ABSTRACT

COMFORT, DONALD ANDREW. Discovery, functional genomics and biochemical characterization of α-specific glycosyl hydrolases from hyperthermophilic microorganisms (Under the direction of Dr. Robert M. Kelly).

Glycosyl hydrolases – biocatalysts that are capable of hydrolyzing linkages between sugars – comprise a significant fraction of the current industrial enzyme market and are used in applications ranging from baking and brewing, to bioenergy, to oil/gas recovery. Biocatalysts for these applications often require stability and functional activity at elevated temperatures. As such, heterotrophic hyperthermophilic microorganisms (growth $T_{opt} \geq 80°C$), capable of using a variety of carbohydrates as carbon and energy sources, have been examined as sources of thermally stable enzymes. The biochemical and functional attributes of the glycosyl hydrolases from these and other organisms can often be inferred from primary amino acid sequence information. The current challenges are to fully understand the known glycosyl hydrolases, through biochemical and biophysical investigations, and to identify and characterize glycosyl hydrolases not yet annotated within genome sequences.

Conserved domains allow for classification of glycosyl hydrolases into families, which have conserved structural features and catalytic residues. Clan-D glycoside hydrolases (GH27 and GH36) were compared structurally and shown to shared a conserved retaining reaction mechanism, catalytic residues, and structural features. Structural alignment and biochemical analysis of site-directed mutants of the family 36 α-galactosidase from *Thermotoga maritima* identified the catalytic residues as D327 and D387 for the nucleophile and acid-base residues, respectively. Azide rescue of D327G a concentration dependent increase in activity and Brønsted analysis showed a change in the rate-limiting step at pH 8-9. When the acid-base mutant enzyme, D387G, was rescued with external anions azide and formate, a strong increase of activity was shown for galactoside substrates with a good leaving group (2,4-dinitrophenyl, DNP-Gal) over a poor leaving group (4-nitrophenyl, PNP-Gal). Similar increases in activity were not noted for the wild-type enzyme. pH curves of catalytic rate and catalytic efficiency for D387G shows no decrease in activity at higher pH, indicative of an acid-base mutant. Thermal denaturation of the mutant enzymes showed no change in the melting temperature for D327G and a 5.5°C decrease in melting temperature

for D387G. These structural, biochemical, and biophysical analyses confirm the catalytic residues and the relationship between the clan-D glycoside hydrolases.

In the post-genomics era, functional annotation of biocatalysts encoded in genome sequences represents an important step in biotechnological applications. Despite the large number of sequenced genomes available and extensive characterization of the enzymes encoded therein, roughly 40% of genes are still annotated as 'hypothetical proteins'. Using a functional genomics approach, combined with bioinformatic techniques and biochemical characterization, the genome of the hyperthermophilic archaeon *Pyrococcus furiosus* was probed to identify new glycosyl hydrolases. The open reading frame (ORF) PF0870 was identified by bioinformatic analysis to encode a glycosyl hydrolase that was found to hydrolyze *p*NP-α-maltopyranoside and maltotriose, thereby defining it as a novel β-amylase. Transcriptional response of *P. furiosus* grown on pullulan, starch, glycogen, maltose, cellobiose, trehalose, and maltose/cellobiose suggested an alternative biochemical role for a previously characterized exo-α-glucanase (Costantino *et al.*, J. Bacteriol., 172:3654-3660, 1990), the native form of which had been previously purified from *P. furiosus* cell extracts. Purification and mass spectroscopy indicated that the native α-glucosidase was encoded in PF0132, an ORF annotated as a hypothetical protein, but now known to represent a new family of glycosyl hydrolases. Strong transcriptional response to growth on pullulan suggested activity on α-1,6-glucosidic linkages, and hydrolysis of the compounds maltose, isomaltose, panose, turanose, and maltotriose confirmed activity against this broad range of glucosidic linkages. At 90°C, hydrolytic efficiency on panose (Glc-α-1,6-Glc-α-1,4-Glc; k_{cat} 278 s^{-1}, K_M 3.2 mM, and k_{cat}/K_M 85.6 $s^{-1}*mM^{-1}$) was similar to that on maltose (k_{cat} 1,600 s^{-1}, K_M 7.1 mM, and k_{cat}/K_M 225 $s^{-1}*mM^{-1}$). This broad substrate specificity and high temperature optimum makes this an especially appealing enzyme for the final hydrolysis step in bioethanol production, which converts remaining maltooligosaccharides into glucose for fermentation. This result demonstrates how functional genomics approaches, when combined with bioinformatics analysis and biochemical characterization, can be utilized to determine the role of proteins encoded in unannotated ORFs of genome sequences.

Discovery, functional genomics and biochemical characterization

of α-specific glycosyl hydrolases from hyperthermophilic microorganisms

by

Donald Andrew Comfort

A dissertation submitted to the Graduate Faculty of

North Carolina State University

In partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

Chemical Engineering

Raleigh, NC

2006

Approved by:

John Cavanagh	Saad A. Khan
Jason M. Haugh	Robert M. Kelly

Chair of Advisory Committee

DEDICATION

This is dedicated to my grandparents. You may be gone, but you are not forgotten.

BIOGRAPHY

Donald Andrew Comfort was born in Warren, MI in a snow storm on March 18, 1977 to Donald F. and Judith L. Comfort. He graduated valedictorian of Adlai Stevenson High School (Sterling Heights, MI) in 1995. Don earned his B.S. of chemical engineering from Case Western Reserve University in 2000. The following fall, he matriculated at North Carolina State University, where he earned his M.S. of chemical engineering (2002). On September 4. 2004 he married his beautiful wife, Kristen, who is also pursuing her Ph.D. in chemical engineering at North Carolina State University. Upon completion of his Ph.D. Don will be working at Wyeth Pharmaceuticals – Manufacturing Science and Technology group, at Research Triangle Park/Sanford, NC, specializing in purification processes.

ACKNOWLEDGMENTS

I would like to thank Dr. Robert Kelly for his guidance, encouragement, patience, and support. You have been great to work with, and to grow professionally under your guidance. Thank you to the Kelly Group Members, past and present, who have been invaluable to my learning experience and enjoyment of graduate school. Thank you to my parents who have never set limits on me. To my wife Kristen, thank you for all your support; you're the greatest. I love you, and I'll be there for you when you do this.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

Chapter 1: Introduction and Overview

Introduction

Biocatalysts that are stable at, and function under, extreme conditions are highly desirable for biotechnological applications (39). Extremophilic microorganisms have been examined in this regard, as a source of such biocatalysts, given their capacity to withstand environmental conditions that are normally debilitating to most organisms and biomolecules (5, 265). Hyperthermophiles are those extremophiles which grow optimally at temperatures > 80 °C and code for proteins which are stable and active at these temperatures (265). Of interest here are hyperthermophilic glycosyl hydrolases, enzymes that cleave covalent linkages in carbohydrates, and elucidation of the function of related 'hypothetical proteins' in genome sequences. 'Hypothetical proteins' still represent more than 40% of the open reading frames (ORFs) in most genome sequences. This work focuses on determining the function of these 'hypothetical proteins' using a combination of biochemical analysis, bioinformatic techniques, and functional genomics and characterization of α-active glycosyl hydrolases from hyperthermophilic microorganisms.

Chapter 2, *Strategic biocatalysis with hyperthermophilic enzymes*, addresses biotechnological applications in which hyperthermophilic enzymes may be employed. Examples of high temperature biocatalysis include syntheses that generate optically pure compounds for pharmaceutical uses (248), starch hydrolysis for bioenergy production and production of sweetners (8, 43), and oil and gas well stimulation (190). These are examples in which enzymes serve as environmentally advantageous catalysts, minimizing or eliminating harsh conditions that are required by traditional processing techniques.

In Chapter 3, *Identification of Active-Site Residues in the Glycosyl Hydrolase Family 36 alpha-galactosidase from Thermotoga maritima*, the α-galactosidase from *T. maritima* was cloned, expressed, and mutated for biochemical studies to identify the catalytic residues of this enzyme family. This project was a collaborative effort drawing on expertise from groups at the NCSU Hyperthermophile Lab in Raleigh, NC; Petersburg Nuclear Physics Institute in St. Petersburg, Russia; and the Royal Institute of Technology at AlbaNova University Centre in Stockholm, Sweden.

Chapter 4, *Functional genomics-based approaches for identification and characterization of novel α-glucan acting glycosyl hydrolases in the hyperthermophilic archaeon Pyrococcus furiosus*, utilizes a functional genomics and bioinformatics analysis approach coupled with biochemical characterization to identify and understand glycosyl hydrolases found in *Pyrococcus furiosus*. The goal was to develop a methodology to assign function to enzymes currently annotated as 'hypothetical proteins'. Using this process, a β-amylase enzyme from *P. furiosus* was identified by bioinformatics and activity determined based upon similarity of a conserved glycosyl hydrolase domain. A functional genomics lead from growth of *P. furiosus* on pullulan led to the investigation of the substrate specificity of the previously characterized α-glucosidase (41). A previously undetermined specificity for α-1,6-glucosidic linkages was found. This approach is broadly applicable to any microorganism for which genomic sequence and microarray slides are available.

Hyperthermophilic Microorganisms

Extremophiles are organisms that thrive in extreme conditions, such as low (227) or high temperatures (76), high (128) or low pH (27), high pressure (307), or saline environments (287). Survival in these conditions is based in part on robust enzymes produced by extremophiles, thereby making them industrially important organisms (5, 284). There are over 400 extremophilic organisms that have been isolated in pure culture. The most studied of these organisms are the hyperthermophiles, organisms which grow optimally at temperatures > 80°C (117). Hyperthermophiles have been isolated from all types of marine thermal environments, such as hot springs, continental solfatatras, deep marine sediment, and even thermal industrial environments (117, 119). Despite the number of isolates, culturivation and growth of these organisms have proven difficult. Very few hyperthermophiles can be cultured on solid media so that pure cultures must be obtained by either serial dilution or by single cell isolation with an 'optical tweezer trap' (117).

Hyperthermophiles, thought to be the least evolved of modern organisms and closest to the last common ancestor, are only represented in the Bacteria and Archaea kingdoms (246). The archaea are divided into two phyla: the Crenarchaeota and the

Euryarchaeota. Members of these phyla represent a broad spectrum of physiological types including: methanogens (129), fermentative anaerobes (76), sulfur-reducers (145), sulfate reducers (29), and microaerophiles (292). Hyperthermophilic bacteria are members of the phyla Thermotogae, Aquificae, Deinococcus-Thermus and Firmicutes (118, 120, 215, 230). Archaea are organisms that were originally thought to thrive only in extreme environments, such as the seas of Antarctica, thermal solfatara fields, and deep-sea vents, but environmental rRNA sequences now indicate they are found in less extreme habitats (233). One archaeon, *Nanoarchaeum equitans* KIN4-M, is the smallest organism identified to date, with a diameter of 400 nm, and is found attached to an archaeal host (116). Its genome sequence is the smallest reported to date and is very compact. The genome has 491,885 bases, with 95% of the genome encoding for machinery and information processing, but does not contain genes for synthesis of lipids, cofactors, or amino acids (297). The archaeon *Pyrolobus fumarii* was isolated from a hydrothermally heated black smoker at the Mid-Atlantic Ridge and is capable of growth at 113°C, the highest record growth temperature (20). The growth temperature upper limit for hyperthermophilic bacteria is lower than for the archaea, reaching maximums of about 90-95°C (120). *P. furiosus* is the most studied archaea and is in the Thermococcaceae family, which also includes numerous *Thermococcus* species. *P. furiosus* grows on peptides in the presence of elemental sulfur and carbohydrates. It grows optimally at 98-100°C (76), and its genome encodes at least 16 glycosyl hydrolases and 25 glycosyl transferases for sugar processing and utilization (232). Still, despite being the most studied archaeon, 43% of the proteins encoded in its genome are annotated as 'hypothetical', 'conserved hypothetical', or have not been assigned a role (222). Archaea are thought to share genetic elements through lateral gene transfer. For example, over a contiguous 16 Kb region of *P. furiosus* and *Thermococcus litoralis* encoding for the maltose/trehalose ABC transport system, only 153 nucleotides were distinct (64). Lateral gene transfer is not limited to the archaea; many regions of the genome of the hyperthermophilic bacterium *Thermotoga maritima* share sequence similarities with hyperthermophilic archaea (7, 207, 302). *T. maritima*, one of the most studied hyperthermophilic bacteria, is a fermentative anaerobe which grows optimally at 80°C and up to 90°C and was isolated from geothermally heated locales on the sea floor

(118). Compared to other microorganisms, its genome encodes a disproportionally high number of carbohydrate active enzymes (34). The transcriptional response of the organism to growth on these various substrates has been investigated for transport related proteins and carbohydrate processing enzymes (34, 40, 208).

Adaptations to Life at High Temperatures

Life at high temperatures requires special biomolecular adaptations to resist the biologically-devastating influnces of the environment. DNA repair mechanisms are required to deal with the increased levels of damage (63). The lipid bi-layer of these organisms has also evolved to be more rugged, with strategies such as membrane-spanning dibiphytanyl diglycerol tetraethers, internal cyclization of dibiphytanyl diglycerol tetraethers, internal covalent cross-linking, or cyclic diphytanyl glycerol ether (32). Hyperthermophilic organisms employ a number of intrinsic and extrinsic techniques to stabilize protein at these extreme temperatures (264). While no unifying strategy has been identified for protein stability, there are a number of trends and common features that have been identified. Hyperthermophilic proteins tend to have more intramolecular interactions, such as hydrogen bonds, electrostatic interactions, hydrophobic interactions, disulfide bonds, and metal binding than proteins from mesophilic organisms (223). These proteins also tend to be more rigid with higher packing efficiency, reduced entropy of unfolding, conformation strain release, and stability of α-helix (153, 169, 223).

To deal with environmental perturbations, hyperthermophilic organisms synthesize and/or accumulate small organic molecules, such as mannosylglycerate, trehalose, and di-myo-inositol 1,1'(3,3')-phosphate, that serve as compatible solutes (osmolytes) and also as energy reserves (183). Compatible solutes can be accumulated in high concentrations without interferring with the central metabolism (28). During periods of osmotic stress, compatible solutes help maintain appropriate turgor pressure and permit normal enzyme function (182). Compatible solutes also play a role in stabilizing enzymes against thermal denaturation; for example, diglycerol phosphate was shown to stabilize and protect proteins against heating, increasing protein half-lives by up to four-fold (157). This stabilization of proteins by compatible solutes is believed to be the result of an increase in the enthalpy of interactions and decrease in the entropy of unfolding (73,

158). Using functional genomics experiments, archaea have been shown to respond to heat-shock by activating genes involved in compatible solute synthesis as well as other chaperone proteins (200, 255). These and other investigations have shown that hyperthermophilic organisms utilize a number of techniques to stabilize proteins and adapt to survival at these extreme temperatures and the fluctuations of environmental conditions.

Glycosyl Hydrolases

Carbohydrate-active enzymes can be viewed as environmentally friendly catalysts that have technological importance. There are many such examples, α-Amylases are used in the baking and brewing industry and in the starch industry for production of high-fructose corn syrup (43). In the dairy industry, β-galactosidase is used to hydrolyze lactose from milk products to minimize effects on lactose-intolerant individuals (94). Chitinases which hydrolyze β-1,4 linkages in chitin, have potential for pesticides, fungicides, and antimalarials (113, 150, 241). Hemicellulases such as xylanases, glucanases, and cellulases have been investigated as means of hydrolyzing wood pulp to glucose for bioenergy conversion (269), bleaching of wood pulp (15, 289), and feed processing for live-stock (14). Galactomannans are utilized in the oil and gas industry for the fracturing of oil wells and are subterraneally hydrolyzed by β-mannanases and α-galactosidase, allowing the oil/gas to flow through the well (190).

Glycosyl hydrolases specifically hydrolyze carbohydrate linkages – recognizing the sugar type, glycosidic linkage, chain length, and anomeric conformation. *In vivo*, they are utilized by organisms for physiological and ecological purposes, such as carbohydrate hydrolysis for energy or release from biofilms (228). These enzymes break down carbohydrates by either endo- or exo-acting reactions. Hydrolysis of the substrate can result in a sugar product with either the same or opposite anomeric conformation as the substrate, proceeding through a retaining or inverting mechanism, respectively (236). Both retaining and inverting mechanisms involve two acidic catalytic residues, the nucleophile and acid-base residues in retaining enzymes and the acid and base catalysts in inverting enzymes (259). The hydrolysis mechanisms are essentially as first predicted by Koshland in 1953 (149) and are shown in Figure 1.1 and Figure 1.2. In the retaining

mechanism, hydrolysis occurs via a two-step process. In the first step, there is nucleophile attack and the nucleophilic residue is glycosylated, creating a covalently bound glycosyl-enzyme intermediate. The second step releases the glycosyl-enzyme intermediate via an oxocarbenium-ion-like transition state, generating a product with the same anomeric conformation as the substrate (259). The inverting mechanism involves both acid- and base-catalytic residues, in a single-step reaction in which there is a general-base-catalyzed attack of water at the anomeric center, concomitantly with general-acid-catalytic assistance of aglycone departure (310). This generates a product of opposite anomeric conformation from the substrate, as shown in Figure 1.2.

In addition to the difference in mechanisms, retaining and inverting glycosyl hydrolases have different structural features. In particular, the distance between the carboxyl catalytic residues is conserved for the different mechanisms. For inverting enzymes, the carboxyls are separated by ~9.5 Å, whereas they are only ~5.5 Å apart in retaining enzymes (187, 300). This feature can be utilized to identify the mechanism in protein crystal structures. Attempts to change the reaction mechanism from a retaining mechanism to an inverting mechanism have met with limited success. MacLeod *et al.* (181) shortened the active site nucleophile (increasing the separation distance) by mutating the glutamate to aspartate (E223D). This reduced the rate of hydrolysis 3,000-4,000-fold, with no apparent change in the reaction mechanism. The active site of the T4 lysozyme was re-engineered into a transglycosidase with hydrolytic activity by a substitution of histidine for threonine (T26H), which also successfully changed the hydrolytic mechanism from inverting to retaining by changing the spatial orientation of the catalytic pocket (154). In the sialidase from *Micromonospora viridifaciens*, mutagenesis of a conserved tyrosine (Y370) changed this retaining enzyme into an inverting mechanism which utilized water as the nucleophile (298).

Classification of Glycosyl Hydrolases

Glycosyl hydrolases are classified by two different schemes. The IUB Enzyme Nomenclature (*i.e.*, EC number) is based upon the substrate specificity. Glycosyl hydrolases are EC 3.2.1.x, where the first three numbers indicate that the enzyme hydrolyzes *O*-linked glycosides and the last position specifies the substrate. The current,

more popular, classification method was proposed by Henrissat in 1991 and is based upon primary sequence similarities as determined by hydrophobic cluster analysis (51, 107-109). These families have grown from 291 glycosyl hydrolase gene sequences and 35 families to 106 families presently, and almost 20,000 gene sequences. The compilation is regularly updated on the CAZY website (http://www.cazy.org/CAZY/) (109). The advantage to this method is the ability to extend structural data to other enzymes in the family.

These families are more loosely grouped into 'clans' which are thought to have evolved from a common ancestral gene, and are recognized by significant similarities in tertiary structure and conserved catalytic residues and catalytic mechanisms (109). There have been attempts in recent years to identify the catalytic residues of the glycosyl hydrolase families (310), see Table 1. Once the catalytic residues have been identified for one enzyme within a family, the catalytic residues of other enzymes within the family can be determined by sequence alignment. There are a number of approaches that have been used for identifying the catalytic residues of glycosyl hydrolases – crystal structure (267), kinetic analysis (25), long-lived inactivator substrate followed by (LC)-MS (188), and chemical rescue (26). The use of crystal structures and detailed kinetic analysis is applicable to both retaining and inverting glycosyl hydrolases. However, the use of inactivators coupled with MS is only possible for retaining enzymes because the inverting reaction does not form a glycosyl-enzyme intermediate (310). The inactivator method has also been the most prevalent method in the identification of the nucleophile residue.

Identification of the Nucleophilic Residue

Inactivator compounds, or mechanism-based inhibitors, are substrate analogs which are not readily de-glycosylated after formation of the glycosyl-enzyme intermediate (203). These inactivators are typically halogenated sugars with fluorine, surrounding the anomeric carbon. Substrates, such as 2-deoxy-2-fluoro glycosides, have been successful at inactivating retaining β-glycosidases, however, success with α-glycosidases inactivation has been limited (189). The catalytic nucleophile of the family 29 α-L-fucosidase from *T. maritima* was identified by trapping the covalent glycosyl enzyme intermediate using 2-deoxy-2-fluoro-α-L-fucosyl fluoride (274). There has been

more success with α-glycosidases using inhibitors such as 2-deoxy-2,2-difluoro-glycosides (105), which was used to identify the nucleophile of the family 27 α-galactosidase, and 5-fluoro-glycosides (186), which were used to identify the nucleophile of the α-glucosidase from the archaeon *Ferroplasma acidophilum* (75) and α-glucosidase from yeast (188).

Chemical rescue has been used to identify the catalytic nucleophile of the family 29 α-L-fucosidase from the archaeon *Solfolobus solfataricus* (38). This technique involves generating a nucleophile mutant using site-directed mutagenesis, thereby eliminating the activity of the enzyme. In the presence of external nucleophiles, such as azide or formate, limited activity is restored. Using the family 1 β-glucosidase nucleophile mutant Glu387Ala from *S. solfataricus*, Moracci *et al.* (197) demonstrated that these external nucleophiles are incorporated into the hydrolysis products. Interestingly, the mechanism of hydrolysis and product anomeric conformation was dependent upon the external nucleophile used. When sodium azide was employed, the hydrolysis product was α-glucosyl azide, indicating that the mechanism proceeded as an inverting enzyme. Whereas when sodium formate was used at the external nucleophile, the hydrolysis product had the β-anomeric conformation and the product was a disaccharide (197). When the *S. solfataricus* α-L-fucosidase nucleophile mutant Asp242Gly was rescued with azide, the hydrolysis product of *p*-nitrophenyl- α-L-fucopyranoside was β-fucosyl azide, with the opposite anomeric conformation of the substrate (38).

Identification of the Acid-Base Residue

The acid-base residue can also be identified through a combination of sequence analysis, crystal structure, use of an inhibitor, and/or through detailed biochemical analysis. Unfortunately, sequence alignment alone is not enough to definitively identify the acid/base residue. Synthesis and application of an acid/base inactivator is considerably more difficult than for the nucleophile residue, such that this approach has been rarely utilized (114, 283). These inactivators are usually bromosugar analogs of natural or synthetic substrates. The acid/base catalyst of the α-glucosidase from

Saccharomyces cerevisiae was labeled using 1'-bromo-3'-(α-D-mannopyranosyl)-2'-propanone and confirmed by LC/MS analysis of the peptic digests of labeled versus unlabeled enzymes samples (114). *N*-Bromoacetyl cellobiosylamine was used to identify Glu127 of the exoglycanase from *Cellulomonas fimi* as the acid/base residue using electrospray ionization mass spectrometry (283).

Kinetic analysis of site-directed mutants has been widely used to identify acid/base residues, more so than inactivators. There are a number of techniques for kinetic identification of the acid/base residue that have been utilized and all involve site-directed mutagenesis. One method is to compare Michaelis-Menten kinetic parameters for substrates with good and poor leaving groups, such as 2,4-dinitrophenyl (DNP) and 4-nitrophenyl (PNP), respectively, in the presence of the acid/base mutant enzyme. For substrates containing DNP there is no reduction in the k_{cat}/K_M, which reflects the first irreversible step, probably formation of the glycosyl-enzyme intermediate (170, 180). For some substrates, such as PNP-sugars, there is a reduction of the k_{cat}/K_M of up to 6000-fold, which is the result of needing protonic assistance for initial bond cleavage (180). The acid/base residue can also be examined using an increasing concentration of azide in a procedure known as "azide rescue". The small azide molecule enters the mutated region normally inhabited by the glutamate or aspartate. In the presence of azide, the acid/base mutant enzyme will form a sugar azide compound with a retained conformation. For instance, the β-exoglucanase/xylanase from *C. fimi* forms β-cellobiosyl azide in the presence of sodium azide (180) and for the β-mannosidase from *C. fimi* which produced β-mannosyl azide (309). When the second step is rate-limiting (good substrate), the increase in the azide accelerates the de-glycosylation step increasing the k_{cat}. This was shown for the β-xylosidase from *Bacillus stearothermophilus* T-6 with the result that the k_{cat}/K_M remained constant for the substrate 2,5-DNP-xylopyranoside because the K_M also increased due to less glycosyl-enzyme intermediate accumulation (26). One can also look at the pH profile of the acid-base mutant enzyme for confirmation of the residue. Wild-type glycosyl hydrolases exhibit the normal bell-shaped curve for pH dependence, however, removal of the protonated group results in the acid-base mutants showing no reduction in the activity at high pH values (26).

Crystal Structures

The crystal structure of the enzyme can also be used to determine the catalytic residues of the enzyme, especially when a substrate in complexed with the structure. Recent structural genomics projects for *T. maritima* (166), *P. furiosus* (1), yeast (168) and others have greatly increased the number of crystal structures available and facilitated the identification of catalytic residues for a wide range of enzymes. The crystal structure of the family 29 α-L-fucosidase from *T. maritima* confirmed the nucleophile as Asp224 and identified the acid-base residue as Glu266 (267). The crystal structure of another glycosidase from *T. maritima*, the β-fructosidase (invertase), was recently solved and identified the nucleophile and acid-base residues as Asp17 and Glu190, respectively (3). The crystal structure of the amylosucrase from *Neisseria polysaccarea*, GHF 13, was used to confirm the nucleophile, however, the crystal structure identified the acid-base residue as Glu328 instead of the previously reported Glu336 (195, 239). In the crystal structure of the family 27 α-N-acetylgalactoaminidase from chicken, the catalytic residues were 5.6 Å apart indicating that the enzyme is a retaining hydrolase (90). Crystal structures are the best means of identifying catalytic residues, especially when complexed with substrates in the catalytic pocket; unfortunately, generating these structures has been a slow process and is only presently starting to become automated and high-through-put (31, 281).

Alpha-Galactosidase

In Chapter 3, biochemical characterization of the α-galactosidase from *T. maritima* is discussed with a focus on identification of the catalytic residues of the family 36 glycosyl hydrolases. α-Galactosidase catalyzes the release of α-linked D-galactopyranosides from compounds such as galactomannan polysaccharide from seeds (58), raffinose and stachyose from legumes and sugar beets (58, 87), pretreatment of animal feeds (100, 164), and in eukaryotic organisms for the de-glyosylation of proteins (99). In humans, reduced activity of the α-galactosidase enzyme results in the lysosomal storage disease known as Fabry's Disease, which is characterized by pain in the extremities, skin and mucous membranes angiokeratoma, hypohydrosis, and corneal opacities (225). Garman and Garboczi (89) compiled more 190 reported point mutations

of α-galactosidase and related enzymes and identified two major protein defects which resulted in decreased activity: active site mutations and folding mutations. During phagocytosis, the lysosome consumes α-galactose glycosylated proteins and glycosphingolipids, which are intended for degradation. In affected individuals, however, the mutant enzyme is unable to degrade these glycoproteins and glycosphingolipids, causing them to accumulate. In particular, the glycosphingolipid globotriaosylceramide builds up in the blood and walls of the blood vessels, leading to blocked arteries (72). Enzyme replacement therapy has been shown to decrease the levels of globotriaosylceramide for those with Fabry's Disease through treatment with drugs such as the FDA approved Fabrazyme from Genzyme Corp. (Cambridge, MA) (301).

α-Galactosidase has also been shown to be useful in the treatment of animal feed for the reduction of raffinose, stachyose, and verbacose. These sugars comprise approximately 6% of soybean meal, a common chick feed, and cannot be digested in the small intestine of monogastric animals (100). They act as anti-nutritive agents which result in fermentation in the hindgut to generate hydrogen and other undesired gases (164). The addition of stachyose or raffinose to soy protein concentrate was shown to significantly reduce total metabolizable energy of feed with a dose-dependent response (165). Pretreatment of the soybean meal with sprayed on α-galactosidase was shown to decrease the raffinose and stachyose content of the soybean meal feed by 69% and 54%, respectively (100). The α-galactosidase from *T. maritima* was also investigated for the hydrolysis of α-galactosides in soybean meal (Levy, Miller, and Kelly, unpublished data). This enzyme has the advantage that in can be incorporated in the soybean meal cracking and de-hulling processes, which occurs at temperatures >60°C.

