. Growth was monitored as described above and aliquots examined by TLC (see procedure below).

DNA cloning, sequencing and analysis. A genomic library of *Paenibacillus* sp. strain JDR-2 DNA, prepared in pUC18 with gel purified 6-9 kb fragments obtained from a partial *Sau3AI* digest, was kindly provided by Ms. Loraine Yomano from the laboratory of Professor Lonnie Ingram. All cloning and general DNA manipulation methods originate from Molecular Cloning: A Laboratory Manual (Sambrook et al., 1989). In addition, DNA purification and gel extraction was performed using kits purchased from Qiagen (Valencia, California). Cloning analysis, planning and image preparation was performed with Clone Manager 6 and Enhance (Scientific and Educational Software, Cary, NC). Analysis of sequences for regulatory elements was conducted using the online tools available through Softberry (http://www.softberry.com/berry.phtml). The pUC18-based 6-9 kb library was transformed into *E. coli* DH5α and screened for xylanase positive clones by plating transformed cells on Remazol Brilliant Blue xylan plates and observing agar clearing after 24 hours (Braun and Rodrigues, 1993). Sequencing of cloned DNA was done in house by subcloning the insert into smaller sizes and using pUC18 M13 priming sites for sequencing of both strands. Primer walking at the ICBR

Genome Sequencing Services Laboratory at the University of Florida filled in gaps and completed 2x coverage. All sequencing employed the Sanger dideoxy chain termination method. The final sequence was assembled using the CAP3 sequence assembly program (Huang and Madan, 1999) located on the Pôle Bio-Informatique Lyonnais server (http://pbil.univ-lyon1.fr/). Sequence analysis was performed with online resources available through the NCBI (http://www.ncbi.nlm.nih.gov) and BCM (http://searchlauncher.bcm.tmc.edu) websites. The main tools employed were BLAST and CD-Search of the CDD (Conserved Domain Database) (Marchler-Bauer et al., 2003) on the NCBI site and the 6 Frame Translation and Readseq utility at the BCM site.

Phylogenetic Analysis of *Paenibacillus* sp. strain JDR-2 XynA$_1$. All presented phylogenetic analyses resulted from sequences that had been trimmed to contain only the highly conserved catalytic domain from the proton donor (WDVVNE) to the catalytic nucleophile (ITELDI). These sequences were aligned using Clustalx and phylogenetic trees constructed using MEGA 2.1 (Molecular Evolutionary Genetics Analysis, Version 2.1.; Kumar et al., 2002). The domain arrangement of the whole xylanase was determined with CDD (Marchler-Bauer et al., 2003) at NCBI (http://www.ncbi.nih.gov/Structure/cdd/wrpsb.cgi). Signal sequences were analyzed by the on-line program Signal-P (Bendtsen et al., 2004) (http://www.cbs.dtu.dk/services/SignalP). Eighty-four bacterial GH 10 xylanases were downloaded from the CAZy(ModO) database (http://afmb.cnrs-mrs.fr/CAZY/) and processed as described above. This processing ensured the strictest comparison between all the bacterial GH 10 xylanases. Four xylanases showed high similarity to *Paenibacillus* sp. strain JDR-2 XynA$_1$. These and eleven randomly chosen sequences from the eighty set were presented in figures for this chapter.

Carbohydrate and protein assays. Total carbohydrate concentrations related to substrate preparations and enzymatic kinetic analysis were determined by the phenol-sulfuric acid assay (Dubois et al., 1956). In conjunction with the total carbohydrate assay, measurements to define the degree of polymerization of substrate and increased reducing terminus levels due to xylanolytic activities were performed by the method of Nelson (Nelson, 1944). Xylose was used as the reference for both assays. Protein levels were determined using Bradford assay reagents (BioRad) with BSA (Fraction V) as the standard (Bradford, 1976).

$XynA_1$ CD cloning, overexpression and purification. The expression vector pET15b+ (Novagen, San Diego, CA) was used to overexpress the catalytic domain of $XynA_1$ independent of other modules. Primers were designed to delimit the CD based on the modular endpoints identified by Pfam (Bateman et al., 2004). The forward primer (5'aagcatatggctccactcaaa) included an *NdeI* site for in-frame fusion with the His-Tag sequence, and the reverse primer (5'tgtgctcagccggatcaat) contained a *BlpI* (*Bpu1102I*) site for directional cloning into pET15b+. This primer selection method added a Gly, Ser, His, Met sequence to the N-terminus just prior to the beginning of the pfam designated sequence. This additional sequence was derived from the pET15 expression ORF coding sequence. There was additional sequence corresponding to Ala, Glu, Gln at the C-terminal end resulting from vector-derived sequence just upstream from the vector-encoded stop codon. The PCR product was generated using Proof Start high fidelity PCR (Qiagen, Valencia, CA). The construct was verified by sequencing. For expression, the pET15$XynA_1$ CD N-terminal His construct was transformed into *E. coli* Rosetta (DE3) chemically competent cells, and grown with selection (see below) for about 24 hours. A single colony was picked and inoculated into 50 ml LB containing 50 µg/ml ampicillin and 34 µg/ml chloramphenicol, grown at 37 °C to an OD 600 nm of about 0.6 to 0.8, and the entire culture

centrifuged for 10 minutes at 5000 x g at 35° C. The pellet was resuspended in 100 ml LB containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and grown for about 1 hr at 37°C. Cells were harvested as above, resuspended in 10 ml LB medium and used to inoculate a 1 L LB batch culture (preequilibrated at 37°C) containing antibiotics as above. This was grown at 37° C to an OD 600 nm of 0.6 to 0.8 and overexpression induced by the addition of IPTG to 1.0 mM final concentration, and incubation continued for no more than 3 hours before cells were harvested. Cell pellets representing the growth of 1 liter of culture were suspended in 35 ml of 20 mM sodium phosphate buffer, pH 7.0 and lysed at 16000 psi with a single pass through a French pressure cell. The total volume was estimated to be 37.5 ml and 1 M $MgCl_2$ was added to a final concentration of 1.5 mM to obtain optimal Benzonase (Novagen, San Diego, California) activity for hydrolysis of nucleic acid. Benzonase was added at 8 units/ml and the cell lysate was incubated with gentle mixing at room temperature for about 45 minutes. The crude cell lysate was centrifuged at 4°C for 30 minutes at 35000 x g, the supernatant filtered through a 0.45 um syringe tip filter, and 5.0 M NaCl added to a final concentration of 0.50 M. Ten ml aliquots of the cell-free extract were affinity-purified using the HiTrap Chelating HP column procedure (Amersham Biosciences, Piscataway, NJ). Each loaded volume yielded a single expected band detected with Coomassie Blue (CB) following SDS-PAGE analysis. Removal of the N-Terminal His-Tag was accomplished using the thrombin cleavage capture kit available through Novagen. $XynA_1$ CD and $XynA_1$ CD N-terminal His were stored for short periods of time at 4°C in 50 mM potassium phosphate buffer, pH 6.5. For longer storage, these stocks were split with equal volumes of glycerol and stored at -20 °C. Enzyme analysis by activity measurement and protein profiles following SDS-PAGE staining with CB were the same after 6 months storage at -20 °C.

Xylanase activity measurements for enzyme optimization and kinetic analysis. The temperature optimum for $XynA_1$ CD xylanase activity was determined by incubation in 0.25 ml reaction mixes containing 1.0% SG $MeGAX_n$ in 0.1 M potassium phosphate, pH 7.0 for 10 to 30 minutes over a 40 °C to 60 °C range. Reactions were halted by the addition of 0.25 ml Nelson's A:B reagent (25:1) (v:v) and the increase in reducing termini determined (Nelson, 1944). The resulting temperature optimum (45 °C) was subsequently used to determine the optimal activity for the enzyme over the pH range from 5.5 to 7.0 in reaction mixes containing 1.0% SG $MeGAX_n$ in 0.1 M potassium phosphate. The optimal conditions from these determinations (pH 6.5 at 45°C) were used in experiments to examine the reaction kinetics of the enzyme with SG $MeGAX_n$ as substrate. Activity units are described as the amount of enzyme producing 1 μmole of reducing termini per minute at 45°C. Production rates were linear through 30 minutes and data obtained are averages of 3 separate experiments performed in triplicate.

Chromatographic resolution and detection of aldouronates and xylooligosaccharides. Standards were obtained by acid and enzymatic hydrolysis of SG $MeGAX_n$. Aldouronate oligomers, $MeGAX_1$ through $MeGAX_5$, were prepared by acid hydrolysis of $MeGAX_n$ in 0.1 N H_2SO_4 at 121 °C for 60 min. The acid hydrolysate was neutralized with $BaCO_3$ and the aldouronates adsorbed onto Bio-Rad AG2-X8 anion exchange resin in the acetate form. Xylose and xylooligosaccharides were eluted with water, and the aldouronates were then eluted with 20% acetic acid. After concentration by flash evaporation, aldouronates were fractionated with 50 mM formic acid eluent on a 2.5 cm x 160 cm BioGel P-2 column (BioRad, Hercules, CA) equilibrated in the same buffer. Identities of $MeGAX_1$ and $MeGAX_2$ were confirmed by ^{13}C and ^{1}H-NMR spectrometry (K. Hasona, unpublished data). Identities of $MeGAX_3$, $MeGAX_4$ and $MeGAX_5$ are based upon the elution profile from the BioGel P-2 column and TLC analysis of

aldouronates resulting from GH 10 and GH 11 catalyzed $MeGAX_n$ hydrolysis. Xylobiose and xylotriose were generated by hydrolysis with a GH 11 xylanase, XynII of *Trichoderma longibrachiatum* (Hampton Research, Aliso Viejo, CA), and fractionated using water based BioGel P-2 column chromatography. These methods allowed the isolation of X_2, X_3 and the aldouronates $MeGAX_1$ through $MeGAX_5$.

To follow the depolymerization of $MeGAX_n$ catalyzed by $XynA_1$ CD, a 250 µl reaction containing 0.5 units of enzyme and 5 mg SG $MeGAX_n$ in potassium phosphate buffer, pH 6.5, was incubated at 30 °C. Samples (5 µL) were removed every 10 min up to 120 min, and spotted on 20 x 20 cm pre-coated 0.25 mm Silica Gel 60 TLC plates (EM Reagents Darmstadt, Germany). An additional 0.5 units of $XynA_1$ CD was added after the initial 120 minutes and incubation was continued for an additional 16 h. A 5 µl sample representing the reaction limit products was also spotted on the plate. Oligomers were separated by ascension with a solvent system containing chloroform: glacial acetic acid: water (6:7:1) (v:v:v) two times for 4 hours each with at least 1 hour of drying time between each solvent presentation. After the second development the plate was allowed to dry for at least 30 minutes and then sprayed with 6.5 mM N-(1-naphthyl)ethylenediamine dihydrochloride in methanol containing 3% sulfuric acid with subsequent heating to detect the carbohydrates (Bounias, 1980).

To compare the ability of *Paenibacillus* sp. strain JDR-2 to utilize $MeGAX_n$, aldouronates and xylooligosaccharides, $MeGAX_n$ was digested with $XynA_1$ CD to generate primarily X_2, X_3, and $MeGAX_3$. For digestions with $XynA_1$ CD, 50 ml of substrate containing 30 mg/ml SG $MeGAX_n$ were prepared with 10 mM sodium phosphate buffer, pH 6.5. Digestions were initiated by addition of 3.5 units $XynA_1$ CD and incubated with rocking at 30 °C for 24 h. An additional 1 unit was added after 24 hours and incubation was continued for 40 h.

Digests were processed by stir-cell filtration under nitrogen pressure through YM-3 ultrafiltration membranes (3 kDa MWCO) (Millipore, Billerica, MA). The filtrates containing oligomers with molecular weights less than 3000 Da were concentrated by flash evaporation and analyzed by the phenol-sulfuric acid assay (Dubois et al., 1956). These concentrated fractions then served as substrates comparing the growth rates and yield of *Paenibacillus* sp. strain JDR-2 on $MeGAX_n$. Cultures were incubated at 30 °C in baffle flasks containing 50 ml medium supplemented with 10 mg/ml of anhydroxylose equivalents (determined by the total carbohydrate assay). Growth was followed by measuring OD 600 nm. Samples of 250 µl were removed at selected times, centrifuged at maximum speed in a microfuge and the supernatant and pellet were separated and saved. The supernatant was incubated at 70 °C for 10 minutes prior to storage. A volume of 6 µl was used for each sample on the TLC plate. TLC plates were developed as described above.

Immunolocalization of $XynA_1$. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (Laemmli, 1970) using a MINI-PROTEAN 3 electrophoresis cell, a 12% Ready Gel and Precision Plus Dual Color pre stained molecular weight standards (Bio-Rad Laboratories, Hercules, CA) following described methods (Mini- PROTEAN 3 Cell Instruction Manual, Bio-Rad Laboratories, Hercules, Ca.). Immunodetection was performed as previously described (Schmidt et al., 2003). XynA1 CD was purified to homogeneity as judged by SDS-PAGE after staining with CB. Chickens were inoculated with XynA1 CD as antigen. An amount of 100 µg was delivered in a total volume of 1 ml PBS. A volume of 700 µl was injected subcutaneously under the wing and 300 ul was injected in the footpad. No adjuvant was used. Eggs were collected from before injection and through the entire process. A boost injection was administered as above about two weeks post

primary injection. The eggs were screened by ELISA in groups of three as crude unprocessed yolk in PBS. Peak fractions were pooled and chicken IgY polyclonal antibody obtained following the method of Polson, et al (Polson et al., 1985).

Immunolocalization studies were performed with cell fractions following growth on SG MeGAXn. *Bacillus subtilis* 168 was cultured as a negative control and compared with *Paenibacillus* sp. strain JDR-2. Colonies of *B. subtilis* and *Paenibacillus* sp. strain JDR-2 were suspended in 2 ml, 1x Z-H and vortexed until cells were fully suspended. The complete 2 ml volume was used to inoculate 50 ml of media in 250 ml baffle flasks containing 0.2% YE, 0.36% SG MeGAX$_n$ in Z-H. Cultures were grown overnight (16 hr) at 30°C with shaking at 150 rpm on a New Brunswick G-2 gyrotory shaker. Cells were harvested at an OD 600 nm of 1.0 (*Paenibacillus* sp. strain JDR-2) and 0.7 (*B. subtilis*). Cultures were centrifuged at 5000 x g for 15 minutes at room temperature and the supernatant was recovered. Cell pellets were resuspended in 50 mM sodium phosphate buffer, pH 6.5, and centrifuged as above. The procedure was repeated with 50 mM sodium phosphate pH 6.5 containing 0.5 M NaCl. The final cell pellet was resuspended in 5 ml, 50 mM sodium phosphate, pH 6.5. Some cell lysis of *Paenibacillus* sp. strain JDR-2 was apparent, observed as increased viscosity, probably due to osmotic shock. A volume of 50 µl Promega DNase RQ1 at 1 unit/µl was added with 1/10 the volume of 10 x DNase RQ1 buffer (0.40 M Tris-HCl, 0.10 M $MgCl_2$, 0.01 M $CaCl_2$, pH 8.0) and the suspension was incubated for 30 minutes at room temperature. Cells were then lysed by two passes at 16,000 psi through the French pressure cell. Lysates were centrifuged at 30,600 x g for 20 min at 4°C. Supernatant was collected as the cell free extract, the pellet was resuspended in 1 ml, 50 mM sodium phosphate pH 6.5 and designated the cell wall suspension. All supernatants were concentrated using YM-10 Centriprep concentrators (Millipore, Billerica, MA) to volumes

less than 4 ml. Samples of the media supernatant concentrate (MSC), NaCl wash (NaCl), cell free extract (CFE) and cell wall suspension (CWS) were analyzed by SDS-PAGE. Reactive antigens were detected on immunoblots using rabbit anti-chicken alkaline phosphatase conjugate (Sigma, St. Louis, Missouri) as previously described , and proteins were detected in gels with CB (Schmidt et al., 2003).

Results

Growth analysis of *Paenibacillus* sp. strain JDR-2. Based on OD 600 nm measurements, the initial growth analysis of *Paenibacillus* sp. strain JDR-2 indicated that the organism utilized $MeGAX_n$ more efficiently compared to glucose or xylose as substrates (Figure 4-1A). More detailed studies (Figure 4-1B-D) using HPLC to follow substrate concentration showed that $MeGAX_n$ is almost completely utilized. Additionally, *Paenibacillus* sp. strain JDR-2 preferentially utilized the $MeGAX_n$ in the presence of glucose or xylose. Under these conditions, the concentrations of glucose and xylose in the medium decreased more slowly, and at a nearly linear rate. Figure 4-1B also shows that xylose accumulates in the medium to a small extent during growth on $MeGAX_n$, indicating that what is produced during the extracellular depolymerization may not be directly assimilated.

Identification and sequencing of *xynA$_1$* encoding a secreted modular GH10 endoxylanase. Analysis of the *Paenibacillus* sp. strain JDR-2 chromosomal DNA library in *E. coli* for xylanases led to the isolation of four clones. Restriction analysis of these clones suggested that the inserts were from the same genomic DNA location. Plasmid pFSJ4 was selected for sequencing which revealed an insert (Figure 4-2) including a large modular xylanase (*xynA$_1$*) of 4401 nt (1467 aa). Sequencing of the compete genomic DNA insert identified genes flanking *xynA$_1$*. In the 5' direction on the same chain there is a *mdep* gene encoding a putative multi-drug efflux permease with 43% amino acid identity to the same in *Bacillus halodurans*

(gene = *BH3482*) determined by blastp. In the 3' direction on the opposite strand there is a putative α-1,6-mannanase gene (*amanA*) that codes for a protein with 67% identity to Aman6 protein (*aman6* gene) from *Bacillus circulans*. Domain analysis revealed that AmanA has the exact modular structure of Aman6 with a GH 76 catalytic module followed by triplicate family 6 CBM. In silico sequence analysis identified a probable promoter region and rho-independent terminator for *xynA1*, but only a terminator for the *mdep* gene and a promoter for the gene encoding AmanA.

Much like many other glycosyl hydrolases, $XynA_1$ is a modular protein composed of 8 separate modules (Figure 4-2). The domains include a triplicate N-terminal set of CBM 22 modules which have previously been shown to bind soluble xylan and β-1,3-1,-4-glucan (Dias et al., 2004; Xie et al., 2001). These modules are followed in sequence by a GH 10 CD and a CBM 9, which has been shown to bind to the reducing end of carbohydrate chains (Boraston et al., 2001; Notenboom et al., 2001). Following CBM 9 is an undefined sequence with high similarity to the same region in Xyn5, a GH 10B xylanase of *Paenibacillus* sp. W-61. This region, as previously reported, has high identity to the lysine-rich region of the SdbA protein of *C. thermocellum*. Xyn5 and $XynA_1$ have 36% and 35% amino acid identity, respectively, to this region of SdbA, and this region in $XynA_1$ has 49% identity to the same in Xyn5. Although these identities to the lysine rich-region of SdbA are relatively high, $XynA_1$ and Xyn5 contain only about 5% and 6.5% lysine, respectively, to the same region of SdbA which has 13% lysine (data not shown) (Ito et al., 2003; Leibovitz et al., 1997). The C-terminal region includes a triplicate set of SLH modules which are predicted to function in surface anchoring (Cava et al., 2004; Kosugi et al., 2002; Mesnage et al., 2000). The *$xynA_1$* coding sequence has been deposited in EMBL with the accession number AJ938162.

Phylogenetic analysis of $XynA_1$. Initial phylogenetic analysis revealed that $XynA_1$ and $XynA_1$ CD amino acid sequences, when subjected to blastp, had high bit scores to the same set of four modular GH 10 xylanases (Figure 4-3). Comparison of the top 9 blastp hits to $XynA_1$ of *Paenibacillus* sp. strain JDR-2 shows the comparative modular structures. Additionally, the bit scores are represented for the whole sequence blastp and the CD sequence blastp. $XynA_1$ and the top four hits were classified as GH 10B and the lower set as GH 10A based on the number of amino acids separating the glutamate residue functioning as the catalytic proton donor from the glutamate functioning as the catalytic nucleophile. Although there are some exceptions, most catalytic domains of GH 10 xylanases have about 105 amino acids separating the two catalytic residues. In the case of sequence group GH 10B, the distance separating the catalytic residues is about 123 amino acids (Table 4-1). For further analysis we reasoned that the catalytic residue bridge sequence was probably the most highly conserved portion of GH 10 xylanases and compared this sequence from many xylanases. Figure 4-4 represents a phylogenetic comparison of the GH 10B subset to eleven randomly selected GH 10A xylanases. Table 4-1 characterizes the modular structures for the xylanases, indicating significant diversity among those represented in subset 10A. In Figure 4-4A the Clustal alignment of the CD region used to prepare the phylogenetic tree revealed three areas in which GH 10B subset differs from GH 10A. The additional sequences accounted for the extra length between the catalytic proton donor and nucleophilic glutamate residues. A phylogenetic tree developed with the neighbor-joining method for the alignment in Figure 4-4A shows sequences within a clade distinct from the others (Figure 4-4B). It should be noted that this comparison set is biased to the extent that it contains five very similar GH 10B sequences with eleven other random GH 10A sequences. However, the presentation of data identifies a relationship that indicates that GH 10B sequences have a

common lineage. Large-scale analysis of 84 bacterial GH 10 xylanases obtained from CAZy identified GH 10B as a subset and allowed few other subsets to be created with a > 95% bootstrap value. Many of these sequences did not place with confidence in any subset potentially allowing for only a few well-defined subgroups of GH 10 xylanases.

$XynA_1$ localization. Chicken polyclonal IgY generated against $XynA_1$ CD (anti-CD) was used to examine the localization of $XynA_1$ in *Paenibacillus* sp. strain JDR-2 cell fractions. CB stained SDS-PAGE bands of both *Paenibacillus* sp. strain JDR-2 and *Bacillus subtilis* 168 proteins were primarily greater than 100 kDa for all fractions (Figure 4-5A). However, anti-CD showed reactivity with *Paenibacillus* sp. strain JDR-2 CWS protein (approx. 150 kDa) which was not apparent with the CB stained gel (Figure 4-5B). The antibody reacted well with $XynA_1$ CD with essentially no cross reactivity towards *B. subtilis* fractions. Size estimation of the reactive *Paenibacillus* sp. strain JDR-2 CWS protein at approximately 150 kDa compares favorably with the MW (154 kDa) obtained from the translated amino acid sequence of the $XynA_1$ modular enzyme. The band identified as $XynA_1$ in the immunoblot is not visible in the CB-stained gel (Figure 4-5A), indicating that $XynA_1$ represents a minor component of the surface protein complement. What is obvious from the CB-stained gel is a band size of approximately 80 kDa. This undoubtedly is the most prominent protein overshadowing all others. Observing that $XynA_1$ is anchored to the surface supports the possibility that *Paenibacillus* sp. strain JDR-2 produces a crystalline surface layer. The size of the prominent band at 80 kDa is roughly the same size as the Sap and 80K surface layer proteins from *Bacillus anthracis* and *Bacillus sphaericus* respectively (Bowditch et al., 1989; Etienne-Toumelin et al., 1995).

Kinetic and product analysis of $XynA_1$ CD. $XynA_1$ CD was overexpressed in pET15b+ and affinity purified using an N-terminus His-Tag. Removal of the affinity tag by thrombin protease treatment resulted in an increased activity against SG $MeGAX_n$ of approximately 50%. Initial characterization showed $XynA_1$ CD to have an optimal pH and temperature of 6.5 and 45°C, respectively (data not shown). Kinetic analysis with SG $MeGAX_n$ as substrate (Figure 4-6) showed $XynA_1$ CD to have V_{max} and K_m values of 8 units/mg and 1.96 mg/ml, respectively, and a k_{cat} of 306.8 /min. Analysis of products by TLC (Figure 4-7) showed that $XynA_1$ CD is a typical GH 10 xylanase hydrolyzing $MeGAX_n$ primarily to X_2 and $MeGAX_3$. Small amounts of X_3 and $MeGAX_4$ were also produced. True limit products of the reaction included xylose, which built up from the seemingly slow conversion of X_3 and $MeGAX_4$ to X_2 and $MeGAX_3$. Thirty minutes after reaction initiation, X_2 and $MeGAX_3$ are the predominant products (Figure 4-7). The small amounts of X_3 and GAX_4 disappeared by 24 hr.