α-Galactosidases have been characterized from all kingdoms of life and are ubiquitous in nature. They are frequently found in plant seeds to process the galactomannan polysaccharide which is used for energy storage storage (59). They are classified into the glycosyl hydrolase families 4, 27, 36, and 57. Family 4 contains only bacterial enzymes with α-galactosidase, maltose-6-phosphate glucosidase, α-glucosidase, 6-phopsho-β-glucosidase, and α-glucuronidase activities. Family 57 only contains only two α-galactosidases, from the archaea *P. furiosus* and *Thermococcus alcaliphilus*

AEDII12RA (201, 285). Most of the α-galactosidases are grouped into families 27 and 36. The α-galactosidases in family 27 are primarily eukaryotic enzymes, however, there are a handful of bacterial α-galactosidases. The *T. maritima* bacterial α-galactosidase characterized herein is found in GHF 36, along with archaeal α-galactosidases from *Sulfolobus solfataricus* P2 and *Sulfolobus tokodaii* 7 and some eukaryotic α-galactosidases (137, 263). The GHF 27 and 36 enzyme group together in a common clan based upon structural similarities in an identified triose phosphate isomerase (TIM) barrel (231). The catalytic nucleophile of the GHF 27 α-galactosidase from green coffee was labeled using 5-fluoro-α-D-galactosyl fluoride which identified Asp145 (176), and the α-galactosidase of *Phanerochaete chrysosporium* was labeled using 2',4',6'-trinitrophenyl 2-deoxy-2,2-difluoro-alpha-D-lyxo-hexopyranoside to identify Asp130 as the nucleophile (105). A number of GHF 27 crystal structures have been solved which have allowed for identification of the acid-base residue (84, 90, 97). These crystal structures will also be used in sequence alignments to identify the catalytic residues of related enzymes in GHF 36.

Most of the α-galactosidases characterized to date have been from GHF 27 and 36. The most thermostable of the family 27 α-galactosidases was cloned and expressed from *Bacillus stearothermophilus* NUB3621, which had a T_{opt} of 75°C and 19 h half-life at 70°C (82). A related α-galactosidase from *B. stearothermophilus* KVE39 was characterized and shown to remove D-raffinose from sugar beet syrup to aid crystallization of sucrose (87). A number of α-galactosidases have been shown to possess transglycosylation activity, including the enzyme from *Trichoderma reesei* (250) and green coffee beans (250, 311). These transglycosylation reactions can be accelerated using donors with good leaving groups such as α-galactosyl fluoride (261).

Many of the α-galactosidases found in GHF 36 have been cloned and expressed from thermophilic bacteria (83, 143, 144, 171, 190), and have have temperature optima of between 90 and 105°C. The *S. solfataricus* α-galactosidase was recently, characterized and found to have a temperature optima of 90°C (263). Using sequence alignments, site-directed mutagenesis, and biochemical characterization of proposed catalytic residues, Stan *et al.* (263) verified the importance of these residues for catalysis, but were unable to

definitely, biochemically support assignment of the role of the amino acid residues in catalysis. This family also contains bacterial α-galactosidases that are not thermophilic such as the enzyme from *Lactobacillus fermentum* and *Bifidobacterium longum* which have maximum activity about 45°C (91, 92). There is also strong similarity between these enzymes with 56% similarity between the *T. maritima* and *Thermus brockianus* α-galactosidases and 94% similarity between the *Thermotoga* enzymes. GHF 57 contains the most thermostable α-galactosidase characterized to date, expressed from *P. furiosus*. This enzyme has optimum activity at 115°C and a 15 hour half-life at 100°C (285).

Beta-Amylases

Organisms capable of processing starch and other α-glucans possess an array of glycolytic enzymes with activities towards maltodextrins of varying lengths. α-Amylases and pullulanases hydrolyze the polysaccharides into maltodextrins, which can then be further hydrolyzed by β-amylases, α-glucosidases, and glucanotransferases. β-Amylases cleave α-1,4 glucosidic linkages liberating maltose from the non-reducing end of starch, glycogen, and maltooligosaccharides, however, the enzyme cannot hydrolyze α-1,6-linkages so conversion of these polysaccharides is incomplete (237). The enzymes are widely distributed in plants and found less frequently in fungi and bacteria. The plant β-amylases have little activity against starch (173), but the bacterial β-amylases are usually capable of degrading starch (60). All β-amylases identified to data are classified into GHF 14, which is dominated by eukaryotic β-amylases. Crystal structures of the β-amylases from *Bacillus cereus* (191, 217), soy bean (192, 193), barley (194), and sweet potato (33) are available.

The recent increase in petroleum costs has brought about an increased interest in fuel ethanol (bioethanol) – ethanol produced from renewable resources such as corn, wheat, sugar beets or cane, or cellulosic by-products (39, 299). Complete hydrolysis of the feed carbohydrate components is essential for economical implementation of fuel ethanol processes (61). β-Amylases play an important role in the final processing stages of maltodextrins generating free maltose, increasing maltose yields up to 20-25% following treatment of α-glucans with α-amylase or pullulanase (57). Saccharification

processes acid treat starch with α-amylases, followed by liquefaction at 105°C, and saccharification at 60°C (43). The saccharification process relies on the enzymes glucoamylase, pullulanse, β-amylase, and α-glucosidase to completely hydrolyze the starch in glucose, and more thermal stable enzymes are being sought which will allow reactors to run at higher temperatures for shorter periods of time (42). The most thermostable β-amylase characterized to date is from *Clostridium thermosulfurogenes* with a temperature optima of 75°C (122, 237, 254). The enzyme is secreted extracellularly and adsorbs onto raw starch, is capable of hydrolyzing raw and soluble starch, and is active on maltooligosaccharides as short at maltotriose. In Chapter 4, a novel β-amylase from *P. furiosus* which has a temperature optimum > 110°C will be discussed.

Carbohydrate Utilization by Pyrococcus furiosus

P. furiosus is a heterotrophic, hyperthermophilic archaeon isolated from solfatara fields off Volcano Island, Italy. It grows optimally at 98-100°C and can utilize carbohydrates and peptides as carbon sources which is consistent with analysis of the genome (76, 232). *P. furiosus* is capable of growing on a number of carbohydrates as determined by the glycosyl hydrolase content including: β-glucosidases (PF0073, PF0442) (13, 135, 293), β-glucanases (PF0854) (13), laminarinase (PF0076) (68, 102), α-glucosidases and α-glucanases (PF0132, PF0272, PF0477, PF1935*) (65, 66, 130, 155, 156, 160), cyclodextrinase and cycloglucanotransferase (PF0478, PF1939) (305), chitinases (PF1233, PF1234) (88, 272), α-galactosidase (PF0444) (285), and β-mannanase (PF1208) (184). These and other carbohydrate active enzymes of the *P. furiosus* genome are found in Table 2. α-Glucan polysaccharides, such as starch and pullulan, are hydrolyzed into smaller maltooligosaccharides by endo-acting pullulanases, α-amylases, debranching enzymes, and glucoamylases. These maltooligosaccharides are further hydrolyzed into glucose for metabolic processes such as glycolysis. *P. furiosus* possesses a pullulanase (PF1935*) (66) and an α-amylase (PF0477) (65, 130) for extracellular hydrolysis of the α-glucans. Lee *et al.* (160) recently proposed a utilization pathway for starch hydrolysis in which both maltodextrins and maltose are transported

into the cell and metabolized. Transport of maltose is by the Mal-I ABC transporter (MalE-like, PF1739-1741, PF1744) (161) and maltodextrins by the Mal-II ABC transporter (PF1933, PF1936-1938) (162). Intracellular processing of maltose is by the α-amylase/glucanotransferase, PF0272, with the products of glucose and maltodextrin$_{n+1}$ (160). Processing of the maltodextrins is by the intracellular α-glucan phosphorylase which generates glucose-1-phosphate and maltodextrin$_{n-1}$ from maltodextrins (308).

The uptake of sugars by organisms is accomplished by either a phosphoenol pyruvate phosphotransferase system (PTS) (95, 226) or via binding with ATP-binding cassette (ABC) transporters (71, 243). Genome analysis of *P. furiosus* and other archaea have not identified PTS in these organisms (148, 296), however, ABC-transporters are prominent in the genomes of the heterotrophic archaea. The genome of *P. furiosus* encodes two ABC-transporter systems for the transport of α-glucans. The Mal-I operon (PF1739-PF1750) contains a malE-like transporter set which transports maltose and trehalose (161). The malE-like operon of *Thermococcus litoralis*, for which the *malE* gene of the *malEFG* gene cluster from *T. litoralis* was cloned and expressed and showed binding activity towards maltose and trehalose (111). Homologs of this operon have also been found in, and have been characterized for, *Escherichia coli* (21), and this group of genes is also present in *P. furiosus* in which it also transports maltose and trehalose. This gene cluster contains *malE*, *malF*, *malG*, *treT*, *trmB*, and *malK* (64, 101, 111, 229). A second operon (PF1933-PF1938) is specific for maltodextrins (162). Both of these transporters are regulated by the dual acting regulator, TrmB which recognizes maltose, maltodextrins, sucrose, and trehalose (162). In the case of the trehalose/maltose binding protein promoter, repression by TrmB is released by the presence of maltose or trehalose, but not sucrose or maltodextrins. Conversely, repression of the maltodextrin promoter was released by sucrose and maltodextrins, but not trehalose or maltose (162).

The uptake of glycogen in *E. coli* was recently proposed by Dauvillée (48) in which glycogen is phosphorylated to phosphorylase-limited glycogen by glycogen phosphorylase (GlgP) followed by formation of maltotetraose by a glycogen debranching enzyme (GlgX). The maltotetraose is then phosphorylated by maltodextrin phosphorylase (MalP) generating maltotriose and glucose-1-phosphate followed by hydrolysis to maltose and glucose by maltodextrin glucosidase (MalZ) (48, 62). These enzymes do not

have direct homologs in *P. furiosus* as determined by BLASTp analysis; GlgX is similar to the α-glucan phosphorylase (PF1535) and the intracellular MalZ is most similar to the intrcellular cyclomaltodextrinase (PF1939). The reaction performed by the enzyme MalZ could also be accomplished by the intracellular α-glucosidase (PF0132).

P. furiosus utilizes sugars for energy through a modified Embden-Meyerhof pathway in which the kinases are ADP-dependent rather than ATP-dependent (138, 139). The presence of ADP-glucokinases and ADP-dependent phosphofructokinases are also present in the other hyperthermophilic archaea *T. litoralis* and *Thermococcus celer* (53, 139, 249). These novel ADP-depenedent kinases have been cloned from *P. furiosus* and *T. litoralis* and enzymatic characterization shows phosphorylation of glucose by ADP (146, 282). Another EM-pathway modification in *P. furiosus* directly converts glyceraldehydes-3-phosphate to 3-phosphoglycerate and reduced ferredoxin through a glyceraldehydes-3-phosphate oxidoreductase (199).

In Chapter 4, a detailed functional genomics study of *P. furiosus* grown on various α-glucan polysaccharides and oligosaccharides is discussed. ORFs identified by bioinformatics techniques and thought to be involved in α-glucan processing were investigated biochemically and the transcriptional response analyzed when applicable. Functional genomics was used to glean hints of enzyme function and subsequent biochemical characterizations were performed. Combining the functional data, biochemical characterization, and bioinformatics analysis, allowed for the proposal of a pullulan utilization pathway which is different from the starch utilization pathway (160).

Genomes and Hypothetical ORFs

In the post-genomics era, determining the function of the open reading frames (ORFs) that have been identified represents a significant challenge. To date, the Comprehensive Microbial Resource (222) lists 285 complete genome sequences with 9 genomes partially sequenced. ORFs are identified with in the genomes and, once identified, functional assignment is completed through a process of sequence comparisons using BLAST (4) followed by a modified Smith-Waterman (260) alignment against the mini-database of BLAST hits. Comparative genome sequence analysis tools, such as BLAST (4), SMART (167, 244), pFam (12) and ClustalW (280), allow for rapid

identification of similar proteins, domains within the proteins, and alignment of multiple proteins. However, the information gleaned from this comparisons is reliant upon accurate assignment of function to other proteins. Challenges arise when there are few or no functional assignments of related proteins, *i.e.* the basis for assigning 'hypothetical protein' or 'conserved hypothetical protein' functional annotations.

In order to completely understand how an organism functions and processes, determination of the role these unknown predicted ORFs is essential (86). Frequently these 'hypothetical proteins' are related to other proteins within only one kingdom, such as 'conserved archaeal proteins'. Interpreting hypothetical proteins or conserved hypothetical proteins that are widely distributed can also be difficult. Information can frequently be gained about hypothetical ORFs using some of the aforementioned tools. However, these are generally broad functional assignments such as ATPases, GTPases, DNA or RNA-binding protein (6). Unfortunately, hypothetical ORFs still comprise 30-40% of sequenced microbial genomes (22), and that number is even higher for archaea. Most challenging in terms of annotation, however, is defining roles for proteins with decreasing levels of amino acid sequence conservation that are found in only a few closely related organisms (257). These proteins are referred to as orphan ORFS, or ORFans (79, 80, 258), and can account for up to 25% of the ORFs of newly sequenced genomes (81). Due to the minimal amount of comparisons that are available for these proteins, establishing the biological function of ORFans ultimately requires biochemical and genetic characterization (85). In *P. furiosus*, there are eight paralogous ORFans and 131 singleton ORFans (256); paralogous ORFans are those related ORFs found in multiple copies in one organism and singleton ORFans exist in only one copy in a single organism.

Functional genomics approaches utilizing transcriptional response patterns can complement biochemical characterization of proteins by helping elucidate the role of proteins within an organism. Combining biochemical characterization with functional genomics can definitively assign function to proteins, which may then translate to related proteins in other genomes. Whole-genome transcriptional analysis has been used to investigate carbohydrate utilization patterns in mesophilic (9) and hyperthermophilic microorganisms (34, 40, 160, 204, 208) to define carbohydrate utilization operons and

regulons. Transcriptional analysis results have confirmed the role of some carbohydrate binding proteins and glycosyl hydrolases from *T. maritima* through biochemical characterization (34, 204). These techniques have been applied to only ORFs with varying degrees of annotation, but not hypothetical proteins. In Chapter 4, an approach based on the combination of bioinformatic analysis, functional genomics, and biochemical characterization are used concomitantly for full characterization of hypothetical proteins.

References

1. **Adams, M. W. W., H. A. Dailey, L. J. Delucas, M. Luo, J. H. Prestegard, J. P. Rose, and B. C. Wang.** 2003. The Southeast Collaboratory for Structural Genomics: A high-throughput gene to structure factory. Accounts Chem. Res. **36:**191-198.
2. **Adhikari, P., S. D. Kirby, A. J. Nowalk, K. L. Veraldi, A. B. Schryvers, and T. A. Mietzner.** 1995. Biochemical characterization of a *Haemophilus influenzae p*eriplasmic iron transport operon. J. Biol. Chem. **270:**25142-25149.
3. **Alberto, F., C. Bignon, G. Sulzenbacher, B. Henrissat, and M. Czjzek.** 2004. The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. J. Biol. Chem. **279:**18903-18910.
4. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic Local Alignment Search Tool. J. Mol. Biol. **215:**403-410.
5. **Antranikian, G., C. E. Vorgias, and C. Bertoldo.** 2005. Extreme environments as a resource for microorganisms and novel biocatalysts, p. 219-262, Adv. Biochem. Eng. Biot., vol. 96.
6. **Aravind, L., and E. V. Koonin.** 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J. Mol. Biol. **287:**1023-1040.
7. **Aravind, L., and E. V. Koonin.** 1999. Novel predicted RNA-binding domains associated with the translation machinery. J. Mol. Evol. **48:**291-302.
8. **Bandlish, R. K., J. M. Hess, K. L. Epting, C. Vieille, and R. M. Kelly.** 2002. Glucose-to-fructose conversion at high temperatures with xylose (glucose) isomerases from *Streptomyces murinus* and two hyperthermophilic *Thermotoga* species. Biotechnol. Bioeng. **80:**185-194.
9. **Barrangou, R., M. A. Azcarate-Peril, T. Duong, S. B. Conners, R. M. Kelly, and T. R. Klaenhammer.** 2006. Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays. P. Natl. Acad. Sci. USA **103:**3816-3821.
10. **Barroga, C. F., H. D. Zhang, N. Wajih, J. H. Bouyer, and M. A. Hermodson.** 1996. The proteins encoded by the rbs operon of *Escherichia coli*. 1. Overproduction, purification, characterization, and functional analysis of RbsA. Protein Sci. **5:**1093-1099.
11. **Bateman, A.** 1999. The SIS domain: a phosphosugar-binding domain. Trends Biochem. Sci. **24:**94-95.
12. **Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. L. Sonnhammer.** 2002. The Pfam Protein Families Database. Nucleic Acids Res. **30:**276-280.
13. **Bauer, M. W., L. E. Driskill, W. Callen, M. A. Snead, E. J. Mathur, and R. M. Kelly.** 1999. An endoglucanase, eglA, from the hyperthermophilic archaeon *Pyrococcus furiosus* hydrolyzes beta-1,4 bonds in mixed-linkage (1 -> 3),(1 -> 4)-beta-D-glucans and cellulose. J. Bacteriol. **181:**284-290.
14. **Bedford, M. R.** 1995. Mechanism of action and potential environmental benefits from the use of feed enzymes. Anim. Feed Sci. Tech. **53:**145-155.

15. **Beg, Q. K., M. Kapoor, L. Mahajan, and G. S. Hoondal.** 2001. Microbial xylanases and their industrial applications: a review. Appl. Microbiol. Biot. **56:**326-338.
16. **Bell, A. W., S. D. Buckel, J. M. Groarke, J. N. Hope, D. H. Kingsley, and M. A. Hermodson.** 1986. The nucleotide sequences of the Rbsd, Rbsa, and Rbsc genes of *Escherichia col* K12. J. Biol. Chem. **261:**7652-7658.
17. **Benach, J., S. Atrian, R. Gonzalez-Duarte, and R. Ladenstein.** 1998. The refined crystal structure of *Drosophila lebanonensis* alcohol dehydrogenase at 1.9 angstrom resolution. J. Mol. Biol. **282:**383-399.
18. **Blanchard, J. E., L. Gal, S. M. He, J. Foisy, R. A. J. Warren, and S. G. Withers.** 2001. The identification of the catalytic nucleophiles of two beta-galactosidases from glycoside hydrolase family 35. Carbohyd. Res. **333:**7-17.
19. **Blankenfeldt, W., M. Asuncion, J. S. Lam, and J. H. Naismith.** 2000. The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). EMBO J. **19:**6652-6663.
20. **Blochl, E., R. Rachel, S. Burggraf, D. Hafenbradl, H. W. Jannasch, and K. O. Stetter.** 1997. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. Extremophiles **1:**14-21.
21. **Boos, W., and H. Shuman.** 1998. Maltose/maltodextrin system of *Escherichia coli*: Transport, metabolism, and regulation. Microbiol. Mol. Biol. R. **62:**204-+.
22. **Bork, P.** 2000. Powers and pitfalls in sequence analysis: The 70% hurdle. Genome Res. **10:**398-400.
23. **Braithwaite, K. L., T. Barna, T. D. Spurway, S. J. Charnock, G. W. Black, N. Hughes, J. H. Lakey, R. Virden, G. P. Hazlewood, B. Henrissat, and H. J. Gilbert.** 1997. Evidence that galactanase A from *Pseudomonas fluorescens* subspecies cellulosa is a retaining family 53 glycosyl hydrolase in which E161 and E270 are the catalytic residues. Biochemistry-US **36:**15489-15500.
24. **Braun, C., T. Lindhorst, N. B. Madsen, and S. G. Withers.** 1996. Identification of Asp 549 as the catalytic nucleophile of glycogen-debranching enzyme via trapping of the glycosyl-enzyme intermediate. Biochemistry-US **35:**5458-5463.
25. **Bravman, T., V. Belakhov, D. Solomon, G. Shoham, B. Henrissat, T. Baasov, and Y. Shoham.** 2003. Identification of the catalytic residues in family 52 glycoside hydrolase, a beta-xylosidase from *Geobacillus stearothermophilus* T-6. J. Biol. Chem. **278:**26742-26749.
26. **Bravman, T., A. Mechaly, S. Shulami, V. Belakhov, T. Baasov, G. Shoham, and Y. Shoham.** 2001. Glutamic acid 160 is the acid-base catalyst of beta-xylosidase from *Bacillus stearothermophilus* T-6: a family 39 glycoside hydrolase. FEBS Lett. **495:**115-119.
27. **Brock, T. D., K. M. Brock, R. T. Belly, and R. L. Weiss.** 1972. *Sulfolobus* - New genus of sulfur-oxidizing bacteria living at low pH and high-temperature. Arch. Mikrobiol. **84:**54-&.
28. **Brown, A. D.** 1976. Microbial water stress. Bacteriol. Rev. **40:**803-846.
29. **Burggraf, S., H. W. Jannasch, B. Nicolaus, and K. O. Stetter.** 1990. *Archaeoglobus profundus* sp-nov, represents a new species within the sulfate-reducing archaebacteria. Syst. Appl. Microbiol. **13:**24-28.

30. **Bushman, J. L., A. I. Asuru, R. L. Matts, and A. G. Hinnebusch.** 1993. Evidence that Gcd6 and Gcd7, translational regulators of Gcn4, are subunits of the guanine-nucleotide exchange factor for Eif-2 in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **13:**1920-1932.
31. **Chandonia, J. M., and S. E. Brenner.** 2006. The impact of structural genomics: Expectations and outcomes. Science **311:**347-351.
32. **Charlier, D., and L. Droogmans.** 2005. Microbial life at high temperature, the challenges, the strategies. Cell. Mol. Life Sci. **62:**2974-2984.
33. **Cheong, C. G., S. H. Eom, C. S. Chang, D. H. Shin, H. K. Song, K. S. Min, J. H. Moon, K. K. Kim, K. Y. Hwang, and S. W. Suh.** 1995. Crystallization, molecular replacement solution, and refinement of tetrameric beta-amylase from sweet-potato. Proteins **21:**105-117.
34. **Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger, and R. M. Kelly.** 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J. Biol. Chem. **278:**7540-7552.
35. **Chir, J., S. Withers, C. F. Wan, and Y. K. Li.** 2002. Identification of the two essential groups in the family 3 beta-glucosidase from *Flavobacterium meningosepticum* by labelling and tandem mass spectrometric analysis. Biochem. J. **365:**857-863.
36. **Clarke, A. J., J. Drummelsmith, and M. Yaguchi.** 1997. Identification of the catalytic nucleophile in the cellulase from *Schizophyllum commune* and assignment of the enzyme to Family 5, subtype 5 of the glycosidases. FEBS Lett. **414:**359-361.
37. **Cobucci-Ponzano, B., M. Mazzone, M. Rossi, and M. Moracci.** 2005. Probing the catalytically essential residues of the alpha-L-fucosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. Biochemistry-US **44:**6331-6342.
38. **Cobucci-Ponzano, B., A. Trincone, A. Giordano, M. Rossi, and M. Moracci.** 2003. Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. Biochemistry-US **42:**9525-9531.
39. **Comfort, D. A., S. R. Chhabra, S. B. Conners, C. J. Chou, K. L. Epting, M. R. Johnson, K. L. Jones, A. C. Sehgal, and R. M. Kelly.** 2004. Strategic biocatalysis with hyperthermophilic enzymes. Green Chem. **6:**459-465.
40. **Conners, S. B., C. I. Montero, D. A. Comfort, K. R. Shockley, M. R. Johnson, S. R. Chhabra, and R. M. Kelly.** 2005. An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. J. Bacteriol. **187:**7267-7282.
41. **Costantino, H. R., S. H. Brown, and R. M. Kelly.** 1990. Purification and characterization of an alpha-glucosidase from a hyperthermophilic archaebacterium, *Pyrococcus furiosus*, exhibiting a temperature optimum of 105 degrees C to 115 degrees C. J. Bacteriol. **172:**3654-3660.
42. **Crabb, W. D., and C. Mitchinson.** 1997. Enzymes involved in the processing of starch to sugars. Trends Biotechnol. **15:**349-352.

43. **Crabb, W. D., and J. K. Shetty.** 1999. Commodity scale production of sugars from starches. Curr. Opin. Microbiol. **2:**252-256.
44. **Dabrowski, S., G. Sobiewska, J. Maciunska, J. Synowiecki, and J. Kur.** 2000. Cloning, expression, and purification of the His(6)-tagged thermostable beta-galactosidase from *Pyrococcus woesei* in *Escherichia coli* and some properties of the isolated enzyme. Protein Expres. Purif. **19:**107-112.
45. **Dahl, M. K., and M. D. Manson.** 1985. Interspecific reconstitution of maltose transport and chemotaxis in *Escherichia coli w*ith maltose-binding protein from various *Entericbacteria*. J. Bacteriol. **164:**1057-1063.
46. **Dan, S., I. Marton, M. Dekel, B. A. Bravdo, S. M. He, S. G. Withers, and O. Shoseyov.** 2000. Cloning, expression, characterization, and nucleophile identification of family 3, *Aspergillus niger* beta-glucosidase. J. Biol. Chem. **275:**4973-4980.
47. **Dassa, E., and M. Hofnung.** 1985. Sequence of gene Malg in *Escherichia coli* K12 - Homologies between integral membrane components from binding protein dependent transport systems. EMBO J. **4:**2287-2293.
48. **Dauvillee, D., I. S. Kinderf, Z. Y. Li, B. Kosar-Hashemi, M. S. Samuel, L. Rampling, S. Ball, and M. K. Morell.** 2005. Role of the *Escherichia coli* glgX gene in glycogen metabolism. J. Bacteriol. **187:**1465-1473.
49. **Davidson, A. L., H. A. Shuman, and H. Nikaido.** 1992. Mechanism of maltose transport in *Escherichia coli* - Transmembrane signaling by periplasmic binding proteins. P. Natl. Acad. Sci. USA **89:**2360-2364.
50. **Davies, G. J., M. Dauter, A. M. Brzozowski, M. E. Bjornvad, K. V. Andersen, and M. Schulein.** 1998. Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 angstrom and its cellobiose complex at 2.0 angstrom resolution. Biochemistry-US **37:**1926-1932.
51. **Davies, G. J., K. S. Wilson, and B. Henrissat.** 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. **321:**557-559.
52. **de Geus, D., A. P. Hartley, S. E. Sedelnikova, S. E. Glynn, P. J. Baker, C. H. Verhees, J. van der Oost, and D. W. Rice.** 2003. Cloning, purification, crystallization and preliminary crystallographic analysis of galactokinase from *Pyrococcus furiosus*. Acta Crystallog. D **59:**1819-1821.
53. **de Vos, W. M., S. W. M. Kengen, W. G. B. Voorhorst, and J. van der Oost.** 1998. Sugar utilization and its control in hyperthermophiles. Extremophiles **2:**201-205.
54. **Dean, M., A. Rzhetsky, and R. Allikmets.** 2001. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. **11:**1156-1166.
55. **Debeche, T., C. Bliard, P. Debeire, and M. J. O'Donohue.** 2002. Probing the catalytically essential residues of the alpha-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. Protein Eng. **15:**21-28.
56. **Degrassi, G., M. Kojic, G. Ljubijankic, and V. Venturi.** 2000. The acetyl xylan esterase of *Bacillus pumilus* belongs to a family of esterases with broad substrate specificity. Microbiol-UK **146:**1585-1591.
57. **Dessouki, A. M., G. I. Issa, and K. S. Atia.** 2001. Pullulanase immobilization on natural and synthetic polymers. J. Chem. Technol. Biot. **76:**700-706.

58. **Dey, P. M., S. Patel, and M. D. Brownleader.** 1993. Induction of alpha-galactosidase in *Penicillium ochrochloron* by guar (*Cyamopsis tetragonobola*) gum. Biotechnol. Appl. Bioc. **17:**361-371.
59. **Dey, P. M., and J. B. Pridham.** 1972. Biochemistry of alpha-galctosidases. Adv. Enzymol. RAMB **36:**91-130.
60. **Dicko, M. H., M. J. F. Searle-van Leeuwen, G. Beldman, O. G. Ouedraogo, R. Hilhorst, and A. S. Traore.** 1999. Purification and characterization of beta-amylase from *Curculigo pilosa*. Appl. Microbiol. Biot. **52:**802-805.
61. **Dien, B. S., R. J. Bothast, N. N. Nichols, and M. A. Cotta.** 2002. The US corn ethanol industry: An overview of current technology and future prospects. Int. Sugar J. **104:**204-+.
62. **Dippel, R., T. Bergmiller, A. Bohm, and W. Boos.** 2005. The maltodextrin system of *Escherichia coli*: Glycogen-derived endogenous induction and osmoregulation. J. Bacteriol. **187:**8332-8339.
63. **DiRuggiero, J., J. R. Brown, A. P. Bogert, and F. T. Robb.** 1999. DNA repair systems in Archaea: Mementos from the last universal common ancestor? J. Mol. Evol. **49:**474-484.
64. **DiRuggiero, J., D. Dunn, D. L. Maeder, R. Holley-Shanks, J. Chatard, R. Horlacher, F. T. Robb, W. Boos, and R. B. Weiss.** 2000. Evidence of recent lateral gene transfer among hyperthermophilic Archaea. Mol. Microbiol. **38:**684-693.
65. **Dong, G. Q., C. Vieille, A. Savchenko, and J. G. Zeikus.** 1997. Cloning, sequencing, and expression of the gene encoding extracellular alpha-amylase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. Appl. Environ. Microb. **63:**3569-3576.
66. **Dong, G. Q., C. Vieille, and J. G. Zeikus.** 1997. Cloning, sequencing, and expression of the gene encoding amylopullulanase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. Appl. Environ. Microb. **63:**3577-3584.
67. **Drewke, C., and M. Ciriacy.** 1988. Overexpression, purification and properties of alcohol dehydrogenase-IV from *Saccharomyces cerevisiae*. Biochim. Biophys. Acta **950:**54-60.
68. **Driskill, L. E., M. W. Bauer, and R. M. Kelly.** 1999. Synergistic interactions among beta-laminarinase, beta-1,4-glucanase, and beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* during hydrolysis of beta-1,4-, beta-1,3-, and mixed-linked polysaccharides. Biotechnol. Bioeng. **66:**51-60.
69. **Duplay, P., H. Bedouelle, A. Fowler, I. Zabin, W. Saurin, and M. Hofnung.** 1984. Sequences of the MalE gene and of its product, the maltose binding protein of *Escherichia coli* K12. J. Biol. Chem. **259:**606-613.
70. **Egloff, M. P., J. Uppenberg, L. Haalck, and H. van Tilbeurgh.** 2001. Crystal structure of maltose phosphorylase from *Lactobacillus brevis*: Unexpected evolutionary relationship with glucoamylases. Structure **9:**689-697.
71. **Ehrmann, M., R. Ehrle, E. Hofmann, W. Boos, and A. Schlosser.** 1998. The ABC maltose transporter. Mol. Microbiol. **29:**685-694.
72. **Eng, C. M., N. Guffon, W. R. Wilcox, D. P. Germain, P. Lee, S. Waldek, L. Caplan, G. E. Linthorst, and R. J. Desnick.** 2001. Safety and efficacy of

recombinant human alpha-galactosidase a replacement therapy in Fabry's disease. New Engl. J. Med. **345:**9-16.

73. **Faria, T. Q., S. Knapp, R. Ladenstein, A. L. Macanita, and H. Santos.** 2003. Protein stabilisation by compatible solutes: Effect of mannosylglycerate on unfolding thermodynamics and activity of ribonuclease A. Chembiochem **4:**734-741.
74. **Febbraio, F., R. Barone, S. Dauria, M. Rossi, R. Nucci, G. Piccialli, L. DeNapoli, S. Orru, and P. Pucci.** 1997. Identification of the active site nucleophile in the thermostable beta-glycosidase from the archaeon *Sulfolobus solfataricus* expressed in *Escherichia coli*. Biochemistry-US **36:**3068-3075.
75. **Ferrer, M., O. V. Golyshina, F. J. Plou, K. N. Timmis, and P. N. Golyshin.** 2005. A novel alpha-glucosidase from the acidophilic archaeon *Ferroplasma acidiphilum* strain Y with high transglycosylation activity and an unusual catalytic nucleophile. Biochem. J. **391:**269-276.
76. **Fiala, G., and K. O. Stetter.** 1986. *Pyrococcus furiosus* sp. nov represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 degrees C. Arch. Microbiol. **145:**56-61.
77. **Fiedler, G., M. Pajatsch, and A. Bock.** 1996. Genetics of a novel starch utilisation pathway present in *Klebsiella oxytoca*. J. Mol. Biol. **256:**279-291.
78. **Field, T. L., W. S. Reznikoff, and P. A. Frey.** 1989. Galactose-1-phosphate uridylyltransferase - Identification of histidine-164 and histidine-166 as critical residues by site-sirected mutagenesis. Biochemistry-US **28:**2094-2099.
79. **Fischer, D.** 1999. Rational structural genomics: affirmative action for ORFans and the growth in our structural knowledge. Protein Eng. **12:**1029-1030.
80. **Fischer, D., and D. Eisenberg.** 1999. Finding families for genomic ORFans. Bioinformatics **15:**759-762.
81. **Fraser, C. M., J. A. Eisen, and S. L. Salzberg.** 2000. Microbial genome sequencing. Nature **406:**799-803.
82. **Fridjonsson, O., H. Watzlawick, A. Gehweiler, and R. Mattes.** 1999. Thermostable alpha-galactosidase from *Bacillus stearothermophilus* NUB3621: cloning, sequencing and characterization. FEMS Microbiol. Lett. **176:**147-153.
83. **Fridjonsson, O., H. Watzlawick, A. Gehweiler, T. Rohrhirsch, and R. Mattes.** 1999. Cloning of the gene encoding a novel thermostable alpha-galactosidase from *Thermus brockianus* ITI360. Appl. Environ. Microb. **65:**3955-3963.
84. **Fujimoto, Z., S. Kaneko, M. Momma, H. Kobayashi, and H. Mizuno.** 2003. Crystal structure of rice alpha-galactosidase complexed with D-galactose. J. Biol. Chem. **278:**20313-20318.
85. **Galperin, M. Y.** 2001. Conserved 'hypothetical' proteins: new hints and new puzzles. Compar. Funct. Genom. **2:**14-18.
86. **Galperin, M. Y., and E. V. Koonin.** 2004. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res. **32:**5452-5463.
87. **Ganter, C., A. Bock, P. Buckel, and R. Mattes.** 1988. Production of thermostable, recombinant alpha-galactosidase suitable for raffinose elimination from sugar-beet syrup. J. Biotechnol. **8:**301-310.

88. **Gao, J., M. W. Bauer, K. R. Shockley, M. A. Pysz, and R. M. Kelly.** 2003. Growth of hyperthermophilic archaeon *Pyrococcus futiosus* on chitin involves two family 18 chitinases. Appl. Environ. Microb. **69:**3119-3128.
89. **Garman, S. C., and D. N. Garboczi.** 2002. Structural basis of Fabry disease. Mol. Genet. Metab. **77:**3-11.
90. **Garman, S. C., L. Hannick, A. Zhu, and D. N. Garboczi.** 2002. The 1.9 angstrom structure of alpha-N-acetylgalactosaminidase: Molecular basis of glycosidase deficiency diseases. Structure **10:**425-434.
91. **Garro, M. S., G. S. Degiori, G. F. Devaldez, and G. Oliver.** 1994. Alpha-D-galactosidase (EC-3.2.1.22) from *Bifidobacterium longum*. Lett. Appl. Microbiol. **19:**16-19.
92. **Garro, M. S., G. S. Degiori, G. F. Devaldez, and G. Oliver.** 1993. Characterization of alpha-galactosidase from *Lactobacillus fermentum*. J. Appl. Bacteriol. **75:**485-488.
93. **Gebler, J. C., R. Aebersold, and S. G. Withers.** 1992. Glu-537, not Glu-461, is the nucleophile in the active-site of (Lac-Z) beta-falactosidase From *Escherichia coli*. J. Biol. Chem. **267:**11126-11130.
94. **Gerday, C., M. Aittaleb, M. Bentahir, J. P. Chessa, P. Claverie, T. Collins, S. D'Amico, J. Dumont, G. Garsoux, D. Georlette, A. Hoyoux, T. Lonhienne, M. A. Meuwis, and G. Feller.** 2000. Cold-adapted enzymes: from fundamentals to biotechnology. Trends Biotechnol. **18:**103-107.
95. **Ginsburg, A., and A. Peterkofsky.** 2002. Enzyme I: The gateway to the bacterial phosphoenolpyruvate: Sugar phosphotransferase system. Arch. Biochem. Biophys. **397:**273-278.
96. **Golan, G., D. Shallom, A. Teplitsky, G. Zaide, S. Shulami, T. Baasov, V. Stojanoff, A. Thompson, Y. Shoham, and G. Shoham.** 2004. Crystal structures of *Geobacillus stearothermophilus* alpha-glucuronidase complexed with its substrate and products - Mechanistic implications. J. Biol. Chem. **279:**3014-3024.
97. **Golubev, A. M., R. A. P. Nagem, J. R. B. Neto, K. N. Neustroev, E. V. Eneyskaya, A. A. Kulminskaya, K. A. Shabalin, A. N. Savel'ev, and I. Polikarpov.** 2004. Crystal structure of alpha-galactosidase from *Trichoderma reesei* and its complex with galactose: Implications for catalytic mechanism. J. Mol. Biol. **339:**413-422.
98. **Goodlove, P. E., P. R. Cunningham, J. Parker, and D. P. Clark.** 1989. Cloning and sequence analysis of the fermentative alcohol-dehydrogenase encoding gene of *Escherichia coli*. Gene **85:**209-214.
99. **Grabowski, G. A., and R. J. Hopkin.** 2003. Enzyme therapy for lysosomal storage disease: Principles, practice, and prospects. Annu. Rev. Genom. Hum. G. **4:**403-436.
100. **Graham, K. K., M. S. Kerley, J. D. Firman, and G. L. Allee.** 2002. The effect of enzyme treatment of soybean meal on oligosaccharide disappearance and chick growth performance. Poultry Sci. **81:**1014-1019.
101. **Greller, G., R. Horlacher, J. DiRuggiero, and W. Boos.** 1999. Molecular and biochemical analysis of MalK, the ATP-hydrolyzing subunit of the trehalose maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. J. Biol. Chem. **274:**20259-20264.

102. **Gueguen, Y., W. G. B. Voorhorst, J. vanderOost, and W. M. deVos.** 1997. Molecular and biochemical characterization of an endo-beta-1,3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Biol. Chem. **272:**31258-31264.
103. **Gunn, F. J., C. G. Tate, and P. J. F. Henderson.** 1994. Identification of a novel sugar-H+ symport protein, FucP, for transport of L-fucose into *Escherichia coli*. Mol. Microbiol. **12:**799-809.
104. **Hahn, M., J. Pons, A. Planas, E. Querol, and U. Heinemann.** 1995. Crystal structure of *Bacillus licheniformis* 1,3-1,4-beta-D-glucan 4-glucanohydrolase at 1.8 Angstrom resolution. FEBS Lett. **374:**221-224.
105. **Hart, D. O., S. M. He, C. J. Chany, S. G. Withers, P. F. G. Sims, M. L. Sinnott, and H. Brumer.** 2000. Identification of Asp-130 as the catalytic nucleophile in the main alpha-galactosidase from *Phanerochaete chrysosporium*, a family 27 glycosyl hydrolase. Biochemistry-US **39:**9826-9836.
106. **Hashimoto, Y., T. Yamamoto, S. Fujiwara, M. Takagi, and T. Imanaka.** 2001. Extracellular synthesis, specific recognition, and intracellular degradation of cyclomaltodextrins by the hyperthermophilic archaeon *Thermococcus* sp strain B1001. J. Bacteriol. **183:**5050-5057.
107. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochem. J. **280:**309-316.
108. **Henrissat, B., and A. Bairoch.** 1993. New families in the classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochem. J. **293:**781-788.
109. **Henrissat, B., and A. Bairoch.** 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. **316:**695-696.
110. **Hinderlich, S., R. Stasche, R. Zeitler, and W. Reutter.** 1997. A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver - Purification and characterization of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. J. Biol. Chem. **272:**24313-24318.
111. **Horlacher, R., K. B. Xavier, H. Santos, J. DiRuggiero, M. Kossmann, and W. Boos.** 1998. Archaeal binding protein-dependent ABC transporter: Molecular and biochemical analysis of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. J. Bacteriol. **180:**680-689.
112. **Hou, Y. M., D. J. Vocadlo, A. Leung, S. G. Withers, and D. Mahuran.** 2001. Characterization of the Glu and Asp residues in the active site of human beta-hexosaminidase B. Biochemistry-US **40:**2201-2209.
113. **Houston, D. R., K. Shiomi, N. Arai, S. Omura, M. G. Peter, A. Turberg, B. Synstad, V. G. H. Eijsink, and D. M. F. van Aalten.** 2002. High-resolution structures of a chitinase complexed with natural product cyclopentapeptide inhibitors: Mimicry of carbohydrate substrate. P. Natl. Acad. Sci. USA **99:**9127-9132.
114. **Howard, S., and S. G. Withers.** 1998. Labeling and identification of the postulated acid/base catalyst in the alpha-glucosidase from *Saccharomyces cerevisiae* using a novel bromoketone C-glycoside. Biochemistry-US **37:**3858-3864.

115. **Hrmova, M., E. A. MacGregor, P. Biely, R. J. Stewart, and G. B. Fincher.** 1998. Substrate binding and catalytic mechanism of a barley beta-D-glucosidase/(1,4)-beta-D-glucan exohydrolase. J. Biol. Chem. **273:**11134-11143.
116. **Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter.** 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature **417:**63-67.
117. **Huber, R., H. Huber, and K. O. Stetter.** 2000. Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. FEMS Microbiol. Rev. **24:**615-623.
118. **Huber, R., T. A. Langworthy, H. Konig, M. Thomm, C. R. Woese, U. B. Sleytr, and K. O. Stetter.** 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90-degrees-C. Arch. Microbiol. **144:**324-333.
119. **Huber, R., and K. O. Stetter.** 2001. Discovery of hyperthermophilic microorganisms, p. 11-24, Hyperthermophilic Enzymes, Pt A, vol. 330.
120. **Huber, R., T. Wilharm, D. Huber, A. Trincone, S. Burggraf, H. Konig, R. Rachel, I. Rockinger, H. Fricke, and K. O. Stetter.** 1992. *Aquifex pyrophilus* gen-nov sp-nov represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. Syst. Appl. Microbiol. **15:**340-351.
121. **Hulett, M. D., J. R. Hornby, S. J. Ohms, J. Zuegg, C. Freeman, J. E. Gready, and C. R. Parish.** 2000. Identification of active-site residues of the pro-metastatic endoglycosidase heparanase. Biochemistry-US **39:**15659-15667.
122. **Hyun, H. H., and J. G. Zeikus.** 1985. General biochemical characterization of thermostable extracellular beta-amylase from *Clostridium thermosulfurogenes*. Applied And Environmental Microbiology **49:**1162-1167.
123. **Imamura, H., S. Fushinobu, B. S. Jeon, T. Wakagi, and H. Matsuzawa.** 2001. Identification of the catalytic residue of *Thermococcus litoralis* 4-alpha-glucanotransferase through mechanism-based labeling. Biochemistry-US **40:**12400-12406.
124. **Imamura, H., S. Fushinobu, M. Yamamoto, T. Kumasaka, B. S. Jeon, T. Wakagi, and H. Matsuzawa.** 2003. Crystal structures of 4-alpha-glucanotransferase from *Thermococcus litoralis* and its complex with an inhibitor. J. Biol. Chem. **278:**19378-19386.
125. **Islam, M. R., S. Tomatsu, G. N. Shah, J. H. Grubb, S. Jain, and W. S. Sly.** 1999. Active site residues of human beta-glucuronidase - Evidence for Glu(540) as the nucleophile and Glu(451) as the acid-base residue. J. Biol. Chem. **274:**23451-23455.
126. **Jensen, S. O., and P. R. Reeves.** 1998. Domain organisation in phosphomannose isomerases (types I and II). BBA-Protein Struct. M. **1382:**5-7.
127. **Jiang, H. W., W. M. Dian, F. Y. Liu, and P. Wu.** 2003. Isolation and characterization of two fructokinase cDNA clones from rice. Phytochemistry **62:**47-52.
128. **Jones, B. E., W. D. Grant, A. W. Duckworth, and G. G. Owenson.** 1998. Microbial diversity of soda lakes. Extremophiles **2:**191-200.

129. **Jones, W. J., J. A. Leigh, F. Mayer, C. R. Woese, and R. S. Wolfe.** 1983. *Methanococcus jannaschii* sp-nov, an extremely thermophilic methanogen from a submarine hydrothermal vent. Arch. Microbiol. **136:**254-261.
130. **Jorgensen, S., C. E. Vorgias, and G. Antranikian.** 1997. Cloning, sequencing, characterization, and expression of an extracellular alpha-amylase from the hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis*. J. Biol. Chem. **272:**16335-16342.
131. **Joshi, M. D., G. Sidhu, I. Pot, G. D. Brayer, S. G. Withers, and L. P. McIntosh.** 2000. Hydrogen bonding and catalysis: A novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J. Mol. Biol. **299:**255-279.
132. **Juncosa, M., J. Pons, T. Dot, E. Querol, and A. Planas.** 1994. Identification of active-site carboxylic residues in *Bacillus licheniformis* 1,3-1,4-beta-D-glucan 4-glucanohydrolase by site-directed mutagenesis. J. Biol. Chem. **269:**14530-14535.
133. **Kang, S., C. Vieille, and J. G. Zeikus.** 2005. Identification of *Pyrococcus furiosus* amylopullulanase catalytic residues. Appl. Microbiol. Biot. **66:**408-413.
134. **Kaper, T., H. H. van Heusden, B. van Loo, A. Vasella, J. van der Oost, and W. M. de Vos.** 2002. Substrate specificity engineering of beta-mannosidase and beta-glucosidase from *Pyrococcus* by exchange of unique active site residues. Biochemistry-US **41:**4147-4155.
135. **Kaper, T., C. H. Verhees, J. H. G. Lebbink, J. F. T. van Lieshout, L. D. Kluskens, D. E. Ward, S. W. M. Kengen, M. M. Beerthuyzen, W. M. de Vos, and J. van der Oost.** 2001. Characterization of beta-glycosylhydrolases from *Pyrococcus furiosus*, p. 329-346, Hyperthermophilic Enzymes, Pt A, vol. 330.
136. **Karras, G. I., G. Kustatscher, H. R. Buhecha, M. D. Allen, C. Pugieux, F. Sait, M. Bycroft, and A. G. Ladurner.** 2005. The macro domain is an ADP-ribose binding module. EMBO J. **24:**1911-1920.
137. **Kawarabayasi, Y., Y. Hino, H. Horikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshima, and H. Kikuchi.** 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. DNA Res. **8:**123-140.
138. **Kengen, S. W. M., F. A. M. Debok, N. D. Vanloo, C. Dijkema, A. J. M. Stams, and W. M. Devos.** 1994. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. J. Biol. Chem. **269:**17537-17541.
139. **Kengen, S. W. M., A. J. M. Stams, and W. M. deVos.** 1996. Sugar metabolism of hyperthermophiles. FEMS Microbiol. Rev. **18:**119-137.
140. **Keresztessy, Z., K. Brown, M. A. Dunn, and M. A. Hughes.** 2001. Identification of essential active-site residues in the cyanogenic beta-glucosidase (linamarase) from cassava (*Manihot esculenta* Crantz) by site-directed mutagenesis. Biochem. J. **353:**199-205.
141. **Keresztessy, Z., L. Kiss, and M. A. Hughes.** 1994. Investigation of the active-site of the cyanogenic beta-D-glucosidase (linamarase) from *Manihot Esculenta*

Crantz (Cassava).2. Identification of Glu-198 as an active-site carboxylate group with acid catalytic function. Arch. Biochem. Biophys. **315:**323-330.

142. **Khademi, S., L. A. Guarino, H. Watanabe, G. Tokuda, and E. F. Meyer.** 2002. Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. Acta Crystallog. D **58:**653-659.
143. **King, M. R., B. A. White, H. P. Blaschek, B. M. Chassy, R. I. Mackie, and I. K. O. Cann.** 2002. Purification and characterization of a thermostable alpha-galactosidase from *Thermoanaerobacterium polysaccharolyticum*. J. Agr. Food Chem. **50:**5676-5682.
144. **King, M. R., D. A. Yernool, D. E. Eveleigh, and B. M. Chassy.** 1998. Thermostable alpha-galactosidase from *Thermotoga neapolitana*: cloning, sequencing and expression. FEMS Microbiol. Lett. **163:**37-42.
145. **Klages, K. U., and H. W. Morgan.** 1994. Characterization of an extremely thermophilic sulfur-metabolizing archaebacterium belonging to the *Thermococcales*. Arch. Microbiol. **162:**261-266.
146. **Koga, S., I. Yoshioka, H. Sakuraba, M. Takahashi, S. Sakasegawa, S. Shimizu, and T. Ohshima.** 2000. Biochemical characterization, cloning, and sequencing of ADP-dependent (AMP-forming) glucokinase from two hyperthermophilic archaea, *Pyrococcus furiosus* and *Thermococcus litoralis*. J. Biochem. **128:**1079-1085.
147. **Koivula, A., L. Ruohonen, G. Wohlfahrt, T. Reinikainen, T. T. Teeri, K. Piens, M. Claeyssens, M. Weber, A. Vasella, D. Becker, M. L. Sinnott, J. Y. Zou, G. J. Kleywegt, M. Szardenings, J. Stahlberg, and T. A. Jones.** 2002. The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: The roles of aspartic acids D221 and D175. J. Am. Chem. Soc. **124:**10015-10024.
148. **Koning, S. M., M. G. L. Elferink, W. N. Konings, and A. J. M. Driessen.** 2001. Cellobiose uptake in the hyperthermophilic archaeon *Pyrococcus furiosus* is mediated by an inducible, high-affinity ABC transporter. J. Bacteriol. **183:**4979-4984.
149. **Koshland, D. E.** 1953. Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. **28:**416-436.
150. **Kramer, K. J., and S. Muthukrishnan.** 1997. Insect chitinases: Molecular biology and potential use as biopesticides. Insect Biochem. Molec. **27:**887-900.
151. **Krug, M., S.-J. Lee, K. Diederichs, W. Boos, and W. Welte.** 2006. Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. J. Biol. Chem. **281:**published on-line.
152. **Kumar, A., C. E. Larsen, and J. Preiss.** 1986. Biosynthesis of bacterial glycogen - Primary structure of *Escherichia coli* Adp-Glucose - alpha-1,4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the GlgA gene. J. Biol. Chem. **261:**6256-6259.
153. **Kumar, S., and R. Nussinov.** 2001. How do thermophilic proteins deal with heat? Cell. Mol. Life Sci. **58:**1216-1233.
154. **Kuroki, R., L. H. Weaver, and B. W. Matthews.** 1999. Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. P. Natl. Acad. Sci. USA **96:**8949-8954.

155. **Laderman, K. A., K. Asada, T. Uemori, H. Mukai, Y. Taguchi, I. Kato, and C. B. Anfinsen.** 1993. Alpha-amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus* - Cloning and sequencing of the gene and expression in *Escherichia coli*. J. Biol. Chem. **268:**24402-24407.

156. **Laderman, K. A., B. R. Davis, H. C. Krutzsch, M. S. Lewis, Y. V. Griko, P. L. Privalov, and C. B. Anfinsen.** 1993. The purification and characterization of an extremely thermostable alpha-amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. J. Biol. Chem. **268:**24394-24401.

157. **Lamosa, P., A. Burke, R. Peist, R. Huber, M. Y. Liu, G. Silva, C. Rodrigues-Pousada, J. LeGall, C. Maycock, and H. Santos.** 2000. Thermostabilization of proteins by diglycerol phosphate, a new compatible solute from the hyperthermophile Archaeoglobus fulgidus. Appl. Environ. Microb. **66:**1974-1979.

158. **Lamosa, P., D. L. Turner, R. Ventura, C. Maycock, and H. Santos.** 2003. Protein stabilization by compatible solutes - Effect of diglycerol phosphate on the dynamics of *Desulfovibrio gigas* rubredoxin studied by NMR. Eur. J. Biochem. **270:**4606-4614.

159. **Lawson, S. L., W. W. Wakarchuk, and S. G. Withers.** 1997. Positioning the acid/base catalyst in a glycosidase: Studies with *Bacillus circulans* xylanase. Biochemistry-US **36:**2257-2265.

160. **Lee, H.-S., K. R. Shockley, G. J. Schut, S. B. Conners, C. I. Montero, M. R. Johnson, C.-J. Chou, S. L. Bridger, N. Wigner, S. D. Brehm, F. E. Jenney Jr., D. A. Comfort, R. M. Kelly, and M. W. W. Adams.** 2006. Transcriptional and biochemical analysis of starch metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **188:**2115-2125.

161. **Lee, S. J., A. Engelmann, R. Horlacher, Q. H. Qu, G. Vierke, C. Hebbeln, M. Thomm, and W. Boos.** 2003. TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. J. Biol. Chem. **278:**983-990.

162. **Lee, S. J., C. Moulakakis, S. M. Koning, W. Hausner, M. Thomm, and W. Boos.** 2005. TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers. Mol. Microbiol. **57:**1797-1807.

163. **Lee, S. S., S. M. He, and S. G. Withers.** 2001. Identification of the catalytic nucleophile of the Family 31 alpha-glucosidase from *Aspergillus niger* via trapping of a 5-fluoroglycosyl-enzyme intermediate. Biochem. J. **359:**381-386.

164. **Leske, K. L., and C. N. Coon.** 1999. Hydrogen gas production of broiler chicks in response to soybean meal and alpha-galactoside free, ethanol-extracted soybean meal. Poultry Sci. **78:**1313-1316.

165. **Leske, K. L., C. J. Jevne, and C. N. Coon.** 1993. Effect of oligosaccharide additions on nitrogen-corrected true metabolizable energy of soy protein-concentrate. Poultry Sci. **72:**664-668.

166. **Lesley, S. A., P. Kuhn, A. Godzik, A. M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H. E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L. S. Brinen, M. D. Miller, T. M. McPhillips, M. A. Miller, D. Scheibe, J. M. Canaves, C. Guda, L. Jaroszewski, T. L. Selby, M. A. Elsliger, J. Wooley, S.**

S. Taylor, K. O. Hodgson, I. A. Wilson, P. G. Schultz, and R. C. Stevens. 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. P. Natl. Acad. Sci. USA **99:**11664-11669.

167. **Letunic, I., R. R. Copley, S. Schmidt, F. D. Ciccarelli, T. Doerks, J. Schultz, C. P. Ponting, and P. Bork.** 2004. SMART 4.0: towards genomic data integration. Nucleic Acids Res. **32:**D142-D144.
168. **Leulliot, N., L. Tresaugues, M. Bremang, I. Sorel, N. Ulryck, M. Graille, I. Aboulfath, A. Poupon, D. Liger, S. Quevillon-Cheruel, J. Janin, and H. van Tilbeurgh.** 2005. High-throughput crystal-optimization strategies in the South Paris Yeast Structural Genomics project: one size fits all? Acta Crystallog. D **61:**664-670.
169. **Li, W. F., X. X. Zhou, and P. Lu.** 2005. Structural features of thermozymes. Biotechnol. Adv. **23:**271-281.
170. **Li, Y. K., J. Chir, S. Tanaka, and F. Y. Chen.** 2002. Identification of the general acid/base catalyst of a family 3 beta-glucosidase from *Flavobacterium meningosepticum*. Biochemistry-US **41:**2751-2759.
171. **Liebl, W., B. Wagner, and J. Schellhase.** 1998. Properties of an alpha-galactosidase, and structure of its gene galA, within an alpha- and beta-galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. Syst. Appl. Microbiol. **21:**1-11.
172. **Little, S., P. Cartwright, C. Campbell, A. Prenneta, J. McChesney, A. Mountain, and M. Robinson.** 1989. Nucleotide-sequence of a thermostable beta-galactosidase from *Sulfolobus solfataricus*. Nucleic Acids Res. **17:**7980-7980.
173. **Lizotte, P. A., C. A. Henson, and S. H. Duke.** 1990. Purification and characterization of pea epicotyl beta-amylase. Plant Physiol. **92:**615-621.
174. **Lovering, A. L., S. S. Lee, Y. W. Kim, S. G. Withers, and N. C. J. Strynadka.** 2005. Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. J. Biol. Chem. **280:**2105-2115.
175. **Lu, M., and N. Kleckner.** 1994. Molecular cloning and characterization of the Pgm gene encoding phosphoglucomutase of *Escherichia coli*. J. Bacteriol. **176:**5847-5851.
176. **Ly, H. D., S. Howard, K. Shum, S. M. He, A. Zhu, and S. G. Withers.** 2000. The synthesis, testing and use of 5-fluoro-alpha-D-galactosyl fluoride to trap an intermediate on green coffee bean alpha-galactosidase and identify the catalytic nucleophile. Carbohyd. Res. **329:**539-547.
177. **Ma, Y. F., J. A. Mills, J. T. Belisle, V. Vissa, M. Howell, K. Bowlin, M. S. Scherman, and M. McNeil.** 1997. Determination of the pathway for rhamnose biosynthesis in mycobacteria: Cloning, sequencing and expression of the *Mycobacterium tuberculosis* gene encoding alpha-D-glucose-1-phosphate thymidylyltransferase. Microbiol-UK **143:**937-945.
178. **Mackenzie, L. F., G. S. Brooke, J. F. Cutfield, P. A. Sullivan, and S. G. Withers.** 1997. Identification of Glu-330 as the catalytic nucleophile of *Candida albicans* exo-beta-(1,3)-glucanase. J. Biol. Chem. **272:**3161-3167.