***Paenibacillus* sp. strain JDR-2 utilization of aldouronates and xylooligosaccharides in comparison to $MeGAX_n$.** Xylooligosaccharides and aldouronates were generated by hydrolysis of SG $MeGAX_n$ with $XynA_1$ CD. A 3 kDa molecular weight cutoff ultrafiltration filtrate product was used to evaluate growth of *Paenibacillus* sp. strain JDR-2 on xylanase generated aldouronates and xylooligosaccharides. Growth was compared for SG $MeGAX_n$ and $XynA_1$ CD filtrate. Data presented in Figure 4-8 shows the aldouronates and xylooligosaccharides resolved by TLC during the growth of *Paenibacillus* sp. strain JDR-2. Through the time course for growth on SG $MeGAX_n$, neither aldouronates nor xylooligosaccharides were detected in the media during exponential growth. In contrast to growth observed on the $XynA_1$ CD-generated products, higher growth rates and yields were observed with $MeGAX_n$ as substrate, indicating preferred utilization of polymeric glucuronoxylan compared to aldouronates and

xylooligosaccharides generated by the in vitro $XynA_1$ CD-catalyzed depolymerization of $MeGAX_n$.

Discussion

Based upon growth and substrate utilization analysis, *Paenibacillus* sp. strain JDR-2 has been shown to more efficiently utilize the biomass polymer $MeGAX_n$ compared to simple sugars such as glucose and xylose. In addition, growth on $MeGAX_n$ with competing simple sugars does not seem to affect its utilization of $MeGAX_n$ (Figure 4-1). This observation stands in contrast to a similar xylanolytic system from *Paenibacillus* sp. W-61 in which the investigators found that glucose strongly repressed xylanase activity (Viet et al., 1991). Although there appear to be metabolic differences, *Paenibacillus* sp. W-61 produces Xyn5, a GH 10B xylanase that is the top blastp hit of $XynA_1$. With 51% identity the two full sequences are very similar with Xyn5 differing; i.e., having only 2 CMB 22 modules rather than three. Kinetic properties of the two xylanases are similar but the generation of aldouronates by *Paenibacillus* sp. W-61 was not determined, precluding a comparison to $XynA_1$ secreted by *Paenibacillus* sp. strain JDR-2 (Ito et al., 2003).

Even though *Paenibacillus* sp. strain JDR-2 utilizes $MeGAX_n$ very efficiently, it is probable that $XynA_1$ is the only extracellular xylanase responsible for this ability. Genomic library screening led to the isolation of four xylanolytic clones with identical restriction profiles, each containing the same *xynA1* coding sequence. Only one other xylanase gene has been identified from this organism during intensive cosmid library screening in *E. coli*, and this encodes a 40 kDa GH 10 catalytic domain designated $XynA_2$. The primary amino acid sequence for $XynA_2$ does not have a detectable secretion signal sequence and is expected to be localized to the cytosol. The *xynA2* gene sequence is located within an operon including *aguA*, encoding a GH 67 α-glucuronidase, and encodes a GH 10 xylanase that may be involved in the intracellular

processing of aldouronates and xylooligosaccharides generated by the action of $XynA_1$ on the cell surface (G. Nong, V. Chow, J. D. Rice, F. St. John, J. F. Preston, Abstr. 105th ASM General Meeting, abstr.O-055, 2005). $MeGAX_3$, the primary aldouronate limit product of GH 10 xylanases (Biely et al., 1997), is presumably efficiently assimilated as it has been identified as an inducer for genes involved in hydrolysis and catabolism of glucuronoxylan in *Geobacillus stearothermophilus* (Shulami et al., 1999).

Phylogenetic characterization of $XynA_1$ placed the sequence with a highly similar set of GH 10 xylanases referred to in this paper as the GH 10B subset (Figure 4-3). This classification is supported by differences observed in the CD coding sequence. Specifically, the area bridging the two catalytic glutamate residues contains three areas of additional sequence that are not observed in other xylanases. Although GH 10B xylanases are modular, there are many similarly modular xylanases that may not be classified as GH 10B. This suggests a unique mode of action for GH 10B xylanases. It is interesting to note that this subset includes GH 10 xylanases in anaerobic *Clostridium* spp. and aerobic *Paenibacillus* spp., all of which are found in soil environments. The common modular architecture (Figures 4-3 and 4-4) that, in this case, includes anchoring motifs, suggests a positive role in niche development of these bacteria. The variability in the number of CBM and SLH modules suggests these may be mobile elements that may be combined from different genes during evolution.

$XynA_1$ CD analysis identified $XynA_1$ as a typical GH 10 endoxylanase, producing primarily X_2, X_3 and $MeGAX_3$ in the early stages of the reaction (Figure 4-7) (Biely et al., 1997; Preston et al., 2003). Hydrolysis seemed to proceed in two stages. The first stage resulted in formation of X_2 and $MeGAX_3$ with small amounts xylose, X_3 and $MeGAX_4$. The second stage included the slow conversion of the minor products to X_2 and $MeGAX_3$ with increased formation

of xylose. The $XynA_1$ CD K_m for SG $MeGAX_n$ was within a comparable range to other GH 10 xylanases but the rate of catalysis was significantly lower than that found in other reports (Figure 4-6). This may in part be due to the special attention given to xylanases showing high activity. Wild type purification of Xyn5 from *Paenibacillus* sp. W-61 yielded an enzyme showing a similar specific activity as $XynA_1$. (Roy et al., 2000).

Xylan binding subsite analysis revealed that $XynA_1$ contains a CD that has four well-conserved subsites. Subsites -2 through +2 are highly conserved in GH 10 xylanases and $XynA_1$ is no exception. In addition, by alignment analysis (data not included), $XynA_1$ does not appear to have a +3 subsite and subsites -3 and +4 do not exist as defined in some other GH 10 xylanases (Charnock et al., 1998; Pell et al., 2004a). Analysis of these subsites as they may impact the catalysis of $MeGAX_n$ seems to support the results of product analysis by TLC. A xylanase with strong binding –2 through +2 subsites should yield X_2 as a primary product of $MeGAX_n$ hydrolysis. Accumulation of xylose and small odd numbered xylooligomers (X, X_3) would only result from processing through odd numbered oligomers such as X_5 and X_7. Structural studies have also identified a glucuronic acid pocket in the +1 subsite that would facilitate hydrolysis of $MeGAX_n$, leaving the MeGA substitution on the nonreducing terminus (Pell et al., 2004b). Since $MeGAX_3$ is a primary limit product of GH 10 xylanases, it follows that large xylooligomers containing MeGA as a nonreducing end substituted residue can only be further processed by positioning into the -3 subsite yielding $MeGAX_3$ (Fujimoto et al., 2004).

$XynA_1$ is the largest GH 10 xylanase so far identified from a *Paenibacillus* sp. The net modular architecture is similar to other *Paenibacillus* sp.; however, the triplicate N-terminal CBM 22 is unique to *Paenibacillus* sp. strain JDR-2. Although at least one other bacterial GH 10 xylanase has been identified with a triplicate set of N-terminal CBM 22 modules (XynB from

Caldicellulosiruptor sp. Rt69B.1), it classifies as a GH 10A subset member according to this analysis (Figure 4-3) (Morris et al., 1999). Published reports that consider the role of carbohydrate binding modules with respect to the function of the catalytic domain suggest that these modules accentuate activity by increasing the localized substrate concentration (Boraston et al., 2004). It is difficult to imagine hydrolysis of $MeGAX_n$ by $XynA_1$ being potentiated by the addition of a third CBM 22 module. Two publications concerned with a set of CBM 22 modules (not necessarily in tandem) from different xylanases have shown that while one seems to function to bind a potential carbohydrate substrate, the other does not (Charnock et al., 2000; Meissner et al., 2000). Additionally, in some cases these modules have been shown to have better binding to β-1,3-1,4-glucan (barley β-glucan) than to xylan (Araki et al., 2004; Meissner et al., 2000). Based on these inconsistent findings it is impossible to assume any precise functionality of these CBM 22 modules. Analysis of this system by expression of each module separately and the application of native affinity polyacrylamide gel electrophoresis (NAPAGE) would clarify the role of each CBM 22 as they may function to accentuate CD catalytic ability (Meissner et al., 2000).

It seems likely that there is a competitive advantage in colonizing a niche for an organism that utilizes surface anchored enzymes to hydrolyze biomass polymers. The proximity of the resulting hydrolysis products would decrease diffusion-dependent assimilation rates. This strategy has been attributed to *Clostridium* spp. that produce the cell surface-localized cellulosome (Bayer et al., 2004; Doi and Kosugi, 2004; Doi et al., 2003). However in anaerobic ecosystems in which *Clostridium* spp. are the primary utilizers of biomass it has been suggested that the main product of crystalline cellulose hydrolysis, cellobiose, actually decreases cellulolytic activity and that excess cellobiose and other free sugars are utilized by other

members of the ecosystem. This relationship is truly communal in that these other bacteria receive carbon substrates and may return the favor in the form of vitamins or other beneficial growth factors for the cellulolytic organisms (Bayer et al., 1994). Inclusion of xylanolytic enzymes in the cellulosome does not establish that these organisms can utilize the products resulting from hydrolysis of $MeGAX_n$. It's probable that hemicellulolytic activities are associated with the cellulosome to increase cellulase access to cellulose by removing associated non-cellulose polymers (Bayer et al., 2004). It has been reported that the mesophilic *Clostridium cellulovorans* can ferment xylan, but there is no clear analysis of hydrolysis products and the extent to which the xylan is utilized (Kosugi et al., 2001; Sleat et al., 1984). Additionally, although there is increasing evidence of GH 10 xylanases in the *Clostridium*, there is no evidence of accessory enzymes such as an α-glucuronidase which is thought to be required for complete utilization of $MeGAX_n$ (Han et al., 2004). All of these characteristics pertaining to the cellulosomal systems seem to stand in contrast to the $MeGAX_n$ hydrolytic system of *Paenibacillus* sp. strain JDR-2. This system does not utilize the hydrolytic products of $XynA_1$ CD efficiently and seems to require the activity of $XynA_1$ anchored to the cell surface for efficient utilization of $MeGAX_n$. This would suggest that the $XynA_1$ anchoring/ vectoral transport mechanism has evolved to yield almost complete recovery of hydrolytic products as an advantage against potential niche competitors. This may be further supported by the fact that *Paenibacillus* sp. strain JDR-2 requires no nutritional supplement for growth on $MeGAX_n$, but growth of *C. thermocellum* and *C. cellulovorans* requires medium supplemented with yeast extract (Bayer et al., 1983; Quinn et al., 1963; Sleat et al., 1984).

Paenibacillus sp. strain JDR-2 $XynA_1$ anchoring to the cell surface may be considered somewhat analogous to the surface anchoring of the cellulosome in the *Clostridia* (Bayer et al.,

2004). An important distinction between these genera is that *Clostridia* are strictly anaerobic organisms and *Paenibacillus* sp. strain JDR-2 requires oxygen for growth. This physiological difference spatially separates these two genera in environmental niche development. In addition, the cell surface associations of cellulases and other enzymes in *Clostridium* spp. is mediated through the cellulosome. This complex of enzymes is maintained via interactions between dockerin modules on individual enzyme proteins and cohesion modules on a scaffoldin protein, which is then anchored to the cell surface (Doi, Kosuge et al., 2003). Surface anchoring of biomass degradative enzymes may provide a strategy for efficient hydrolysis and transport of resulting products, yielding a distinct advantage over organisms with free enzyme lignocellulose degradative systems. In the example of *C. thermocellum* cellulose, cellobiose and cellodextrins (degrees of polymerization ≤ 4) are directly transported by a cellodextrin ABC transporter. Transport of these oligosaccharides conserves ATP by the action of intracellular phosphorylase yielding a significant growth advantage (Zhang and Lynd, 2005). *Paenibacillus* sp. strain JDR-2 may utilize a similar mechanism for the efficient utilization of X_2 and $MeGAX_3$. Additionally, *Paenibacillus* sp. strain JDR-2 seems to couple the action of substrate hydrolysis to product uptake. This may be a secondary method for the conservation of energy. Once internalized, the $MeGAX_3$ is apparently processed by α-glucuronidase (AguA)-mediated hydrolysis to MeGA and X_3, with subsequent hydrolysis of xylotriose by intracellular GH 10 xylanases ($XynA_2$) with specificity for small xylooligosaccharides and β-xylosidase (Gallardo et al., 2003; Pell et al., 2004a; Preston et al., 2003). Although this may be the case, it is also possible that *Paenibacillus* sp. strain JDR-2 utilizes an intracellular phosphorylase as described for several cellulolytic organisms (Lou et al., 1996; Lou et al., 1997; Reichenbecher et al., 1997).

Considering the efficient utilization of $MeGAX_n$ by *Paenibacillus* sp. strain JDR-2, this organism may provide a platform for future biocatalyst development. Under conditions of low oxygen, *Paenibacillus* sp. strain JDR-2 produces succinate and acetate as fermentation products (unpublished data). Alternatively, the genes encoding the cell surface anchored $XynA_1$, as well as those involved in the assimilation and metabolism of $XynA_1$-generated products, may be used to engineer other bacterial platforms to efficiently convert $MeGAX_n$ to desired fermentation products. The aggressive utilization of $MeGAX_n$ by *Paenibacillus* sp. strain JDR-2 supports its further development and genetic exploitation for the conversion of lignocellulosic biomass to alternative fuels and bio-based products.

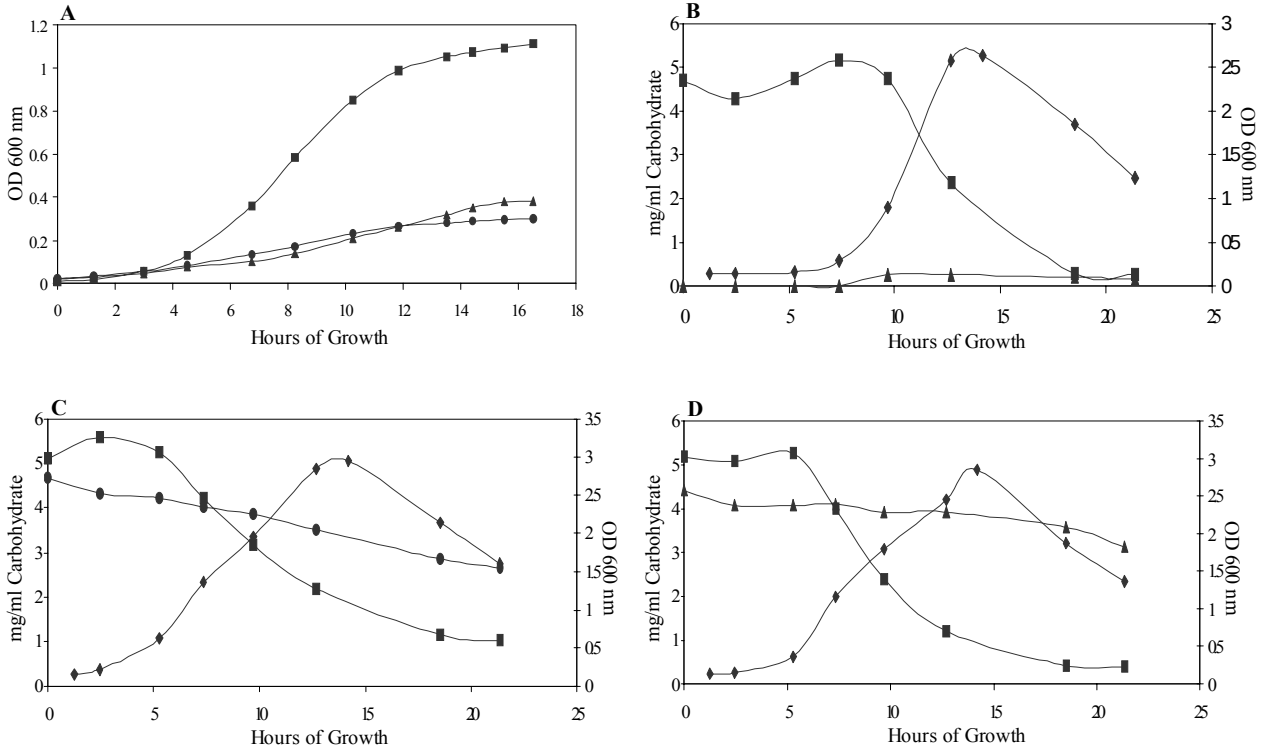

Figure 4-1. Growth of *Paenibacillus* sp. strain JDR-2. (A) *Paenibacillus* sp. strain JDR-2 growth characterization on individual sugar substrates. OD measurements were at 600 nm for 4 ml cultures. For these cultures 4 ml of 1.0% carbohydrate in Z-H mineral salts was inoculated with 200 µl from an overnight culture of 1% YE in Z-H mineral salts. This was started from a single colony of *Pb* sp. JDR-2 from a 2 day 1.0% YE, 0.5% oat spelt xylan, Z-H agar plate. (B) *Paenibacillus* sp. strain JDR-2 growth on 0.5% SG MeGAX$_n$. (C) *Pb* sp. JDR-2 growth on 0.5% SG MeGAX$_n$ and 0.5% glucose. (D) *Paenibacillus* sp. strain JDR-2 growth on 0.5 % SG MeGAX$_n$ and 0.5% xylose. (B-D) OD measurements at 600 nm of 50 ml cultures in 125 ml baffle flasks. For these 50 ml baffle flask cultures, 4 ml of inoculum was used from an overnight culture in 1.0% YE in Z-H mineral salts. HPLC was used to quantify carbohydrate concentrations. (A) Squares, growth on SG MeGAX$_n$; triangles, growth on xylose; circles, growth on glucose. (B-D) Diamond, OD 600 nm; square, SG MeGAX$_n$; triangle, xylose; circle, glucose.

Figure 4-2. Genetic map of *xynA1* and surrounding sequence resulting from sequencing of the *Paenibacillus* sp. strain JDR-2 genomic DNA insert of pFSJ4. Graphic textures refer to indicated modules as identified by pfam. Putative promoters and rho-independent terminators are identified by an arrow and the symbol Ω, respectively.

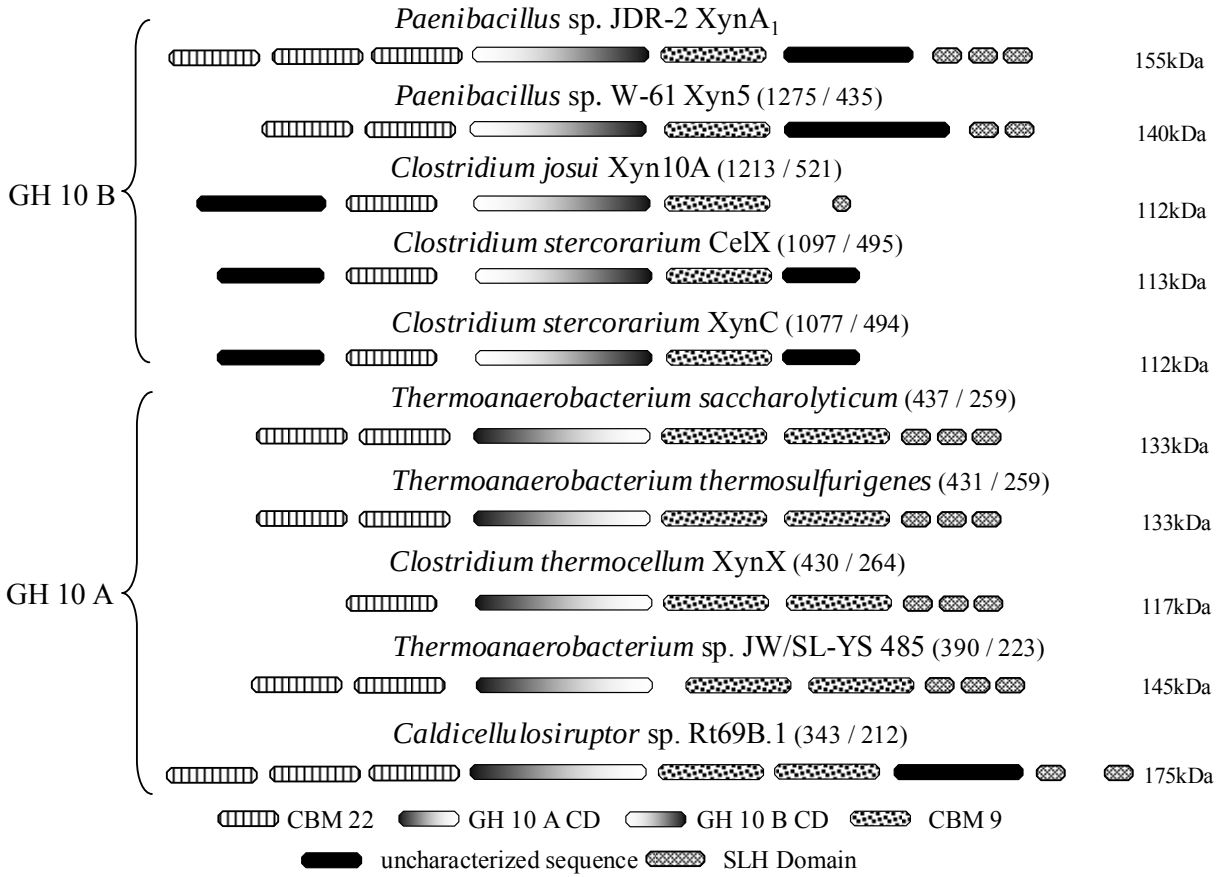

Figure 4-3. Domain alignment of GH 10 subset B and subset A sequences. The Conserved Domain Database was used to predict the domains from the 9 most similar sequences identified through a BLAST search. Similarities relative to *Paenibacillus* sp. strain JDR-2 are arranged in descending order. Next to each organism name is the BLAST bits score (similarity to *Paenibacillus* sp. stain JDR-2) as (X / Y). X = whole $XynA_1$ sequence blastp bits score, Y = pfam designated CD module blastp bits score.

Table 4-1. Source and characteristics of sequences used for phylogenetic comparison

#	Organism	Swiss-Prot/TrEMBL /GenPept	Acid/Nucleophile Distance	Domain Arrangement[a]	Predicted Signal Sequence[b]
1	*Clostridium stercorarium*	Q9XDV5	123	CBM4-9/CD/CBM9	yes
2	*Clostridium stercorarium*	Q8GJ37	123	CBM4-9/CD/CBM9	yes
3	*Paenibacillus sp. JDR-2*	Q53I45	123	3CBM4-9/CD/CBM9/3SLH	yes
4	*Clostridium josui*	Q9F1V3	123	CBM4-9/CD/CBM9/tSLH	yes
5	*Paenibacillus sp. W-61*	BAC45001.1	122	2CBM4-9/CD/CBM9/2SLH	yes
6	*Bacillus stearothermophilus*	P45703	106	CD	no
7	*Thermoanaerobacter thermosulfurogenes*	Q60046	104	2CBM4-9/CD/2CBM9/3SLH	yes
8	*Cellvibrio mixtus*	O68541	104	CD	yes
9	*Bacteroides ovatus*	P49942	104	CD	yes
10	*Cellvibrio japonicus*	P14768	118	CBM2/CBM6/CD	yes
11	*Streptomyces lividans*	P26514	107	CD/RBT	yes
12	*Cellulomonas fimi*	P07986	105	CD/CBM2	yes
13	*Clostridium stercorarium*	P40942	107	CD	yes
14	*Caldicellulosiruptor saccharolyticus*	O30427	107	2CBM4-9/CD	yes
15	*Clostridium thermocellum*	O32374	116	CBM4-9/CD/DCK	yes
16	*Clostridium acetobutylicum*	Q97TP5	106	CD	yes

[a]Domain identification performed through NCBI Conserved Domain Database.
[b]Signal sequence determined using Signal-P on-line tool.

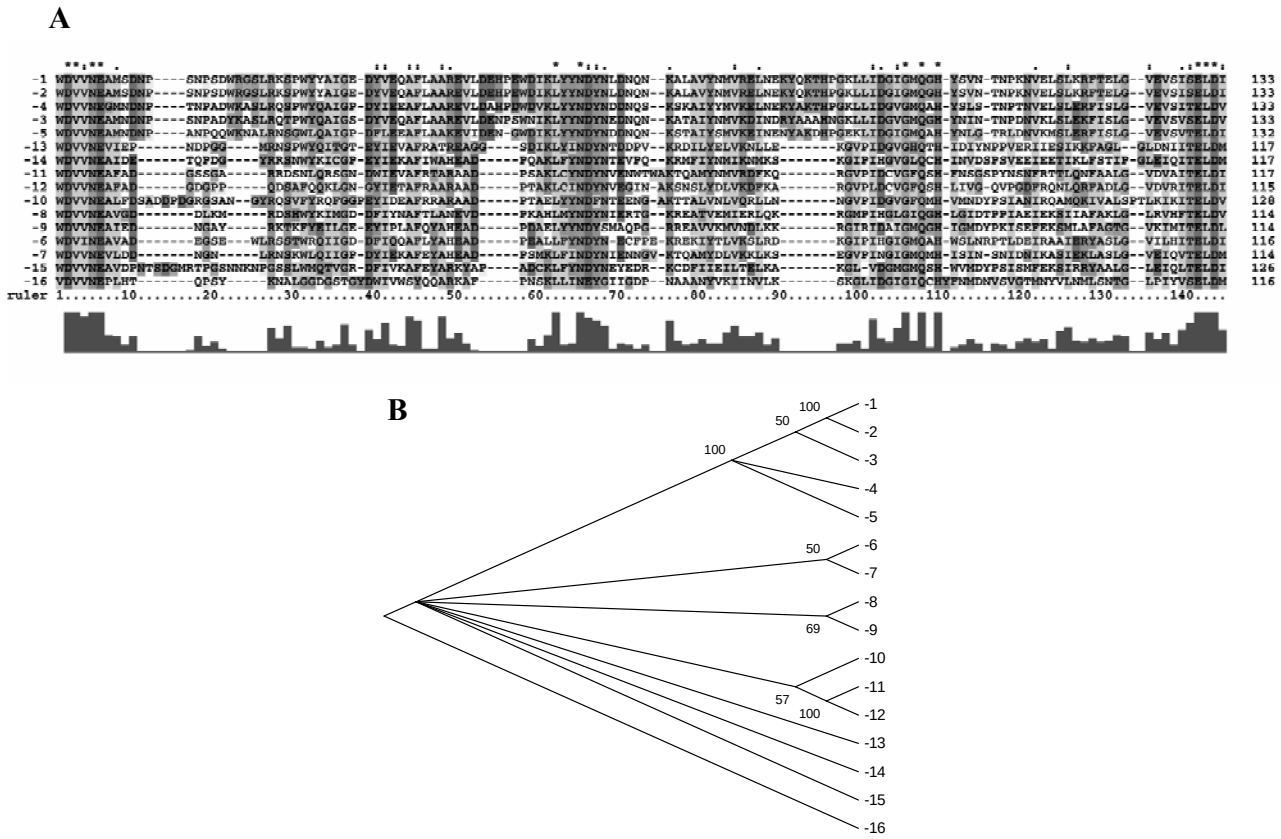

Figure 4-4. Phylogenetic analysis of a randomly selected set of GH 10 xylanases with respect to the $XynA_1$ CD GH 10B subset (Table 3-1). Sequence for comparison consists of the highly conserved bridge between the catalytic nucleophile and proton donor glutamate residues. Sequences 1-5 represent the top four blastp hits to *Paenibacillus* sp. strain JDR-2 $XynA_1$ (Table 1). All other sequences were randomly selected from a list of bacterial xylanases. (A) Clustal alignment of the analyzed sequences. (B) Neighbor-Joining/Bootstrap phylogenetic tree analysis.

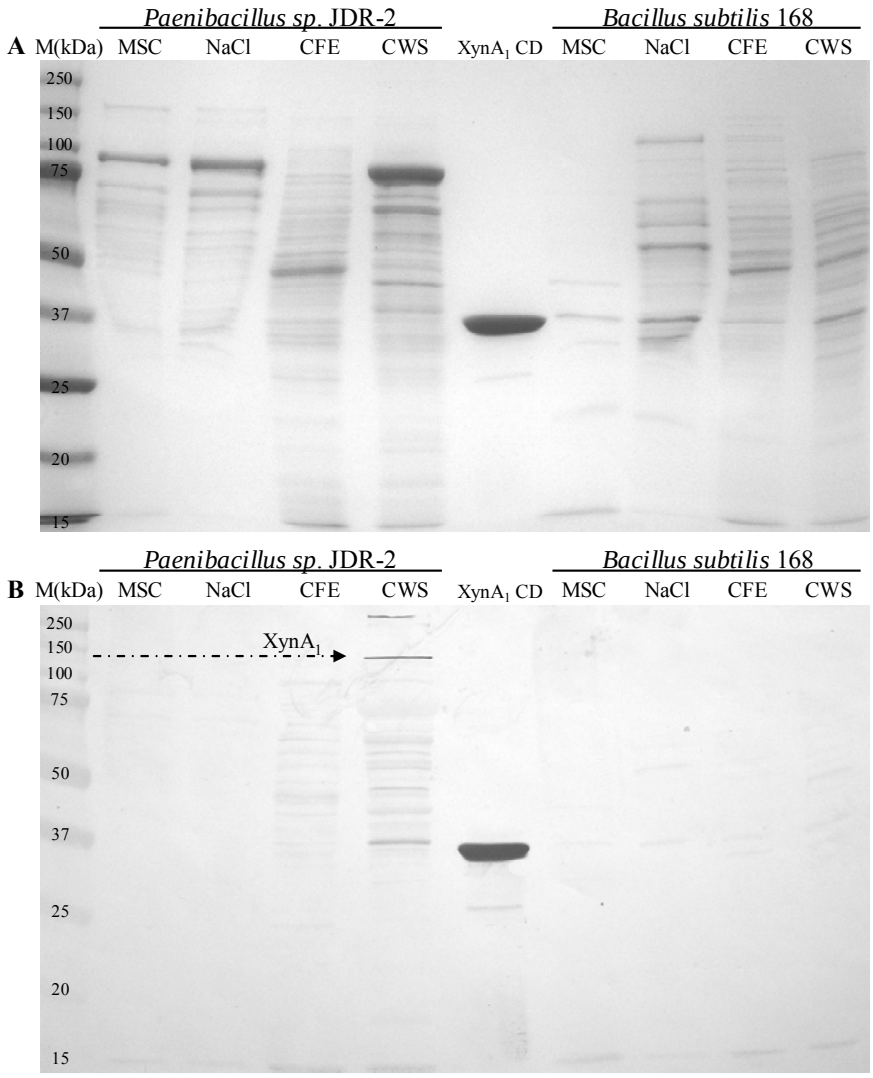

Figure 4-5. Localization of modular $XynA_1$ in subcellular fractions. (A) SDS-PAGE analysis of protein content stained with Coomassie blue. (B) Immunodetection of $XynA_1$ in companion gel blot using anti-$XynA_1$ CD IgY polyclonal preparation. Cells were grown in medium containing 1% YE, 0.5% SG xylan in Z-H mineral salts. Recombinant $XynA_1$ CD was used as a positive control. *Bacillus subtilis* 168 was grown in the same manner for use as a Gram positive negative control. MSC, media supernatant concentrate; NaCl, concentrate from 0.5 M NaCl wash of cells; CFE, supernatant of French press lysate; CWS, cell wall suspension. Protein amounts (ug) loaded for SDS-PAGE were as follows: *Paenibacillus sp.*: MSC, 3.0; NaCl, 3.0; CFE, 4.0; CWS, 4.0. *B. subtilis*: MSC, 2.5; NaCl, 3.0; CFE, 3.0; CWS, 3.0. For immunoblot: *Paenibacillus sp.*: MSC, 1.0; NaCl, 1.0; CFE, 1.0; CWS, 1.0. *B. subtilis*: MSC, 1.0; NaCl, 1.0; CFE, 1.0; CWS, 1.0. $XynA_1$ CD positive control, 3.0 ug, was loaded for SDS-PAGE and detected with CB; 1.0 ug was loaded for the Western Blot.

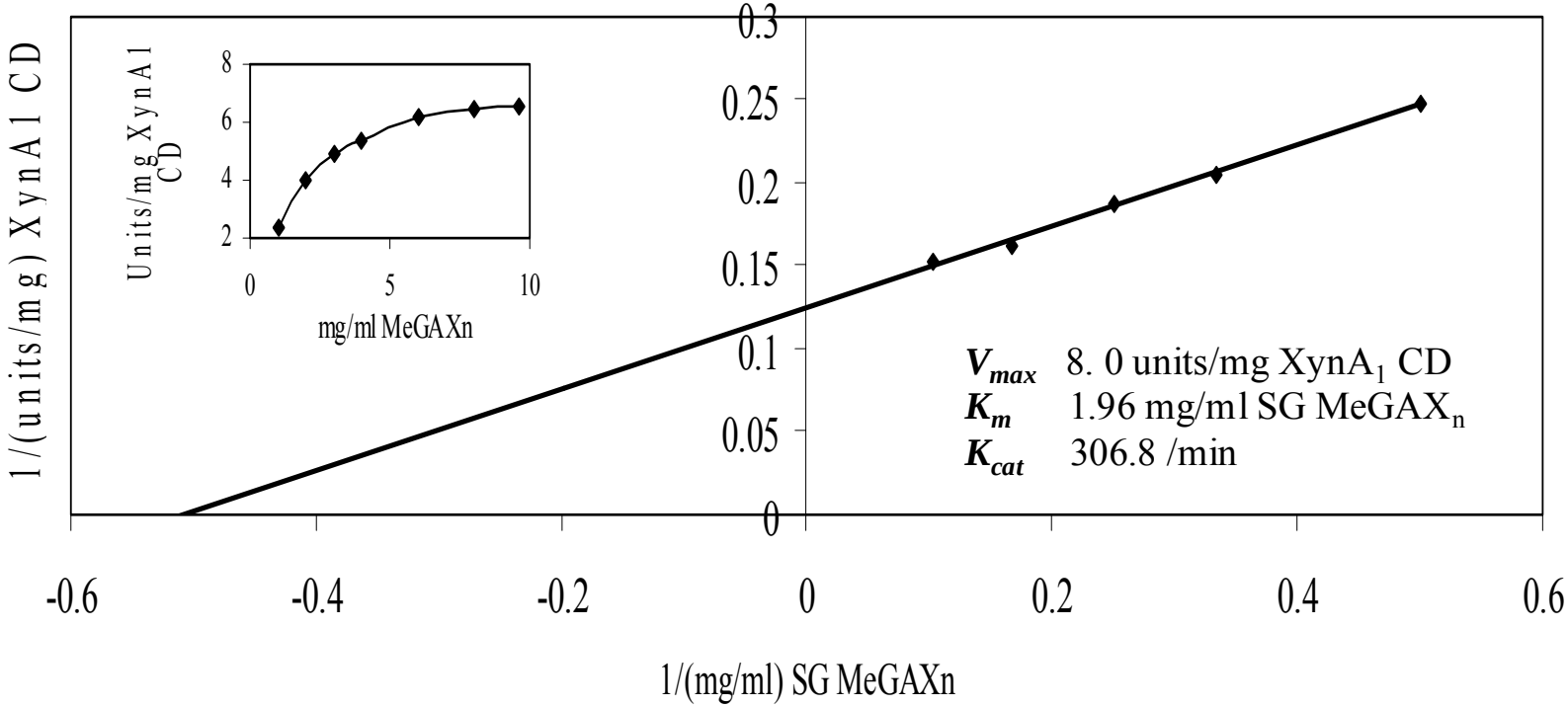

Figure 4-6. Lineweaver - Burk kinetic analysis of $XynA_1$ CD. Inset represents a graph of $XynA_1$ CD velocity vs. SG $MeGAX_n$ concentration. This data was used to prepare the double reciprocal plot. $XynA_1$ CD was analyzed using SG $MeGAX_n$ as substrate and measuring the increase in reducing terminus by the Nelson's test. All samples were performed in triplicate and the final data represent the average of three separate assays.

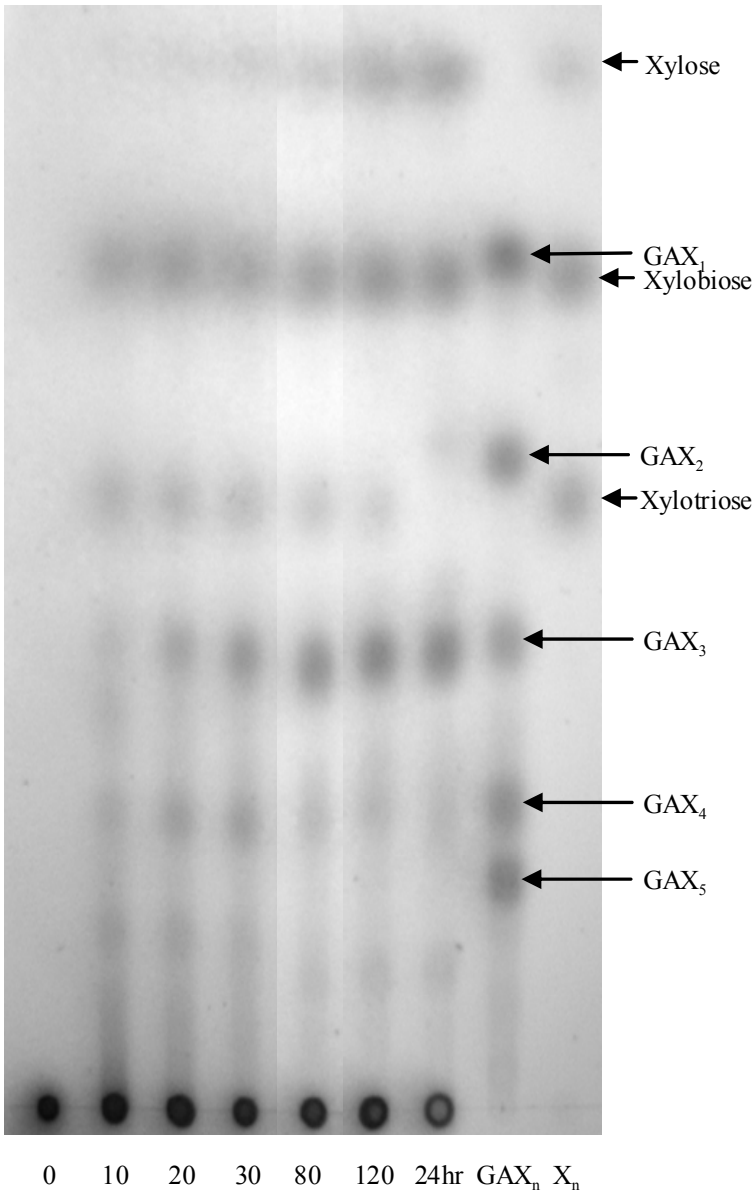

Figure 4-7. Kinetic analysis of product formation catalyzed by $XynA_1$ CD hydrolysis of SG $MeGAX_n$. A 250 µl reaction containing 20 mg/ml SG $MeGAX_n$ in 10 mM KH_2PO_4, pH 6.5 was digested with 0.5 units of $XynA_1$ CD. The reaction was incubated at 30 ºC with sampling every 10 minutes. Another 0.5 unit $XynA_1$ CD was added after 2 hours then incubated overnight. Samples (5 µl) were subjected to TLC at indicated times and resolved products detected as described in materials and methods.

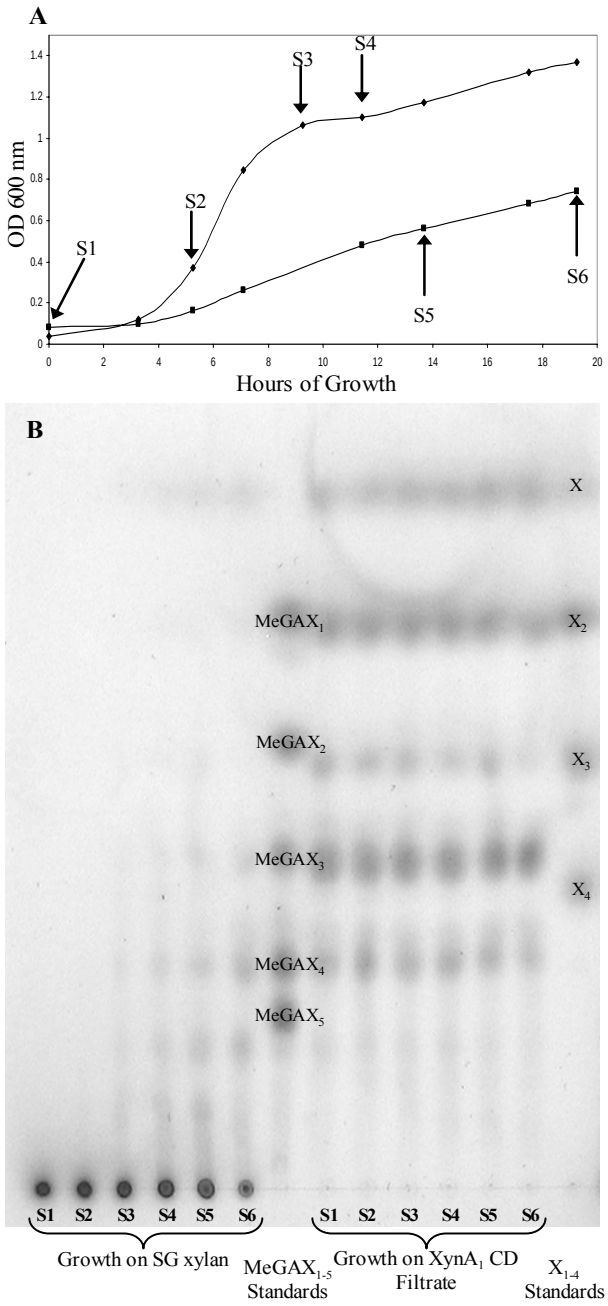

Figure 4-8. Differential carbohydrate utilization by *Paenibacillus* sp. strain JDR-2. Growth of *Paenibacillus* sp. strain JDR-2 was compared using SG $MeGAX_n$ and the concentrated, filter sterilized, YM-3 filtrate from an overnight $XynA_1$ SG $MeGAX_n$ digest as 1% substrates in Z-H mineral salts. (A) Growth of 50 ml cultures after inoculation with 4 ml from an overnight culture of *Paenibacillus* sp. strain JDR-2 in 0.5 % SG $MeGAX_n$ in Z-H mineral salts. Incubation at 30°C at 150 rpm in 125 ml baffle flasks and 0.25 ml samples (S1 through S6) removed at the times indicated. (B) TLC analysis of 6 µl aliquots of S1 through S6 supernatants after microcentrifugation and heat inactivation of xylanase activity at 70°C for 10 minutes.

CHAPTER 5
CHARACTERIZATION OF XynC FROM *Bacillus subtilis* SUBSPECIES *subtilis* STRAIN 168 AND ANALYSIS OF ITS ROLE IN DEPOLYMERIZATION OF GLUCURONOXYLAN

Introduction

A version of this chapter has been accepted for publication as a peer reviewed manuscript in a December 2006 issue of Journal of Bacteriology.

Biocatalyst production of value-added products and fuel ethanol from lignocellulosics is being developed as an alternative to traditional chemical synthesis methods. Despite the obvious environmental advantages, these 'green' processes need to be made less expensive to become the method of choice for chemical production (Arato et al., 2005; Ghosh and Ghose, 2003; Ingram et al., 1999; Lynd et al., 2005; Sedlak and Ho, 2004; Sun and Cheng, 2002; Zhou et al., 2003b). Hardwood and crop residues are attractive underutilized lignocellulosic resources which could be used for the generation of fermentable hexoses and pentoses, the former resulting from the cellulose fraction as glucose and the latter from the hemicellulose fraction as xylose with minor amounts of arabinose. The primary component of hemicellulose in hardwood and crop residues is 4-O-methyl-D-glucuronoxylan ($MeGAX_n$) in which xylose residues in the β-D-1,4-xylan polymer are substituted periodically with α-1,2-linked 4-O-methyl-D-glucuronopyranosyl residues (MeGA) (Jacobs et al., 2001; Preston et al., 2003; Timell, 1967). Significant limitations with current lignocellulose bioconversion lie in the pretreatment process required to liberate utilizable sugars from the complex hemicellulose fraction. Industrial protocols for fermentation of plant biomass generally require solubilization and hydrolysis of the hemicellulose fraction in dilute acid (0.5%-1.5% sulfuric acid) at high temperature (120°C to 140 °C) (Lloyd and Wyman, 2005). In addition to generating fermentation-inhibiting compounds such as furfural (Martinez et al., 2000; Zaldivar et al., 1999), this process does not efficiently cleave the α-1,2 glycosidic bond linking MeGA to xylose , and the resulting aldobiuronate, 4-O-methyl-D-

glucuronopyranosyl-α-1,2-D-xylose ($MeGAX_1$), is not fermented by ethanologenic bacterial biocatalysts currently used to convert lignocellulosic biomass to ethanol (Jones et al., 1961; Preston et al., 2003). These limitations may be overcome if a milder pretreatment process is developed along with a biocatalyst that has abilities for the depolymerization and assimilation of complex sugars resulting from partial depolymerization of $MeGAX_n$ (Qian et al., 2003). Enzyme systems have been identified in gram-positive spore forming bacteria, e.g., *Geobacillus stearothermophilus* (Shulami et al., 1999) and *Paenibacillus* sp. strain JDR-2 (St. John et al., 2006) (G. Nong, V. Chow, J. D. Rice, F. St. John, J. F. Preston, Abstr. 105th ASM General Meeting, abstr.O-055, 2005) that allow efficient depolymerization and assimilation of $MeGAX_n$, and complete catabolism of xylose and MeGA. The expression of these $MeGAX_n$-utilization systems in fermentative bacteria is needed to develop biocatalysts for the efficient conversion of the hemicellulose fraction to biobased products and alternative fuels.

Of the Gram-positive endospore-forming bacteria, *B. subtilis* has been the most extensively characterized, and serves as a model, both for bacterial sporulation processes as well as a general Gram-positive characteristic phenotype (Piggot and Hilbert, 2004; Volker and Hecker, 2005). Originally thought to be a strict aerobe, recent studies have shown that *B. subtilis* is a facultative anaerobe having limited growth under nitrate/nitrite respiration and fermentative growth conditions (Nakano et al., 1997; Nakano and Zuber, 1998). Although aerobic respiration allows for greater growth rate and cell yield than anaerobic respiration or fermentation, *B. subtilis* has broad metabolic potential as judged by fermentation pathway assessment (Cruz Ramos et al., 2000). Additionally, this organism is well known for robust secretion of extracellular proteins and several strains have been used for commercial enzyme production (Schallmey et al., 2004). Genetic manipulation of *B. subtilis* is relatively direct and it is well studied for its ability to take-

up and assimilate foreign DNA when flanked by sequences homologous to the target integration site (Anagnostopoulos and Spizizen, 1961; Niaudet et al., 1982; Spizizen, 1958). Further, there are now several stable theta-replicating vectors available for plasmid manipulation (Nguyen et al., 2005; Titok et al., 2003). The completed genome sequence of *Bacillus subtilis* subsp. *subtilis* str. 168 (*B. subtilis* 168) (Kunst et al., 1997) has provided a blueprint with which to determine its genetic potential and develop cogent strategies to engineer desired metabolic potential. With these properties, *B. subtilis* is a candidate to serve as a platform with which to develop biocatalysts for the direct conversion of lignocellulosic biomass to biobased products.