179. **Mackenzie, L. F., G. J. Davies, M. Schulein, and S. G. Withers.** 1997. Identification of the catalytic nucleophile of endoglucanase I from *Fusarium oxysporum* by mass spectrometry. Biochemistry-US **36:**5893-5901.
180. **MacLeod, A. M., T. Lindhorst, S. G. Withers, and R. A. J. Warren.** 1994. The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid-127 - Evidence from detailed kinetic-studies of mutants. Biochemistry-US **33:**6371-6376.
181. **MacLeod, A. M., D. Tull, K. Rupitz, R. A. J. Warren, and S. G. Withers.** 1996. Mechanistic consequences of mutation of active site carboxylates in a retaining beta-1,4-glycanase from *Cellulomonas fini*. Biochemistry-US **35:**13165-13172.
182. **Martins, L. O., L. S. Carreto, M. S. DaCosta, and H. Santos.** 1996. New compatible solutes related to di-myo-inositol-phosphate in members of the order Thermotogales. J. Bacteriol. **178:**5644-5651.
183. **Martins, L. O., R. Huber, H. Huber, K. O. Stetter, M. S. DaCosta, and H. Santos.** 1997. Organic solutes in hyperthermophilic Archaea. Appl. Environ. Microb. **63:**896-902.
184. **Matsui, I., K. Ishikawa, H. Ishida, and Y. Kosugi.** August 6, 1999 1999. Methods for making and using a thermophilic enzyme as a β-glycosidase. United States Patent.
185. **McCarter, J. D., D. L. Burgoyne, S. C. Miao, S. Q. Zhang, J. W. Callahan, and S. G. Withers.** 1997. Identification of Glu-268 as the catalytic nucleophile of human lysosomal beta-galactosidase precursor by mass spectrometry. J. Biol. Chem. **272:**396-400.
186. **McCarter, J. D., and S. G. Withers.** 1996. 5-fluoro glycosides: A new class of mechanism-based inhibitors of both alpha- and beta-glucosidases. J. Am. Chem. Soc. **118:**241-242.
187. **McCarter, J. D., and S. G. Withers.** 1994. Mechanisms of enzymatic glycoside hydrolysis. Curr. Opin. Struc. Biol. **4:**885-892.
188. **McCarter, J. D., and S. G. Withers.** 1996. Unequivocal identification of Asp-214 as the catalytic nucleophile of *Saccharomyces cerevisiae* alpha-glucosidase using 5-fluoro glycosyl fluorides. J. Biol. Chem. **271:**6889-6894.
189. **McCarter, J. D., W. Yeung, J. Chow, D. Dolphin, and S. G. Withers.** 1997. Design and synthesis of 2'-deoxy-2'-fluorodisaccharides as mechanism-based glycosidase inhibitors that exploit aglycon specificity. J. Am. Chem. Soc. **119:**5792-5797.
190. **McCutchen, C. M., G. D. Duffaud, P. Leduc, A. R. H. Petersen, A. Tayal, S. A. Khan, and R. M. Kelly.** 1996. Characterization of extremely thermostable enzymatic breakers (alpha-1,6-galactosidase and beta-1,4-mannanase) from the hyperthermophilic bacterium *Thermotoga neapolitana* 5068 for hydrolysis of guar gum. Biotechnol. Bioeng. **52:**332-339.
191. **Mikami, B., M. Adachi, T. Kage, E. Sarikaya, T. Nanmori, R. Shinke, and S. Utsumi.** 1999. Structure of raw starch-digesting *Bacillus cereus* beta-amylase complexed with maltose. Biochemistry-US **38:**7050-7061.
192. **Mikami, B., M. Degano, E. J. Hehre, and J. C. Sacchettini.** 1994. Crystal structures of soybean beta-amylase reacted with beta-maltose and maltal - active-

site components and their apparent roles in catalysts. Biochemistry-US **33:**7779-7787.

193. **Mikami, B., E. J. Hehre, M. Sato, Y. Katsube, M. Hirose, Y. Morita, and J. C. Sacchettini.** 1993. The 2.0-Angstrom resolution structure of soybean beta-amylase complexed with alpha-cyclodextrin. Biochemistry-US **32:**6836-6845.
194. **Mikami, B., H. J. Yoon, and N. Yoshigi.** 1999. The crystal structure of the sevenfold mutant of barley beta-amylase with increased thermostability at 2.5 angstrom resolution. J. Mol. Biol. **285:**1235-1243.
195. **Mirza, O., L. K. Skov, M. Remaud-Simeon, G. P. de Montalk, C. Albenne, P. Monsan, and M. Gajhede.** 2001. Crystal structures of amylosucrase from *Neisseria polysaccharea* in complex with D-glucose and the active site mutant Glu328Gln in complex with the natural substrate sucrose. Biochemistry-US **40:**9032-9039.
196. **Moracci, M., L. Capalbo, M. Ciaramella, and M. Rossi.** 1996. Identification of two glutamic acid residues essential for catalysis in the beta-glycosidase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Protein Eng. **9:**1191-1195.
197. **Moracci, M., A. Trincone, G. Perugino, M. Ciaramella, and M. Rossi.** 1998. Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from *Sulfolobus solfataricus*: Dependence of the mechanism on the action of external nucleophiles. Biochemistry-US **37:**17262-17270.
198. **Mourez, M., N. Hofnung, and E. Dassa.** 1997. Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. EMBO J. **16:**3066-3077.
199. **Mukund, S., and M. W. W. Adams.** 1995. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Biol. Chem. **270:**8389-8392.
200. **Muller, V., R. Spanheimer, and H. Santos.** 2005. Stress response by solute accumulation in archaea. Curr. Opin. Microbiol. **8:**729-736.
201. **Murphy, D., and J. Reid.** September 28, 1999 1999.alpha.-galactosidase. USA patent 5,958,751.
202. **Nakayama, A., K. Yamamoto, and S. Tabata.** 2001. Identification of the catalytic residues of bifunctional glycogen debranching enzyme. J. Biol. Chem. **276:**28824-28828.
203. **Namchuk, M., C. Braun, J. D. McCarter, and S. G. Withers.** 1996. Fluorinated sugars as probes of glycosidase mechanisms, p. 279-293, Biomedical Frontiers of Fluorine Chemistry, vol. 639.
204. **Nanavati, D. A., K. Thirangoon, and K. A. Noll.** 2006. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. Appl. Environ. Microb. **72:**1336-1345.
205. **Nardini, M., and B. W. Dijkstra.** 1999. Alpha/beta hydrolase fold enzymes: the family keeps growing. Curr. Opin. Struc. Biol. **9:**732-737.
206. **Naumoff, D. G.** 2001. beta-Fructosidase superfamily: Homology with some alpha-L-arabinases and beta-D-xylosidases. Proteins **42:**66-76.

207. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, L. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. Nature **399:**323-329.

208. **Nguyen, T. N., A. D. Ejaz, M. A. Brancieri, A. M. Mikula, K. E. Nelson, S. R. Gill, and K. M. Noll.** 2004. Whole-genome expression profiling of *Thermotoga maritima* in response to growth on sugars in a chemostat. J. Bacteriol. **186:**4824-4828.

209. **Nieman, C. E., A. W. Wong, S. M. He, L. Clarke, J. J. Hopwood, and S. G. Withers.** 2003. Family 39 alpha-L-iduronidases and beta-D-xylosidases react through similar glycosyl-enzyme intermediates: Identification of the human iduronidase nucleophile. Biochemistry-US **42:**8054-8065.

210. **Numao, S., S. M. He, G. Evjen, S. Howard, O. K. Tollersrud, and S. G. Withers.** 2000. Identification of Asp197 as the catalytic nucleophile in the family 38 alpha-mannosidase from bovine kidney lysosomes. FEBS Lett. **484:**175-178.

211. **Okada, H., K. Mori, K. Tada, M. Nogawa, and Y. Morikawa.** 2000. Identification of active site carboxylic residues in *Trichoderma reesei* endoglucanase Cel12A by site-directed mutagenesis. J. Mol. Catal. B-Enzym. **10:**249-255.

212. **Ollis, D. L., E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K. H. G. Verschueren, and A. Goldman.** 1992. The Alpha/Beta hydrolase fold. Protein Eng. **5:**197-211.

213. **Oriol, R., I. Martinez-Duncker, I. Chantret, R. Mollicone, and P. Codogno.** 2002. Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. Mol. Biol. Evol. **19:**1451-1463.

214. **Ose, T., T. Fujie, M. Yao, N. Watanabe, and I. Tanaka.** 2004. Crystal structure of the ATP-binding cassette of multisugar transporter from *Pyrococcus horikoshii* OT3. Proteins **57:**635-638.

215. **Oshima, T., and K. Imahori.** 1974. Description Of Thermus-Thermophilus (Yoshida And Oshima) Comb-Nov, A Nonsporulating Thermophilic Bacterium From A Japanese Thermal Spa. Int. J. Syst. Bacteriol. **24:**102-112.

216. **Overduin, P., W. Boos, and J. Tommassen.** 1988. Nucleotide sequence of the Ugp genes of *Escherichia coli* K-12 - Homology to the maltose system. Mol. Microbiol. **2:**767-775.

217. **Oyama, T., M. Kusunoki, Y. Kishimoto, Y. Takasaki, and Y. Nitta.** 1999. Crystal structure of beta-amylase from *Bacillus cereus* var. mycoides at 2.2 angstrom resolution. J. Biochem. **125:**1120-1130.

218. **Paal, K., M. Ito, and S. G. Withers.** 2004. *Paenibacillus* sp TS12 glucosylceramidase: kinetic studies of a novel sub-family of family 3 glycosidases and identification of the catalytic residues. Biochem. J. **378:**141-149.

219. **Parker, C. T., E. Pradel, and C. A. Schnaitman.** 1992. Identification and sequences of the lipopolysaccharide core biosynthetic genes Rfaq, Rfap, and Rfag of *Escherichia coli* K-12. J. Bacteriol. **174:**930-934.
220. **Pearce, S. R., M. L. Mimmack, M. P. Gallagher, U. Gileadi, S. C. Hyde, and C. F. Higgins.** 1992. Membrane topology of the integral membrane components, Oppb and Oppc, of the oligopeptide permease of *Salmonella typhimurium*. Mol. Microbiol. **6:**47-57.
221. **Pego, J. V., and S. C. M. Smeekens.** 2000. Plant fructokinases: a sweet family get-together. Trends Plant Sci. **5:**531-536.
222. **Peterson, J. D., L. A. Umayam, T. Dickinson, E. K. Hickey, and O. White.** 2001. The Comprehensive Microbial Resource. Nucleic Acids Res. **29:**123-125.
223. **Petsko, G. A.** 2001. Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat", p. 469-478, Hyperthermophilic Enzymes, Pt C, vol. 334.
224. **Planas, A., M. Juncosa, J. Lloberas, and E. Querol.** 1992. Essential catalytic role of Glu-134 in endo-beta-1,3-1,4-D-glucan 4-glucanohydrolase from *Bacillus licheniformis a*s determined by site-directed mutagenesis. FEBS Lett. **308:**141-145.
225. **Poeppl, A. G., G. J. Murray, and J. A. Medin.** 2005. Enhanced filter paper enzyme assay for high-throughput population screening for Fabry disease. Anal. Biochem. **337:**161-163.
226. **Postma, P. W., J. W. Lengeler, and G. R. Jacobson.** 1993. Phosphoenolpyruvate - carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. **57:**543-594.
227. **Preston, C. M., K. Y. Wu, T. F. Molinski, and E. F. DeLong.** 1996. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen nov, sp, nov. P. Natl. Acad. Sci. USA **93:**6241-6246.
228. **Pysz, M. A., S. B. Conners, C. I. Montero, K. R. Shockley, M. R. Johnson, D. E. Ward, and R. A. Kelly.** 2004. Transcriptional analysis of biofilm formation processes in the anaerobic, hyperthermophilic bacterium *Thermotoga maritima*. Appl. Environ. Microb. **70:**6098-6112.
229. **Qu, Q. H., S. J. Lee, and W. Boos.** 2004. TreT, a novel trehalose glycosyltransferring synthase of the hyperthermophilic archaeon *Thermococcus litoralis*. J. Biol. Chem. **279:**47890-47897.
230. **Rainey, F. A., A. M. Donnison, P. H. Janssen, D. Saul, A. Rodrigo, P. L. Bergquist, R. M. Daniel, E. Stackebrandt, and H. W. Morgan.** 1994. Description of *Caldicellulosiruptor saccharolyticus* Gen-Nov, Sp-Nov - An obligately anaerobic, extremely thermophilic, cellulolytic bacterium. FEMS Microbiol. Lett. **120:**263-266.
231. **Rigden, D. J.** 2002. Iterative database searches demonstrate that glycoside hydrolase families 27, 31, 36 and 66 share a common evolutionary origin with family 13. FEBS Lett. **523:**17-22.
232. **Robb, F. T., D. L. Maeder, J. R. Brown, J. DiRuggiero, M. D. Stump, R. K. Yeh, R. B. Weiss, and D. M. Dunn.** 2001. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: Implications for physiology and enzymology, p. 134-157, Hyperthermophilic Enzymes, Pt A, vol. 330.

233. **Robertson, C. E., J. K. Harris, J. R. Spear, and N. R. Pace.** 2005. Phylogenetic diversity and ecology of environmental Archaea. Curr. Opin. Microbiol. **8:**638-642.
234. **Rojas, A. L., R. A. P. Nagem, K. N. Neustroev, M. Arand, M. Adamska, E. V. Eneyskaya, A. A. Kulminskaya, R. C. Garratt, A. M. Golubev, and I. Polikarpov.** 2004. Crystal structures of beta-galactosidase from *Penicillium* sp. and its complex with galactose. J. Mol. Biol. **343:**1281-1292.
235. **Rydberg, E. H., G. Sidhu, H. C. Vo, J. Hewitt, H. C. F. Cote, Y. L. Wang, S. Numao, R. T. A. MacGillivray, C. M. Overall, G. D. Brayer, and S. G. Withers.** 1999. Cloning, mutagenesis, and structural analysis of human pancreatic alpha-amylase expressed in *Pichia pastoris*. Protein Sci. **8:**635-643.
236. **Rye, C. S., and S. G. Withers.** 2000. Glycosidase mechanisms. Curr. Opin. Chem. Biol. **4:**573-580.
237. **Saha, B. C., G. J. Shen, and J. G. Zeikus.** 1987. Behavior of a novel thermostable beta-amylase on raw starch. Enzyme Microb. Technol. **9:**598-601.
238. **Salleh, H. M., J. Mullegger, S. P. Reid, W. Y. Chan, J. Y. Hwang, R. A. J. Warren, and S. G. Withers.** 2006. Cloning and characterization of *Thermotoga maritima* beta-glucuronidase. Carbohyd. Res. **341:**49-59.
239. **Sarcabal, P., M. Remaud-Simeon, R. M. Willemot, G. P. de Montalk, B. Svensson, and P. Monsan.** 2000. Identification of key amino acid residues in *Neisseria polysaccharea* amylosucrase. FEBS Lett. **474:**33-37.
240. **Saurin, W., W. Koster, and E. Dassa.** 1994. Bacterial binding protein dependent permeases - Characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins. Mol. Microbiol. **12:**993-1004.
241. **Schlumbaum, A., F. Mauch, U. Vogeli, and T. Boller.** 1986. Plant chitinases are potent inhibitors of fungal growth. Nature **324:**365-367.
242. **Schmees, G., A. Stein, S. Hunke, H. Landmesser, and E. Schneider.** 1999. Functional consequences of mutations in the conserved 'signature sequence' of the ATP-binding-cassette protein MalK. Eur. J. Biochem. **266:**420-430.
243. **Schmitt, L., and R. Tampe.** 2002. Structure and mechanism of ABC transporters. Curr. Opin. Struc. Biol. **12:**754-760.
244. **Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting.** 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. P. Natl. Acad. Sci. USA **95:**5857-5864.
245. **Schutzbach, J. S., J. W. Zimmerman, and W. T. Forsee.** 1993. The purification and characterization of recombinant yeast dolichyl-phosphate-mannose synthase - Site-directed mutagenesis of the putative dolichol recognition sequence. J. Biol. Chem. **268:**24190-24196.
246. **Schwartzman, D. W., and C. H. Lineweaver.** 2004. The hyperthermophilic origin of life revisited. Biochem. Soc. T. **32:**168-171.
247. **Seemann, J. E., and G. E. Schulz.** 1997. Structure and mechanism of L-fucose isomerase from *Escherichia coli*. J. Mol. Biol. **273:**256-268.
248. **Sehgal, A. C., and R. M. Kelly.** 2002. Enantiomeric resolution of 2-aryl propionic esters with hyperthermophilic and mesophilic esterases: Contrasting thermodynamic mechanisms. J. Am. Chem. Soc. **124:**8190-8191.

249. **Selig, M., K. B. Xavier, H. Santos, and P. Schonheit.** 1997. Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. Arch. Microbiol. **167:**217-232.
250. **Shabalin, K. A., A. A. Kulminskaya, A. N. Savel'ev, S. M. Shishlyannikov, and K. N. Neustroev.** 2002. Enzymatic properties of alpha-galactosidase from *Trichoderma reesei* in the hydrolysis of galactooligosaccharides. Enzyme Microb. Technol. **30:**231-239.
251. **Shallom, D., V. Belakhov, D. Solomon, S. Gilead-Gropper, T. Baasov, G. Shoham, and Y. Shoham.** 2002. The identification of the acid-base catalyst of alpha-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. FEBS Lett. **514:**163-167.
252. **Shallom, D., V. Belakhov, D. Solomon, G. Shoham, T. Baasov, and Y. Shoham.** 2002. Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. J. Biol. Chem. **277:**43667-43673.
253. **Shallom, D., M. Leon, T. Bravman, A. Ben-David, G. Zaide, V. Belakhov, G. Shoham, D. Schomburg, T. Baasov, and Y. Shoham.** 2005. Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from *Geobacillus stearothermophilus* T-6. Biochemistry-US **44:**387-397.
254. **Shen, G. J., B. C. Saha, Y. E. Lee, L. Bhatnagar, and J. G. Zeikus.** 1988. Purification and characterization of a novel thermostable beta-amylase from *Clostridium thermosulphurogenes*. Biochem. J. **254:**835-840.
255. **Shockley, K. R., D. E. Ward, S. R. Chhabra, S. B. Conners, C. I. Montero, and R. M. Kelly.** 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microb. **69:**2365-2371.
256. **Siew, N., Y. Azaria, and D. Fischer.** 2004. The ORFanage: an ORFan database. Nucleic Acids Res. **32:**D281-D283.
257. **Siew, N., and D. Fischer.** 2003. Analysis of singleton ORFans in fully sequenced microbial genomes. Proteins **53:**241-251.
258. **Siew, N., and D. Fischer.** 2003. Twenty thousand ORFan microbial protein families for the biologist? Structure **11:**7-9.
259. **Sinnott, M. L.** 1990. Catalytic mechanisms of enzymatic glycosyl transfer. Chem. Rev. **90:**1171-1202.
260. **Smith, T. F., and M. S. Waterman.** 1981. Identification Of Common Molecular Subsequences. J. Mol. Biol. **147:**195-197.
261. **Spangenberg, P., C. Andre, V. Langlois, M. Dion, and C. Rabiller.** 2002. Alpha-galactosyl fluoride in transfer reactions mediated by the green coffee beans alpha-galactosidase in ice. Carbohyd. Res. **337:**221-228.
262. **Staedtler, P., S. Hoenig, R. Frank, S. G. Withers, and W. Hengstenberg.** 1995. Identification of the active-site nucleophile in 6-phospho-beta-galactosidase from *Staphylococcus aureus* by labeling with synthetic inhibitors. Eur. J. Biochem. **232:**658-663.

263. **Stan, J. J. B., N. Smits, H. Wu, A. P. L. Snijders, P. C. Wright, W. M. de Vos, and J. van der Oost.** 2006. Identification of a novel alpha-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. J. Bacteriol. **188:**2392-2399.
264. **Sterner, R., and W. Liebl.** 2001. Thermophilic adaptation of proteins. Crit. Rev. Biochem. Mol. **36:**39-106.
265. **Stetter, K. O.** 1999. Extremophiles and their adaptation to hot environments. FEBS Lett. **452:**22-25.
266. **Stoll, D., S. M. He, S. G. Withers, and R. A. J. Warren.** 2000. Identification of Glu-519 as the catalytic nucleophile in beta-mannosidase 2A from *Cellulomonas fimi*. Biochem. J. **351:**833-838.
267. **Sulzenbacher, G., C. Bignon, T. Nishimura, C. A. Tarling, S. G. Withers, B. Henrissat, and Y. Bourne.** 2004. Crystal structure of *Thermotoga maritima* alpha-L-fucosidase - Insights into the catalytic mechanism and the molecular basis for fucosidosis. J. Biol. Chem. **279:**13119-13128.
268. **Sulzenbacher, G., L. F. Mackenzie, K. S. Wilson, S. G. Withers, C. Dupont, and G. J. Davies.** 1999. The crystal structure of a 2-fluorocellotriosyl complex of the *Streptomyces lividans* endoglucanase CelB2 at 1.2 angstrom resolution. Biochemistry-US **38:**4826-4833.
269. **Sun, Y., and J. Y. Cheng.** 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresource Technol. **83:**1-11.
270. **Tachibana, Y., A. Kuramura, N. Shirasaka, Y. Suzuki, T. Yamamoto, S. Fujiwara, M. Takagi, and T. Imanaka.** 1999. Purification and characterization of an extremely thermostable cyclomaltodextrin glucanotransferase from a newly isolated hyperthermophilic archaeon, a *Thermococcus* sp. Appl. Environ. Microb. **65:**1991-1997.
271. **Takeda, K., C. Akimoto, and M. Kawamukai.** 2001. Effects of the *Escherichia coli sfsA* gene on mal genes expression and a DNA binding activity of SfsA. Biosci. Biotech. Bioch. **65:**213-217.
272. **Tanaka, T., T. Fukui, H. Atomi, and T. Imanaka.** 2003. Characterization of an exo-beta-D-glucosaminidase involved in a novel chitinolytic pathway from the hyperthermophilic archaeon *Thermococcus kodakamensis* KOD1. J. Bacteriol. **185:**5175-5181.
273. **Tanaka, T., T. Fukui, S. Fujiwara, H. Atomi, and T. Imanaka.** 2004. Concerted action of diacetylchitobiose deacetylase and exo-beta-D-glucosaminidase in a novel chitinolytic pathway in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J. Biol. Chem. **279:**30021-30027.
274. **Tarling, C. A., S. M. He, G. Sulzenbacher, C. Bignon, Y. Bourne, B. Henrissat, and S. G. Withers.** 2003. Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from *Thermotoga maritima* through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis. J. Biol. Chem. **278:**47394-47399.
275. **Tatara, Y., B. R. Lee, T. Yoshida, K. Takahashi, and E. Ichishima.** 2003. Identification of catalytic residues of Ca2+-independent 1,2-alpha-D-mannosidase from *Aspergillus saitoi* by site-directed mutagenesis. J. Biol. Chem. **278:**25289-25294.

276. **Teplyakov, A., G. Obmolova, M. A. Badet-Denisot, B. Badet, and I. Polikarpov.** 1998. Involvement of the C terminus in intramolecular nitrogen channeling in glucosamine 6-phosphate synthase: evidence from a 1.6 angstrom crystal structure of the isomerase domain. Struct. Fold. Des. **6:**1047-1055.
277. **Teste, M. A., B. Enjalbert, J. L. Parrou, and J. M. Francois.** 2000. The *Saccharomyces cerevisiae* YPR184w gene encodes the glycogen debranching enzyme. FEMS Microbiol. Lett. **193:**105-110.
278. **Thoden, J. B., A. D. Hegeman, G. Wesenberg, M. C. Chapeau, P. A. Frey, and H. M. Holden.** 1997. Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry-US **36:**6294-6304.
279. **Thomas, P. J., and J. F. Hunt.** 2001. A snapshot of Nature's favorite pump. Nat. Struct. Biol. **8:**920-923.
280. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673-4680.
281. **Todd, A. E., R. L. Marsden, J. M. Thornton, and C. A. Orengo.** 2005. Progress of structural genomics initiatives: An analysis of solved target structures. J. Mol. Biol. **348:**1235-1260.
282. **Tuininga, J. E., C. H. Verhees, J. van der Oost, S. W. M. Kengen, A. J. M. Stams, and W. M. de Vos.** 1999. Molecular and biochemical characterization of the ADP-dependent phosphofructokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Biol. Chem. **274:**21023-21028.
283. **Tull, D., D. L. Burgoyne, D. T. Chow, S. G. Withers, and R. Aebersold.** 1996. A mass spectrometry-based approach for probing enzyme active sites: Identification of Glu 127 in *Cellulomonas fimi* exoglycanase as the residue modified by N-bromoacetyl cellobiosylamine. Anal. Biochem. **234:**119-125.
284. **van den Burg, B.** 2003. Extremophiles as a source for novel enzymes. Curr. Opin. Microbiol. **6:**213-218.
285. **Van Lieshout, J. F. T., C. H. Verhees, T. J. G. Ettema, S. Van der Sar, H. Imamura, H. Matsuzawa, J. Van der Oost, and W. M. De Vos.** 2003. Identification and molecular characterization of a novel type of alpha-galactosidase from *Pyrococcus furiosus*. Biocat. Biotransfor. **21:**243-252.
286. **van Santen, Y., J. A. E. Benen, K. H. Schroter, K. H. Kalk, S. Armand, J. Visser, and B. W. Dijkstra.** 1999. 1.68-angstrom crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis. J. Biol. Chem. **274:**30474-30480.
287. **Ventosa, A., J. J. Nieto, and A. Oren.** 1998. Biology of moderately halophilic aerobic bacteria. Microbiol. Mol. Biol. R. **62:**504-+.
288. **Verhees, C. H., D. G. M. Koot, T. J. G. Ettema, C. Dijkema, W. M. De Vos, and J. Van Der Oost.** 2002. Biochemical adaptations of two sugar kinases from the hyperthermophilic archaeon *Pyrococcus furiosus*. Biochem. J. **366:**121-127.
289. **Viikari, L., A. Kantelinen, J. Sundquist, and M. Linko.** 1994. Xylanases in bleaching - From an idea to the industry. FEMS Microbiol. Rev. **13:**335-350.
290. **Viladot, J. L., E. de Ramon, O. Durany, and A. Planas.** 1998. Probing the mechanism of *Bacillus* 1,3-1,4-beta-D-glucan 4-glucanohydrolases by chemical

rescue of inactive mutants at catalytically essential residues. Biochemistry-US **37:**11332-11342.