Genomic review of *B. subtilis* 168 reveals many enzymes specific for the degradation of plant cell wall polysaccharides (http://afmb.cnrs-mrs.fr/CAZY/). To characterize the ability of *B. subtilis* 168 to utilize $MeGAX_n$ and determine a starting point for engineering of $MeGAX_n$ utilization enzyme systems, the genomic data was evaluated for possible $MeGAX_n$ hydrolytic enzymes. We found no protein homolog for a GH 10 β-xylanase or a GH 67 α-glucuronidase, both of which are presumed to be required for complete utilization of $MeGAX_n$ (Preston et al., 2003; Shulami et al., 1999; St. John et al., 2006). The *xynA* gene in *B. subtilis* 168 has been shown to encode a GH 11 xylanase, XynA (Hastrup, 1988; Lindner et al., 1994; Miyazaki et al., 2006; Wolf et al., 1995). Members of this family are known to generate xylobiose and xylotriose, along with the aldopentauronate, 4-O-methylglucuronosyl-α-1,2-xylotetraose ($MeGAX_4$), in which the MeGA is linked to a xylose residue penultimate to the nonreducing terminus (Biely et al., 1997). $MeGAX_4$ is not known to be assimilated and metabolized without first removing the xylose residue at the nonreducing terminus (Nagy et al., 2002; Nurizzo et al., 2002; St. John et al., 2006). Further genome review also identified the *ynfF* gene whose translated protein product has 40% identity and 60% similarity to XynA of *Erwinia*

chrysanthemi D1, a GH 5 endoxylanase that has been well characterized (Hurlbert and Preston, 2001; Larson et al., 2003; Preston et al., 2003).

GH 5 xylanases are not as prevalent as GH 10 and GH 11 xylanases which have been extensively studied (Biely et al., 1997; Clarke et al., 1997; Henrissat and Davies, 1997; Pell et al., 2004b). Just as GH 10 and GH 11 xylanases, they are presumed to function through a pair of glutamate residues catalyzing hydrolysis by a double displacement mechanism with retention of anomeric configuration (Gebler et al., 1992; Larson et al., 2003). Although their general protein fold is the same as other GH 5 categorized glycosyl hydrolases (Henrissat and Bairoch, 1993; Henrissat et al., 1995; Henrissat and Davies, 1997), GH 5 xylanases are more similar in primary sequence to GH 30 hydrolases (Collins et al., 2005). Recently, the first crystal structure of a GH 5 xylanase was published (Larson et al., 2003). The catalytic domain (CD) contains an α/β_8 barrel similar to that in the GH 10 xylanases but closely associated with this domain is a putative carbohydrate binding module (CBM). The β-sheet structure of the CBM is formed with N-terminal and C-terminal regions suggesting that the CD and CBM may function synchronously together, rather than having a simple spatial relationship that has been suggested for many other CD and CBM associations (Boraston et al., 2004).

Unlike GH 10 and GH 11 xylanases, GH 5 xylanases have not been linked to the process of $MeGAX_n$ degradation and utilization in microbial ecosystems. In only one report has the role of a GH 5 xylanase been postulated. XynA from the well characterized pectinolytic phytopathogen *Erwinia chrysanthemi* D1 was suggested to facilitate access to the pectin component of biomass (Braun and Rodrigues, 1993; Hurlbert and Preston, 2001), although a *xynA* knockout showed no detectable decrease in measurable virulence on corn leaves (Keen et al., 1996). Additional studies with XynA led to the first substrate and product characterization of

a GH 5 xylanase showing XynA to have an activity that correlated directly to the degree of substitution of the glucuronosyl moiety, and that reduction of the carboxyl carbon of MeGA greatly reduced XynA activity (Hurlbert and Preston, 2001). In this previous work, all hydrolysis products resulting from sweetgum (SG) $MeGAX_n$ hydrolysis by XynA contained a single MeGA moiety as determined by ^{13}C NMR and biochemical analysis, but early MALDI MS results were difficult to reconcile and interpretation predicted two MeGA substitutions per hydrolysis product (Hurlbert and Preston, 2001). Previous reports of xylanases with sequences homologous to the GH 5 defined in *Erwinia chrysanthemi* have been reported, although characterization has been limited with respect to substrate specificities and product formation (Ito et al., 2003; Okai et al., 1998; Suzuki et al., 1997). An earlier report (Nishitani and Nevins, 1991) concerned a 42 kDa xylanase (XynC is 43.9 kDa) purified from a commercial *B. subtilis* enzyme preparation (Novo Ban L-120). This publication preceded the current classification system of Carbohydrate-Active Enzymes (CAZY-http://afmb.cnrs-mrs.fr/CAZY/) (Coutinho and Henrissat, 1999; Henrissat and Bairoch, 1993; Henrissat et al., 1995; Henrissat and Davies, 1997). However, hydrolysis product analysis comparisons to our own observations suggest that it may have been a GH 5 xylanase with similar properties to XynC of *B. subtilis* 168. Based on thorough product analysis, this may have been the first GH 5 xylanase purified and characterized. However, because there is no sequence data available and no identified *B. subtilis* strain associated, we were unable to directly compare their findings to our own.

In this study, I expressed and characterized the *ynfF* gene product (XynC) from *B. subtilis* 168. This is the second characterization of a true GH 5 xylanase showing no β-1,4-glucanase activity on carboxymethyl cellulose (Collins et al., 2005), with the first complete characterization of an endoxylanase classified as a member of glycosyl hydrolase family 5 from a gram-positive

bacterium. The secretion of both XynA and XynC and the utilization of xylooligosaccharides for growth make *B. subtilis* an attractive candidate to genetically transform and define the requirements for the complete utilization of $MeGAX_n$ for conversion to biobased products.

Materials and Methods

***B. subtilis* 168 *ynfF* cloning and XynC Purification**. *Bacillus subtilis* subsp. *subtilis* str. 168 (*B. subtilis* 168) was obtained from the Bacillus Genetic Stock Center (http://www.bgsc.org). Genomic DNA from overnight cultures grown in Luria-Bertani (LB) broth at 37°C with shaking was extracted using a DNeasy Tissue Kit (Qiagen, Valencia, CA), and the *ynfF* gene was amplified using the ProofStart DNA polymerase PCR kit (Qiagen) for directional cloning into the NcoI and XhoI restriction sites (highlighted by underlined region in primer sequence) of the pET 41 expression vector (EMD Biosciences). The 5' primer (ttccatggcagcaagtgatgtaacagttaatg) was designed to truncate the XynC secretion signal sequence as predicted using SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) (Bendtsen et al., 2004) and create an in frame fusion behind the affinity purification tags of pET 41. The 3' primer (gttctcgagttaacgatttacaacaaatgttgt) contained the last few nucleotides of *ynfF* including the stop codon. The expression construct (pET41ynfF#1) contains a GST-Tag, a His-Tag, a thrombin cleavage site, an S-Tag, an enterokinase cleavage site, and the *ynfF* gene which was verified by sequencing. The pET41ynfF#1 construct was introduced into chemically competent *Escherichia coli* Rosetta (DE3) (EMD Biosciences) and grown overnight at 37°C on LB agar plates containing 30 µg kanamycin/ml and 34 µg chloramphenicol/ml (LBKC medium). Transformants grown on LBKC medium at 35°C were induced with 1 mM IPTG when the OD 600 nm reached 0.6 to 0.7, and were processed for XynC purification as recommended in the pET System Manuel, 10th Edition (EMD Biosciences). Enzyme was isolated using a HiTrap HP metal chelating column in the nickel form as per the method outlined in the HiTrap Chelating HP

instruction manual (GE Healthcare Bio-Sciences). Xylanase activity was localized in the elution buffer fraction containing 500 mM sodium chloride, 500 mM imidazole and 20 mM sodium phosphate, pH 7.4. To minimize protein precipitation associated with rapid desalting of this particular protein, the fraction was dialyzed in 10,000 MWCO tubing (Pierce Biotechnology) in a five step series in which the elution buffer was successively diluted 2-fold with 50 mM Tris-HCl, pH 7.4, at 45 minute intervals. The resulting fraction was further dialyzed with 2, 2-fold volume dilutions from the 50 mM Tris-HCl, pH 7.4 with enterokinase cleavage buffer (20 mM Tris-HCl, pH 7.4, 50 mM NaCl and 2 mM $CaCl_2$) and finally with undiluted enterokinase cleavage buffer. Following overnight digestion with enterokinase at room temperature and filtration through a 0.45 μm filter, the preparation was dialyzed against GST-Bind buffer, the GST-Tag removed with GST-Bind agarose beads, and site-specific protease enterokinase removed with EKapture agarose (EMD Biosciences) equilibrated in GST-Bind buffer. At each step protein concentration was determined (Bradford, 1976) using BSA (fraction V) as the standard, purity was determined by SDS-PAGE analysis (Laemmli, 1970) and xylanase activity was determined as described below. The final product was stored in GST-Bind buffer at 4°C on ice and exhibited no loss of activity after 11 months.

MeGAX$_n$ substrate preparation and carbohydrate analysis. MeGAX$_n$ was isolated from sweetgum wood (SG MeGAX$_n$) as previously described (Jones et al., 1961) and characterized by 13C NMR (Kardosova et al., 1998). Birch and beech wood xylans were obtained from Sigma-Aldrich. MeGAX$_n$ was prepared from these xylans by solubilization of 20 mg dry weight xylan/ml deionized water for 5 hours at 50°C to 60°C, centrifuged at 30,000 g for 25 minutes at room temperature, and supernatants decanted for carbohydrate analysis and use as reaction substrates. Xylose standards were used for total carbohydrate measurements (Dubois et

al., 1956) and determination of reducing termini (Nelson, 1944). D-Glucuronic acid standards were used for total uronic acid determination (Blumenkrantz and Asboe-Hansen, 1973). The degree of polymerization (DP) and the degree of uronic acid substitution (DS) for each substrate were determined. Xylanolytic activities of XynC toward these substrates were compared using the optimized reaction conditions for SG MeGAX$_n$ described below.

Optimization of activity and kinetic evaluation of XynC. XynC activity was determined by measuring the increase in reducing termini (Nelson, 1944) resulting from depolymerization of SG MeGAX$_n$ substrate (10 mg substrate/ml). Reactions containing 250 µl substrate in 50 mM potassium phosphate or 50 mM sodium acetate with 200 ng XynC were run from pH 5.0 through pH 6.5 at 37°C. The temperature optimum was determined with reactions in 50 mM sodium acetate, pH 6.0, over the range from 50°C through 70°C. Specific activities are given as units/mg XynC, where one unit equals one µmole of reducing termini generated per minute. The kinetic analyses were in 250 µl reaction mixes containing 200 ng XynC in 50 mM sodium acetate, pH 6.0, with SG MeGAX$_n$ as substrate at 37°C. Thermal stability assays were conducted with 250 µl reaction mixes containing 2.5 mg SG MeGAX$_n$ and 150 ng XynC in 50 mM sodium acetate, pH 6.0, between 30°C and 60°C. Aliquots were sampled four times in the first hour with decreasing sampling frequency through 60 hours of incubation.

XynC catalyzed depolymerization of MeGAX$_n$ for product analysis. A batch reaction of 20 ml was prepared consisting of 50 mg SG MeGAX$_n$/ml in 50 mM sodium acetate, pH 6.0. The reaction was initiated by the addition of 1 unit of XynC in a volume of 100 µl and maintained for 18 hours at 30°C with slow vertical rotation on a roller drum type rotator. The digest was processed by collecting consecutive filtrates from a Centriprep YM-3 centrifugal filter device (Millipore, Billerica, MA) by centrifugation at 2000 x g at 20°C in one hour intervals.

After every other centrifugation, the volume in the filter unit was refilled to 15 ml with deionized water and filtrates combined to a total volume of 15 ml. Four, 15 ml filtrate fractions, were collected as well as a final, 6 ml filtrate fraction. The resulting retentate was heated to 70ºC for 15 minutes to inactivate XynC and was further processed twice by diluting to 50 ml with deionized water and concentrating in a 50 ml stir cell (Millipore, Billerica, MA) with a YM-3 membrane. The first recovered 15 ml filtrate (Filtrate) from the Centiprep YM-3 concentrator and the final washed retentate (Retentate) are the primary focus of the studies reported here. To verify compatibility of XynC with YM-3 membranes, which are made of regenerated cellulose acetate, XynC was checked for endogluconase activity by using carboxymethyl cellulose as substrate. No activity was detected.

MALDI TOF MS analysis of XynC-generated MeGAX$_n$ hydrolysis products. Samples for MALDI TOF MS analysis were prepared as described previously (Rydlund and Dahlman, 1997). The matrix, 2,5-dihydroxybenzoic acid (DHBA), was dissolved to 5 mg/ml in filtered deionized water and diluted with an equal volume of 100% acetonitrile. Carbohydrate solutions, 1 to 2 mg/ml containing 0.1% trifluoroacetic acid (TFA), were prepared just prior to analysis by adding 1 µl 10% TFA to 100 µl carbohydrate solutions. Ten µl of the carbohydrate-0.1% TFA solution was mixed with 10 µl of the matrix acetonitrile solution, and 1 µl was spotted on the MALDI TOF MS plate. Mass spectrum data was collected using a Voyager-DE Pro (Applied Biosystems, Foster City, CA) at the University of Florida Protein Core facility. The mass spectrometer was set for positive polarity in the reflector mode and the acceleration voltage set to 18000V with a laser intensity of 2500. Data was converted into ASCII format and analyzed in Excel.

NMR analysis of XynC-generated $MeGAX_n$ hydrolysis products. Samples for ^{1}H-NMR were prepared by three successive dissolutions in 4 ml 99.9 atom percent D_2O (Sigma-Aldrich), each with subsequent lyophilization. Each time the lyophilized $MeGAX_n$ XynC hydrolysate residue was dissolved in D_2O, it was warmed to 50ºC for 15 minutes to enhance ^{1}H displacement with deuterium. Final carbohydrate samples were prepared by dissolving the lyophilized carbohydrate powder to a concentration of 15 mg/ml in D_2O. To 750 µl of these preparations, 2.5 µl of acetone was added as reference and the final samples transferred to Wilmad 505-PS NMR tubes (Wilmad, Buena, NJ). ^{1}H-NMR data collection was performed using a Bruker Avance 500 MHz spectrometer with a 5 mm TXI probe at the Advanced Magnetic Resonance Imaging and Spectroscopy (AMRIS) facility at the McKight Brain Institute, University of Florida. NMR data was analyzed and images prepared using Nuts Lite (Acorn NMR Inc., Livermore, CA).

Fractionation and analysis of xylanase activities secreted by *B. subtilis* 168. A single colony of *B. subtilis* 168 from an overnight culture on LB agar plate was inoculated into 25 ml of 1% YE, Spizizen salts (YES medium)(Spizizen, 1958) in a 250 ml baffle flask and grown for 15 hours at 37°C with rotating on a New Brunswick G-2 gyrotory shaker at 200 rpm. Twenty ml of this culture was inoculated into a Fernbach flask containing 1 liter of prewarmed YES medium and incubated as above until an OD 600 of approximately 1.5 absorbance units was obtained. Cells were removed by centrifugation in a Sorvall GSA rotor at 13,200 x g for 20 minutes at 4°C. The supernatant was filtered through a 0.45 µm filter and 1 ml of protease inhibitor cocktail for bacterial cell extracts (Sigma, St. Louis, MO) added. This preparation was concentrated/dialyzed using a 350 ml Amicon stir cell with a YM-3 membrane (Millipore, Billerica, MA) with 20 mM sodium phosphate, pH 6.0, containing 150 mM NaCl. The YM-3

retentate (BSC) was concentrated to less than 5 ml and loaded on a calibrated BioGel P-60 chromatography column (60 cm by 0.75 cm) in the above buffer system. Fractions comprising two peaks of xylanase activity (Fraction A and Fraction B) were concentrated/dialyzed separately to less than 1 ml against 20 mM sodium phosphate, pH 6.0, with Amicon YM-3 centrifugal filtration devices (Millipore, Billerica, MA). Equal volumes of each fraction were reacted with 20 mg SG $MeGAX_n$ /ml in 50 mM sodium acetate, pH 6.0, and incubated overnight at 37°C. TLC was performed as previously described (Bounias, 1980; St. John et al., 2006). Controls for TLC included undigested SG $MeGAX_n$, SG $MeGAX_n$ digested with recombinantly expressed XynC (330 nmoles xylose equivalents loaded for samples and controls), $MeGAX_{1-4}$ oligomers and X_{1-4} xylooligomers (25 nmole xylose equivalents for each oligomer).

Quantification of levels of xylanase transcripts by quantitative reverse transcriptase PCR. Primers for quantitative reverse transcriptase PCR (Q-RT-PCR) were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky, 2000), with simultaneous secondary structure analysis using mfold (http://www.bioinfo.rpi.edu/applications/mfold/) to avoid DNA secondary structure in priming sites (Zuker, 2003). Primers of approximately 20 nucleotides were designed to have an annealing temperature of 60ºC and mfold secondary structure modeling was performed at 59ºC under simulated conditions normal for PCR (50 mM Na+ and 3 mM Mg++). Several other primers were designed without such purpose but yielded good amplification efficiencies. All primer pairs yielded efficiencies between 85% and 95%, and all amplicons were less than 200 bp. High quality genomic DNA of *B. subtilis* 168 was obtained for use in making primer set standard curves (Tsai and Olson, 1991). Standard curves were prepared from 102 to 106 gene copies for all genes. The reference genes were extended to 108 copies based on their possible

relative levels to the genes of interest. All RT data collections were performed using a Bio-Rad iCycler (Bio-Rad Laboratories, Hercules, CA). For primer set optimization and standard curve analysis, Bio-Rad iQ SYBR Green Supermix was used, and for Q-RT-PCR Bio-Rad iScript One-Step RT-PCR kit (Bio-Rad Laboratories) was used. Reaction volumes were 16 µl for both kits. Cultures for RNA extraction (25 ml) were grown in 250 ml baffle flasks at 37°C rotating on a New Brunswick G-2 gyrotory shaker at 225 rpm in media with glucose, arabinose, arabinose and xylose, SG $MeGAX_n$ or birch $MeGAX_n$ at 0.5% (xylose was at 0.25%) supplemented with 0.1% YE and 0.005% tryptophan in Spizizen minimal salt base (Spizizen, 1958). RNA was extracted in early to mid log phase as judged from previous growth curves, using an RNeasy RNA extraction kit (Qiagen, Valencia, CA). Extracted RNA was subsequently DNase treated using RQ1 DNase (Promega, Madison WI) and re-purified using an RNeasy column. As a control, 10 ng of RNA was used to amplify the *rpoA* gene with no reverse transcriptase to check for DNA contamination. In most cases the DNase treatment had to be repeated before the negative control showed insignificant DNA contamination. All data obtained from Q-RT-PCR was based on 10 ng RNA starting quantity as determined by absorbance at 260 nm. For relative expression analysis *rpoA* and *atpA* genes were used. Selection of these genes as reference housekeeping genes for relative analysis was judged from triplicate measurements of each growth condition averaged together. Threshold values for *atpA* in all conditions averaged 17.37 CT with a standard deviation of 0.31. For the *rpoA* gene, the average was 15.53 CT with a standard deviation of 0.53. Based on this, these genes were used as references to analyze all transcript data in Gene Expression Macro V1.1 software available through Bio-Rad (Bio-Rad Laboratories) (http://www.bio-rad.com/LifeScience/jobs/2004/04-0684/genex.xls) (Livak and Schmittgen, 2001; Schmittgen and Zakrajsek, 2000; Vandesompele et al., 2002).

Results

Cloning, overexpression and characterization of XynC. An initial cloning attempt was designed for directional cloning of *xynC* into pET15 with an N-terminal His-Tag fusion. The cloning was successful as revealed by DNA sequencing and activity analysis but the approach failed in that the correctly fused His-Tag failed to allow affinity purification on a nickel chelating column. Denaturation with 6 M guanidine-HCl allowed purification of the gene product containing the His-Tag by nickel affinity chromatography, indicating that the affinity tag was rendered inaccessible during maturation of the recombinant active enzyme. This problem was solved by re-cloning XynC using pET41 as described in the Materials and Methods.

The second cloning attempt was successful and resulted in pure protein in amounts sufficient for enzymatic characterization. The N-terminal His-Tag was used for affinity purification to verify proper expression product conformation. Purification resulted in a protein which was relatively insoluble at low ionic strengths. Qualitatively, it maintained solubility best in tertiary amine buffers, such as 50 mM Tris-HCl or imidazole, containing at least a 20 mM strong electrolyte such as NaCl. After enterokinase protease release of the affinity tag fusion product from XynC, there were no further apparent problems with solubility. Although the method employed here provided adequate quantities of XynC for characterization, better yields may be obtained in future purifications by performing the enterokinase reaction with higher buffer concentrations to limit the precipitation of the GST-XynC fusion protein.

Optimization and kinetic evaluation of XynC. Optimization of XynC was performed in standard reactions with 50 mM potassium phosphate or 50 mM sodium acetate buffers ranging from pH 5.0 to 6.5. Figure 5-1A shows that highest activity was observed using potassium phosphate buffer at pH 6.0. Nearly the same activity was obtained in 50 mM sodium acetate buffer, pH 6.0. Since previous studies for the characterization of the GH 5 XynA of *Erwinia*

chrysanthemi D1 used 50 mM sodium acetate, pH 6.0, this buffer was selected for further studies with the XynC from *B. subtilis* 168. The relationship between temperature and activity (Figure 5-1B) showed the maximum amount of reducing termini formation for a 15 min digestion was obtained at 65°C. The relationship between stability and temperature was evaluated with half-life stability analysis (Figure 5-1C) as described in Materials and Methods. This suggested that although XynC appears to be relatively thermo-tolerant, the most reliable kinetic data would be obtained at lower temperatures. From this, all other reactions including kinetic analyses were performed at 37ºC in 50 mM sodium acetate, pH 6.0, for 15 minutes with 0.012 units (200 ng) XynC.

Kinetic analysis (Figure 5-1D) of XynC using SG $MeGAX_n$ for substrate showed it to have a K_m of 1.63 mg/ml and a V_{max} of 59.5 units/mg XynC, which corresponds to a k_{cat} of 2635/ minute. As seen in Table 5-1, measurements of XynC activity on different $MeGAX_n$ sources indicate that activity is directly correlated to the degree of substitution of the glucuronosyl moiety on the xylan chain.

Mass analysis of SG $MeGAX_n$ XynC hydrolysis products. YM-3 Filtrate and Retentate fractions of the SG $MeGAX_n$ XynC hydrolysate were analyzed by MALDI TOF MS. As shown in Figure 5-2, clusters of peaks were observed, with each cluster being a defined aldouronate with the individual peaks being some salt adduct form of the specific aldouronate. Table 5-2 lists the mass to charge ratios and designated products along with the ion adducts. In each case, the species with a single Na adduct was most prominent. Based upon these assignments, the YM-3 filtrate and YM-3 retentate fractions contain detectable xylooligosaccharides ranging in DP from 4 to 12, and 4 to 20, respectively, each with a single MeGA substitution. There was no evidence for the presence of unsubstituted xylooligosaccharides in these fractions.

NMR analysis of SG MeGAX$_n$ XynC hydrolysis products. Analysis of the filtrate by ^{1}H-NMR (Figure 5-3A) yielded the expected distribution of peaks based on previous proton shift assignments for similar glucuronosyl substituted xylooligomers (Cavagna et al., 1984; Excoffier et al., 1986; Kardosova et al., 1998). The integration ratio of the signals for the proton on carbon five of the MeGA (U5), the proton on carbon five of the non-reducing xylose (nr-X5) and the proton on carbon one of the reducing xylose (α and β configurations) (α/β-X1) was 1.0: 1.1: 0.9, establishing that the products of XynC depolymerization contained a single MeGA moiety α-1,2-linked to β-1,4-xylooligosaccharides. The signals ascribed to the proton of carbon one for the MeGA residue (α/β-U1) appears as two doublets, the first at 5.31 and 5.30 and the second at 5.29 and 5.28 ppm. The integrated ratio for these two doublets is 0.23:0.77. This is nearly identical to the α/β intensity split of 0.26:0.74 for the integrated signals from the proton on the α anomer of carbon one (α,r-X1) (doublet at 5.18 and 5.17 ppm) and the proton on the β anomer of carbon one (β,r-X1) (doublet at 4.58 ppm and 4.63 ppm) for the reducing terminal xylose. Thus, the doublet shift values for the proton of carbon one of the MeGA (α/β-U1) that is α-1,2-linked to a xylose residue are split to a ratio that reflects the equilibrium of the α and β anomers of the xylose residue at the reducing terminus. As stated above, this interpretation is consistent with other published interpretations of ^{1}H-NMR spectra of xylooligomers substituted with MeGA, and indicates the substitution is penultimate to the reducing terminal xylose (Cavagna et al., 1984; Excoffier et al., 1986). Figure 5-3B presents the predicted limit products based on the above ^{1}H NMR observations showing the expected positioning of the MeGA moiety with respect to the reducing end xylose and some number (n) of β-1,4-linked xylose residues.

Fractionation of *B. subtilis* 168 spent medium concentrate and TLC analysis of xylanolytic peak fractions. The spent medium was processed as described in the Materials and

Methods and the resulting total xylanase yield was approximately 20 units. Following concentration by YM-3 filtration, 2.9 units of the *B. subtilis* spent medium concentrate (BSC) were loaded onto a BioGel P-60 column. Two xylanase positive peaks, Faction A and Fraction B, were collected and processed as described above. TLC analysis (Figure 5-4) of the SG $MeGAX_n$ digestions with these fractions revealed two unique hydrolysis patterns. Fraction A eluted at a position expected for a globular protein with a molecular mass of approximately 31 kDa. The activity in this fraction catalyzed the depolymerization of SG $MeGAX_n$ to form products resolved and detected by TLC that were identical to those generated by the recombinantly expressed XynC protein. Hydrolysis products of XynC did not include any detectable neutral sugars (Figure 5-4). Fraction B eluted at a position that corresponded to a molecular mass, extrapolated from the relationship of log M_r vs. elution position for standards, of approximately three kDa. This activity catalyzed the generation of products from SG $MeGAX_n$ that would be expected for a typical GH 11 xylanase, releasing xylobiose, xylotriose and aldopentauronate ($MeGAX_4$) as primary limit products (Biely et al., 1997). Based upon the translated nucleotide sequences, the expected M_r values for XynA and XynC are 20.4 kDa and 43.9 kDa, respectively, indicating that interactions with the BioGel P-60 significantly affect the elution that would be expected for resolution by gel permeation based upon molecular size. SG $MeGAX_n$ digestion with the BSC fraction generated products that were primarily a result of XynA hydrolysis, but also included aldotetrauronate ($MeGAX_3$) (Figure 5-4).

Relative levels of xylanase transcripts determined by Q-RT-PCR. To study xylanase gene expression in *B. subtilis* 168 the genes *xynA*, *xynB*, *xynC*, *abnA* and *gapA* were analyzed. The *abnA* gene codes for an extracellular arabinofuranosidase and is well studied, being highly expressed while *B. subtilis* is growing on arabinose and repressed by glucose (Raposo et al.,

2004). The *gapA* gene which codes for the glycolytic glyceraldehyde-3-phosphate dehydrogenase in *B. subtilis*, was reported to be upregulated while *B. subtilis* was growing on glucose (Fillinger et al., 2000; Moreno et al., 2001). The *abnA* and *gapA* genes served as valid internal controls. This was confirmed in our studies where *abnA* showed the greatest dynamic range of all genes analyzed (Table 5-3). Our results also support previous findings for the *xynB* gene, showing that expression was greatest while growing on xylose. However, the same effect was not observed for *xynA* (Lindner et al., 1994) and *xynC*. Changes in transcript levels for *xynA* and especially *xynC* were modest with respect to *abnA* (Figure 5-5). The *xynA* transcript was in greatest quantity while cells were growing on $MeGAX_n$ as substrate and lowest while growing on glucose, suggesting that *xynA* expression may be activated by growth on $MeGAX_n$. In contrast, the *xynC* transcript was constitutively expressed without significant changes in regulation observed while growing on the different sugars.

Discussion

XynC from *B. subtilis* 168 is the first GH 5 xylanase from a Gram-positive bacterium to be fully characterized with respect to substrate requirements and resulting hydrolysis products. This report is the first to present kinetic constants for a xylanase in this family. The turnover rate of 2635/ minute is close to that observed by the GH 5 xylanase (XynA) from the Gram-negative bacterium *Erwinia chrysanthemi* str. PI (data not shown), suggesting that xylanases in this family may be highly conserved in function. A review of publications concerning GH 10 and GH 11 xylanases in which a V_{max} or k_{cat} is given reveals a very broad range of catalytic rates (Charnock et al., 1998; Chaudhary and Deobagkar, 1997; Elegir et al., 1994; Gupta et al., 2000; Khasin et al., 1993; Pell et al., 2004a; Preston et al., 2003). To some extent, these differences may result from differences in assay conditions used in different laboratories. It can only be stated that GH 5 xylanases have a catalytic rate which is at the lower range of the rates reported in the literature

for GH 10 and GH 11 xylanases. Although all kinetic studies were performed at 37°C, half-life analysis showed that XynC has a $t_{1/2}$ at 50°C of greater than five hours, a property that supports its application in preprocessing of lignocellulosic biomass.

Both the MALDI TOF MS data and the ^{1}H-NMR data indicate that the GH 5 xylanase, XynC, catalyzes the depolymerization of $MeGAX_n$ to release products in which β-1,4-linked xylooligosaccharides are substituted with a single α-1,2-linked MeGA. MALDI TOF MS analysis of products generated by XynC revealed an array of peak clusters, each differing by a single xylose residue. Tabulation of mass data showed that single salt adducts differed from double salt adducts by a single mass unit, indicating that formation of adducts results from proton displacement. Studies in which this technique for carbohydrate analysis was developed (Stahl et al., 1991) observed single salt adducts for neutral polysaccharides but no double salt adducts. The occurrence of double salt adducts may result from the presence of the carboxylic acid. Each detected species in MALDI TOF MS consisted of some number of xylose residues substituted with a single MeGA substitution. This interpretation calls into question the interpretation of the MALDI TOF MS data for products generated from the action of the GH 5 endoxylanase from *Erwinia chrysanthemi* str. D1 (Hurlbert and Preston, 2001). Recent studies with the GH 5 endoxylanase from *Erwinia chrysanthemi* str. PI have applied MALDI TOF MS to identify the products generated from the depolymerization of SG $MeGAX_n$ along with potassium ion supplements (J. Rice, G. Nong, A. Ragunathan, F. St. John, J. P. Preston, Abstr. 105th ASM General Meeting, Abstr. B-138, 2005). Their results support the assignments made in Table 5-2 and the interpretation provided here. Further support for this interpretation comes from the ratio, 1.0: 1.1: 0.9, of the integrated signals for the proton on carbon five of the MeGA

(U5), proton on carbon five of the nonreducing terminal xylose (nr-X5) and the proton on carbon one of the reducing terminal xylose (α/β X1).

The evidence for the position of the xylose residue that is substituted with MeGA in XynC-generated products is circumstantial in that it is based upon the NMR shift assignment peak intensities ascribed to an induction of an α/β resonance split in the α-1,2 linked MeGA moiety. This induction can be rationalized by a direct interaction with the proton on carbon one of the reducing terminal xylose residue, which is in anomeric equilibrium between α and β configurations. Such an effect has been interpreted to explain the observations obtained from the ^{1}H-NMR and 2D -^{1}H/^{13}C- NMR analyses of aldouronates in which the nonreducing terminal xylose in β-1,4-xylobiose, and the internal xylose in β-1,4-xylotriose, are substituted with α-1,2-linked MeGA (Cavagna et al., 1984; Excoffier et al., 1986). The products generated by a xylanase purified from a commercial preparation of *Bacillus subtilis* were previously characterized by methylation analysis that identified the MeGA substitution on the xylose residue penultimate to xylose at the reducing terminus (Nishitani and Nevins, 1991), and this activity may have been encoded by a gene homologous to *xynC*. This data identifies a site of cleavage that is different from that indicated from previous studies of the GH 5 endoxylanase secreted by the D1 strain of *Erwinia chrysanthemi* (Hurlbert and Preston, 2001), which was based upon ^{13}C-NMR and limited digestion by β-xylosidase. Two dimensional HMQC (Heteronuclear Multiple-Quantum Coherence) NMR spectra of the products generated by the GH 5 xylanase secreted by *Erwinia chrysanthemi* PI (J. Rice, G. Nong, A. Ragunathan, F. St. John, J. P. Preston, Abstr. 105th ASM General Meeting, Abstr. B-138, 2005) support the interpretation provided for the products generated by the XynC GH 5 endoxylanase secreted by *B. subtilis* 168. From this it may be concluded that GH 5 endoxylanases from both *B. subtilis*

168 and the *Erwinia chrysanthemi* strains catalyze the exclusive cleavage of a β-1,4-xylosidic bond penultimate to that linking carbon one of the xylose residue that is substituted with α-1,2-linked MeGA as depicted in Figure 5-3B.

Most work concerning xylanase activity in *B. subtilis* has involved the *xynA* gene whose protein product (XynA) has long been thought to be the primary extracellular xylanase (Wolf et al., 1995). Studies have found that *xynA* is constitutively expressed even while cells are growing on glucose (Lindner et al., 1994), suggesting that *xynA* regulation may not be subject to catabolite repression (Moreno et al., 2001) (supplemental data). The *xynC* gene was originally identified by genome sequencing of *B. subtilis* 168 (Kunst et al., 1997; Rose and Entian, 1996). Analysis of the fractions derived from the BioGel P-60 chromatography of BSC established that *B. subtilis* 168 produces XynC as a xylanase activity separable from that of XynA. These results are supported by a recent proteomic evaluation of the *B. subtilis* secretome which identifies XynA and XynC (and XynD) to be present as secreted protein products (Tjalsma et al., 2004).

Transcript quantification of the internal control genes, *abnA* and *gapA*, as measured by Q-RT-PCR while *B. subtilis* 168 was growing on various carbohydrates, reflect our expectations based on previous studies. The *abnA* gene is known to be activated by arabinose and repressed by glucose via a CcpA/*cre* interaction (Raposo et al., 2004) and the dynamic range of expression levels for the *abnA* transcript in this study support these previous reports. Further, the *gapA* gene, a glycolytic glyceraldehyde-3-phosphate dehydrogenase from *B. subtilis*, has been reported to be indirectly upregulated by CcpA (Fillinger et al., 2000) and our studies confirm these findings, revealing a 30-fold increase in *gapA* transcript quantity during growth with glucose over the growth containing arabinose and xylose (Table 5-3).

In our analysis, *xynC* is unresponsive to metabolite-mediated induction or repression, being expressed at a constant level while *B. subtilis* 168 is growing on a variety of carbohydrates, and *xynA* expression appears to be activated by growth on $MeGAX_n$. Glucose repression attributed to a *cre* motif has been identified in the GH 11 xylanase (XynA) from *Geobacillus stearothermophilus* str. No. 236 (Cho and Choi, 1999). Putative *cre* sites were identified and glucose repression was shown in their studies within *B. subtilis*. Alignment of this GH 11 shows high similarity to the nucleotide sequence of XynA from *B. subtilis* 168, including the regions of the predicted *cre* motif. Although this may suggest a possible role for glucose mediated CcpA repression of *xynA* in *B. subtilis* 168, results in our study do not support this possibility. Transcriptome analyses of *xynA* and *xynC* in *B. subtilis* wild type (str. ST100) and a *ccpA* mutant (str. ST101) showed little change with and without glucose as substrate (Moreno et al., 2001)(supplemental data). Unlike the xylose utilization operon of *B. subtilis* (Hastrup, 1988; Jacob et al., 1991; Kraus et al., 1994) and the above mentioned Q-RT-PCR control genes, neither of the two secreted xylanases of *B. subtilis* 168 seems to be repressed by glucose via the CcpA/*cre*-mediated mechanism and in contrast, *xynA* is upregulated 3-fold while growing on $MeGAX_n$ by a process as yet undefined, and *xynC* appears to be strictly constitutive.

GH 11 xylanases such as XynA are known to release xylobiose as a major limit product and it is thought that *B. subtilis* can utilize this xylooligomer (Hastrup, 1988; Lindner et al., 1994). However, the products of XynC hydrolysis of $MeGAX_n$ are too large for direct utilization. As can be observed in Figure 5-4, the hydrolysis of $MeGAX_n$ by XynA and XynC together (in the BSC) releases a smaller unique aldouronate which is defined by the activity these xylanases display as individual xylanases toward $MeGAX_n$ (Figures 5-6). By deduction, aldotetrauronate ($MeGAX_3$) released by this combined hydrolysis has a MeGA substitution on

the second xylose of xylotriose, positioning it penultimate from the reducing and nonreducing terminal xylose residues. This aldotetrauronate is unlike the well studied $MeGAX_3$ limit product of a GH 10 xylanase which is substituted with an MeGA directly on the nonreducing terminal xylose of β-1,4-xylotriose (Biely et al., 1997; Pell et al., 2004b), and which is suggested to be readily assimilated and metabolized by some Gram-positive bacteria (Shulami et al., 1999; St. John et al., 2006). This observation further supports the predicted limit product of XynC hydrolysis of $MeGAX_n$ as presented in Figures 5-3A and 5-6.

Genomic analysis of *xynC* revealed that it is the latter part of a bicistronic message. The upstream gene putatively codes for an extracellular GH 43 β-xylosidase, XynD. Analysis of this operon (in silico) predicts each open reading frame to have a promoter but only *xynC* to have a terminator, suggesting some level of coordinate expression. As mentioned above XynC and XynD have both been detected in the secretome of *B. subtilis*. Although only small amounts of xylose were observed in the BSC digestion of SG $MeGAX_n$ (Figure 5-4) (spot for xylose not observable in this print), their genomic localization, the characterized limit product of XynC (Figure 5-3B), and the putative activity of XynD (GH 43 nonreducing terminal β-xylosidase) suggest a combined role in hydrolysis of $MeGAX_n$ by these two enzymes.

In summary, XynC of *B. subtilis* 168 is a GH 5 endoxylanase that, based on substrate and product analysis, has specificity for the MeGA substitution on the $MeGAX_n$ chain. Hydrolysis products are decorated with a single MeGA moiety penultimate to the reducing end xylose as proposed by Nishitani and Nevins (Nishitani and Nevins, 1991), and there are no detectable neutral xylooligosaccharides released. With its unique mode of action and moderate thermo-tolerance, XynC may be applied towards the characterization of hemicellulose from different biomass sources (Jacobs et al., 2001) and possibly integrated with regimes for pretreatment of

lignocellulosics to facilitate efficient biocatalyst utilization of $MeGAX_n$. With these existing activities, *B. subtilis* 168 may be further engineered for the formation of $MeGAX_n$-utilization enzyme systems. A recombinant system for complete utilization of $MeGAX_n$ in *B. subtilis* should contribute to the development of microbial systems for efficient pretreatment of the hemicellulose fractions of currently underutilized resources of lignocellulosic biomass. Selection or engineering of fermentative strains of *B. subtilis* 168 may also allow development of next-generation Gram-positive biocatalysts for efficient conversion of $MeGAX_n$ to 'green' chemicals and fuel ethanol.

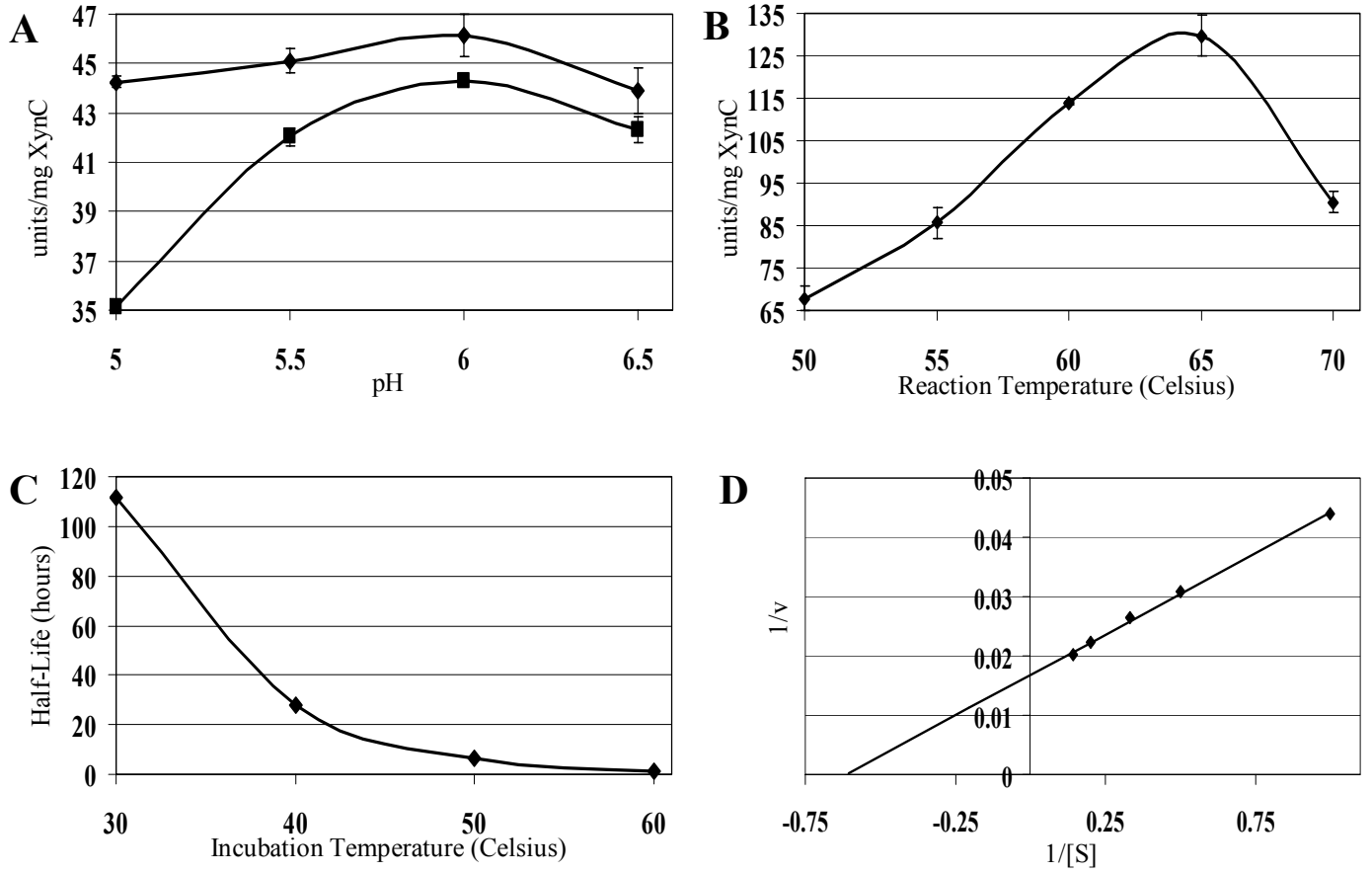

Figure 5-1. Optimization of XynC activity. Buffer and pH conditions were optimized and were applied for determination of the optimal reaction temperature and half-life analyses. These results were used to define the reaction conditions for the kinetic analysis. (A) Buffer and pH optimization using 50 mM buffers with a pH between 5.0 and 6.5: diamonds, potassium phosphate; squares, sodium acetate. (B) Temperature optimum in 50 mM sodium acetate, pH 6.0. (C) Half-life analysis determined by pre-incubating XynC at the specified temperatures and measuring remaining activity over time. Data is presented as the half-life obtained from the linear regression of inactivation at each temperature. (D) Lineweaver-Burk kinetic analysis was based on a reaction velocity vs. substrate concentration data set fit to a logarithmic equation.

Table 5-1. Relationship of XynC activity to the degree of MeGA substitution on $MeGAX_n$

$MeGAX_n$ source[a]	DP[b]	DS[c]	Specific[d] Activity
Sweetgum wood	231	6.8	43.7 ± 1.4
Beech wood	156	7.2	36.5 ± 1.5
Birch wood	229	10	21.9 ± 0.4

[a] Sweetgum xylan was prepared as described in materials and methods; all others were purchased from Sigma-Aldrich.

[b] DP equals the molar ratio of total anhydroxylose to total reducing terminal xylose.

[c] DS equals the molar ratio of total anhydroxylose to total uronic acid.

[d] Specific activity is given as units/mg XynC, where 1 unit is equal to 1 μmole of reducing terminal xylose equivalents generated per minute.

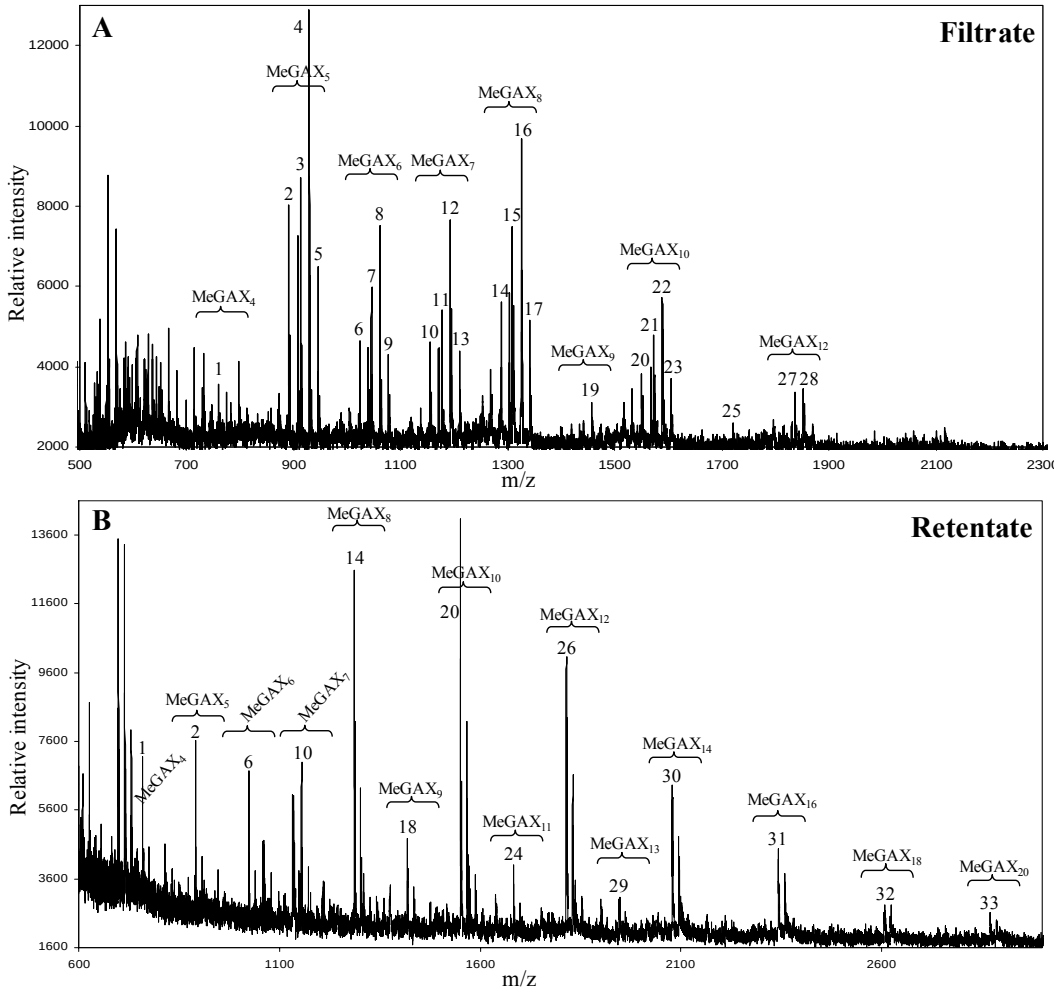

Figure 5-2. MALDI-TOF MS analysis of the Filtrate (A) and Retentate (B) resulting from 3 kDa ultrafiltration of a SG $MeGAX_n$ XynC digest. Peak m/z values were tabulated and show that each cluster of peaks is composed of various single and double salt adducts which differ from the previous cluster by a single xylose residue. Each chemical species is composed of the designated number of xylose residues containing a single MeGA residue. Complements of different sodium and/or potassium adducts comprise designated clusters.