291. **Vocadlo, D. J., C. Mayer, S. M. He, and S. G. Withers.** 2000. Mechanism of action and identification of Asp242 as the catalytic nucleophile of *Vibrio furnisii* N-acetyl-beta-D-glucosaminidase using 2-acetamido-2-deoxy-5-fluoro-alpha-L-idopyranosyl fluoride. Biochemistry-US **39:**117-126.
292. **Volkl, P., R. Huber, E. Drobner, R. Rachel, S. Burggraf, A. Trincone, and K. O. Stetter.** 1993. *Pyrobaculum aerophilum* sp-nov, a novel nitrate-reducing hyperthermophilic archaeum. Appl. Environ. Microb. **59:**2918-2926.
293. **Voorhorst, W. G. B., I. L. Rik, E. J. Luesink, and W. M. Devos.** 1995. Characterization of the celb gene coding for beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. J. Bacteriol. **177:**7105-7111.
294. **Walter, K. A., G. N. Bennett, and E. T. Papoutsakis.** 1992. Molecular characterization of 2 *Clostridium acetobutylicum* ATCC 824 butanol dehydrogenase isozyme genes. J. Bacteriol. **174:**7149-7158.
295. **Wang, Q. P., D. Tull, A. Meinke, N. R. Gilkes, R. A. J. Warren, R. Aebersold, and S. G. Withers.** 1993. Glu280 is the nucleophile in the active-site of *Clostridium thermocellum* CelC, a family-A endo-beta-1,4-glucanase. J. Biol. Chem. **268:**14096-14102.
296. **Warner, J. B., and J. S. Lolkema.** 2003. CcpA-dependent carbon catabolite repression in bacteria. Microbiol. Mol. Biol. R. **67:**475-+.
297. **Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Y. Lin, E. Mathur, J. W. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Soll, K. O. Stetter, J. M. Short, and M. Noordewier.** 2003. The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. P. Natl. Acad. Sci. USA **100:**12984-12988.
298. **Watson, J. N., V. Dookhun, T. J. Borgford, and A. J. Bennet.** 2003. Mutagenesis of the conserved active-site tyrosine changes a retaining sialidase into an inverting sialidase. Biochemistry-US **42:**12682-12690.
299. **Wheals, A. E., L. C. Basso, D. M. G. Alves, and H. V. Amorim.** 1999. Fuel ethanol after 25 years. Trends Biotechnol. **17:**482-487.
300. **White, A., S. G. Withers, N. R. Gilkes, and D. R. Rose.** 1994. Crystal structure of the catalytic domain of the Beta-1,4-glycanase Cex from *Cellulomonas fimi*. Biochemistry-US **33:**12546-12552.
301. **Wilcox, W. R., M. Banikazemi, N. Guffon, S. Waldek, P. Lee, G. E. Linthorst, R. J. Desnick, and D. P. Germain.** 2004. Long-term safety and efficacy of enzyme replacement therapy for Fabry disease. Am. J. Hum. Genet. **75:**65-74.
302. **Worning, P., L. J. Jensen, K. E. Nelson, S. Brunak, and D. W. Ussery.** 2000. Structural analysis of DNA sequence: evidence for lateral gene transfer in *Thermotoga maritima*. Nucleic Acids Res. **28:**706-709.
303. **Yamamoto, T., K. Shiraki, S. Fujiwara, M. Takagi, K. Fukui, and T. Imanaka.** 1999. In vitro heat effect on functional and conformational changes of

cyclodextrin glucanotransferase from hyperthermophilic archaea. Biochem. Bioph. Res. Co. **265:**57-61.

304. **Yanase, M., H. Takata, T. Takaha, T. Kuriki, S. M. Smith, and S. Okada.** 2002. Cyclization reaction catalyzed by glycogen debranching enzyme (EC 2.4.1.25/EC 3.2.1.33) and its potential for cycloamylose production. Appl. Environ. Microb. **68:**4233-4239.
305. **Yang, S. J., H. S. Lee, C. S. Park, Y. R. Kim, T. W. Moon, and K. H. Park.** 2004. Enzymatic analysis of an amylolytic enzyme from the hyperthermophilic archaeon *Pyrococcus furiosus* reveals its novel catalytic properties as both an alpha-amylase and a cyclodextrin-hydrolyzing enzyme. Appl. Environ. Microb. **70:**5988-5995.
306. **Yao, Z. J., and M. A. Valvano.** 1994. Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (Rfb) of *Escherichia coli*-K-12-W3110 - Identification of genes that confer group-6 specificity to *Shigella flexneri* serotype-Y and serotype-4A. J. Bacteriol. **176:**4133-4143.
307. **Yayanos, A. A.** 1995. Microbiology to 10,500 meters in the deep-sea. Annu. Rev. Microbiol. **49:**777-805.
308. **Yu, F., Y. Jen, E. Takeuchi, M. Inouye, H. Nakayama, M. Tagaya, and T. Fukui.** 1988. Alpha-glucan phosphorylase from *Escherichia coli* - Cloning of the gene, and purification and characterization of the protein. J. Biol. Chem. **263:**13706-13711.
309. **Zechel, D. L., S. P. Reid, D. Stoll, O. Nashiru, R. A. J. Warren, and S. G. Withers.** 2003. Mechanism, mutagenesis, and chemical rescue of beta-mannosidase from *Cellulomonas fimi*. Biochemistry-US **42:**7195-7204.
310. **Zechel, D. L., and S. G. Withers.** 2001. Dissection of nucleophilic and acid-base catalysis in glycosidases. Current Opinion in Chemical Biology **5:**643-649.
311. **Zhu, A., and J. Goldstein.** 1994. Cloning and functional expression of a cDNA-encoding coffee bean alpha-galactosidase. Gene **140:**227-231.
312. **Zona, R., F. Chang-Pi-Hin, M. J. O'Donohue, and S. Janecek.** 2004. Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. Eur. J. Biochem. **271:**2863-2872.
313. **Zuccotti, S., D. Zanardi, C. Rosano, L. Sturla, M. Tonetti, and M. Bolognesi.** 2001. Kinetic and crystallographic analyses support a sequential-ordered BiBi catalytic mechanism for *Escherichia coli* glucose-1-phosphate thymidylyltransferase. J. Mol. Biol. **313:**831-843.
314. **Zufferey, R., R. Knauer, P. Burda, I. Stagljar, S. T. Heesen, L. Lehle, and M. Aebi.** 1995. STT3, A highly conserved protein required for yeast oligosaccharyl transferase-activity *In vivo*. EMBO J. **14:**4949-4960.

TABLE 1.1 – CATALYTIC RESIDUES OF GLYCOSYL HYDROLASES REPRESENTED BY FAMILY ASSIGNMENT

GHF	Enzyme	Species	Inverting/ Retaining	Method	Nucleophile/Acid-Base	Reference
1	β-glucosidase	*Sulfolobus solfataricus*	R	Inactivator/MS Kinetic analysis	Asp387 (N) Glu206 (AB)	(74, 196)
1	β-glucosidase	*Hordeum vulgare* L.	R	Kinetic analysis	Glu391 (N) Glu181 (AB)	(115)
1	β-mannosidase	*Pyrococcus horikoshii*	R	Kinetic analysis	Glu399 (N) Glu207 (AB)	(134)
1	β-D-glucosidase (laminarase)	*Manihot Esculenta* Crantz	R	Kinetic analysis Inactivator/MS	Glu413 (N) Glu198 (AB)	(140, 141)
1	6-phospho-β-galactosidase	*Staphylococcus aureus*	R	Inactivator/MS	Glu375 (N)	(262)
2	β-galactosidase	*Escherichia coli*	R	Inactivator/MS	Glu537 (N)	(93)
2	β-mannosidase	*Cellulomonas fimi*	R	Inactivator/MS Kinetic analysis	Glu519 (N) Glu429 (AB)	(266, 309)
2	β-glucuronidase	*Homo sapiens*	R	Kinetic analysis	Glu540 (N) Glu451 (AB)	(125)
2	β-glucuronidase	*Thermotoga maritima*	R	Inactivator/MS Kinetic analysis	Glu476 (N) Glu383 (AB)	(238)
3	β-glucosidase	*Aspergillus niger*	R	Inactivator/MS	Asp291 (N)	(46)
3	β-glucosidase	*Flavobacterium meningosepticum*	R	Kinetic analysis/inactivator MS Inactivator/MS	Asp247 (N) Glu473 (AB)	(35, 170)
3	β-glucosylceramidase	*Paenibacillus* sp. TS12	R	Inactivator/MS Kinetic analysis	Asp223 (N) Glu411 (AB)	(218)
3	N-acetyl-β-D-glucosaminidase	*Vibrio furnissii*	R	Inactivator/MS	Asp242 (N)	(291)
5	exo-β-(1,3)-glucanase	*Candida albicans*	R	Inactivator/MS	Glu330 (N)	(178)
5	endo-β-(1,4)-glucanase	*Bacillus agaradherans*	R	Crystal structure	Glu228 (N) Glu139 (AB)	(50)
5	cellulase	*Schizophyllum commun*	R	Inactivator/MS		(36)
5	β-(1,3)-(1,4)-endoglucanase	*Bacillus licheniformis*	R	Kinetic analysis, crystal structure	Glu134 (N) Glu138 (AB)	(104, 132, 224, 290)

TABLE 1.1 – CATALYTIC RESIDUES OF GLYCOSYL HYDROLASES REPRESENTED BY FAMILY ASSIGNMENT (continued)

GHF	Enzyme	Species	Inverting/ Retaining	Method	Nucleophile/Acid-Base	Reference
5	endo-β-(1,4)-glucanase C	*Clostridium thermocellum*	R	Inactivator/MS	Glu280 (N)	(295)
6	cellobiohydrolase	*Trichoderma reesei*	I	Crystal structure/Kinetic analysis	Asp221 (A) Asp175 (B)	(147)
7	endoglucanase	*Fusarium oxysporum*	R	Inactivator/MS	Glu197 (N)	(179)
9	endoglucanase	*Nasutitermes takasagoensis*	R	Crystal structure	Asp54/Asp57 (N) Glu412 (AB)	(142)
10	endoglucanase/xylanase	*Cellolomonas fimi*	R	Kinetic analysis Inactivator/MS	Glu233 (N) Glu127 (AB)	(180, 181, 283)
11	xylanase	*Bacillus circulans*	R	Kinetic analysis	Glu78 (N) Asp35/Glu172 (AB)	(131, 159)
12	endoglucanase	*Streptomyces lividans*	R	Crystal structure	Glu120 (N)	(268)
12	endoglucanase	*Trichoderma reesei*	R	Kinetic analysis	Glu116 (N) Gu200 (AB)	(211)
13	α-glucosidase	*Saccharomyces cerevisiae*	R	Inactivator/MS Inactivator/MS	Asp412 (N) Glu276 (AB)	(114, 188)
13	Glycogen debranching enzyme	*Homo sapiens*	R	Inactivator/MS	Asp549 (N)	(24)
13	α-amylase	*Homo sapiens*	R	Crystal structure/kinetic analysis	Asp197 (N)	(235)
13	amylosucrase	*Neisseria polysaccharea*	R	Kinetic analysis Crystal structure	Asp294 (N) Glu328 (AB)	(195, 239)
14	β-amylase	*Bacillus cereus*	I	Crystal structure	Glu172 (A) Glu367 (B)	(191, 217)
14	β-amylase	Soy bean	I	Crystal structure	Glu186 (A) Glu380 (B)	(193)
20	β-hexosaminidase B	*Homo sapiens*	R	Kinetic analysis	Glu355 (AB)	(112)
27	α-galatosidase	*Phanerochaete chrysosporium*	R	Inactivator/MS	Asp130 (N)	(105)
27	α-galactosidase	*Coffea arabica*	R	Inactivator/MS	Asp145 (N)	(176)

TABLE 1.1 – CATALYTIC RESIDUES OF GLYCOSYL HYDROLASES REPRESENTED BY FAMILY ASSIGNMENT (continued)

GHF	Enzyme	Species	Inverting/ Retaining	Method	Nucleophile/Acid-Base	Reference
27	α-N-acetylgalactosaminidase	Chicken	R	Crystal structure	Asp140 (N) Asp201 (AB)	(90)
28	endopolygalacturonase	*Aspergillus niger*	I	Crystal structure	Asp180/Asp202 (A) Asp201 (B)	(286)
29	α-L-fucosidase	*Sulfolobus solfataricus*	R	Chemical rescue of inactive mutant Kinetic analysis	Asp242 (N) Gly58/Glu292 (AB)	(37, 38)
29	α-L-fucosidase	*Thermotoga maritima*	R	Inactivator/MS Crystal structure/Kinetic analysis	Asp224 (N) Glu266 (AB)	(274) (267)
31	α-glucosidase	*Aspergillus niger*	R	Inactivator/MS	Asp224 (N)	(163)
31	α-glucosidase	*Escherichia coli*	R	Crystal structure	Asp416 (N)	(174)
32	β-fructosidase (invertase)	*Thermotoga maritima*	R	Crystal structure	Asp17 (N) Glu190 (AB)	(3)
35	β-galactosidase	*Xanthomonas manihotis*	R	Inactivator/MS	Glu260 (N)	(18)
35	β-galactosidase	*Bacillus circulans*	R	Inactivator/MS	Glu233 (N)	(18)
35	β-galactosidase	*Homo sapiens*	R	Inactivator/MS	Glu268 (N)	(185)
35	β-galactosidase	*Penicillum* sp.	R	Crystal structure	Glu299 (N) Glu200 (AB)	(234)
38	α-mannosidase	Bovine kidney lysosomes	R	Inactivator/MS	Asp197 (N)	(210)
38	α-mannosidase	*Canavalia ensiformis*	R	Inactivator/MS	Asp of IDPFGH (N)	(74)
39	α-L-iduronidase	*Homo sapien*	R	Inactivator/MS	Glu299 (N)	(209)
39	β-xylosidase	*Bacillus stearothermophilus*	R	Kinetic analysis	Glu160 (AB)	(26)
39	β-xylosidase	*Thermoanaerobacterium saccharolyticum*	R	Inactivator/MS	Glu277 (N)	(291)
43	β-D-xylosidase	*Geobacillus stearothermophilus*	I	Kinetic analysis Crystal structure	Glu187 (A) Asp15 (B)	(253)
47	α-(1,2)-mannosidase	*Aspergillus saitoi*	I	Kinetic analysis	Glu124 (A)	(275)

TABLE 1.1 – CATALYTIC RESIDUES OF GLYCOSYL HYDROLASES REPRESENTED BY FAMILY ASSIGNMENT (continued)						
GHF	**Enzyme**	**Species**	**Inverting/ Retaining**	**Method**	**Nucleophile/Acid-Base**	**Reference**
51	α-L-arabinofuranosidase	*Geobacillus stearothermophilus*	R	Kinetic analysis	Glu294 (N) Glu175 (AB)	(251, 252)
51	α-L-arabinofuranosidase	*Thermobacillus xylanilyticus*	R	Kinetic analysis	Glu298 (N) Glu176 (AB)	(55)
52	β-xylosidase	*Geobacillus stearothermophilus*	R	Kinetic analysis	Glu335 (N) Asp495 (AB)	(25)
53	endo- β-(1,4)-galactanase	*Pseudomonas fluorescens*	R	Kinetic analysis	Asp161 (AB) Asp270 (N)	(23)
57	4- α-glucanotransferase	*Thermococcus litoralis*	R	Inactivator/MS crystal structure	Glu123 (N) Asp214 (AB)	(123, 124)
57	amylopullulanase	*Pyrococcus furiosus*	R	Kinetic analysis	Glu291 (N)	(133)
57	α-galactosidase	*Pyrococcus furiosus*	R	Kinetic analysis	Glu117 (N)	(285)
65	maltose phosphorylase	*Lactobacillus brevis*	I	Crystal structure	Glu487 (A)	(70)
67	α-glucuronidase	*Geobacillus stearothermophilus*	I	Crystal structure	Glu285 (A) Asp364/Glu392 (B)	(96)
79	β-D-endoglucuronidase (heparanase)	*Homo sapiens*	R	Kinetic analysis/crystal threading	Glu343 (N) Glu225 (AB)	(121)

TABLE 1.2 – CARBOHYDRATE PROCESSING PROTEINS/ENZYMES IN THE *PYROCOCCUS FURIOSUS* GENOME

ORF	Enzyme/Function	Enzyme Family	Size	TIGR Annotation	Signal Peptide	Reference
PF0044	4-amino-4-deoxy-L-arabinose transferase	GTF2,39	53.3	hypothetical protein	N	(213)
PF0058	glycosyl transferase	GTF-2	40.1	dolichol monophosphate mannose synthase	N	(245)
PF0073*	β-glucosidase (CelB)	GHF-1	54.7	beta-glucosidase	N	(293)
PF0074	β-alcohol dehyrdrogenase	ADH	28.0	alcohol dehydrogenase, short chain	N	(17)
PF0075	β-alcohol dehydrogenase, Fe-containing	Fe-ADH	42.9	alcohol dehydrogenase	N	(67, 98, 294)
PF0076*	β-1,3-glucanase (LamA)	GHF-16	33.8	endo-beta-1,3-glucanase	Y	(68, 102)
PF0116	β-ABC ATP binding protein, MalK-like	MalK	41.4	putative multiple sugar transport protein	N	(49, 214, 242)
PF0117	β-ABC permease, UgpE	MalG-Like	32.0	putative sugar transport protein	Y?	(198, 216, 240)
PF0118	β-ABC permease, UgpA	MalG-Like	36.0	putative ABC transporter permease protein	Y?	(198, 216, 240)
PF0119	put. β ABC solute binding protein	MalE-like	61.3	putative perplasmic sugar binding protein	Y	(21, 45, 69)
PF0124	Related to TrmB	TrmB	39.4	hypothetical protein	N	(77, 161, 162)
PF0132*	α-glucosidase	GHF-new	54.8	hypothetical protein	N	(41), Comfort *et al.*(this work)
PF0133	hyp. protein		21.3	hypothetical protein	N	
PF0156	put. glycosyl transferase	GTF-66/STT	108.6	oligosaccharyl transferase stt3 subunit related protein	Y?	(314)
PF0190	typtone/S^0 periplasmic binding protein	SBP	94.8	hypothetical protein	Y	
PF0191	typtone/S^0 ABC permease; ABC-type dipeptide permease	BPD	39.0	oligopeptide transport system permease protein	Y?	(220, 240)
PF0192	typtone/S^0 ABC permease; ABC-type dipeptide permease	BPD	52.5	hypothetical oligopeptide transport system permease protein appc	N?	(220, 240)
PF0193	typtone/S^0 ABC ATP-binding protein	NBD	35.9	putative ABC transport ATP binding protein	N?	(10, 54, 242, 279)
PF0194	typtone/S^0 ABC ATP-binding protein	NBD	38.3	dipeptide ABC transporter, ATP-binding protein	N	(10, 54, 242, 279)
PF0235	glucose-1-phosphate thymidylyltransferase, rhamnose biosynthesis pathway?		46.6	glucose-1-phosphate thymidylyltransferase	N	(19, 177)
PF0261	S-layer domain containing protein		102.8	hypothetical protein	Y	
PF0272*	α-amylase (AmyA)	GHF-57	77.1	alpha-amylase	N	(155, 156)

TABLE 1.2 – CARBOHYDRATE PROCESSING PROTEINS/ENZYMES IN THE *PYROCOCCUS FURIOSUS* GENOME (continued)						
ORF	**Enzyme/Function**	**Enzyme Family**	**Size**	**TIGR Annotation**	**Signal Peptide**	**Reference**
PF0307	put. glycosyl transferase, RfgA	GTF-4	37.5	glycosyltransferase	N	(219)
PF0312*	ADP-dependent glucokinase	KIN	51.3	ADP-dependent glucokinase	N	(52, 282, 288)
PF0318	put. carbohydrate esterase; dipeptidyl aminopeptidase	CE-10	73.0	similar to acylaminoacyl-peptidase	N	(56)
PF0354	diacetylchitobiose deacetylase		31.0	hypothetical protein	N	(273)
PF0355	put. chitin regulator; PIN domain		17.2	hypothetical protein	N	
PF0356	β-mannosidase/β-glucosidase	GHF-1	56.3	beta-galactosidase	N	(44, 134, 172)
PF0357	chitin binding protein, ChtA		71.2	dipeptide-binding protein	Y	
PF0358	Chitin ABC permease, ChtB		37.5	ABC transporter (oppbc family)	N?	(220, 240)
PF0359	Chitin ABC permease, ChtC		34.0	oligopeptide ABC transport system permease	Y?	(220)
PF0360	Chitin ABC ATP-binding protein, ChtD		36.5	oligopeptide ABC transporter (ATP-binding protein)	N	(54, 279)
PF0361	Chitin ABC ATP-binding protein, ChtF		36.7	oligopeptide ABC transporter (ATP-binding protein)	N	(54, 279)
PF0362	Chitin glucosamine-fructose-6-phosphate aminotransferase/GlcN-P isomerase	SIS	36.7	glucosamine-fructose-6-phosphate aminotransferase	N	(11, 276)
PF0363	β-glucosaminidase	GHF-35	90.2	beta-galactosidase precursor	N	(272)
PF0411	oligosaccharyl transferase, STT3 subunit		84.3	protein-export membrane protein	N	(314)
PF0441	Galactose-1-phosphate uridylyltransferase		38.5	galactose-1-phosphate uridylyltransferase	N	(78)
PF0442*	β-glucosidase (BglA)	GHF-1	49.8	beta-glucosidase	N	(135)
PF0443	fucose or arabinose sugar permease		49.4	putative membrane transport protein	Y	(103)
PF0444*	α-galactosidase (GalA)	GHF-57	41.5	hydrolase	N	(285)
PF0445*	GalK, Galactokinase		39.4	galactokinase	N	(52, 288)
PF0446	glucose-1 phosphate transferase		17.3	NDP-sugar synthase	N	(30)
PF0448	putative mannose-1-phosphate guanyltransferase		25.2	putative mannose-1-phosphate guanyltransferase	Y?	(30)
PF0465	put glycosyl transferase	GTF-2	40.0	putative glycosyl transferase	N	(245)
PF0477*	α-amylase	GHF-13	54.4	alpha-amylase	Y	(65, 130)
PF0478	cyclomaltodextrinase	GHF-13	79.2	alpha-amylase	Y	(106, 270, 303)
PF0496	Related to TrmB		30.6	hypothetical protein	N	(151, 161, 162)

TABLE 1.2 – CARBOHYDRATE PROCESSING PROTEINS/ENZYMES IN THE *PYROCOCCUS FURIOSUS* GENOME (continued)						
ORF	**Enzyme/Function**	**Enzyme Family**	**Size**	**TIGR Annotation**	**Signal Peptide**	**Reference**
PF0508	Dolichyl-phosphate-mannose-protein mannosyltransferase, domain		76.9	integral membrane glycosyltransferase	?	
PF0588	Phosphosugar mutase		49.6	Phosphosugar mutase	N	(175)
PF0589	Mannose-1-phosphate; guanylyltransferase/mannose-6-phosphate isomerase		52.6	mannose-6-phosphate isomerase/mannose-1-phosphate guanylyl transferase	N	(19, 126)
PF0591	mannosyl-3-phosphoglycerate synthase	GTF-55	45.4	hypothetical protein	N	
PF0608	Alcohol dehydrogenase		41.4	alcohol dehydrogenase	N	
PF0661	Related to TrmB	TrmB	30.2	hypothetical protein	N	(151, 161, 162)
PF0765	UDP-N-acetyl-D-mannosaminuronate dehydrogenase		47.2	NSP-sugar dehydrogenase	N	
PF0769	put. glycosyl transferase, UDP-sugar	GTF-2	34.5	glycosyl transferase	N?	
PF0770	UDP-glucose pyrophosphorylase		33.1	glucose-1-phosphate uridylyltransferase	N	(19, 313)
PF0771	UDP-glucose-6-dehydrogenase		43.8	NDP-sguar dehydrogenase	N	
PF0777	polysaccharide biosynthesis related protein		23.4	hypothetical protein	N?	
PF0782	polysaccharide biosynthesis related protein		12.2	hypothetical protein	N	
PF0784	put. glycosyl transferase, RfgA or GlgA	GTF-4	22.3	hypothetical protein	N	(152, 219)
PF0788	put. glycosyl transferase, RfgA or GlgA	GTF-4	24.8	glycosyl transferase	N	(152, 219)
PF0791	put. glycosyl transferase, RfgA or GlgA	GTF-4	45.0	hypothetical protein	N	(152, 219)
PF0794	UDP-N-acetylglucosamine 2-epimerase		40.6	UDP-n-acetylglucosamine 2-epimerase	N	
PF0795	put. glycosyl transferase, RfgA or GlgA	GTF-4	46.0	capsular polysaccharide biosynthesis protein	N	(152, 219)
PF0796	galactoside o-acetyltransferase	GTF	18.4	galactoside o-acetyltransferase	N	
PF0798	put. glycosyl transferase	GTF-2	34.6	glycosyl transferase	N	(245)
PF0799	UDP-glucose/GDP-mannose dehydrogenase Related protein		7.1	unknown	N	
PF0811	transmembrane oligosaccharyl transferase, putative		16.8	hypothetical protein	N?	
PF0837	Possible ABC-2 integral membrane protein		4.2	cobalamin (5-phosphate) synthase, c-terminal fragment	N	
PF0854*	endo-1,4-glucanase A (EglA)	GHF-12	36.0	endo-1,4-beta-glucanase b	Y	(13)

TABLE 1.2 – CARBOHYDRATE PROCESSING PROTEINS/ENZYMES IN THE *PYROCOCCUS FURIOSUS* GENOME (continued)						
ORF	**Enzyme/Function**	**Enzyme Family**	**Size**	**TIGR Annotation**	**Signal Peptide**	**Reference**
PF0861	mannomutase		49.8	phospho-sugar mutase	N	(175)
PF0868	sugar-phosphate nucleotydyl transferase, RfbA or GlgA		47.2	NDP-sugar synthase	N	(19, 30, 177)
PF0919	β-glycosidase fragment		8.4	hypothetical protein	N	
PF0928	β-glucosidase fragment		6.3	hypothetical protein	N	
PF1108	carbohydrate esterase; alpha/beta hydrolase	CE-1	46.9	putative alpha-dextrin endo-1, 6-alpha-glucosidase	Y	(205, 212)
PF1109	hypothetical protein		106.5	hypothetical protein	Y	
PF1110	hypothetical protein		19.2	hypothetical protein	N	
PF1111	hypothetical protein; methyltransferase		40.3	hypothetical protein	N	
PF1198	sugar fermentation stimulation protein, maltose regulatory protein		26.1	sugar fermentation stimulation protein	N	(271)
PF1208*	β-mannosidase (BmnA)	GHF-1	59.1	beta-mannosidase	N	(184)
PF1209	Cellobiose binding protein, βCbtA		71.6	oligopeptide ABC transporter (oligopeptide-binding protein)	Y	
PF1210	Cellobiose ABC permease, βCbtB		37.3	dipeptide ABC transporter, permease protein	N?	
PF1211	Cellobiose ABC permease, βCbtC		33.8	dipeptide ABC transporter, permease protein	?	
PF1212	Cellobiose ABC ATP-binding protein, βCbtD		36.2	putative peptide ABC uptake protein	N	(54, 279)
PF1213	Cellobiose ABC ATP-binding protein, βCbtF		36.1	oligopeptide ABC transporter (ATP-binding protein)	N	(54, 279)
PF1217	possible glycosyl transferase fragment		5.3	hypothetical protein	N	
PF1218	glycosyltransferase, family 1, fragment		7.0	hypothetical protein	N	
PF1219	oligosaccharyl transferase, STT3 subunit		14.2	hypothetical protein	N	
PF1233*	chitinase (ChiB)	GHF-18	78.6	putative chitinase	N	(88)
PF1234*	chitinase (ChiA)	GHF-18	39.7	putative chitinase	Y	(88)
PF1256	Related to T. litoralis a-glucosidase		27.9	hypothetical protein	N	
PF1287	SufC, ABC-type transport system involved in Fe-S clus		27.2	put. ABC transporter (ATP-binding protein)	N	

TABLE 1.2 – CARBOHYDRATE PROCESSING PROTEINS/ENZYMES IN THE *PYROCOCCUS FURIOSUS* GENOME (continued)						
ORF	**Enzyme/Function**	**Enzyme Family**	**Size**	**TIGR Annotation**	**Signal Peptide**	**Reference**
PF1354	polysaccharide biosynthesis protein, RfbX		47.4	putative oligosaccharide transporter	N	(306)
PF1355	UDP-glucose 6-dehydrogenase OR UDP-N-acetyl-D-mannosaminuronate dehydrogenase		48.0	NDP-sugar dehydrogenase	N	
PF1356	UTP--glucose-1-phosphate uridylyltransferase, RfbA or GlgA		32.3	glucose-1-phosphate uridylyltransferase	N	(19, 30, 177)
PF1357	UDP- or dTTP-glucose 4-epimerase or 4-6-dehydratase		38.1	UDP- or dTTP-glucose 4-epimerase or 4-6-dehydratase	N	(278)
PF1358	put. glycosyl transferase	GTF-4	34.7	glycosyl transferase	N	
PF1359	put. glycosyl transferase	GTF-2	34.8	glycosyl transferase	N	(245)
PF1360	put. glycosyl transferase	GTF-4	42.0	hypothetical protein	N	
PF1361	put. glycosyl transferase	GTF-4	44.9	glycosyl transferase	N	
PF1362	put. glycosyl transferase/glycogen synthase	GTF-4	47.4	hypothetical protein	N	(152)
PF1363	put. glycosyl transferase	GTF-4	37.5	glycosyl transferase	N	
PF1364	put. glycosyl transferase	GTF-4	43.3	hypothetical protein	N	
PF1393	α-amylase (put.)	GHF-57	74.5	hypothetical protein	N	(156, 312)
PF1458	Fructokinase	KIN	34.5	sugar kinase	N	(127, 221)
PF1459	L-fucose isomerase		54.3	hypothetical protein	N	(247)
PF1460	glycogen debranching enzyme (put.)	GHF-?	65.1	hypothetical protein	N	(202, 277)
PF1535	α-glucan phosphorylase	GTF-35	97.7	alpha-glucan phosphorylase	N	(308)
PF1536	put. ABC ribose binding	BAL-like	19.9	hypothetical protein	N	(136)
PF1589	Ribokinase		28.3	hypothetical protein	N	
PF1594	put. glycosyl transferase	GTF-2	32.0	dolichol-phosphate mannose synthase	N	(245)
PF1606	put. glycosyl transferase	GTF-2	24.4	dolichol-phosphate mannose synthase	N	(245)
PF1616	Myo-inositol-1-phosphate synthase		42.4	myo-inositol-1-phosphate synthase	N	
PF1630	UDP-n-acetylglucosamine 2-epimerase		42.8	UDP-n-acetylglucosamine 2-epimerase	N	(110)
PF1695	β-ABP PBP lipoprotein	BMP	44.2	hypothetical lipoprotein	N?	
PF1696	putative β-ABC sugar transporter		55.9	putative ABC sugar transporter	N	(54, 279)
PF1697	put. β-rib/gal ABC permease	RbsA	35.3	putative ribose/galactose ABC transporter	N?	(10)

TABLE 1.2 – CARBOHYDRATE PROCESSING PROTEINS/ENZYMES IN THE *PYROCOCCUS FURIOSUS* GENOME (continued)						
ORF	**Enzyme/Function**	**Enzyme Family**	**Size**	**TIGR Annotation**	**Signal Peptide**	**Reference**
PF1698	put. β-rib/gal ABC permease		32.3	putative ribose ABC transporter	N?	(16)
PF1728	mannose-1-phosphate guanyltransferase	NTPT	40.1	NDP-sugar synthase	N	(126)
PF1738	fructokinase	KIN	35.0	sugar kinase	N	(127, 221)
PF1739	trehalose/maltose binding protein, MalE	MalE	50.4	trehalose/maltose binding protein	Y	(21, 45, 69)
PF1740	trehalose/maltose ABC permease, MalF	UgpA	33.2	trehalose/maltose transport inner membrane protein	N?	(216)
PF1741	trehalose/maltose ABC permease, MalG	MalG/UgbE	30.8	trehalose/maltose transport inner membrane protein	Y?	(47, 198)
PF1742	glycosyltransferase, RfaG/trehalose synthase	GTF-4	48.0	putative trehalose synthase	N	(219)
PF1743	Maltose regulator, TrmB	TrmB	38.8	hypothetical protein	N	(161, 162) (151)
PF1744	trehalose/maltose ATP binding	MalK	41.7	trehalose/maltose transport ATP-hydrolyzing	N	(49, 242)
PF1745	put. L-fucose isomerase	SIS	56.4	hypothetical protein	N	(247)
PF1746	glycogen debranching enzyme (put.)	GHF-?	72.7	hypothetical protein	Y?	(304)
PF1747	β-fructosidase (put.)	GHF-?	38.1	hypothetical protein	N	(206)
PF1748	ABC-type Fe3+ transport system permease, ThiP	MalG	36.7	putative sulfate transport system permease protein, ABC	Y	(2)
PF1749	ABC-type Fe3+ transport system permease	MalG	25.4	putative sulfate transport integral membrane protein	N	(2)
PF1750	ABC-type sugar transport systems ATPase, MalK	MalK	38.4	putative sugar-binding transport ATP-binding protein	N	(49, 242)
PF1751	ABC type, sugar transport system, TbpA		40.7	putative solute binding lipoprotein	Y?	
PF1782	putative membrane-bound dolichyl-phosphate-mannose-protein mannosyltransferase	GTF-39?	60.0	hypothetical protein	N	
PF1885	put. glycosyl transferase, RfaG	GTF-4	40.0	glycosyl transferase	N	(219)
PF1886	ribokinase-like subgroup A		32.4	sugar kinase	N	
PF1916	put. glycosyl transferase	GTF-2	43.6	hypothetical protein	N	
PF1933	ABC ATP-binding protein, MalK		41.3	putative sugar transport ATP-hydrolyzing	N	(49, 242)

TABLE 1.2 – CARBOHYDRATE PROCESSING PROTEINS/ENZYMES IN THE *PYROCOCCUS FURIOSUS* GENOME (continued)						
ORF	**Enzyme/Function**	**Enzyme Family**	**Size**	**TIGR Annotation**	**Signal Peptide**	**Reference**
PF1935*	amylopullulanase (Apu)	GHF-57	113.3	amylopullulanase	Y	(66)
PF1936	maltose ABC permease, MalG-like	MalG	45.2	putative sugar transport inner membrane protein (malg-like)	N	(47, 198)
PF1937	maltose ABC permease, MalF-like/UgpA	MalF	33.9	putative sugar transport inner membrane protein (malf-like)	Y	(216)
PF1938	maltose-binding protein, MalE-like	MalE	48.2	putative sugar binding protein (malE-like)	Y	(21, 45, 69)
PF1939*	cyclodextrinase (α-amylase)	GHF-13	76.1	neopullulanase (alpha-amylase II)	N	(305)
PF1967	ABC type sugar transport system, UgpB	MalE-like	47.5	hypothetical protein	Y?	(216, 240)
PF1968	ABC type sugar transport system, UgpA		31.8	sugar transport membrane protein	N?	(216, 240)
PF1969	ABC type sugar transport system, UgpE	MalG-Like	30.4	sugar ABC transporter, permease protein	Y	(198, 216)
PF1970	ABC Type sugar transport, MalK	MalK	41.9	putative multiple sugar-binding transport ATP-binding protein	N	(49, 242)
PF1981	pred. nucleotidyl transferase		27.5	hypothetical protein	N	
PF2004	glycerol or sugar kinase		55.0	glycerol kinase	N	
PF2044	UDP-Glc/ADP-Glc glycogen synthase	GTF-5	50.9	glycogen synthase	Y?	(152)

*glycosyl hydrolase enzyme
Ambiguous presense of a signal peptide is noted with a '?', with the best prediction 'Y/N'

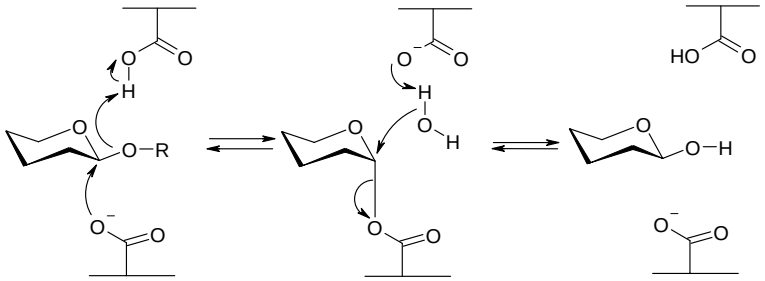


Figure 1.1 – Retaining glycosyl hydrolase mechanism. A double displacement mechanism in which the sugar becomes covalently bound to the enzyme as a glycosyl enzyme intermediate. Nucleophile is on bottom and acid-base residue is above.

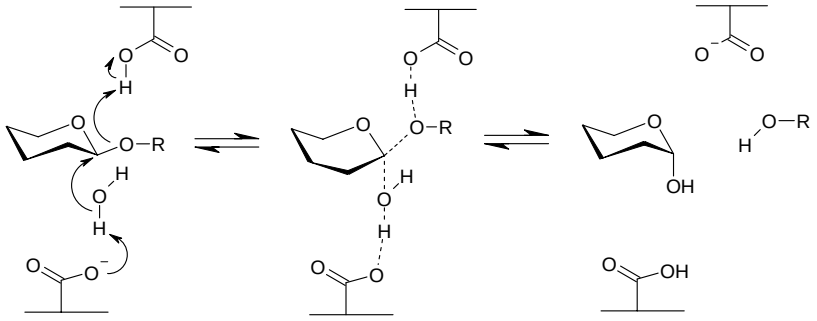


Figure 1.2 – Inverting glycosyl hydrolase mechanism. A single displacement reaction generating a product with opposite anomeric conformation.

Chapter 2: Strategic Biocatalysis with Hyperthermophilic Enzymes

Summary

With the advent of genome sequence information, in addition to capabilities for cloning and expressing genes of interest in foreign hosts, a wide range of hyperthermophilic enzymes have become accessible for potential applications for biocatalytic processes. Not only can these enzymes be useful for strategic opportunities at high temperatures, but there may also be advantages that derive from their relatively low activity at suboptimal temperatures. Examples of several possible ways in which hyperthermophilic enzymes could be used are presented, including cases where they could serve as environmentally benign alternatives in existing industrial processes.

Published as: Comfort, Donald A.; Chhabra, Swapnil R.; Conners, Shannon B.; Chou, Chung-Jung; Epting, Kevin L.; Johnson, Matthew R.; Jones, Kristen L.; Sehgal, Amitabh C.; Kelly, Robert M. **Strategic biocatalysis with hyperthermophilic enzymes.** Green Chemistry (2004), 6(9), 459-465.

Introduction

Enzymes have become an increasingly attractive alternative to chemical catalysts because of their environmentally benign attributes (61). Not only are they highly specific to the molecular features of substrates, enzymes are also highly catalytic in terms of reaction rate acceleration. Unlike chemical catalysts, enzymes rarely require toxic metals ions for functionality and, while in some cases have been found to catalyze reactions in organic solvents (49, 58, 94), they typically work best under aqueous conditions at moderate temperatures. If enzymes were able to operate over a wider range of conditions than has historically been considered for biocatalysis, the opportunities for their utilization could be expanded (64). Even though methodologies exist for expanding the functional ranges of enzymes (e.g., directed evolution) (14, 22, 79), it is often desirable to identify wild-type enzymes for particular applications as starting points for optimization efforts and because of regulatory concerns. In this regard, enzymes from extremophiles, which are intrinsically stable in otherwise biologically-challenging environments, should be considered.

Extremophilic enzymes are derived from extremophiles – organisms that grow optimally under otherwise biologically harsh conditions (e.g. high temperature, low temperature, high acidity or alkalinity, high salinity, etc.) (10, 80, 89). Extremophilic enzymes are frequently resistant to the environmental situation in which the host organism thrives, offering novel sources of robust biocatalysts that can withstand the demands of industrial processes (20, 69). Thus, extremophilic enzymes offer the advantage that they could be integrated into existing processes with little need to change normal operating parameters. Hyperthermophiles, extremophiles growing at temperatures of 80°C and above, produce biocatalysts that are intrinsically thermostable and thermoactive (93). In some instances they also offer another unique advantage: at suboptimal temperatures, hyperthermophilic enzymes either have reduced activity or essentially no activity.

Discussed here are opportunities for biocatalysis that make strategic use of hyperthermophilic enzymes which function at high temperatures, in addition to examples that capitalize on their reduced or modified function at suboptimal temperatures.

Oil and Gas Well Stimulation – Enzyme Breakers

Galactomannans are used in the pharmaceutical industry for drug delivery (33, 34, 74), in the food industry as thickening agents (23), and in the oil industry for oil/gas well fracturing (41, 42, 50). Galactomannan is a polysaccharide frequently found in the seed endosperm of leguminous plants, and is comprised of a mannose backbone linked by β-1,4 glycosidic bonds and decorated with α-1,6-bonded galactose residues (25). The rheological properties and solubility of galactomannan are determined by the degree of backbone decoration with galactose and the length of the mannan backbone (82). These properties can be modified through the enzymatic action of glycoside hydrolases, enzymes that break glycosidic linkages, liberating sugar molecules or saccharide chains. Enzymes known to act on galactomannan are endo-β-D-mannanase (EC 3.2.1.78), β-mannosidase (EC 3.2.1.25), and α-galactosidase (EC 3.2.1.22). β-Mannosidases, are exo-acting β-mannanases, which typically hydrolyze the terminal glycoside group on the polymer backbone, as shown in Figure 2.1. Endo-acting β-mannanases cleave the internal β-1,4-mannosidic bond, reducing the chain length, while α-galactosidase removes the side chains from the mannan backbone. While α-galactosidase does not reduce the viscosity of galactomannan, it enables complete hydrolysis of the galactomannan polysaccharide by β-mannanase and β-mannosidase (50).

The enzymatic modification of galactomannan can be applied to the oil industry to fracture oil/gas well and stimulate flow for petroleum extraction (41, 42, 50, 86). Oil and gas well fracturing stimulates the well by creating cracks through which oil and gas can seep through the bedrock perpendicular to the well bore (17). To stimulate the well, a fracturing fluid, consisting of proppant (i.e., sand particles) suspended in a matrix, is pumped into the well at sufficient pressure to fracture the well, as shown in Figure 2.1. After fracturing, the fluid must be broken down so that it can be pumped out, and the well can start to produce. One common fracturing fluid matrix is guar gum, a common galactomannan, which can be hydrolyzed either by chemical oxidation or enzymatic action. The viscosity in the guar solution is needed to suspend the proppant in the fluid being injected into the well. Enzymes or chemical oxidizers used for this purpose are referred to as “breakers”(50).

Chemical oxidizers, such as persulfates, can react before the fracturing process is complete, resulting in reduced well stimulation (50). Alternatively, enzymatic breakers can be incorporated into the fracturing fluid and the guar hydrolyzed *in situ*. The proppant remains in the well to hold open crevices and allow the oil or gas to flow. As wells are drilled deeper, temperatures at the lower reaches will be hotter, with temperature gradients over 30°C/km (19). Therefore, the biophysical properties of the enzymatic breakers are important because of the increased temperature that can occur in the deeper reaches of the well. Enzymes from hyperthermophilic organisms that are active at high temperatures are ideal for use in these environments. In addition, it is desirable to have enzymes that have little activity at surface temperatures (35-40°C), because this would result in premature hydrolysis of the guar. To this end, hyperthermophilic enzymes are potentially important reagents for the oil/gas well stimulation process. Hyperthermophilic enzymes have extremely low activity at low temperatures, but their thermoactivity would be activated at higher temperatures such that efficient hydrolysis of guar in deep wells occurs (41, 42). In fact, hyperthermophilic organisms have been isolated from deep well environments. The hyperthermophile *Thermococcus sibiricus* has been cultured from a high temperature oil well in Siberia (53), indicating that the temperature range within the wells is within the range of hyperthermophilic enzymes.

From genome sequence information as well as from more conventional enzyme discovery approaches, hyperthermophilic galactomannanases and α-galactosidases have been identified (52, 55). There are a number of hyperthermophilic β-mannanases that are active over a broad range of temperature and pH. For instances, the β-mannanases from *Thermotoga maritima* and *Thermotoga neapolitana* both have optimal activity at 90°C and pH 7.0 (28, 55). Another thermophilic β-mannanase from *Rhodothermus marinus* is optimally at 85°C, but under acidic conditions (pH 5.4) (57). There are also a number of α-galactosidases with optimal temperatures ranging between 90 and 95°C from the hyperthermophiles *Thermus brockianus* ITI360 (32), *T. maritima* (46), and *T. neapolitana* (43). Furthermore, the pH optima vary from 5.0 for *T. maritima* to 7.0 for *T.*

neapolitana, thus providing a wide range of operating conditions for these α-galactosidases.

The potential for use of hyperthermophilic enzymes for oil and gas recovery has been examined in laboratory settings. *T. neapolitana* β-mannosidase, β-mannanase and α-galactosidase have all been evaluated for viscosity reduction of galactomannan solutions at elevated temperatures (50). It was shown that β-mannanase substantially reduced the viscosity of galactomannan solutions. When β-mannanase was combined with α-galactosidase the residual particulate material was minimized, indicating that these enzymes could be used as enzymatic breakers for oil and gas well recovery (50). Based on these results, opportunities for hyperthermophilic enzymes for this application have been proposed (41, 42). As the demand of oil and gas increases and these resources become more difficult to recover, hyperthermophilic enzyme-based approaches for oil and gas recovery efforts should be examined more closely for both environmental and production reasons.

Bioethanol Production

Oil is expected to be the primary energy source world-wide through 2025, during which time the consumption of oil is expected to rise by 54% (1). During this same period, the greenhouse gas carbon dioxide (CO_2) is expected to increase from 23.9 to 37.1 billion metric tons per year (1). In order to reduce CO_2 emissions and save natural resources, alternative energy sources must be utilized in place of fossil fuels. One possible alternative is an ethanol-based fuel economy. Hydrous ethanol as an auto fuel was used extensively in Brazil in the early 1990's, but gasoline has since replaced it there as the most common auto fuel. Still, there are more than 3.5 million operating vehicles in Brazil that use hydrous ethanol as fuel (95). In the U.S., ethanol is currently used as an oxygenate in gasoline, which allows it to burn cleaner and improves the octane rating (96). With recent concern over the health effect of methyl tert-butyl ether (MTBE) as a ground water contaminant, alternatives oxygenates including ethanol are being utilized (75, 96). Ethanol has also been proposed as a stable source of protons for fuel cells (27). Ethanol as an auto fuel or fuel cell proton source is environmentally attractive. The

carbon dioxide released during ethanol combustion is recently fixed CO_2 and part of the active carbon cycle, so new greenhouse gas release can be significantly reduced (95).

In order to be a green process, ethanol must be produced from a renewable resource. Fermentation of sugars to ethanol is one such method. There are many current processes for producing ethanol via fermentation using a variety of carbohydrate sources. The most common source in the U.S. is corn, which is processed by milling, liquefaction, saccharification by enzyme treatment, fermentation, and distillation, as shown in Figure 2.2. Another sugar source that has been studied is potatoes and potato waste (44). The potatoes are ground into small particles, heated and treated with enzymes to hydrolyze the starch to dextrin, fermented by yeast to ethanol, which is then recovered. A similar strategy can potentially be used to ferment starch from other tubers, such as sweet potatoes. The common sweet potato (*Iponea babatas*), in fact, has an intrinsically greater potential for hydrolysis, over 40% for *I. batatas* as compared to under 10% for potatoes (*Solanum tuberosum)* (38).

The enzymes necessary for complete hydrolysis of the starch to glucose or maltose include α-amylase (EC 3.2.1.1), pullulanase (EC 3.2.1.41), α-glucosidase (EC 3.2.1.20), and glucoamylase (EC 3.2.1.3). These enzymes work synergistically to remove the branched chains, hydrolyze the linear polysaccharide, and hydrolyze maltooligosaccharides to glucose. The starch is hydrolyzed through the processes of liquefaction and saccharification, shown in Figure 2.2. The starch hydrolysis process can be improved by using hyperthermophilic glycosidases, allowing the processing to be carried out at elevated temperatures where the solubility of the starch is greater and substrate and product diffusional rates are higher. Higher temperatures are also beneficial in reducing contamination of the sugar mixture, which can serve as an excellent medium for unwanted organisms incapable of producing ethanol. To date, many of these required amylolytic enzymes needed for starch hydrolysis have already been identified and/or characterized from hyperthermophilic organisms (6, 13, 39, 45, 81).

Another property of hyperthermophilic enzymes is that they have significantly reduced activity at low temperatures. One method to increase the processing efficiency of tuber and root starch into ethanol involves developing transgenic plants, which produce the hyperthermophilic amyolytic enzymes *in situ*. For example, sweet potatoes grow

below ground level at about 18°C, a temperature at which hyperthermophilic enzymes are virtually inactive. Thus, there should be minimal interference with normal plant development. Furthermore, hydrolytic activity of the hyperthermophilic enzymes can be activated as the sweet potatoes are mashed and heated. By incorporating genes encoding hyperthermophilic enzymes directly into the plant, there are considerable potential savings of time and money since the enzymes are produced as a consequence of plant formation. Using this technique, no enzymes need to be added, in contrast to traditional starch processing shown in Figure 2.2. Ultimately, the liquefaction and saccharification steps could be combined to hydrolyze the starch in a one-pot process.

Production of α-amylase using genetically modified potatoes has been demonstrated: a fusion gene of α-amylase from *Bacillus stearothermophilus* and glucose isomerase from *Thermus thermophilus* has been inserted into a potato (5). The starch from the potato was hydrolyzed to glucose and converted to fructose by glucose isomerase. The insertion of thermophilic archaeal genes into plants has also been shown to be possible for the expression of α-glucosidase from *Sulfolobus solfataricus* in a transgenic tobacco plant that produced the active enzyme capable of surviving tissue preservation (54).

The final step in the production of ethanol is the fermentation of glucose into ethanol followed by distillation to purify the ethanol. To reduce the purification expenses, fermenting the saccharification mixture to generate the greatest concentration of ethanol is desirable. One method of increasing the ethanol concentration is very high gravity fermentation, which has produced 23.8% v/v ethanol from wheat mash containing 38% dissolved solids (85). Continuing research in very high gravity fermentation has looked at supplementing the media with nitrogen additives (2) and utilizing multistage continuous fermenters (4), both of which were able to generate about 17% v/v ethanol. While no hyperthermophilic organism has yet been identified that could be used in the fermentation of glucose to ethanol under these conditions, the potential of operating this process closer to distillation temperatures is intriguing and could improve process economics.

High Fructose Corn Syrup Production

In addition to production of fuel ethanol from transgenic plants, other processes that take advantage of hyperthermophilic enzymes can be envisioned, such as the production of high fructose corn syrup (HFCS), a sweetener used in soft drinks, teas, condiments and baked goods. The HFCS process currently utilizes corn as a source of starch for the production of fructose. While other sources of starch, such as potatoes, wheat, and tapioca can be used for producing fructose syrup, corn is the most economical choice (90). The production of HFCS was one of the first large scale applications of immobilized enzymes (93).

The HFCS process consists of three main steps, liquefaction, saccharification, and isomerization. The starch is processed by liquefaction and saccharification at high temperatures to produce low glucose, as shown in Figure 2.2. In the liquefaction process, starch granules are gelatinized in a jet cooker at 105 to 110°C and pH 5.8-6.5 for 5-8 minutes and then treated with a thermostable α-amylase at 95°C for 1-2 hours. After treatment, the average oligosaccharide chain length is 10-13 glucose residues (21). This process must be run at low pH in order to prevent byproduct and color contamination (93). The saccharification process utilizes glucoamylase and pullulanase to complete the hydrolysis of starch to glucose. This step occurs at 55 to 60°C and a lower pH (4.2-5.0), which requires acidification of the mixture. In the last step, isomerization, the glucose is converted into fructose by immobilized xylose isomerase (often referred to as glucose isomerase because it can convert glucose to fructose). The immobilized xylose isomerase columns typically operate at 60°C and pH 7.5-8.0 for times of 1-4 hours (21). Under these operating conditions, a mixture containing 42% fructose is produced (90). However, 55 % fructose is required to have the same sweetness as sucrose from cane sugar. Enriched syrup of 90% fructose is prepared using strong acid cation-exchange chromatography, which is then combined with the 42% fructose to achieve the desired 55% fructose blend (63).

There has been much work in recent years attempting to discover, or create through mutagenesis techniques, thermally stable, low pH active xylose/glucose isomers enzymes (11, 12, 40, 47, 51, 76-78, 99). This work is driven by the desire to eliminate the chromatography purification step from the HFCS process by achieving 55% fructose

content through adjustment of the equilibrium concentration of fructose by raising the temperature. Because the glucose: fructose equilibrium favors fructose as temperature increases (84), operation at higher temperatures could potentially eliminate the need for chromatographic enrichment (37). At the current operating temperature of 60°C, the theoretical equilibrium concentration of fructose is 50.7%, whereas if the process is operated at 90°C, the equilibrium rises to 55.6% fructose (93). A comparison of the *Thermotoga* xylose isomerases and the xylose isomerase from *Streptomyces murinus* (an enzyme currently used in the industrial production of fructose from glucose) determined that for production at 90°C, the estimated productivity for *S. murinus* isomerase dropped to 200 kg fructose per kg enzyme, whereas the *T. neapolitana* isomerase had a lifetime productivity of 1000 kg fructose per kg enzyme (3). One concern when operating at these elevated temperatures are byproducts resulting from the Maillard reaction. In order to reduce these byproducts, the reaction would need to be operated under acidic conditions and at rapid processing rates to minimize enzyme-substrate exposure.

There are a number of possibilities for streamlining the HFCS process by incorporating extremophilic enzymes. The liquefaction step occurs at pH 6.2, which has to be obtained by neutralizing the starch solution from the natural pH of about 4.5 (21). If a thermally stable α-amylase that has activity at pH 4.5 can be directly incorporated into the process, it will eliminate the need to change the pH to 6.2 and then back to 4.5 for the saccharification step. One such α-amylase has been generated by directed evolution via DNA shuffling of three enzymes, and was screened from over 19,000 possible candidates (59). These candidates were screened for activity at pH 4.5 and 100°C without Ca^{2+} and compared to current wild-type enzymes. The enhanced thermostability of this enzyme resulted in a greater degree of starch hydrolysis at 115°C at pH 4.5 (59). If acidophilic, thermostable pullulanases and glucoamylases are discovered or evolved, the saccharification step can be performed at the same temperature and pH as the liquefaction step that will eliminate the need to cool the liquefied stream or saccharified stream. This assumes that a thermally stable xylose isomerase can also be integrated into the process. There are already several potential glucose isomerase candidates for thermal starch processing, including enzymes from *T. thermophilus* (pH_{opt} 7.0 and T_{opt} 95°C) (26), *T. maritima* (pH_{opt} 6.5-7.5 and T_{opt} 105°C) (7), and *T. neapolitana* (pH_{opt} 7.1 and T_{opt} 95°C)

(92) More work is needed to determine if a single reaction vessel could be used to produce fructose from a starch starting material.

Thermophilic Esterases

The resolution of stereoisomers is one of the most difficult purification challenges (18, 88) and essential drugs marketed as pure enantiomers, such as Naproxen. Naproxen is a popular nonsteroidal anti-inflammatory drug (NSAID) which is distributed solely as the *(S)*-enantiomer (24). This is unique within the NSAID family because most others are marketed as racemic mixtures (36). Resolution of Naproxen is accomplished by combinations of asymmetric synthesis, chromatography, and diasteriometric crystallization (36).

To simplify purification of chiral molecules, the synthetic reaction can sometimes be designed to favor one enantiomer by the addition of protective groups (29). The dielectric constant of the solvent used in chiro-selective molecular recognition may also have a profound effect, allowing resolution of one chiral selector by changing solvent (62). For example, the use of ionic solvents was able to improve enantioselectivity of the lipase-catalyzed reaction of 1-phenylethanol (65). An alternative method to purifying racemic mixtures under environmentally benign conditions is to exploit the natural enantioselectivity of enzymes (48). Esterases, for example, can discriminate between the (*R*)- and (*S*)-enantiomers of a particular ester and selectively hydrolyze only one form, thereby creating a separation potential. The selectivity of these enzymes can be dependent upon the presence of organic solvents in the resolution of racemic mixtures (87).

Enzyme flexibility has been shown to be a function of an enzyme's reaction environment (30). By relaxing conformational restraints within an enzyme by addition of a co-solvent, there is often a corresponding loss of enantioselectivity (30). Temperature is another means of affecting the stereoselectivity of an enzymatic reaction (9). One would expect the stereoselectivity of a reaction to increase at lower temperatures because of decreased conformational flexibility, but this is not always the case (56). For the secondary alcohol dehydrogenase (SADH) from *Thermoanaerobacter ethanolicus*, the conversion of 2-butanone to *(R)*-2-butanol was favored when temperatures were above

26°C while *(S)*-2-butanol was favored at temperatures below 26°C (98). In another example, increasing the temperature from 28°C to 55°C increased the selectivity ratio (E) from 24 to 130, respectively, for the kinetic resolution of racemic 1-N-(propen-1-yl)-3-N-(phenylacetoxy)-aminoazetidin-2-one by Penicillin G-Acylase (8).

Hyperthermophilic proteins offer the unique advantage that they have a large soluble temperature range below the optimum activity temperature and that they are frequently resistant to denaturation by organic chemicals (20). Sehgal and Kelly (66, 68) showed that enantioselectivity with hyperthermophilic esterase could be manipulated in the presence co-solvent or by running reactions at suboptimal temperatures. When the reaction of racemic Naproxen methyl ester was conducted at 50°C using an esterase from *S. solfataricus* (*Sso* EST1, T_{opt} >95°C) the enzyme was able to selectively hydrolyze the *(S)*-enantiomer of Naproxen methyl ester, producing 95% *(S)*-Naproxen (67), as illustrated in Figure 2.3. It was also shown that a co-solvent increased the activity of a hyperthermophilic esterase at suboptimal temperatures (68), probably the result of increased conformational flexibility. Using circular dichroism and NMR, the *Sso* EST1 showed increased flexibility when the co-solvent DMSO was added at 3.5 % (v/v) compensating for the thermal activation of the enzyme.