Table 5-2. MALDI-TOF peak assignments

Xylooligomer	Peak no.	Adduct	MW
$MeGAX_4$	1	Na	759.23
$MeGAX_5$	2	Na	891.18
	3	Na+Na	913.18
	4	Na+K	929.08
	5	K+K	945.11
$MeGAX_6$	6	Na	1023.22
	7	Na+Na	1045.07
	8	Na+K	1061.08
	9	K+K	1076.97
$MeGAX_7$	10	Na	1155.08
	11	Na+Na	1177.11
	12	Na+K	1192.92
	13	K+K	1208.97
$MeGAX_8$	14	Na	1287.00
	15	Na+Na	1308.89
	16	Na+K	1324.87
	17	K+K	1340.81
$MeGAX_9$	18	Na	1419.06
	19	Na+K	1456.77
$MeGAX_{10}$	20	Na	1550.83
	21	Na+Na	1572.73
	22	Na+K	1588.73
	23	K+K	1604.65
$MeGAX_{11}$	24	Na	1682.83
	25	Na+K	1721.83
$MeGAX_{12}$	26	Na	1814.73
	27	Na+Na	1836.63
	28	Na+K	1852.63
$MeGAX_{13}$	29	Na	1947.75
$MeGAX_{14}$	30	Na	2078.54
$MeGAX_{16}$	31	Na	2342.39
$MeGAX_{18}$	32	Na	2607.07
$MeGAX_{20}$	33	Na	2871.04

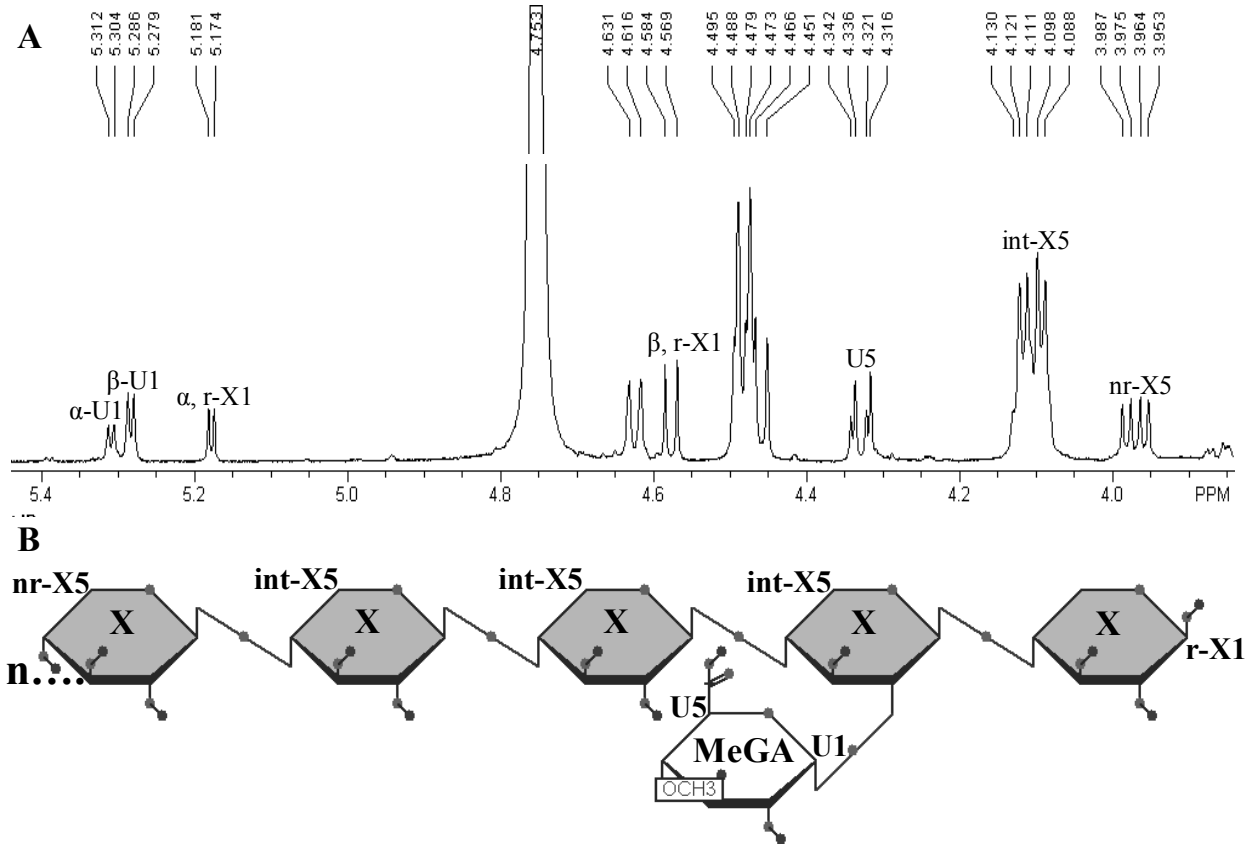

Figure 5-3. ^{1}H-NMR of SG MeGAX$_n$ 3 kDa filtrate revealing the general action of XynC hydrolysis of MeGAX$_n$ and the predicted limit product of XynC MeGAX$_n$ digestion. Integrated intensity values for specific shift positions have been used to determine the product of a XynC digestion, establishing that there is a single MeGA substitution for every reducing terminal xylose and every nonreducing terminal xylose, and that this substitution is penultimate to the reducing terminal xylose. (A) Shift assignments are labeled as: α,β-U1, 4-O-methylglucuronic acid carbon one hydrogen; U5, 4-O-methylglucuronic acid carbon five hydrogen; α,βr-X1, reducing terminal xylose carbon one hydrogen; nr-X5, nonreducing terminal xylose carbon five hydrogen; int-X5, internal xylose carbon five hydrogen. (B) Limit product generated by XynC catalyzed hydrolysis of MeGAX$_n$: X, xylose; MeGA, 4-O-methylglucuronic acid.

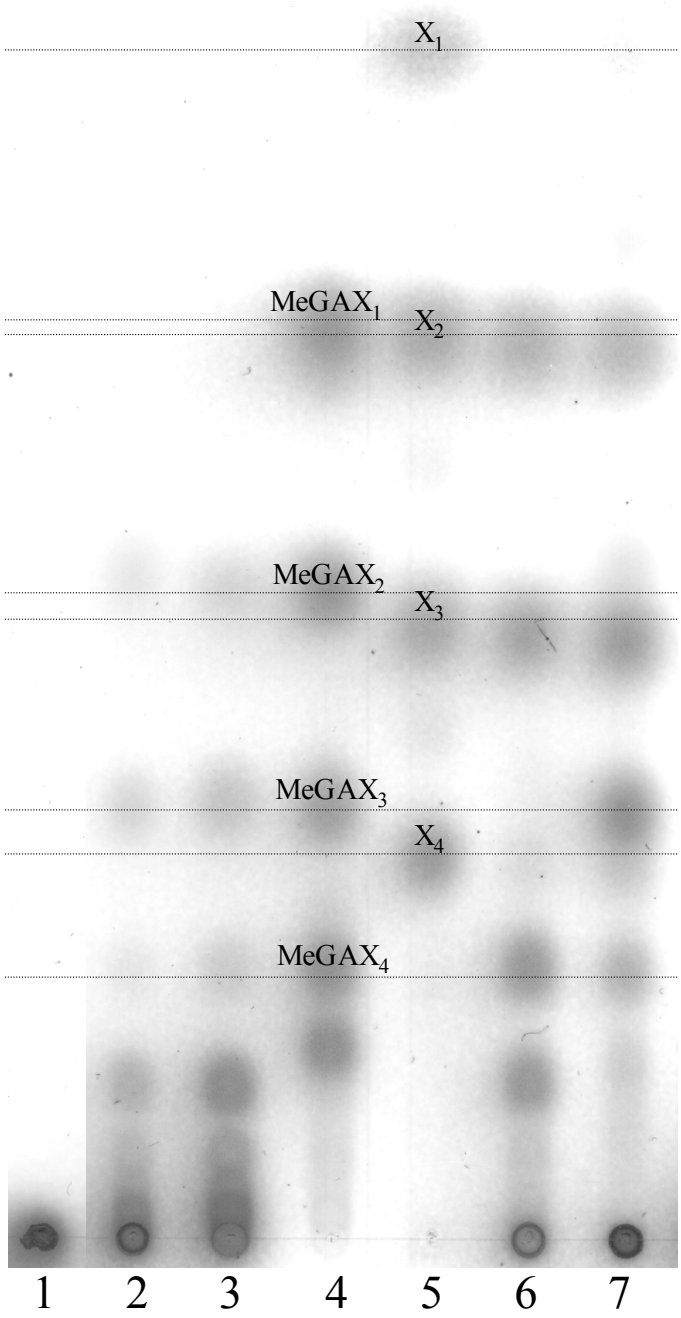

Figure 5-4. Identification of products generated by XynA (GH 11) and XynC (GH 5) secreted by *B. subtilis* 168. Spent medium from a mid-log phase culture of *B. subtilis* 168 was concentrated by YM-3 filtration to provide the BSC fraction. This was fractionated using a BioGel P-60 column to provide Fractions A and B that were used to digest SG $MeGAX_n$ and identify the xylanase hydrolysis pattern by TLC. SG $MeGAX_n$ and a SG $MeGAX_n$ digested with recombinantly expressed XynC were used as controls. 1) SG $MeGAX_n$; 2) SG $MeGAX_n$ digestion with Fraction A; 3) SG $MeGAX_n$ digestion with recombinant XynC; 4) $MeGAX_{1-4}$ aldouronate standards; 5) X_{1-4} xylooligomer standards; 6) SG $MeGAX_n$ digestion with Fraction B; 7) SG $MeGAX_n$ digestion with BSC.

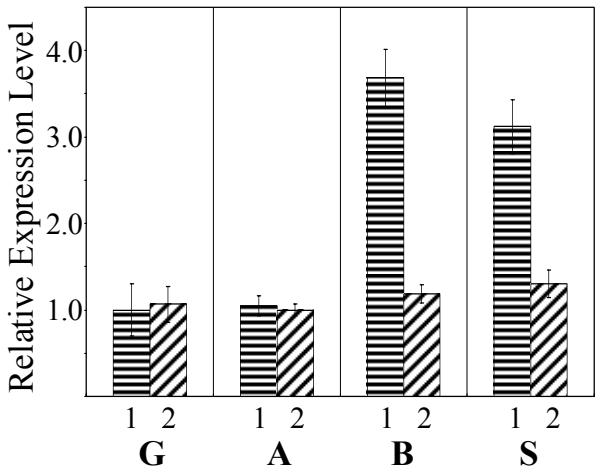

Figure 5-5. Regulation of expression of *xynA* and *xynC* genes in early- to mid-exponential phase growth cultures of *B. subtilis* 168 with different sugars as substrate, measured using Q-RT-PCR. G, glucose; A, arabinose; B, birch wood $MeGAX_n$; S, sweetgum wood $MeGAX_n$; 1, *xynA*; 2, *xynC*.

Table 5-3. Relative transcript quantity[a] measured by Q-RT-PCR for *gapA*, *abnA*, *xynA* and *xynC* genes

Growth Substrate[b]	*gapA*[c]	*abnA*[d]	*xynB*[e]	*xynC*[f]	*xynA*[g]
Glucose	32.6 ± 5.0	1.0 ± 0.3	1.2 ± 0.3	1.1 ± 0.2	1.0 ± 0.3
Arabinose	1.3 ± 0.2	611.9 ± 105.9	1.0 ± 0.2	1.0 ± 0.1	1.1 ± 0.1
Arabinose/Xylose	1.0 ± 0.1	621.3± 68.0	132.3 ± 13.4	1.4 ± 0.1	1.6 ± 0.2
SG $MeGAX_n$	3.6 ± 0.4	40.7 ± 4.7	1.4 ± 0.4	1.3 ± 0.2	3.1 ± 0.3
Birch $MeGAX_n$[h]	2.8 ± 0.3	57.1 ± 11.8	2.2 ± 0.6	1.2 ± 0.1	3.7 ± 0.3

[a] C_T values were used to calculate transcript copy number which was normalized to the lowest transcript level in each conditional gene set using BioRad Gene Expression Macro V1.1.
[b] Media for growth consisted of 0.1% YE in Spizizen salts with the specified sugar added to a final concentration of 0.5% except for xylose which was added to a final concentration of 0.25%.
[c] Value of 1 for *gapA* equals $2.3*10^5$ transcript copies per 10 ng total RNA
[d] Value of 1 for *abnA* equals $3.1*10^2$ transcript copies per 10 ng total RNA
[e] Value of 1 for *xynB* equals $1.4*10^3$ transcript copies per 10 ng total RNA
[f] Value of 1 for *xynC* equals $1.9*10^4$ transcript copies per 10 ng total RNA
[g] Value of 1 for *xynA* equals $1.6*10^5$ transcript copies per 10 ng total RNA
[h] Birch wood xylan used as received from Sigma-Aldrich.

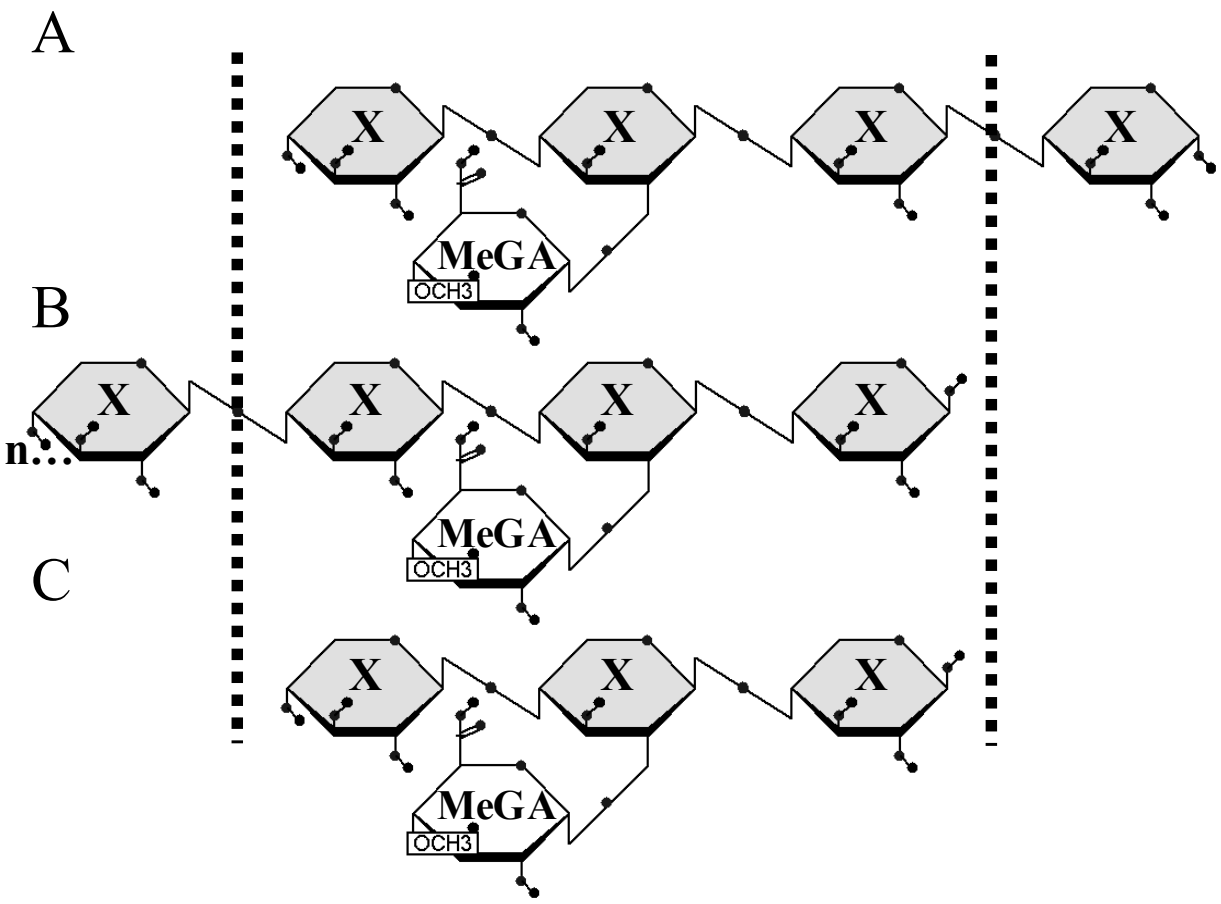

Figure 5-6. Limit aldouronates expected from a SG $MeGAX_n$ digestion with a GH 11 xylanase and a GH 5 xylanase co-secreted in the growth medium of *B. subtilis* 168. A) $MeGAX_4$ with a MeGA substitution penultimate to the nonreducing terminal xylose, the smallest aldouronate product resulting from a GH 11 hydrolysis of $MeGAX_n$; B) The predicted hydrolysis limit product of a GH 5 xylanase as presented in this chapter, having a single MeGA substitution penultimate to the reducing terminal xylose; C) $MeGAX_3$ with a MeGA substitution on the second of three xylose residues positioned penultimate to the reducing and nonreducing ends.

CHAPTER 6
SUMMARY DISCUSSION

Current Research Directions

Ethanol produced through bioconversion of lignocellulosic biomass is not currently economically competitive with gasoline or corn starch based ethanol. High costs are attributed to the multiple preprocessing steps required to liberate the simple sugars for microbial bioconversion. To reduce the expense of this preprocessing step, research has primarily targeted development of more efficient enzyme systems and development of more robust biocatalysts. In the former endeavor, advances occur with reduced cost of enzyme production, more efficient hydrolytic enzyme mixtures and enzymes with improved thermo and catalytic stability for potential enzyme reuse. Research is directed toward these goals to develop novel enzymatic methods for hydrolysis of carbohydrates and development of efficient organisms for protein expression.

Advancements resulting from these studies may directly impact the current pretreatment methods by allowing reduced preprocessing requirements prior to enzymatic hydrolysis. When considering dilute acid pretreatment, cost savings could be obtained from reduced acid use and/or reduced temperature and pressure. All of these factors could be fine tuned to reduce the overall processing input. Application of enzyme technologies for efficient hydrolysis of cellulose and hemicellulose should be applied with accessory enzymes such as esterases to facilitate detachment of lignin from hemicellulose and acetyl esterases to deacetylate the hemicellulose. Non enzymatic proteins also may have a role. Researchers have identified the fungal protein swollenin that can disrupt the crystalline structure of cellulose (Saloheimo et al., 2002). Catalytic activities could be engineered with non catalytic carbohydrate binding modules for a targeted hydrolysis. Bacterial enzymes make extensive use of associated noncatalytic

modules and are therefore prime candidates for development for enzymatic preprocessing methods.

Development of more robust biocatalysts is being pursued through technologies which expand the current capabilities of the organism. Potential improvements include increasing sugar substrate range, maximizing the metabolic potential through genetic engineering and generally increasing the fitness of the biocatalyst for the desired bioconversion. Since high bioconversion efficiency is achieved by using high substrate concentrations, robust biocatalysts are required to perform well under osmotically and chemically stressful conditions. Endeavors in this direction are to develop biocatalysts which are capable of efficient utilization of all sugars resulting from the enzymatic hydrolysis of lignocellulose, and condition the biocatalyst through directed evolution for optimized product yields.

Gram-positive Biocatalysts

The research initiatives discussed above may significantly impact that cost of lignocellulose-derived ethanol and allow it to compete with fossil fuels. Further cost reduction could be realized by development of bacterial biocatalysts which are capable of efficient secretion of hydrolytic enzymes and assimilation of complex hydrolysis products. Many researchers consider this the ultimate goal. In theory there would be little to no required enzyme addition to fully degrade the pretreated biomass. Development of the biocatalysts that secrete hydrolytic enzymes would increase process simplicity and greatly reduce the cost of ethanol production, making it competitive with fossil fuels.

To achieve this goal, Gram-positive bacteria are considered as candidate organisms. As briefly reviewed in Chapter 5, Gram-positive bacterium such as *Bacillus subtilis* exhibit characteristics that make them particularly attractive for development of second-generation biocatalysts. These characteristics include robust protein secretion systems, an attribute required

for efficient recombinant protein production. Other Gram-positive bacteria also have unique characteristics that may facilitate utilization of complex pretreated lignocellulosics.

Utilization of $MeGAX_n$ by *Paenibacillus* sp. strain JDR-2 exemplifies the potential for development of a Gram-positive bacterium as a biocatalyst. As presented in Chapter 4, this organism displays a unique ability to take-up hydrolytic fragments of glucuronoxylan resulting from hydrolysis with $XynA_1$, a surface anchored large modular GH 10 xylanase. Evidence suggests that efficient vectoral transport is somehow coupled to surface localized hydrolysis of polymeric $MeGAX_n$ by $XynA_1$. The organism did not grow efficiently when presented with the limit hydrolysis products of $XynA_1$ CD.

A similar system, characterized best in *Clostridium thermocellum* but also found in other organisms, allows efficient utilization of cellooligosaccharides up to a degree of polymerization of four. Hydrolysis of crystalline cellulose localized to the cell surface by the cellulosome complex releases cellobiose and higher cellooligosaccharides. These large oligomers (up to DP4) are transported and hydrolyzed within the cell by phosphorolytic cleavage (Lou et al., 1996; Lou et al., 1997; Reichenbecher et al., 1997; Zhang and Lynd, 2005). Decoupling glucose transport from substrate level ATP production may add greater flexibility for metabolic engineering. Recombinant application of these systems for complex carbohydrate utilization in Gram-positive biocatalysts may greatly increase efficiency of complex lignocellulose utilization.

$XynA_1$ of *Paenibacillus* sp. strain JDR-2 is a modular xylanase that may have three chemical features that complement its catalytic abilities. The modules associated with $XynA_1$ CD may facilitate targeting of the catalytic domain to soluble hemicellulosic methylglucuronoxylan, the reducing terminus of cellulosic glucan, and the cell surface. As reviewed in Chapter 3, this architectural modular arrangement localizes the catalytic activity and

substrate to the cell surface. This example of a family 10 glycosyl hydrolase represents a model module arrangement. There are several other less complex common modular arrangements associated with GH 10 xylanases that should have unique qualities. Although many of these modules have been biochemically characterized in terms of substrate binding, some have not been characterized with respect to catalytic activity, and more importantly with respect to the growth of the native bacteria. Although it may easily be shown that removal of a xylanase decreases or prevents the ability of an organism to utilize xylan, how does altering the carbohydrate binding module change the ability of the xylanase to support growth on glucuronoxylan? To harness the abundant tools of bacterial glycosyl hydrolases, research may be directed to engineer Gram-positive bacteria as was done in the past for development of *Escherichia coli* as an ethanologenic biocatalyst.

Use of *B. subtilis* or some comparable Gram-positive organism as a model bacterium for bioengineering is long overdue. In addition to defined systems for their genetic manipulation, these organisms have many natural attributes which make them ideal for protein secretion and utilization of complex sugars. Successful bioengineering of many of the metabolic and physiological features presented throughout this dissertation would require such a host.

Efficient processing of the methylglucuronoxylan component is necessary to realize maximum efficiency for any bioconversion process. *Paenibacillus* sp. strain JDR-2 is known to have the enzyme system, transporters and the catabolic pathway for efficient utilization of this biomass fraction. GH 10 xylanases such as $XynA_1$, are central for the utilization of this polymer in Gram-positive bacteria. As briefly discussed in Chapter 2 and Chapter 4, the smallest glucuronic acid limit product of glucuronoxylan hydrolysis by a GH 10 xylanase is

aldotetrauronic acid. This substituted oligomer is thought to be transported into the cell where it acts to induce genes involved in aldouronate/xylan utilization (Shulami et al., 1999).

B. subtilis already has enzymes for utilization of lignocellulosics. Besides several endoglucanases and arabinofuranosidases, it has the well studied glycosyl hydrolase family 11 xylanase XynA. Work presented in Chapter 5 characterized a second xylanase from *B. subtilis* 168. This family 5 glycosyl hydrolase has high similarity to the family 5 xylanases from *Erwinia chrysanthemi* strains. It showed specificity for glucuronoxylan and resulted in products substituted penultimate to the reducing terminal xylose with a single glucuronosyl moiety. An enzyme with this specificity may contribute significantly to the degradation of lignocellulosics.

Engineering *B. subtilis* for secretion and anchoring of a multimodular GH 10 xylanase and aldouronate-utilization system, both of which are produced in *Paenibacillus* sp. strain JDR-2, may result in a biocatalyst capable of efficient and complete utilization of the hemicellulose fraction derived from hardwood and crop residues. Further engineering of *B. subtilis* to make specific fermentation products may provide a biocatalyst for direct conversion of these biomass resources to the desired bio-based products.

LIST OF REFERENCES

Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxa, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P., and Holst, O. (2000). Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. Biochem J *345*, 53-60.

Akin, D. E., Morrison, W. H., 3rd, Rigsby, L. L., Barton, F. E., 2nd, Himmelsbach, D. S., and Hicks, K. B. (2006). Corn stover fractions and bioenergy: chemical composition, structure, and response to enzyme pretreatment. Appl Biochem Biotechnol *129-132*, 104-116.

Aldhous, P. (2005). Energy: China's burning ambition. Nature *435*, 1152-1154.

Ali, E., Araki, R., Zhao, G., Sakka, M., Karita, S., Kimura, T., and Sakka, K. (2005a). Functions of family-22 carbohydrate-binding modules in *Clostridium josui* Xyn10A. Biosci Biotechnol Biochem *69*, 2389-2394.

Ali, E., Zhao, G., Sakka, M., Kimura, T., Ohmiya, K., and Sakka, K. (2005b). Functions of family-22 carbohydrate-binding module in *Clostridium thermocellum* Xyn10C. Biosci Biotechnol Biochem *69*, 160-165.

Ali, M. K., Hayashi, H., Karita, S., Goto, M., Kimura, T., Sakka, K., and Ohmiya, K. (2001a). Importance of the carbohydrate-binding module of *Clostridium stercorarium* Xyn10B to xylan hydrolysis. Biosci Biotechnol Biochem *65*, 41-47.

Ali, M. K., Kimura, T., Sakka, K., and Ohmiya, K. (2001b). The multidomain xylanase Xyn10B as a cellulose-binding protein in *Clostridium stercorarium*. FEMS Microbiol Lett *198*, 79-83.

Anagnostopoulos, C., and Spizizen, J. (1961). Requirements for transformation in *Bacillus subtilis*. J Bacteriol *81*, 741-746.

Araki, R., Ali, M. K., Sakka, M., Kimura, T., Sakka, K., and Ohmiya, K. (2004). Essential role of the family-22 carbohydrate-binding modules for beta-1,3-1,4-glucanase activity of *Clostridium stercorarium* Xyn10B. FEBS Lett *561*, 155-158.

Arato, C., Pye, E. K., and Gjennestad, G. (2005). The lignol approach to biorefining of woody biomass to produce ethanol and chemicals. Appl Biochem Biotechnol *121-124*, 871-882.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L.*, et al.* (2004). The Pfam protein families database. Nucleic Acids Res *32*, 138-141.

Bayer, E. A., Belaich, J. P., Shoham, Y., and Lamed, R. (2004). The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol *58*, 521-554.

Bayer, E. A., Kenig, R., and Lamed, R. (1983). Adherence of *Clostridium thermocellum* to cellulose. J Bacteriol *156*, 818-827.

Bayer, E. A., Morag, E., and Lamed, R. (1994). The cellulosome--a treasure-trove for biotechnology. Trends Biotechnol *12*, 379-386.

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J Mol Bio *340*, 783-795.

Berlin, A., Balakshin, M., Gilkes, N., Kadla, J., Maximenko, V., Kubo, S., and Saddler, J. (2006). Inhibition of cellulase, xylanase and beta-glucosidase activities by softwood lignin preparations. J Biotechnol *125*, 198-209.

Biely, P., Vrsanska, M., Tenkanen, M., and Kluepfel, D. (1997). Endo-beta-1,4-xylanase families: differences in catalytic properties. J Biotechnol *57*, 151-166.

Black, G. W., Hazlewood, G. P., Xue, G. P., Orpin, C. G., and Gilbert, H. J. (1994). Xylanase B from *Neocallimastix patriciarum* contains a non-catalytic 455-residue linker sequence comprised of 57 repeats of an octapeptide. Biochem J *299*, 381-387.

Black, G. W., Rixon, J. E., Clarke, J. H., Hazlewood, G. P., Ferreira, L. M., Bolam, D. N., and Gilbert, H. J. (1997). Cellulose binding domains and linker sequences potentiate the activity of hemicellulases against complex substrates. J Biotechnol *57*, 59-69.

Black, G. W., Rixon, J. E., Clarke, J. H., Hazlewood, G. P., Theodorou, M. K., Morris, P., and Gilbert, H. J. (1996). Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates. Biochem J *319*, 515-520.

Blumenkrantz, N., and Asboe-Hansen, G. (1973). New method for quantitative determination of uronic acids. Anal Biochem *54*, 484-489.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J *382*, 769-781.

Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A., and Kilburn, D. G. (2001). Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. Biochemistry *40*, 6240-6247.

Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R., and Davies, G. (2003). Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals

similarities with fucose-specific lectins and "galactose-binding" domains. J Mol Biol *327*, 659-669.

Boraston, A. B., Tomme, P., Amandoron, E. A., and Kilburn, D. G. (2000). A novel mechanism of xylan binding by a lectin-like module from *Streptomyces lividans* xylanase 10A. Biochem J *350*, 933-941.

Bounias, M. (1980). N-(1-Naphthyl)ethylenediamine dihydrochloride as a new reagent for nanomole quantification of sugars on thin-layer plates by a mathematical calibration process. Anal Biochem *106*, 291-295.

Bowditch, R. D., Baumann, P., and Yousten, A. A. (1989). Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer protein from *Bacillus sphaericus* 2362 and of a related cryptic gene. J Bacteriol *171*, 4178-4188.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quanities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, 248-254.

Braun, E. J., and Rodrigues, C. A. (1993). Purification and properties of an endoxylanase from a corn stalk rot strain of *Erwinia chrysanthemi*. Phytopathology *83*, 332-337.

Brun, E., Gans, P., Marion, D., and Barras, F. (1995). Overproduction, purification and characterization of the cellulose-binding domain of the *Erwinia chrysanthemi* secreted endoglucanase EGZ. Eur J Biochem *231*, 142-148.

Brun, E., Johnson, P. E., Creagh, A. L., Tomme, P., Webster, P., Haynes, C. A., and McIntosh, L. P. (2000). Structure and binding specificity of the second N-terminal cellulose-binding domain from *Cellulomonas fimi* endoglucanase C. Biochemistry *39*, 2445-2458.

Brun, E., Moriaud, F., Gans, P., Blackledge, M. J., Barras, F., and Marion, D. (1997). Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*. Biochemistry *36*, 16074-11686.

Carpita, N. C., Defernez, M., Findlay, K., Wells, B., Shoue, D. A., Catchpole, G., Wilson, R. H., and McCann, M. C. (2001). Cell wall architecture of the elongating maize coleoptile. Plant Physiol *127*, 551-565.

Carrard, G., Koivula, A., Soderlund, H., and Beguin, P. (2000). Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. Proc Natl Acad Sci U S A *97*, 10342-10347.

Cava, F., de Pedro, M. A., Schwarz, H., Henne, A., and Berenguer, J. (2004). Binding to pyruvylated compounds as an ancestral mechanism to anchor the outer envelope in primitive bacteria. Mol Microbiol *52*, 677-690.

Cavagna, F., Deger, H., and Jurgen, P. (1984). 2D-N.M.R. analysis of the structure of an aldotriouronic acid obtained from birch wood. Carbohydr Chem *129*, 1-8.

Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000). The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. Biochemistry *39*, 5013-5021.

Charnock, S. J., Spurway, T. D., Xie, H., Beylot, M. H., Virden, R., Warren, R. A., Hazlewood, G. P., and Gilbert, H. J. (1998). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. J Biol Chem *273*, 32187-32199.

Chaudhary, P., and Deobagkar, D. N. (1997). Characterization of cloned endoxylanase from *Cellulomonas* sp. NCIM 2353 expressed in *Escherichia coli*. Curr Microbiol *34*, 273-279.

Cho, S. G., and Choi, Y. J. (1999). Catabolite repression of the xylanase gene (*xynA*) expression in *Bacillus stearothermophilus* no. 236 and *B. subtilis*. Biosci Biotechnol Biochem *63*, 2053-2058.

Clarke, J. H., Davidson, K., Gilbert, H. J., Fontes, C. M., and Hazlewood, G. P. (1996). A modular xylanase from mesophilic *Cellulomonas fimi* contains the same cellulose-binding and thermostabilizing domains as xylanases from thermophilic bacteria. FEMS Microbiol Lett *139*, 27-35.

Clarke, J. H., Rixon, J. E., Ciruela, A., Gilbert, H. J., and Hazlewood, G. P. (1997). Family-10 and family-11 xylanases differ in their capacity to enhance the bleachability of hardwood and softwood paper pulps. Appl Microbiol Biotechnol *48*, 177.

Collins, T., Gerday, C., and Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol Rev *29*, 3-23.

Coutinho, P. M., and Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach, In Recent Advances in Carbohydrate Bioengineering, H. J. Gilbert, G. Davies, B. Henrissat, and B. Svensson, eds. (Cambridge: The Royal Society of Chemistry), pp. 3-12.

Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., and Haynes, C. A. (1996). Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. Proc Natl Acad Sci U S A *93*, 12229-12234.

Cruz Ramos, H., Hoffmann, T., Marino, M., Nedjari, H., Presecan-Siedel, E., Dreesen, O., Glaser, P., and Jahn, D. (2000). Fermentative metabolism of *Bacillus subtilis*: physiology and regulation of gene expression. J Bacteriol *182*, 3072-3080.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853-859.

Davies, G. J., Wilson, K. S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem J *321*, 557-559.

Devillard, E., Bera-Maillet, C., Flint, H. J., Scott, K. P., Newbold, C. J., Wallace, R. J., Jouany, J. P., and Forano, E. (2003). Characterization of Xyn10B, a modular xylanase from the ruminal protozoan *Polyplastron multivesiculatum*, with a family 22 carbohydrate-binding module that binds to cellulose. Biochem J *373*, 495-503.

Dewulf, J., Van Langenhove, H., and Van De Velde, B. (2005). Exergy-based efficiency and renewability assessment of biofuel production. Environ Sci Technol *39*, 3878-3882.

Dias, F. M., Goyal, A., Gilbert, H. J., Jose, A. M. P., Ferreira, L. M., and Fontes, C. M. (2004). The N-terminal family 22 carbohydrate-binding module of xylanase 10B of *Clostridium themocellum* is not a thermostabilizing domain. FEMS Microbiol Lett *238*, 71-78.

Dien, B. S., Cotta, M. A., and Jeffries, T. W. (2003). Bacteria engineered for fuel ethanol production: current status. Appl Microbiol Biotechnol *63*, 258-266.

Dien, B. S., Nichols, N. N., and Bothast, R. J. (2002). Fermentation of sugar mixtures using *Escherichia coli* catabolite repression mutants engineered for production of L-lactic acid. J Ind Microbiol Biotechnol *29*, 221-227.

Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A., and Kilburn, D. G. (1994). C1-Cx revisited: intramolecular synergism in a cellulase. Proc Natl Acad Sci U S A *91*, 11383-11387.

Ding, S. Y., and Himmel, M. E. (2006). The maize primary cell wall microfibril: a new model derived from direct visualization. J Agric Food Chem *54*, 597-606.

Doblin, M. S., Kurek, I., Jacob-Wilk, D., and Delmer, D. P. (2002). Cellulose biosynthesis in plants: from genes to rosettes. Plant Cell Physiol *43*, 1407-1420.

Doi, R. H., and Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol *2*, 541-551.

Doi, R. H., Kosugi, A., Murashima, K., Tamaru, Y., and Han, S. O. (2003). Cellulosomes from mesophilic bacteria. J Bacteriol *185*, 5907-5914.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956). Colorimetric method for the determination of sugars and related substances. Anal Chem *28*, 350-356.

Eggeman, T., and Elander, R. T. (2005). Process and economic analysis of pretreatment technologies. Bioresour Technol *96*, 2019-2025.

Elegir, G., Szakacs, G., and Jeffries, T. W. (1994). Purification, characterization, and substrate specificities of multiple xylanases from *Streptomyces* sp. Strain B-12-2. Appl Environ Microbiol *60*, 2609-2615.

Energy Information Administration, D. O. E. (2006[a]). World proved reserves of oil and natural gas, most recent estimates, In Internet Explorer, (http://www.eia.doe.gov/emeu/international/reserves.html), ed. (EIA).

Energy Information Administration, D. O. E. (2006[b]). International energy outlook, world oil consumption by region, reference case, 1990-2030, In Adobe Acrobat, (http://www.eia.doe.gov/oiaf/ieo/pdf/ieoreftab_4.pdf), ed. (EIA).

Erickson, D. (2003). Minnesota's energy future (http://www.mnforsustain.org/erickson_dell_minnesotas_energy_future_table_of_contents.htm).

Etienne-Toumelin, I., Sirard, J. C., Duflot, E., Mock, M., and Fouet, A. (1995). Characterization of the *Bacillus anthracis* S-layer: cloning and sequencing of the structural gene. J Bacteriol *177*, 614-620.

Excoffier, G., Nardin, R., and Vignon, M. R. (1986). Determination de la structure primaire d'un acide tri-$_D$-xylo-4-O-methyl-$_D$-glucuronique par R. M. N. a deux dimensions. Carbohydr Chem *149*, 319-328.

Farrell, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O'Hare, M., and Kammen, D. M. (2006). Ethanol can contribute to energy and environmental goals. Science *311*, 506-508.

Fengel, D. (1971). Ideas on the ultrastructural organization of the cell wall components. J Poly Sci C *36*, 383-392.

Ferreira, L. M., Durrant, A. J., Hall, J., Hazlewood, G. P., and Gilbert, H. J. (1990). Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. Biochem J *269*, 261-264.

Fillinger, S., Boschi-Muller, S., Azza, S., Dervyn, E., Branlant, G., and Aymerich, S. (2000). Two glyceraldehyde-3-phosphate dehydrogenases with opposite physiological roles in a nonphotosynthetic bacterium. J Biol Chem *275*, 14031-14037.

Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H., and Gilbert, H. J. (1995). Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. Biochem J *307*, 151-158.

Fromm, J., Rockel, B., Lautner, S., Windeisen, E., and Wanner, G. (2003). Lignin distribution in wood cell walls determined by TEM and backscattered SEM techniques. J Struct Biol *143*, 77-84.

Fujimoto, Z., Kaneko, S., Kuno, A., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2004). Crystal structures of decorated xylooligosaccharides bound to a family 10 xylanase from *Streptomyces olivaceoviridis* E-86. J Biol Chem *279*, 9606-9614.

Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002). Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. J Mol Biol *316*, 65-78.

Fujimoto, Z., Kuno, A., Kaneko, S., Yoshida, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2000). Crystal structure of *Streptomyces olivaceoviridis* E-86 beta-xylanase containing xylan-binding domain. J Mol Biol *300*, 575.

Fujita, M., and Harada, H. (2001). Ultrastructure and formation of wood cell wall, In Wood and Cellulosic Chemistry, D. N.-S. Hon, and N. Shiraishi, eds. (New York: Marcel Dekker), pp. 1-49.

Gallardo, O., Diaz, P., and Pastor, F. I. (2003). Characterization of a *Paenibacillus* cell-associated xylanase with high activity on aryl-xylosides: a new subclass of family 10 xylanases. Appl Microbiol Biotechnol *61*, 226-233.

Gebler, J., Gilkes, N., Claeyssens, M., Wilson, D., Beguin, P., Wakarchuk, W., Kilburn, D., Miller, R., Warren, R., and Withers, S. (1992). Stereoselective hydrolysis catalyzed by related beta-1,4-glucanases and beta-1,4-xylanases. J Biol Chem *267*, 12559-12561.

Ghosh, P., and Ghose, T. K. (2003). Bioethanol in India: recent past and emerging future. Adv Biochem Eng Biotechnol *85*, 1-27.

Gilad, R., Rabinovich, L., Yaron, S., Bayer, E. A., Lamed, R., Gilbert, H. J., and Shoham, Y. (2003). CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. J Bacteriol *185*, 391-398.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. (1991). Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. Microbiol Rev *55*, 303-315.

Gupta, S., Bhushan, B., and Hoondal, G. S. (2000). Isolation, purification and characterization of xylanase from *Staphylococcus* sp. SG-13 and its application in biobleaching of kraft pulp. J Appl Microbiol *88*, 325-334.

Hall, J., Hazlewood, G. P., Huskisson, N. S., Durrant, A. J., and Gilbert, H. J. (1989). Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subspecies *cellulosa*: internal signal sequence and unusual protein processing. Mol Microbiol *3*, 1211-1219.

Hammerschlag, R. (2006). Ethanol's energy return on investment: a survey of the literature 1990-present. Environ Sci Technol *40*, 1744-1750.

Han, S. O., Yukawa, H., Inui, M., and Doi, R. H. (2004). Isolation and expression of the *xynB* gene and its product, XynB, a consistent component of the *Clostridium cellulovorans* cellulosome. J Bacteriol *186*, 8347-8355.

Hastrup, S. (1988). Analysis of the *Bacillus subtilis* xylose regulon, In Genetics and Biotechnology of Bacilli, A. T. Ganesan, and J. A. Hoch, eds. (New York: Academic Press, Inc.), pp. 79-83.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J *280*, 309-316.

Henrissat, B., and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J *293*, 781-788.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc Natl Acad Sci U S A *92*, 7090-7094.

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol *7*, 637-644.

Henshaw, J. L., Bolam, D. N., Pires, V. M., Czjzek, M., Henrissat, B., Ferreira, L. M., Fontes, C. M., and Gilbert, H. J. (2004). The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. J Biol Chem *279*, 21552-21559.

Hill, J., Nelson, E., Tilman, D., Polasky, S., and Tiffany, D. (2006). Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci U S A *103*, 11206-11210.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genom Res *9*, 868-877.

Hurlbert, J. C., and Preston, J. F. (2001). Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi*. J Bacteriol *183*, 2093-2100.

Ingram, L. O., Aldrich, H. C., Borges, A. C., Causey, T. B., Martinez, A., Morales, F., Saleh, A., Underwood, S. A., Yomano, L. P., York, S. W., *et al.* (1999). Enteric bacterial catalysts for fuel ethanol production. Biotechnol Prog *15*, 855-866.

Ingram, L. O., Gomez, P. F., Lai, X., Moniruzzaman, M., Wood, B. E., Yomano, L. P., and York, S. W. (1998). Metabolic engineering of bacteria for ethanol production. Biotechnol Bioeng *58*, 204-214.

Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998). Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. J Bacteriol *180*, 1709-1714.

Ito, Y., Tomita, T., Roy, N., Nakano, A., Sugawara-Tomita, N., Watanabe, S., Okai, N., Abe, N., and Kamio, Y. (2003). Cloning, expression, and cell surface localization of *Paenibacillus* sp. strain W-61 xylanase 5, a multidomain xylanase. Appl Environ Microbiol *69*, 6969-6978.

Jacob, S., Allmansberger, R., Gartner, D., and Hillen, W. (1991). Catabolite repression of the operon for xylose utilization from *Bacillus subtilis* W23 is mediated at the level of transcription and depends on a cis site in the *xylA* reading frame. Mol Gen Genet *229*, 189-196.

Jacobs, A., Larsson, P. T., and Dahlman, O. (2001). Distribution of uronic acids in xylans from various species of soft- and hardwood as determined by MALDI mass spectrometry. Biomacromolecules *2*, 979-990.

Jeffries, T. W. (2005). Ethanol fermentation on the move. Nat Biotechnol *23*, 40-41.

Jeffries, T. W., and Jin, Y. S. (2004). Metabolic engineering for improved fermentation of pentoses by yeasts. Appl Microbiol Biotechnol *63*, 495-509.

Jervis, E. J., Haynes, C. A., and Kilburn, D. G. (1997). Surface diffusion of cellulases and their isolated binding domains on cellulose. J Biol Chem *272*, 24016-24023.

Jiang, Z. Q., Deng, W., Li, L. T., Ding, C. H., Kusakabe, I., and Tan, S. S. (2004). A novel, ultra-large xylanolytic complex (xylanosome) secreted by *Streptomyces olivaceoviridis*. Biotechnol Lett *26*, 431-436.

Jin, Y. S., Laplaza, J. M., and Jeffries, T. W. (2004). *Saccharomyces cerevisiae* engineered for xylose metabolism exhibits a respiratory response. Appl Environ Microbiol *70*, 6816-6825.

Jindou, S., Xu, Q., Kenig, R., Shulman, M., Shoham, Y., Bayer, E. A., and Lamed, R. (2006). Novel architecture of family-9 glycoside hydrolases identified in cellulosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum*. FEMS Microbiol Lett *254*, 308-316.

Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G., and McIntosh, L. P. (1996a). Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. Biochemistry *35*, 14381-14394.

Johnson, P. E., Tomme, P., Joshi, M. D., and McIntosh, L. P. (1996b). Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC 2. NMR and ultraviolet absorption spectroscopy. Biochemistry *35*, 13895-13906.

Jones, J. K. N., Purves, C. B., and Timell, T. E. (1961). Constitution of 4-O-methylglucuronoxylan from the wood of trembling aspen (*Populus tremuloides Michx.*). Can J Chem *39*, 1059-1066.

Kardosova, A., Matulova, M., and Malovikova, A. (1998). (4-O-Methyl-alpha-D-glucurono)-D-xylan from *Rudbeckia fulgida*, var. *sullivantii* (Boynton et Beadle). Carbohydr Res *308*, 99-105.