Enzyme Discovery

The need for enzymes which meet ever-expanding processing criteria (e.g., thermostability, specificity, etc.) continues to drive the search for new enzymes either from traditional sources, newly discovered organisms, or through directed evolution. One approach that is being utilized is cloning the metagenome (35), *i.e.*, a collection of genes of all the organisms present in an environmental sample. Until recently, this concept had been combined with screening to explore genetic diversity and discover new enzymes (60). This approach has recently been expanded in the Sargasso Sea project in which surface water samples were collected, genomic DNA extracted, and cloned into genomic libraries with insert sizes of 2 to 6 kb (91). Whole-genome shotgun sequencing techniques were applied to this genomic library and led to identification of an estimated 1800 genomic species based on sequence similarity, of which 148 were unknown bacteria. The sequencing project also revealed over 1.2 million previously unknown

genes (91). Approaches such as this one promise to be important for discovering enzymes from unculturable organisms.

For organisms that are culturable, advances in molecular biology and increased amounts of bioinformatic data provide tools to identify gene products that can be used for processes. Prior to recombinant expression of proteins, researchers were limited to discovering proteins one at a time from organism cultures. With the advances of PCR and molecular cloning, the rate of protein discovery has been increased by the ability of high-throughput robotics and screening (70-72). Advances in molecular biology have also led to methodologies for determining differential expression of gene using microarrays or real-time PCR. Both techniques quantitatively determine the amount of messenger ribonucleic acid (mRNA) that is produced for a particular gene under various conditions. Full-genome microarrays have been utilized to put together sugar utilization pathways (15), identify novel putative interactive sites for anticancer therapy (73), and discover new wood processing enzymes (83).

The utility of microarrays for identifying promising new biocatalysts is just starting to be recognized. In some cases, differential gene expression data has revealed unexpected enzyme activities on substrates of interest. For example, BLAST searches (97) showed that the *T. maritima* enzyme Cel5A (TM1751) was most similar to the characterized endo-1,4-glucanase B (EngB) from *Clostridium cellulovorans* (28% identity) (31). Based on sequence alignments alone, *Tm*Cel5A would be expected to be most active on polysaccharides containing the β-glucan linkage; however, the purified enzyme displayed comparable activity on glucomannan, galactomannan, and β-glucan (16). Subsequent experiments analyzing the differential gene expression in *T. maritima* on various saccharide growth substrates confirmed the expression of *Tm*Cel5A on galactomannan and glucomannan (15), as shown in Figure 2.1. If not for the information obtained through differential expression experiments, potentially important galactomannanases for application such as oil and gas discovery would have gone unnoticed. As additional microbial genomes are sequenced, microarrays should become an increasingly important tool for identifying new enzymes for biotechnological applications.

Final Comments

There are many areas in which enzymes from hyperthermophiles can be considered for use in technologically important biocatalytic processes. Some examples are presented here. The combination of power of biocatalysis and the expanded functional temperature range intrinsic to hyperthermophilic enzymes should open up numerous opportunities as the production of chemicals and biochemicals moves increasingly towards environmentally benign processes.

Acknowledgments

The authors acknowledge support from the U.S. National Science Foundation Biotechnology Program for the work described here. S.B.C. acknowledges support from an NIEHS Bioinformatics Traineeship. M.R.J. and D.A.C. acknowledge support from GAANN Fellowships.

References

1. 2004. International energy outlook 2004. Energy Information Administration, US Department of Energy.
2. **Bafrncova, P., D. Smogrovicova, I. Slavikova, J. Patkova, and Z. Domeny.** 1999. Improvement of very high gravity ethanol fermentation by media supplementation using *Saccharomyces cerevisiae*. Biotechnol. Lett. **21:**337-341.
3. **Bandlish, R. K., J. M. Hess, K. L. Epting, C. Vieille, and R. M. Kelly.** 2002. Glucose-to-fructose conversion at high temperatures with xylose (glucose) isomerases from *Streptomyces murinus* and two hyperthermophilic *Thermotoga* species. Biotechnol. Bioeng. **80:**185-194.
4. **Bayrock, D. P., and W. M. Ingledew.** 2001. Application of multistage continuous fermentation for production of fuel alcohol by very-high-gravity fermentation technology. J. Ind. Microbiol. Biot. **27:**87-93.
5. **Beaujean, A., C. Ducrocq-Assaf, R. S. Sangwan, G. Lilius, L. Bulow, and B. S. Sangwan-Norreel.** 2000. Engineering direct fructose production in processed potato tubers by expressing a bifunctional alpha-amylase/glucose isomerase gene complex. Biotechnol. Bioeng. **70:**9-16.
6. **Bibel, M., C. Brettl, U. Gosslar, G. Kriegshauser, and W. Liebl.** 1998. Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*. FEMS Microbiol. Lett. **158:**9-15.
7. **Brown, S. H., C. Sjoholm, and R. M. Kelly.** 1993. Purification and characterization of a highly thermostable glucose-isomerase produced by the extremely thermophilic eubacterium, *Thermotoga maritima*. Biotechnol. Bioeng. **41:**878-886.
8. **Cainelli, G., V. De Matteis, P. Galletti, D. Giacomini, and P. Orioli.** 2000. Temperature and solvent effects on enzyme stereoselectivity: inversion temperature in kinetic resolutions with lipases. Chem. Commun.**:**2351-2352.
9. **Cainelli, G., D. Giacomini, and P. Galletti.** 1999. Temperature and solvent effects in facial diastereoselectivity of nucleophilic addition: entropic and enthalpic contribution. Chem. Commun.**:**567-572.
10. **Cavicchioli, R., K. S. Siddiqui, D. Andrews, and K. R. Sowers.** 2002. Low-temperature extremophiles and their applications. Curr. Opin. Biotech. **13:**253-261.
11. **Cha, J. H., and C. A. Batt.** 1998. Lowering the pH optimum of D-xylose isomerase: the effect of mutations of the negatively charged residues. Mol. Cells **8:**374-382.
12. **Chang, C. S., B. C. Park, L. D. S., and S. S. W.** 1999. Crystal structures of thermostable xylose isomerases from *Thermus caldophilus* and *Thermus thermophilus*: Possible structural determinants of thermostability. J. Mol. Biol. **288:**623-634.
13. **Chang, S. T., K. N. Parker, M. W. Bauer, and R. M. Kelly.** 2001. alpha-Glucosidase from *Pyrococcus furiosus*, p. 260-269, Hyperthermophilic Enzymes, Pt a, vol. 330.
14. **Chen, K. Q., and F. H. Arnold.** 1993. Tuning the activity of an enzyme for unusual environments - Sequential random mutagenesis of subtilisin-E for catalysis in dimethylformamide. Proc. Natl. Acad. Sci. USA **90:**5618-5622.
15. **Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger, and R. M. Kelly.** 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J. Biol. Chem. **278:**7540-7552.

16. **Chhabra, S. R., K. R. Shockley, D. E. Ward, and R. M. Kelly.** 2002. Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium *Thermotoga maritima* grown on glucan- and mannan-based polysaccharides. Appl. Environ. Microb. **68:**545-554.
17. **Clark, J. B.** 1949. A hydraulic process for increasing the productivity of wells. T. Am. I. Min. Met. Eng. **186:**1-8.
18. **Collet, A.** 1999. Separation and purification of enantiomers by crystallisation methods. Enantiomer **4:**157-172.
19. **Corry, D., and C. Brown.** 1998. Temperature and heat flow in the Celtic Sea basins. Petrol. Geosci. **4:**317-326.
20. **Cowan, D. A.** 1997. Thermophilic proteins: Stability and function in aqueous and organic solvents. Comp. Biochem. Phys. A **118:**429-438.
21. **Crabb, W. D., and J. K. Shetty.** 1999. Commodity scale production of sugars from starches. Curr. Opin. Microbiol. **2:**252-256.
22. **Dalby, P. A.** 2003. Optimising enzyme function by directed evolution. Curr. Opin. Struc. Biol. **13:**500-505.
23. **Daniel, J., and R. Whistler.** 1984. Guar gum: Production, structure, and uses, p. 159-184. *In* V. F. Rasper (ed.), Cereal Polysaccharides in Technology and Nutrition.
24. **Davies, N. M.** 1997. Methods of analysis of chiral non-steroidal anti-inflammatory drugs. J. Chromatogr. B **691:**229-261.
25. **Dea, I. C. M., and A. Morrison.** 1975. Chemistry and interactions of seed galactomannans. Adv. Carbohyd. Chem. Bi. **31:**241-312.
26. **Dekker, K., H. Yamagata, Sakaguchik, and S. Udaka.** 1991. Xylose (glucose) isomerase gene from the thermophile *Thermus thermophilus* - Cloning, sequencing, and comparison with other thermostable xylose isomerases. J. Bacteriol. **173:**3078-3083.
27. **Deluga, G. A., J. R. Salge, L. D. Schmidt, and X. E. Verykios.** 2004. Renewable hydrogen from ethanol by autothermal reforming. Science **303:**993-997.
28. **Duffaud, G. D., C. M. McCutchen, P. Leduc, K. N. Parker, and R. M. Kelly.** 1997. Purification and characterization of extremely thermostable beta-mannanase, beta-mannosidase, and alpha-galactosidase from the hyperthermophilic eubacterium *Thermotoga neapolitana* 5068. Appl. Environ. Microb. **63:**169-177.
29. **Faber, K., and M. C. R. Franssen.** 1993. Prospects for the increased application of biocatalysts in organic transformations. Trends Biochem. Sci. **11:**461-470.
30. **Fitzpatrick, P. A., and A. M. Klibanov.** 1991. How can the solvent affect enzyme enantioselectivity. J. Am. Chem. Soc. **113:**3166-3171.
31. **Foong, F., T. Hamamoto, O. Shoseyov, and R. H. Doi.** 1991. Nucleotide-sequence and characteristics of endoglucanase gene engB from *Clostridium cellulovorans*. J. Gen. Microbiol. **137:**1729-1736.
32. **Fridjonsson, O., H. Watzlawick, A. Gehweiler, T. Rohrhirsch, and R. Mattes.** 1999. Cloning of the gene encoding a novel thermostable alpha-galactosidase from *Thermus brockianus* ITI360. Appl. Environ. Microb. **65:**3955-3963.
33. **Gliko-Kabir, I., B. Yagen, M. Baluom, and A. Rubinstein.** 2000. Phosphated crosslinked guar for colon-specific drug delivery II. In vitro and *in vivo* evaluation in the rat. J. Control. Release **63:**129-134.
34. **Gliko-Kabir, I., B. Yagen, A. Penhasi, and A. Rubinstein.** 2000. Phosphated crosslinked guar for colon-specific drug delivery I. Preparation and physicochemical characterization. J. Control. Release **63:**121-127.
35. **Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman.** 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. **5:**R245-R249.

36. **Harrington, P. J., and E. Lodewijk.** 1997. Twenty years of naproxen technology. Org. Process Res. Dev. **1:**72-76.
37. **Hartley, B. S., N. Hanlon, R. J. Jackson, and M. Rangarajan.** 2000. Glucose isomerase: Insights into protein engineering for increased thermostability. Biochim. Biophys. Acta **1543:**294-335.
38. **Hoover, R.** 2001. Composition, molecular structure, and physicochemical properties of tuber and root starches: a review. Carbohyd. Polym. **45:**253-267.
39. **Jorgensen, S., C. E. Vorgias, and G. Antranikian.** 1997. Cloning, sequencing, characterization, and expression of an extracellular alpha-amylase from the hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis*. J. Biol. Chem. **272:**16335-16342.
40. **Kaneko, T., S. Takahashi, and K. Saito.** 2000. Characterization of acid-stable glucose isomerase from *Streptomyces* sp. and development of single-step processes for high-fructose corn sweetener (HFCS) production. Biosci. Biotech. Bioch. **64:**940-947.
41. **Kelly, R. M., S. A. Khan, P. Leduc, A. Tayal, and R. K. Prud'homme.** Aug 6, 2002 2002. Compositions for fracturing subterranean formations. US Patent #6,428,995.
42. **Kelly, R. M., S. A. Khan, P. Leduc, A. Tayal, and R. K. Prud'homme.** June 6, 1995 1995. Methods and compositions for fracturing subterranean formations. US Patent, #5,421,412.
43. **King, M. R., D. A. Yernool, D. E. Eveleigh, and B. M. Chassy.** 1998. Thermostable alpha-galactosidase from *Thermotoga neapolitana*: cloning, sequencing and expression. FEMS Microbiol. Lett. **163:**37-42.
44. **Kobayashi, F., T. Sawada, Y. Nakamura, M. Ohnaga, M. Godliving, and T. Ushiyama.** 1998. Saccharification and alcohol fermentation in starch solution of steam-exploded potato. Appl. Biochem. Biotech. **69:**177-189.
45. **Kwak, Y. S., T. Akiba, and T. Kudo.** 1998. Purification and characterization of alpha-amylase from hyperthermophilic archaeon *Thermococcus profundus*, which hydrolyzes both alpha-1,4 and alpha-1,6 glucosidic linkages. J. Ferment. Bioeng. **86:**363-367.
46. **Liebl, W., B. Wagner, and J. Schellhase.** 1998. Properties of an alpha-galactosidase, and structure of its gene galA, within an alpha- and beta-galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. Syst. Appl. Microbiol. **21:**1-11.
47. **Liu, S. Y., J. Wiegel, and F. C. Gherardini.** 1996. Purification and cloning of a thermostable xylose (glucose) isomerase with an acidic pH optimum from *Thermoanaerobacterium* strain JW/SL-YS 489. J. Bacteriol. **178:**5938-5945.
48. **Manco, G., G. Carrea, E. Giosue, G. Ottolina, G. Adamo, and M. Rossi.** 2002. Modification of the enantioselectivity of two homologous thermophilic carboxylesterases from *Alicyclobacillus acidocaldarius* and *Archaeoglobus fulgidus* by random mutagenesis and screening. Extremophiles **6:**325-331.
49. **Marhuenda-Egea, F. C., and M. J. Bonete.** 2002. Extreme halophilic enzymes in organic solvents. Curr. Opin. Biotech. **13:**385-389.
50. **McCutchen, C. M., G. D. Duffaud, P. Leduc, A. R. H. Petersen, A. Tayal, S. A. Khan, and R. M. Kelly.** 1996. Characterization of extremely thermostable enzymatic breakers (alpha-1,6-galactosidase and beta-1,4-mannanase) from the hyperthermophilic bacterium *Thermotoga neapolitana* 5068 for hydrolysis of guar gum. Biotechnol. Bioeng. **52:**332-339.
51. **Meng, M., C. Lee, M. Bagdasarian, and J. G. Zeikus.** 1991. Switching substrate preference of thermophilic xylose isomerase from D-xylose to D-glucose by redesigning the substrate binding pocket. Proc. Natl. Acad. Sci. USA **88:**4015-4019.

52. **Miller, E. S., K. N. Parker, W. Liebl, D. Lam, W. Callen, M. A. Snead, E. J. Mathur, J. M. Short, and R. M. Kelly.** 2001. alpha-D-galactosidases from *Thermotoga* species, p. 246-260, Hyperthermophilic Enzymes, Pt A, vol. 330.
53. **Miroshnichenko, M. L., H. Hippe, E. Stackebrandt, N. A. Kostrikina, N. A. Chernyh, C. Jeanthon, T. N. Nazina, S. S. Belyaev, and E. A. Bonch-Osmolovskaya.** 2001. Isolation and characterization of *Thermococcus sibiricus* sp nov from a Western Siberia high-temperature oil reservoir. Extremophiles **5:**85-91.
54. **Montalvo-Rodriguez, R., C. Haseltine, K. Huess-LaRossa, T. Clemente, J. Soto, P. Staswick, and P. Blum.** 2000. Autohydrolysis of plant polysaccharides using transgenic hyperthermophilic enzymes. Biotechnol. Bioeng. **70:**151-159.
55. **Parker, K. N., S. R. Chhabra, D. Lam, W. Callen, G. D. Duffaud, M. A. Snead, J. M. Short, E. J. Mathur, and R. M. Kelly.** 2001. Galactomannanases man2 and man5 from *Thermotoga* species: Growth physiology on galactomannans, gene sequence analysis, and biochemical properties of recombinant enzymes. Biotechnol. Bioeng. **75:**322-333.
56. **Phillips, R. S.** 1996. Temperature modulation of the stereochemistry of enzymatic catalysis: Prospects for exploitation. Trends Biotechnol. **14:**13-16.
57. **Politz, O., M. Krah, K. K. Thomsen, and R. Borriss.** 2000. A highly thermostable endo-(1,4)-beta-mannanase from the marine bacterium *Rhodothermus marinus*. Appl. Microbiol. Biot. **53:**715-721.
58. **Reetz, M. T.** 2002. Lipases as practical biocatalysts. Curr. Opin. Chem. Biol. **6:**145-150.
59. **Richardson, T. H., X. Q. Tan, G. Frey, W. Callen, M. Cabell, D. Lam, J. Macomber, J. M. Short, D. E. Robertson, and C. Miller.** 2002. A novel, high performance enzyme for starch liquefaction - Discovery and optimization of a low pH, thermostable alpha-amylase. J. Biol. Chem. **277:**26501-26507.
60. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microb. **66:**2541-2547.
61. **Rozzell, J. D.** 1999. Commercial scale biocatalysis: Myths and realities. Bioorgan. Med. Chem. **7:**2253-2261.
62. **Sakai, K., R. Sakurai, and N. Hirayama.** 2004. Chiral discrimination controlled by the solvent dielectric constant. Tetrahedron: Assymetry **15:**1073-1076.
63. **Schenck, F. W.** 2000. High fructose syrups - a review. Int. Sugar J. **102:**285-288.
64. **Schmid, A., J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, and B. Witholt.** 2001. Industrial biocatalysis today and tomorrow. Nature **409:**258-268.
65. **Schofer, S. H., N. Kaftzik, P. Wasserscheid, and U. Kragl.** 2001. Enzyme catalysis in ionic liquids: lipase catalyzed kinetic resolution of 1-phenylethanol with improved enantioselectivity. Chem. Commun. **5:**425-426.
66. **Sehgal, A. C., and R. M. Kelly.** 2002. Enantiomeric resolution of 2-aryl propionic esters with hyperthermophilic and mesophilic esterases: Contrasting thermodynamic mechanisms. J. Am. Chem. Soc. **124:**8190-8191.
67. **Sehgal, A. C., and R. M. Kelly.** 2003. Strategic selection of hyperthermophilic esterases for resolution of 2-arylpropionic esters. Biotechnol. Progr. **19:**1410-1416.
68. **Sehgal, A. C., R. Tompson, J. Cavanagh, and R. M. Kelly.** 2002. Structural and catalytic response to temperature and cosolvents of carboxylesterase EST1 from the extremely thermoacidophilic archaeon *Sulfolobus solfataricus* P1. Biotechnol. Bioeng. **80:**784-793.

69. **Sellek, G. A., and J. B. Chaudhuri.** 1999. Biocatalysis in organic media using enzymes from extremophiles. Enzyme Microb. Technol. **25:**471-482.
70. **Short, J. M.** Apr 9, 2002 2002. Screening for novel bioactivities. US Patent #6,368,798.
71. **Short, J. M.** Sept 24, 2002 2002. Sequence based screening. US Patent #6,455,254.
72. **Short, J. M., and M. Keller.** Jan 16, 2001 2001. High throughput screening for novel enzymes. US Patent #6,174,673.
73. **Simon, R., M. Mirlacher, and G. Sauter.** 2003. Tissue microarrays in cancer diagnosis. Expert Rev. Mol. Diagn. **3:**421-430.
74. **Sinha, V. R., and R. Kumria.** 2001. Polysaccharides in colon-specific drug delivery. Int. J. Pharm. **224:**19-38.
75. **Squillace, P. J., M. J. Moran, W. W. Lapham, C. V. Price, R. M. Clawges, and J. S. Zogorski.** 1999. Volatile organic compounds in untreated ambient groundwater of the United States, 1985-1995. Environ. Sci. Technol. **33:**4176-4187.
76. **Srih-Belghith, K., and S. Bejar.** 1998. A thermostable glucose isomerase having a relatively low optimum pH: study of activity and molecular cloning of the corresponding gene. Biotechnol. Lett. **20:**553-556.
77. **Sriprapundh, D., C. Vieille, and J. G. Zeikus.** 2003. Directed evolution of *Thermotoga neapolitana* xylose isomerase: high activity on glucose at low temperature and low pH. Protein Eng. **16:**683-690.
78. **Sriprapundh, D., C. Vieille, and J. G. Zeikus.** 2000. Molecular determinants of xylose isomerase thermal stability and activity: analysis of thermozymes by site-directed mutagenesis. Protein Eng. **13:**259-265.
79. **Stemmer, W. P. C.** 1994. DNA shuffling by random fragmentation and reassembly - *in vitro* recombination for molecular evolution. Proc. Natl. Acad. Sci. USA **91:**10747-10751.
80. **Stetter, K. O.** 1996. Hyperthermophilic procaryotes. FEMS Microbiol. Rev. **18:**149-158.
81. **Tachibana, Y., M. M. Leclere, S. Fujiwara, M. Takagi, and T. Imanaka.** 1996. Cloning and expression of the alpha-amylase gene from the hyperthermophilic archaeon *Pyrococcus* sp KOD1, and characterization of the enzyme. J. Ferment. Bioeng. **82:**224-232.
82. **Tayal, A., R. M. Kelly, and S. A. Khan.** 1999. Rheology and molecular weight changes during enzymatic degradation of a water-soluble polymer. Macromolecules **32:**294-300.
83. **Teeri, T. T., and H. Brumer.** 2003. Discovery, characterisation and applications of enzymes from the wood-forming tissues of poplar: Glycosyl transferases and xyloglucan endotransglycosylases. Biocatal. Biotransfor. **21:**173-179.
84. **Tewari, Y. B., and R. N. Goldberg.** 1984. Thermodynamics of the conversion of aqueous glucose to fructose. J. Solution Chem. **13:**523-547.
85. **Thomas, K. C., S. H. Hynes, A. M. Jones, and W. M. Ingledew.** 1993. Production of fuel alcohol from wheat by VHG technology - Effect of sugar concentration and fermentation temperature. Appl. Biochem. Biotech. **43:**211-226.
86. **Tjon-Joe-Pin, R. M.** April 13, 1993 1993. Enzyme breaker for galactomannan based fracturing fluid. US Patent #5,201,370.
87. **Tsai, S. W., and H. J. Wei.** 1994. Enantioselective esterification of racemic naproxen by lipases in organic solvent. Enzyme Microb. Technol. **16:**328-333.
88. **Turner, N. J.** 2003. Controlling chirality. Curr. Opin. Biotech. **14:**401-406.
89. **van den Burg, B.** 2003. Extremophiles as a source for novel enzymes. Curr. Opin. Microbiol. **6:**213-218.

90. **Venkatasubramanian, K., and L. S. Harrow.** 1979. Design and operation of a commercial immobilized glucose isomerase reactor system. Ann. NY Acad. Sci. **326:**141-153.
91. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Y. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66-74.
92. **Vieille, C., J. M. Hess, R. M. Kelly, and J. G. Zeikus.** 1995. Xyla cloning and sequencing and biochemical characterization of xylose isomerase from *Thermotoga neapolitana*. Appl. Environ. Microb. **61:**1867-1875.
93. **Vieille, C., and G. J. Zeikus.** 2001. Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. Microbiol. Mol. Biol. R. **65:**1-+.
94. **Wescott, C. R., and A. M. Klibanov.** 1994. The solvent dependence of enzyme specificity. BBA-Protein Struct. M. **1206:**1-9.
95. **Wheals, A. E., L. C. Basso, D. M. G. Alves, and H. V. Amorim.** 1999. Fuel ethanol after 25 years. Trends Biotechnol. **17:**482-487.
96. **Williams, P. R. D., C. A. Cushing, and P. J. Sheehan.** 2003. Data available for evaluating the risks and benefits of MTBE and ethanol as alternative fuel oxygenates. Risk Anal. **23:**1085-1115.
97. **Zhang, J. H., and T. L. Madden.** 1997. PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation. Genome Res. **7:**649-656.
98. **Zheng, C. S., V. T. Pham, and R. S. Phillips.** 1994. Asymmetric reduction of aliphatic and cyclic-ketones with secondary alcohol-dehydrogenase from *Thermoanaerobacter ethanolicus* - Effects of substrate structure and temperature. Catal. Today **22:**607-620.
99. **Zhu, G. P., C. Xu, M. K. Teng, L. M. Tao, X. Y. Zhu, C. J. Wu, J. Hang, L. W. Niu, and Y. Z. Wang.** 1999. Increasing the thermostability of D-xylose isomerase by introduction of a proline into the turn of a random coil. Protein Eng. **12:**635-638.

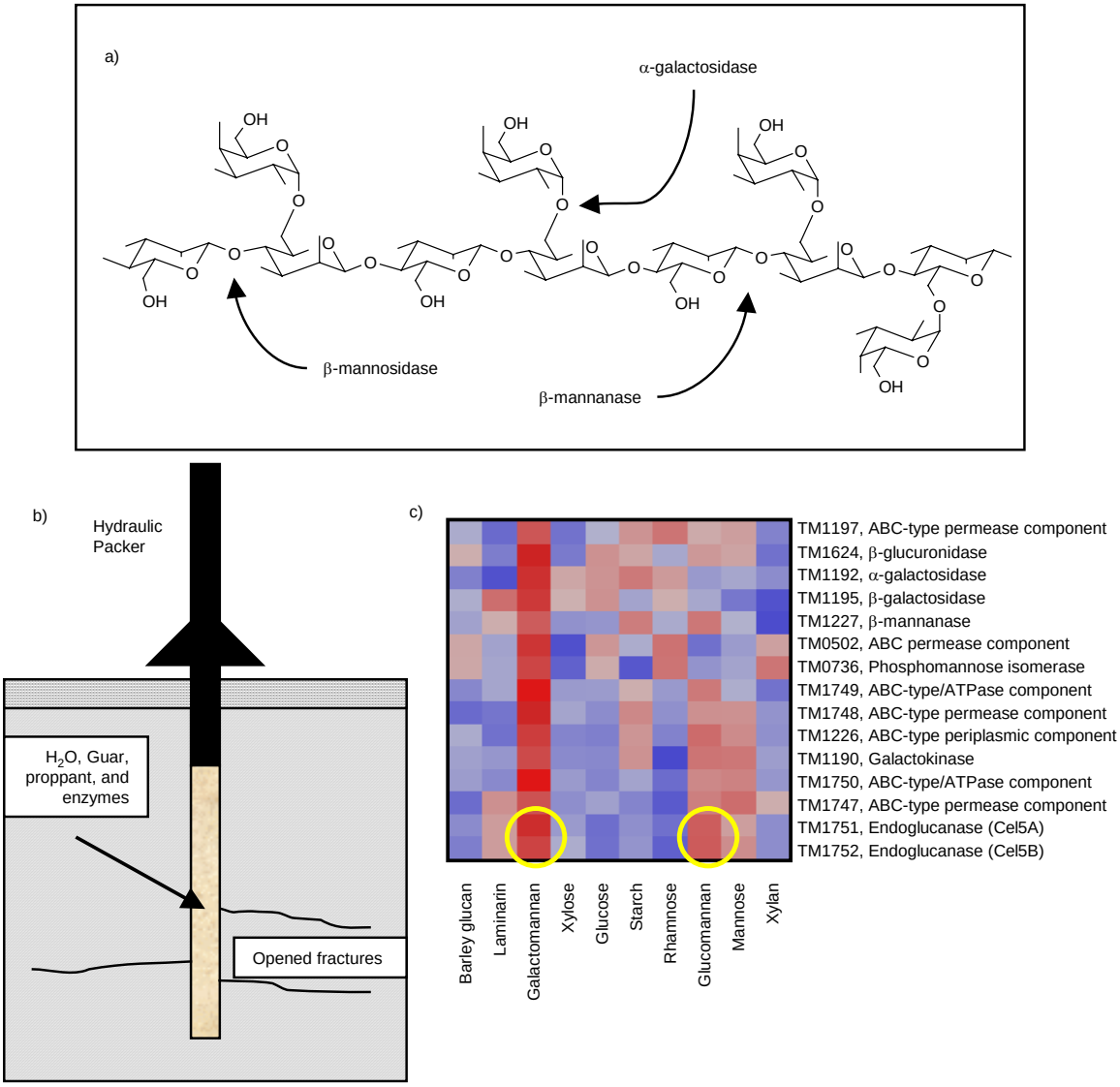


Figure 2.1 a) Galactomannan is hydrolysed by the action of the galactomannanases α-galactosidase, β-mannanase, and β-mannosidase, reducing the galactomannan to oligo- and monosaccharides. Hydroxyl groups at positions 2, 3, and 4 have been omitted for figure clarity. **b)** Galactomannan can be used as the supporting matrix for hydraulic fracturing of oil/gas wells to prop open crevices allowing oil and gas to flow. **c)** Microarrays can be used to identify differentially expressed genes corresponding to desirable biocatalysts. Shown is a differential expression heat plot for the hyperthermophile *T. maritima* grown on a variety of carbohydrates, including galactomannan. In addition to the expected up-regulation (red) of genes encoding β-mannanase (TM1227) and α-galactosidase (TM1192), mixed model statistical analysis of differential expression data also showed up-regulation of two genes encoding putative endoglucanases *Tm*Cel5A (TM1751) and *Tm*Cel5B (TM1752) on mannan polysaccharides (circled in yellow) (15). These genes were cloned and expressed in *E. coli* and found to be highly active on galactomannan.