Keen, N. T., Boyd, C., and Henrissat, B. (1996). Cloning and characterization of a xylanase gene from corn strains of *Erwinia chrysanthemi*. Mol Plant Microb Interac *9*, 651-657.

Khasin, A., Alchanati, I., and Shoham, Y. (1993). Purification and characterization of a thermostable xylanase from *Bacillus stearothermophilus* T-6. Appl Environ Microbiol *59*, 1725-1730.

Kim, S., and Holtzapple, M. T. (2005). Lime pretreatment and enzymatic hydrolysis of corn stover. Bioresour Technol *96*, 1994-2006.

Kim, S., and Holtzapple, M. T. (2006a). Delignification kinetics of corn stover in lime pretreatment. Bioresour Technol *97*, 778-785.

Kim, S., and Holtzapple, M. T. (2006b). Effect of structural features on enzyme digestibility of corn stover. Bioresour Technol *97*, 583-591.

Kim, T. H., and Lee, Y. Y. (2005). Pretreatment and fractionation of corn stover by ammonia recycle percolation process. Bioresour Technol *96*, 2007-2013.

Kim, T. H., Lee, Y. Y., Sunwoo, C., and Kim, J. S. (2006). Pretreatment of corn stover by low-liquid ammonia recycle percolation process. Appl Biochem Biotechnol *133*, 41-57.

Kolenova, K., Vrsanska, M., and Biely, P. (2006). Mode of action of endo-beta-1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. J Biotechnol *121*, 338-345.

Kosugi, A., Murashima, K., and Doi, R. H. (2001). Characterization of xylanolytic enzymes in *Clostridium cellulovorans*: expression of xylanase activity dependent on growth substrates. J Bacteriol *183*, 7037-7043.

Kosugi, A., Murashima, K., Tamaru, Y., and Doi, R. H. (2002). Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. J Bacteriol *184*, 884-888.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989). Determination of the three-dimensional solution structure of the c-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hydrid distance geometry-dynamical simulated annealing. Biochemistry *28*, 7241-7257.

Kraus, A., Hueck, C., Gartner, D., and Hillen, W. (1994). Catabolite repression of the *Bacillus subtilis xyl* operon involves a cis element functional in the context of an unrelated sequence, and glucose exerts additional *xylR*-dependent repression. J Bacteriol *176*, 1738-1745.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature *390*, 249-256.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Larson, S. B., Day, J., Barba de la Rosa, A. P., Keen, N. T., and McPherson, A. (2003). First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. Biochemistry *42*, 8411-8422.

Leibovitz, E., Ohayon, H., Gounon, P., and Beguin, P. (1997). Characterization and subcellular localization of the *Clostridium thermocellum* scaffoldin dockerin binding protein SdbA. J Bacteriol *179*, 2519-2523.

Lindner, C., Stulke, J., and Hecker, M. (1994). Regulation of xylanolytic enzymes in *Bacillus subtilis*. Microbiology *140*, 753-757.

Liu, C., and Wyman, C. E. (2005). Partial flow of compressed-hot water through corn stover to enhance hemicellulose sugar recovery and enzymatic digestibility of cellulose. Bioresour Technol *96*, 1978-1985.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods *25*, 402-408.

Lloyd, T. A., and Wyman, C. E. (2005). Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. Bioresour Technol *96*, 1967-1977.

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

Lou, J., Dawson, K. A., and Strobel, H. J. (1997). Cellobiose and cellodextrin metabolism by the ruminal bacterium *Ruminococcus albus*. Curr Microbiol *35*, 221-227.

Lynd, L. R., van Zyl, W. H., McBride, J. E., and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol *16*, 577-583.

Lynd, L. R., Wyman, C. E., and Gerngross, T. U. (1999). Biocommodity Engineering. Biotechnol Prog *15*, 777-793.

Marchler-Bauer, A., Anderson, J. B., Weese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J., *et al.* (2003). CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res *31*, 383-387.

Martinez, A., Rodriguez, M. E., York, S. W., Preston, J. F., and Ingram, L. O. (2000). Use of UV absorbance to monitor furans in dilute acid hydrolysates of biomass. Biotechnol Prog *16*, 637-641.

McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., and Kilburn, D. G. (2000). Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. Protein Eng *13*, 801-809.

Meissner, K., Wassenberg, D., and Liebl, W. (2000). The thermostabilizing domain of the modular xylanase XynA of *Thermotoga maritima* represents a novel type of binding domain with affinity for soluble xylan and mixed-linkage beta-1,3/beta-1, 4-glucan. Mol Microbiol *36*, 898-912.

Mellerowicz, E. J., Baucher, M., Sundberg, B., and Boerjan, W. (2001). Unravelling cell wall formation in the woody dicot stem. Plant Mol Biol *47*, 239-274.

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J *19*, 4473-4484.

Millward-Sadler, S. J., Davidson, K., Hazlewood, G. P., Black, G. W., Gilbert, H. J., and Clarke, J. H. (1995). Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas fluorescens* subsp. *cellulosa* and *Cellvibrio mixtus*. Biochem J *312*, 39-48.

Miyazaki, K., Takenouchi, M., Kondo, H., Noro, N., Suzuki, M., and Tsuda, S. (2006). Thermal stabilization of *Bacillus subtilis* family-11 xylanase by directed evolution. J Biol Chem *281*, 10236-10242.

Moreno, M. S., Schneider, B. L., Maile, R. R., Weyler, W., and Saier, M. H., Jr. (2001). Catabolite repression mediated by the CcpA protein in *Bacillus subtilis*: novel modes of regulation revealed by whole-genome analyses. Mol Microbiol *39*, 1366-1381.

Morris, D. D., Gibbs, M. D., Ford, M., Thomas, J., and Bergquist, P. L. (1999). Family 10 and 11 xylanase genes from *Caldicellulosiruptor* sp. strain Rt69B.1. Extremophiles *3*, 103-111.

Mosier, N., Hendrickson, R., Ho, N., Sedlak, M., and Ladisch, M. R. (2005a). Optimization of pH controlled liquid hot water pretreatment of corn stover. Bioresour Technol *96*, 1986-1993.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., and Ladisch, M. (2005b). Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol *96*, 673-686.

Nagy, T., Emami, K., Fontes, C. M., Ferreira, L. M., Humphry, D. R., and Gilbert, H. J. (2002). The membrane-bound alpha-glucuronidase from *Pseudomonas cellulosa* hydrolyzes 4-O-methyl-D-glucuronoxylooligosaccharides but not 4-O-methyl-D-glucuronoxylan. J Bacteriol *184*, 4925-4929.

Nagy, T., Nurizzo, D., Davies, G. J., Biely, P., Lakey, J. H., Bolam, D. N., and Gilbert, H. J. (2003). The alpha-glucuronidase, GlcA67A, of *Cellvibrio japonicus* utilizes the carboxylate and methyl groups of aldobiouronic acid as important substrate recognition determinants. J Biol Chem *278*, 20286-20292.

Nakano, M. M., Dailly, Y. P., Zuber, P., and Clark, D. P. (1997). Characterization of anaerobic fermentative growth of *Bacillus subtilis*: identification of fermentation end products and genes required for growth. J Bacteriol *179*, 6749-6755.

Nakano, M. M., and Zuber, P. (1998). Anaerobic growth of a "strict aerobe" (*Bacillus subtilis*). Annu Rev Microbiol *52*, 165-190.

Nelson, N. (1944). A photometric adaptation of the Somogyi method for the determination of glucose. J Biol Chem *153*, 375-380.

Nguyen, H. D., Nguyen, Q. A., Ferreira, R. C., Ferreira, L. C., Tran, L. T., and Schumann, W. (2005). Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability. Plasmid *54*, 241-248.

Niaudet, B., Goze, A., and Ehrlich, S. D. (1982). Insertional mutagenesis in *Bacillus subtilis*: mechanism and use in gene cloning. Gene *19*, 277-284.

Nishitani, K., and Nevins, D. J. (1991). Glucuronoxylan xylanohydrolase. A unique xylanase with the requirement for appendant glucuronosyl units. J Biol Chem *266*, 6539-6543.

Notenboom, V., Boraston, A. B., Kilburn, D. G., and Rose, D. R. (2001). Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. Biochemistry *40*, 6248-6256.

Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G., and Rose, D. R. (2002). 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 *41*, 4246-4254.

Nurizzo, D., Nagy, T., Gilbert, H. J., and Davies, G. J. (2002). The structural basis for catalysis and specificity of the *Pseudomonas cellulosa* alpha-glucuronidase, GlcA67A. Structure *10*, 547-556.

O'sullivan, A. C. (1997). Cellulose: the structure slowly unravels. Cellulose *4*, 173-207.

Okai, N., Fukasaku, M., Kaneko, J., Tomita, T., Muramoto, K., and Kamio, Y. (1998). Molecular properties and activity of a carboxyl-terminal truncated form of xylanase 3 from *Aeromonas caviae* W-61. Biosci Biotechnol Biochem *62*, 1560-1567.

Pell, G., Szabo, L., Charnock, S. J., Xie, H., Gloster, T. M., Davies, G. J., and Gilbert, H. J. (2004a). Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. J Biol Chem *279*, 11777-11788.

Pell, G., Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Fontes, C. M., Ferreira, L. M., Nagy, T., Clark, S. J., Davies, G. J., and Gilbert, H. J. (2004b). The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. J Biol Chem *279*, 9597-9605.

Perlack, R. D., Wright, L. L., Turhollow, A. F., Graham, R. L., Stokes, B. J., and Erbach, D. C. (2005). Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply (USDA, DOE).

Piggot, P. J., and Hilbert, D. W. (2004). Sporulation of *Bacillus subtilis*. Curr Opin Microbiol *7*, 579-586.

Pires, V. M., Henshaw, J. L., Prates, J. A., Bolam, D. N., Ferreira, L. M., Fontes, C. M., Henrissat, B., Planas, A., Gilbert, H. J., and Czjzek, M. (2004). The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. J Biol Chem *279*, 21560-21568.

Polson, A., Coetzer, T., Kruger, J., von Maltzahn, E., and van der Merwe, K. J. (1985). Improvements In the Isolation of IgY From the Yolks of Eggs Laid by Immunized Hens. Immunol Invest *14*, 323-327.

Ponyi, T., Szabo, L., Nagy, T., Orosz, L., Simpson, P. J., Williamson, M. P., and Gilbert, H. J. (2000). Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from *Pseudomonas xylanase* A to insoluble ligands. Biochemistry *39*, 985-891.

Preston, J. F., Hurlbert, J. C., Rice, J. D., Ragunathan, A., and St. John, F. J. (2003). Microbial strategies for the depolymerization of glucuronoxylan: leads to biotechnological applications of endoxylanases, In Applications of Enzymes to Lignocellulosics, S. D. Mansfield, and J. N. Saddler, eds. (Washington D.C.: American Chemical Society), pp. 191-210.

Puls, J. (1997). Chemistry and biochemistry of hemicellulose: relationships between hemicellulose structure and enzymes required for hydrolysis. Macromol Symp *120*, 183-196.

Qian, Y., Yomano, L. P., Preston, J. F., Aldrich, H. C., and Ingram, L. O. (2003). Cloning, characterization, and functional expression of the *Klebsiella oxytoca* xylodextrin utilization operon (*xynTB*) in *Escherichia coli*. Appl Environ Microbiol *69*, 5957-5967.

Quinn, L. Y., Oates, R. P., and Beers, T. S. (1963). Support of cellulose digestion by *Clostridium thermocellum* in a kinetin-supplemented basal medium. J Bacteriol *86*, 1359.

Raghothama, S., Simpson, P. J., Szabo, L., Nagy, T., Gilbert, H. J., and Williamson, M. P. (2000). Solution structure of the CBM10 cellulose binding module from *Pseudomonas xylanase* A. Biochemistry *39*, 978-984.

Raposo, M. P., Inacio, J. M., Mota, L. J., and de Sa-Nogueira, I. (2004). Transcriptional regulation of genes encoding arabinan-degrading enzymes in *Bacillus subtilis*. J Bacteriol *186*, 1287-1296.

Reichenbecher, M., Lottspeich, F., and Bronnenmeier, K. (1997). Purification and properties of a cellobiose phosphorylase (CepA) and a cellodextrin phosphorylase (CepB) from the cellulolytic thermophile *Clostridium stercorarium*. Eur J Biochem *247*, 262-267.

Reis, D., Vian, B., and Roland, J.-C. (1994). Cellulose-glucuronoxylans and plant cell wall structure. Micron *25*, 171-187.

Rose, M., and Entian, K. D. (1996). New genes in the 170 degrees region of the *Bacillus subtilis* genome encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino acid transporter. Microbiology *142 (Pt 11)*, 3097-3101.

Roy, N., Okai, N., Tomita, T., Muramoto, K., and Kamio, Y. (2000). Purification and some properties of high-molecular-weight xylanases, the xylanases 4 and 5 of *Aeromonas caviae* W-61. Biosci Biotechnol Biochem *64*, 408-413.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol *132*, 365-386.

Rydlund, A., and Dahlman, O. (1997). Oligosaccharides obtained by enzymatic hydrolysis of birch kraft pulp xylan: analysis by capillary zone electrophoresis and mass spectrometry. Carbohydr Res *300*, 95-102.

Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997). Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nat Struct Biol *4*, 810-818.

Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyyssonen, E., Bhatia, A., Ward, M., and Penttila, M. (2002). Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. Eur J Biochem *269*, 4202-4211.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd ed (Cold Spring Harbor, N. Y.: Cold Spring Harbor National Laboratory Press).

Schallmey, M., Singh, A., and Ward, O. P. (2004). Developments in the use of *Bacillus* species for industrial production. Can J Microbiol *50*, 1-17.

Scharpf, M., Connelly, G. P., Lee, G. M., Boraston, A. B., Warren, R. A., and McIntosh, L. P. (2002). Site-specific characterization of the association of xylooligosaccharides with the CBM13 lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A by NMR spectroscopy. Biochemistry *41*, 4255-4263.

Schmidt, L., Preston, J., Dickson, D., Rice, J., and Hewlett, T. (2003). Environmental quantification of *Pasteuria penetrans* endospores using in situ antigen extraction and immunodetection with a monoclonal antibody. FEMS Microbiol Ecol *44*, 17-26.

Schmittgen, T. D., and Zakrajsek, B. A. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods *46*, 69-81.

Scurlock, J. (http://bioenergy.ornl.gov/papers/misc/biochar_factsheet.html). Bioenergy feedstock characteristics (Oak Ridge: Oak Ridge National Laboratory, Department of Energy).

Sedlak, M., and Ho, N. W. (2004). Production of ethanol from cellulosic biomass hydrolysates using genetically engineered *Saccharomyces* yeast capable of cofermenting glucose and xylose. Appl Biochem Biotechnol *113-116*, 403-416.

Shimon, L. J., Pages, S., Belaich, A., Belaich, J. P., Bayer, E. A., Lamed, R., Shoham, Y., and Frolow, F. (2000). Structure of a family IIIa scaffoldin CBD from the cellulosome of *Clostridium cellulolyticum* at 2.2 A resolution. Acta Crystallogr D Biol Crystallogr *56*, 1560-1568.

Shulami, S., Gat, O., Sonenshein, A. L., and Shoham, Y. (1999). The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6. J Bacteriol *181*, 3695-3704.

Simpson, P. J., Jamieson, S. J., Abou-Hachem, M., Karlsson, E. N., Gilbert, H. J., Holst, O., and Williamson, M. P. (2002). The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. Biochemistry *41*, 5712-5719.

Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000). The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. J Biol Chem *275*, 41137-41142.

Singh, S., Madlala, A. M., and Prior, B. A. (2003). *Thermomyces lanuginosus*: properties of strains and their hemicellulases. FEMS Microbiol Rev *27*, 3-16.

Sleat, R., Mah, R. A., and Robinson, R. (1984). Isolation and characterization of an anaerobic, cellulytic bacterium, *Clostridium cellulovorans* sp. nov. Appl Environ Microbiol *48*, 88-93.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., *et al.* (2004). Toward a systems approach to understanding plant cell walls. Science *306*, 2206-2211.

Spizizen, J. (1958). Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. Proc Natl Acad Sci U S A *44*, 1072-1078.

St. John, F. J., Rice, J. D., and Preston, J. F. (2006). *Paenibacillus* sp. strain JDR-2 and XynA1: a novel system for methylglucuronoxylan utilization. Appl Environ Microbiol *72*, 1496-1506.

Stahl, B., Steup, M., Karas, M., and Hillenkamp, F. (1991). Analysis of neutral oligosaccharides by matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem *63*, 1463-1466.

Sun, Y., and Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour Technol *83*, 1-11.

Sunna, A., Gibbs, M. D., and Bergquist, P. L. (2001). Identification of novel beta-mannan- and beta-glucan-binding modules: evidence for a superfamily of carbohydrate-binding modules. Biochem J *356*, 791-798.

Suzuki, T., Ibata, K., Hatsu, M., Takamizawa, K., and Kawai, K. (1997). Cloning and expression of a 58-kDa xylanase VI gene (*xynD*) of *Aeromonas caviae* ME-1 in *Escherichia coli* which is not categorized as a family F or family G xylanase. J Ferm Bioeng *84*, 86-89.

Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2001). Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. J Biol Chem *276*, 49061-49065.

Teymouri, F., Laureano-Perez, L., Alizadeh, H., and Dale, B. E. (2005). Optimization of the ammonia fiber explosion (AFEX) treatment parameters for enzymatic hydrolysis of corn stover. Bioresour Technol *96*, 2014-2018.

Timell, T. E. (1967). Recent progress in the chemistry of wood hemicelluloses. Wood Sci Technol *1*, 45-70.

Titok, M. A., Chapuis, J., Selezneva, Y. V., Lagodich, A. V., Prokulevich, V. A., Ehrlich, S. D., and Janniere, L. (2003). *Bacillus subtilis* soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector. Plasmid *49*, 53-62.

Tjalsma, H., Antelmann, H., Jongbloed, J. D., Braun, P. G., Darmon, E., Dorenbos, R., Dubois, J. Y., Westers, H., Zanen, G., Quax, W. J., *et al.* (2004). Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. Microbiol Mol Biol Rev *68*, 207-233.

Tomme, P., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (1996). Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. 1. Binding specificity and calorimetric analysis. Biochemistry *35*, 13885-13894.

Tomme, P., Warren, R. A., Miller, R. C., Kilburn, D. G., and Gilkes, N. R. (1995). Cellulose-binding domains: classification and properties, In Enzymatic Degradation of Insoluble Polysaccharides, J. N. Saddler, and M. Penner, eds. (Washington: American Chemical Society), pp. 142-163.

Tsai, Y. L., and Olson, B. H. (1991). Rapid method for direct extraction of DNA from soil and sediments. Appl Environ Microbiol *57*, 1070-1074.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol *3*, 1-12.

Vardakou, M., Flint, J., Christakopoulos, P., Lewis, R. J., Gilbert, H. J., and Murray, J. W. (2005). A family 10 *Thermoascus aurantiacus* xylanase utilizes arabinose decorations of xylan as significant substrate specificity determinants. J Mol Biol *352*, 1060-1067.

Vardakou, M., Katapodis, P., Samiotaki, M., Kekos, D., Panayotou, G., and Christakopoulos, P. (2003). Mode of action of family 10 and 11 endoxylanases on water-unextractable arabinoxylan. Int J Biol Macromol *33*, 129-134.

Viet, D., Kamio, Y., Abe, N., Kaneko, J., and Izaki, K. (1991). Purification and properties of beta-1,4-xylanase from *Aeromonas caviae* W-61. Appl Environ Microbiol *57*, 445-449.

Volker, U., and Hecker, M. (2005). From genomics via proteomics to cellular physiology of the gram-positive model organism *Bacillus subtilis*. Cell Microbiol *7*, 1077-1085.

Wilkie, K. C. B. (1979). The hemicelluloses of grasses and cereals, In Advances in Carbohydrate Chemistry and Biochemistry, R. S. Tipson, and D. Horton, eds. (New York: Academic Press), pp. 215-264.

Wolf, M., Geczi, A., Simon, O., and Borriss, R. (1995). Genes encoding xylan and beta-glucan hydrolysing enzymes in *Bacillus subtilis*: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase. Microbiology *141*, 281-290.

Wyman, C. E., Dale, B. E., Elander, R. T., Holtzapple, M., Ladisch, M. R., and Lee, Y. Y. (2005a). Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. Bioresour Technol *96*, 2026-2032.

Wyman, C. E., Dale, B. E., Elander, R. T., Holtzapple, M., Ladisch, M. R., and Lee, Y. Y. (2005b). Coordinated development of leading biomass pretreatment technologies. Bioresour Technol *96*, 1959-1966.

Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M., Dias, F. M., Ferreira, L. M., and Bolam, D. N. (2001). *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. Biochemistry *40*, 9167-9176.

Zaldivar, J., Martinez, A., and Ingram, L. O. (1999). Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng *65*, 24-33.

Zhang, Y. H., and Lynd, L. R. (2005). Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. Proc Natl Acad Sci U S A *102*, 7321-7325.

Zhao, G., Ali, E., Sakka, M., Kimura, T., and Sakka, K. (2006a). Binding of S-layer homology modules from *Clostridium thermocellum* SdbA to peptidoglycans. Appl Microbiol Biotechnol *70*, 464-469.

Zhao, G., Li, H., Wamalwa, B., Sakka, M., Kimura, T., and Sakka, K. (2006b). Different binding specificities of S-layer homology modules from *Clostridium thermocellum* AncA, Slp1, and Slp2. Biosci Biotechnol Biochem *70*, 1636-1641.

Zhou, S., Causey, T. B., Hasona, A., Shanmugam, K. T., and Ingram, L. O. (2003a). Production of optically pure D-lactic acid in mineral salts medium by metabolically engineered *Escherichia coli* W3110. Appl Environ Microbiol *69*, 399-407.

Zhou, S., Shanmugam, K. T., and Ingram, L. O. (2003b). Functional replacement of the *Escherichia coli* D-(-)-lactate dehydrogenase gene (ldhA) with the L-(+)-lactate dehydrogenase gene (ldhL) from *Pediococcus acidilactici*. Appl Environ Microbiol *69*, 2237-2244.

Zucker, M., and Hankin, L. (1970). Regulation of pectate lyase synthesis in *Pseudomonas fluorescens* and *Erwinia carotovora*. J Bacteriol *104*, 13-18.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.

Zverlov, V. V., Volkov, I. Y., Velikodvorskaya, G. A., and Schwarz, W. H. (2001). The binding pattern of two carbohydrate-binding modules of laminarinase Lam16A from *Thermotoga neapolitana*: differences in beta-glucan binding within family CBM4. Microbiology *147*, 621-629.

BIOGRAPHICAL SKETCH

I received most of my primary education in the public school system of Florida. After several semesters of part time attendance at Edison Community College, I moved to Gainesville, FL, and finished my Associate of Arts degree at Santa Fe Community College. I was accepted into the College of Agriculture at the University of Florida and began my last two years studying microbiology. Before finishing, I decided to take some time off. When I returned I involved myself within the subject by obtaining a lab assistant position in the laboratory of Dr. James F. Preston. Working as a laboratory assistant and being enrolled in undergraduate research studies during this time I learned the application of various techniques pertaining to the purification of carbohydrate fractions from crude biomass and isolation and characterization of purified complex sugars. I also obtained significant knowledge isolating bacteria and using various instrumentation such as NMR, MALDI-TOF MS, HPLC, etc. This was a valuable experience which contributed to my abilities in graduate school. A year after graduating with my Bachelor of Science, I applied to the University in hopes of continuing my education in this field. As a graduate student I gained valuable experience as a teaching assistant for PCB 5136, Techniques Lab. My research was performed in the laboratory of Dr. James F. Preston. I believe this experience was unique because Dr. Preston continuously encouraged students to find their own path. This allowed our research endeavors to evolve unhindered by expectations and enabled me the time to develop as a scientist.