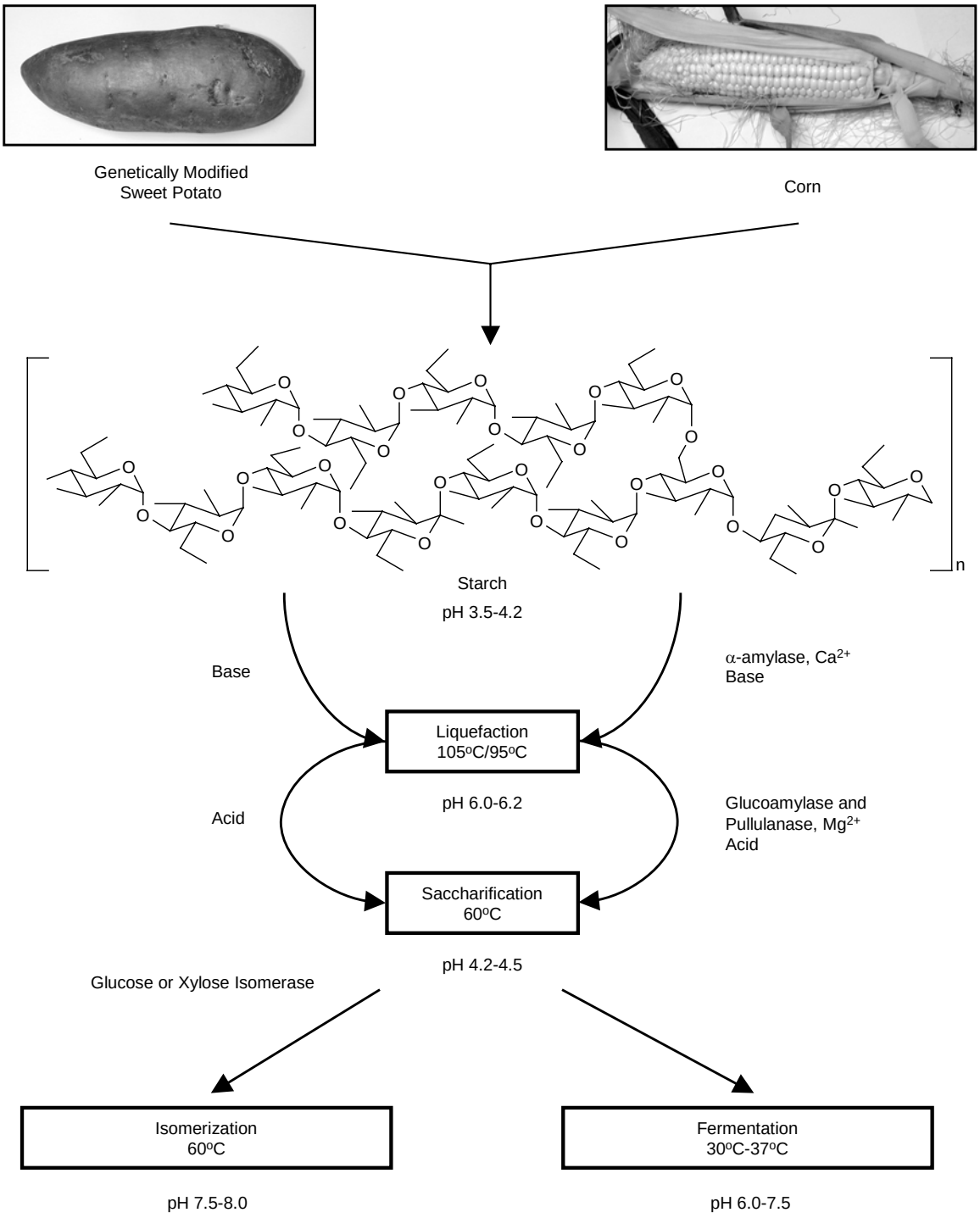


Figure 2.2 The saccharification process is similar for various sources of starch. Starch is hydrolysed by α-amylase during the liquefaction process, which occurs at 95°C. The starch is further hydrolysed by glucoamylase and pullulanase after acidification. The resulting glucose can then be utilized in subsequent processes, such as fermentation to ethanol or isomerization to high fructose corn syrup. If transgenic plants containing genes encoding starch processing enzymes are utilized, addition of biocatalysts from other sources may be unnecessary (21). Hydroxyl groups at positions 2, 3, 4, and 6 have been omitted for clarity.

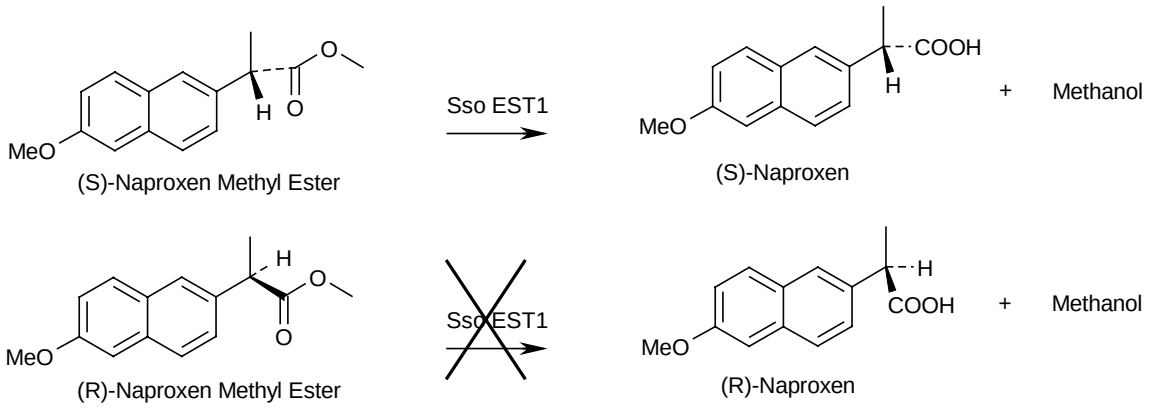


Figure 2.3 The hydrolysis of racemic Naproxen by *Sso* EST1 preferentially produces *(S)*-Naproxen at 50°C (67).

Biochemical analysis of *Thermotoga maritima* GH36 α-galactosidase (*Tm*GalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases

*Donald A. Comfort[1], Kirill S. Bobrov[2], Dina R. Ivanen[2], Konstantin A. Shabalin[2], James M. Harris[1], Anna A. Kulminskaya[2], Harry Brumer[3], and Robert M. Kelly[1]**

[1] Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

[2] Petersburg Nuclear Physics Institute, Russian Academy of Science
Molecular and Radiation Biology Division
Gatchina, St. Petersburg 188300, Russia

[3] School of Biotechnology
Royal Institute of Technology
AlbaNova University Centre, Stockholm, Sweden

Running Title: Thermotoga maritima α-galactosidase biochemical mechanism

*Address correspondence to**:** **Robert M. Kelly**
Dept. of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

Phone: (919) 515-6396
Fax: (919) 515-3465
Email: rmkelly@eos.ncsu.edu

ABSTRACT

Organization of glycoside hydrolase (GH) families into clans expands the utility of information on catalytic mechanisms of member enzymes. This issue was examined for GH27 and GH36 through biochemical analysis of GH36 α-galactosidase from *Thermotoga maritima* (*Tm*GalA). Catalytic residues in *Tm*GalA were inferred through structural homology with GH27 members to facilitate design of site-directed mutants. Conserved acidic residues were confirmed through mutations (D220G, D327G, and D387G) that exhibited reduced activity, leading to the definitive assignment of the nucleophile (D327) and acid-base residue (D387). The wild-type (WT) enzyme was 1,000-fold more active than D387G on *p*-nitrophenyl-α-D-galactopyranoside, but comparably active on 2,4-dinitrophenyl-α-D-galactopyranoside. Furthermore, the WT pH profile was bell-shaped, in contrast to D387G, which exhibited higher activity with increasing pH. The D327G mutant was 200-800 fold less active on aryl galactosides than the WT. A Brønsted plot revealed minimal influence of aglycone-leaving group on activity or catalytic efficiency for the WT, while a change in the rate-limiting step was noted for the D327G mutant. Like the WT, the pH profile of D327G was bell-shaped, although with the optimum shifted up approximately one unit. While WT and D327G mutant both melted at 89°C, the D220G and D387G mutants melted at 83°C, implying that these latter mutations altered structural features. Product analysis confirmed that the WT acted with retention of the anomeric stereochemistry analogous to GH27 enzymes. These results show that the biochemical characteristics of GH36 *Tm*GalA are closely related to GH27 enzymes, confirming the mechanistic commonality of clan GH-D members.

α-D-Galactosidases (EC 3.2.1.22) are exo-acting glycoside hydrolases that cleave α-linked galactose residues from carbohydrates commonly found in legumes and seeds, such as melibiose, raffinose, stachyose, and gluco- or galactomannans (47). α-Galactosidases and related enzymes are of interest for both physiological and biotechnological applications. In humans, Fabry's disease, a lysosomal storage problem, is an X-linked disorder that leads to a mutant α-galactosidase that forms glycosphingolipids with a terminal α-galactose sugar; the disease can be treated by α-galactosidase replacement therapy (15, 22). Similarly, deficiencies in α-*N*-acetylgalactosaminidases cause a related genetic lysosomal storage disorder, Schindler disease (54). α-Galactosidases have been used for biotechnological purposes, including: pre-treatment of animal feed to hydrolyze non-metabolizable sugars, thereby increasing the nutritive value (21), degradation of raffinose to improve the crystallization of sucrose (17), processing of soy molasses and soybean milk (60), improvement of the viscosity and gelling properties of galactomannan (14, 59), stimulation of oil/gas wells through hydrolysis of the propant matrix (46), and conversion of B-type blood antigens to produce type O blood (23, 26, 37, 40). Because of their medical and technological importance, a number of α-galactosidases from eukaryotic and microbial sources have been studied (1, 10, 11, 13, 43).

α-Galactosidases have been classified by substrate specificity (14) and by sequence similarity using hydrophobic cluster analysis (30). With respect to substrate specificity, Group I α-galactosidases hydrolyze oligosaccharides, such as melibiose, stachyose, raffinose, and verbacose; Group II α-galactosidases are active on polysaccharide substrates, such as galactomannan and glucomannan. In the carbohydrate-active enzymes family classification scheme, which groups glycoside hydrolases (GH) on the basis of amino acid sequence and structural similarities (http://afmb.cnrs-mrs.fr/CAZY/index.html (28-30)), α-galactosidases can be found in GH4, GH27, GH36, and GH57. α-Galactosidases from eukaryotic organisms are predominantly grouped into family GH27, whereas those from microbial sources are grouped into families GH4, GH36, and GH57. Of these, families GH27 and GH36 are thought to share a common

ancestral gene, thus forming a glycoside hydrolase clan (GH-D) with a common triose phosphate isomerase (TIM) barrel structure (30, 52).

Seven *apo* and ligand-complex three dimensional structures from four organisms have been determined for GH27 enzymes, including an *N*-acetyl-α-galactosaminidase from *Gallus gallus* (chicken) (20), α-galactosidase A from *Homo sapiens* (human) (19), an α-galactosidase from the fungus *Hypocrea jecorina* (née *Trichoderma reesei*) (24), and an α-galactosidase from *Oryza sativa* (rice) (16, 38). Early work on the main α-galactosidase from the white-rot fungus *Phanerochaete chrysosporium* (7) established that enzymes in GH27 act through a double-displacement mechanism involving retention of the stereochemistry of the anomeric center (Figure 3.1) (53). Specific mechanism-based inactivation, coupled with mass spectrometry, was subsequently used to identify the conserved catalytic nucleophile in this family for both the *P. chrysosporium* (25) and the *Coffea arabica* (coffee) (44) α-galactosidases. The analysis of three-dimensional enzyme-product complexes later facilitated identification of the conserved catalytic general acid/base residue (8, 34, 63).

In contrast to GH27 members, the molecular details governing catalysis by GH36 enzymes are less clear. GH36 are particularly interesting enzymes with respect to their potential for carbohydrate synthesis through protein engineering, since both hydrolytic (α-galactosidase and *N*-acetyl-α-galactosaminidase) and transglycosylation activities (raffinose and stachyose synthase) are represented in this group. Previously, global protein sequence analysis established the relationship of GH27 and GH36 in clan GH-D (29), thus implying a conserved overall $(\beta/\alpha)_8$ tertiary structure and retaining catalytic mechanism. However, sequence similarity within clan GH-D was too low to accurately predict the location of key amino acids. Structural information on the GH36 α-galactosidase from *T. maritima* (*Tm*GalA) enzyme has recently been deposited (41), facilitating efforts to compare catalytic mechanisms of GH27 and GH36. As such, site-directed mutagenesis and detailed kinetic analysis of *Tm*GalA has led to identification of the key catalytic and active site residues in GH36, suggesting that a common mechanism governs enzymes in clan GH-D.

EXPERIMENTAL PROCEDURES

Materials and General Procedures. Chemicals, growth media, and buffer reagents were obtained from Fisher Scientific (Hampton, NH) and Sigma-Aldrich (St. Louis, MO). All protein structure figures were produced using PyMol v0.98 or v0.99, available at URL http://pymol.sourceforge.net/.

Synthesis of α-Galactosides. All *o*-nitrophenyl, *p*-nitrophenyl (Gal-PNP), and 4-methylumbellyferyl glycosides were purchased from Sigma-Aldrich (St. Louis, MO). 4-Bromophenyl, 3,4-dimethylphenyl, 4-methoxyphenyl, phenyl, and *m*-nitrophenyl α-D-galactopyranosides were synthesized from galactose by acetylation, fusion of the acetates with corresponding phenol and anhydrous zinc chloride were done as described (49), and by Zemplén deacetylation. 2,4-Dinitrophenyl (Gal-DNP), 2,5-dinitrophenyl, and 2-naphtyl α-D-galactopyranosides were synthesized from tetra-*O*-acetyl α-D-galactopyranosyl chloride (33) and sodium salts of corresponding phenols in HMPA were prepared as described elsewhere (2, 12). Dinitrophenyl derivatives were deacetylated at 4°C in anhydrous methanol saturated with HCl (7). De-protection of 2-naphtyl α-D-galactoside was achieved by the Zemplén procedure. All products were obtained in crystalline form and characterized by ^{1}H NMR analysis. These data were consistent with the expected structure and literature data (when available) in each case.

Cloning, Mutagenesis, Protein Expression, and Protein Purification. Recombinant *Escherichia coli* strains were grown in Luria-Broth containing 50 mg/L kanamycin. The α-galactosidase gene (*galA*) from *T. maritima* MSB8 (ORF TM1192) was amplified from genomic DNA by PCR using primers (direct: 5'-CGCGCCATGGAAATATTCGGAAAGACCTTCAG-3'; reverse: 5'-GCGCGAATTCTACCTGAGTTCCATCATT-3'), which introduced *Nco*I and *EcoR*I restriction sites (underlined) and, in the process, mutating the first codon of the gene from GTG to ATG. The PCR fragment was inserted into pET24d (Novagen, San Diego, CA) and sequenced to verify cloning. Mutant clones were generated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) with the primer and its

complement listed in Table 3.1. Recombinant α-galactosidase (*Tm*GalA) and its derived mutants were expressed in *E. coli* BL-21(DE3) from cultures grown overnight and the proteins were purified as previously described (48). In the case of the mutants, new column matrices were used for each mutant. Protein concentrations were determined using the Biorad protein assay (Biorad, Hercules, CA) with bovine serum albumin as standard.

NMR Analysis of Reaction Products. All ^{1}H spectra were recorded with an AMX-500 Bruker spectrophotometer. Prior to analysis by NMR, *Tm*GalA wild-type (WT), D327G and D387G mutants, buffer materials, and substrates were freeze-dried twice from D_2O. The ^{1}H NMR measurements were done in 50 mM sodium acetate (pH 5.0) buffer with acetone used as an internal standard (δ ^{1}H 2.225). Product formation and anomerity of Gal-PNP hydrolysis were studied by on-line detection of the total number of the products formed using ^{1}H NMR spectroscopy. Galactose concentrations at different time intervals during the reaction were determined using specific chemical shifts: α-galactose, δ = 5.264; α-galactose, δ = 4.584. NMR analysis of azide rescue experiments for D327G and D387G mutants were carried out as follows. ^{1}H NMR spectra were collected for hydrolysis of 1 mM Gal-PNP with 0.02 U of each enzyme in 50 mM sodium phosphate pH 6.5 with 100 mM NaN_3. Spectra were recorded every 15 min until the reaction was completed.

Reaction Kinetics and Biochemical Characterization. Kinetic studies of TmGalA WT and D327G were performed on a Jasco V-560 UV-Vis spectrometer equipped with a circulating water bath using 1 cm path length quartz cuvettes. The buffer employed for all kinetic experiments with WT enzyme was 50 mM sodium acetate (pH 5.0), containing 0.5 mg/ml of BSA in case of the WT enzyme. Extinction coefficients for phenols were determined by measuring the absorbence of each compound in the same buffer at 37°C. This temperature was chosen to minimize spontaneous hydrolysis of the substrates. The wavelength monitored and molar extinction coefficients used for each substrate were as follows (Δε, mM^{-1} cm^{-1}): 2',4'-dinitrophenyl, 440 nm, 8.48; 2',5'-dinitrophenyl, 440 nm, 3.44; 4'-nitrophenyl, 400 nm, 7.61; 2'-nitrophenyl, 265 nm, 2.15; 4'-methylumbelliferyl,

355 nm, 2.87; 3'-nitrophenyl, 380 nm, 0,49; 4'-bromophenyl, 289 nm, 0.94; β-naphtyl, 330 nm, 1.47; 4-O-methylphenyl, 285 nm, 1.93; 3',4'-dimethylphenyl, 285 nm, 1.29; phenyl, 277 nm, 0.863.

Substrates hydrolysis was initiated with the addition of an appropriate aliquot of enzyme (0.47 μg of WT protein and 39 – 62 μg of D327G mutant enzyme); initial rate of phenolate release was monitored at 37°C at an appropriate wavelength relative to a reference cuvette containing no enzyme. The rate of enzyme-catalyzed hydrolysis was determined by varying substrate concentrations (typically 10-12 points) ranging from 0.5 × 5 - 10 K_m, when possible. The kinetic parameters were estimated by direct fit of the data to the Michaelis-Menten equation using the program Origin 7.5 (OriginLab Corp., Northampton, MA). Logarithmic plots of kinetic parameters (k_{cat} and k_{cat}/K_m) were fitted by linear regression with leaving group pK_a values. The Brønsted coefficients, β_{lg}, were calculated from the slopes of these plots.

Azide rescue reactions of D327G were performed in 50 mM sodium acetate buffer with 0 to 400 mM sodium azide, measuring hydrolysis of 2.5 mM Gal-PNP at 37°C for 20 minutes. Reactions were stopped by the addition of 10% sodium carbonate. Addition of azide resulted in an increase of pH from 5.0 to 6.2 for 400 mM NaN_3. Hydrolysis was monitored using the Jasco spectrometer as described above.

Kinetic studies on the D387G mutant were conducted at 37°C using a Perkin Elmer Lambda Bio 20 spectrophotometer (Wellesley, MA), heated by Perkin Elmer PTP-1 Peltier System. V_0 was determined from initial slope and Michaelis-Menten kinetic parameters were calculated by least-squares regression in Excel (Microsoft, Redmond, WA). Substrates used for D387G were Gal-DNP and Gal-PNP. For Gal-PNP hydrolysis, the activity assay was modified to increase sensitivity by changing to an end point assay, which was stopped after 5 minutes by addition of 1 M Na_2CO_3. This changed the extinction coefficient to 18 $mM^{-1}*cm^{-1}$. Experiments were run in 200 mM sodium citrate buffer (pH 5.0). The influence of external nucleophiles on WT and D387G was studied using Gal-PNP and Gal-DNP at concentrations from 12 μM to 2 mM. Azide rescue was determined in the presence of 2 M sodium azide with 200 mM sodium citrate, pH 5.0 (after addition of azide). Formate rescue was conducted in 2 M sodium formate buffer (pH 5.0).

The pH dependence of enzymatic hydrolysis was investigated by determining k_{cat} and K_m values for a series of pH values between 2.5 and 8.5, continuously monitoring the hydrolysis of Gal-PNP at 37°C at 400 nm. Only pHs for which the enzyme was stable for at least 25 min were used. All buffers (citrate-phosphate buffer solutions (pH 2.8 – 7.5); glycine buffers (pH 8.0 – 10.5)) were at a final concentration 50 mM and contained 0.05 M NaCl. Reaction mixtures were pre-warmed until the reaction was initiated by the addition of 0.4 μg of WT and approximately 60 μg of D327G enzyme. k_{cat}/K_m values were then plotted against pH and fitted to the appropriate curve in Origin 7.5, yielding apparent pK_a of the ionazable groups. Reaction volumes were 600 μL containing 147 μg D387G or 0.93 μg WT.

Differential Scanning Microcalorimetry. Purified proteins were dialyzed overnight against 2 L of filtered 50 mM NaP, pH 7.2 with 150 mM NaCl. Samples of 1.5 to 3.0 mg/ml of protein were run on an N-DSC (Calorimetry Sciences Corp., Lindon, UT) from 25 to 125°C at 0.5°C/sec. Results were analyzed using CpCalc 2.0 (Calorimetry Sciences Corp.), and the maximum excess heat capacity used as the melting temperature of the proteins.

RESULTS

Substrate Specificity of TmGalA on Glycosides. A series of 4-nitrophenyl glycosides and natural galactosylated substrates were used to determine the substrate specificity. Using accepted nomenclature (9), the glycone occupies the -1 sub-site in the enzymatic active-site pocket, while the departing group occupies the +1 sub-site. It appears that *Tm*GalA prefers substrates with galactopyranoside as a glycone, as there was no detectable enzymatic activity on xylo-, fuco-, manno-, or glucopyranosides (data not shown).

Site-Directed Mutagenesis of TmGalA Active Site Residues. The primary sequence of *Tm*GalA is shown in Figure 3.2. Conserved acidic amino acid residues in GH36 were identified by sequence alignments with other GH36 α-galactosidases (61), and alanine mutants were generated for the conserved residues D103, D220, D221, D229, D275,

D299, D327, and D387. Three residues, D220, D327, and D387, exhibited < 1% of the wild-type (WT) enzyme activity on Gal-PNP in sodium acetate at 80°C, implicating their role in catalysis. These residues were mutated to glycine.

Stereochemical Course of the Hydrolytic Reaction. ^{1}H NMR provides a direct method for determining a stereochemical outcome of the glycosidase hydrolysis. Chemical shifts and coupling constants of the anomeric protons in α- and β-glycosides, as well as in the product hemiacetals, are distinct and readily observed. Figure 3.3 shows the kinetics of the formation of α- and β-anomers of galactose in the hydrolysis of Gal-PNP. The α-anomer initially formed in the presence of the WT enzyme and accumulated in amounts sufficient for detection, followed by a noticeable amount of β-anomer which appeared due to mutarotation. Mutarotation proceeded until the normal anomer ratio of 30% α- and 70% β-galactopyranose was established (60 min after the reaction was stopped). The data unequivocally demonstrate that enzymatic hydrolysis proceeds with the retention of the anomeric configuration, presumably as a result of a double displacement reaction (57). The stereochemical outcome of the hydrolysis of Gal-PNP catalyzed by D387G was monitored by NMR. An increase of the formation rate of the α-D-galactose over the β-anomer arising due to the mutarotation was observed indicating that D387G functions as a retaining glycoside hydrolase similar to the WT.

Analysis of Structural Features of TmGalA with respect to Biocatalytic Mechanism. The tertiary structure of *Tm*GalA consists of three domains (Figure 3.4): an N-terminal β-super-sandwich domain (residues 1-179, see Figure 3.2 for primary sequence), a core $(\beta/\alpha)_8$ barrel (residues 180-482), and a C-terminal anti-parallel β-sheet domain (residues 483-525) (41). As expected, based on common clanship in GH-D (29)(URL: http://www.cazy.org/CAZY/), *Tm*GalA from GH36 shares a homologous $(\beta/\alpha)_8$ barrel catalytic domain with all the known three-dimensional structures of GH27 enzymes (Figure 3.5). DaliLite (31) was used for pair-wise comparison of *Tm*GalA with representative structures from *Homo sapiens* α-galactosidase A (PDB 1r46) (19), *Hypocrea jecorina* (née *Trichoderma reesei*) α-galactosidase (PDB 1SZN) (24), *Oryza sativa* α-galactosidase (PDB 1UAS) (16), and *Gallus gallus N*-acetyl-α-galatosaminidase (PDB 1KTB) (20). As shown in Table 3.2, comparison of the complete structures

indicated modest structural overlap, as judged by the RMSD values of the C^{α} atoms, despite very low sequence similarity. Both the RMSD values and Z-scores were slightly improved when only the corresponding $(\beta/\alpha)_8$ barrel domains were compared, which indicates minimal structural homology of the C-terminal β-sheet domain observed in GH27 enzymes. Indeed, there is apparent lack of overlap of the C^{α} trace of *Tm*GalA with that of the representative *O. sativa* α-galactosidase (Figure 3.5). Furthermore, nearly 30 residues at the C-terminus of *Tm*GalA were not observed crystallographically. No binding nor catalytic function has been previously ascribed to this domain in GH27.

The three-dimensional structures of the GH27 enzymes have been rigorously compared and show little variation with regard to tertiary structure (16, 19, 20, 24). The $(\beta/\alpha)_8$ and C-terminal domains are clearly superimposible (18), while only the *T. reesei* α-galactosidase (PDB 1SZN) is peculiar with regard to two loop extensions on the $(\beta/\alpha)_8$ barrel (24). Due to the high structural similarity within GH27, we selected the *O. sativa* α-galactosidase (1UAS) as an illustrative structure for further comparison with *Tm*GalA of GH36 (Figure 3.5). Manual analysis of the superimposed *O. sativa* and *T. maritima* enzymes was used to derive the primary sequence alignment shown in Figure 3.2.

The N-terminal β-super-sandwich has no homologous structure in family GH27 enzymes of clan GH-D (*cf.* Figures 3.2 and 3.5), but bears highest structural similarity to chain D of bovine lysosomal α-mannosidase (bLAM, PDB 1O7D) from GH38 (27). Analysis using Dali v. 2.0 (32), via the Dali server (URL: www.ebi.ac.uk/dali/), indicated that residues 1-179 of *Tm*GalA (PDB 1ZY9) align with PDB 1O7D:chain D with a Z-value of 8.0, an RMSD value of 3.7 Å, and a sequence identity of 11% over 138 of 269 residues. The Secondary Structure Matching (SSM) server (36) (URL: www.ebi.ac.uk/msd-srv/ssm/) similarly returned PDB 1O7D:chain D as the top hit, with an RMSD value of 2.8 Å for 132 aligned residues with 11 aligned secondary structural elements, 14 gaps, and 11% sequence identity. A superposition of the N-terminal domain of *Tm*GalA with the D peptide of bLAM based on the SSM output is shown in Figure 3.6. Interestingly, this domain contributes a key substrate-binding residue (W65) to the active site of *Tm*GalA, which replaces the homologous W164 (Figures 3.2 and 3.7).

Indeed, a high degree of homology was observed between the active sites of *Tm*GalA and the *O. sativa* α-galactosidase (Figure 3.7). Galactose, observed as an equilibrium mixture of α- and β-anomers in the *O. sativa* structure, provides a convenient point of reference for the identification of the corresponding catalytic and substrate-binding residues in the active site of *Tm*GalA, which are summarized in Figure 3.2. Many active site residues are identical between the two enzymes, or are conservative substitutions, while in some cases more radical replacements are observed. As mentioned above, *Tm*W65, which is found on the loop joining β7 and β8 of the N-terminal β-super-sandwich replaces *Os*W164 from loop joining β5 and α5 of the $(\beta/\alpha)_8$ barrel. Likewise, the side chains of F328 and W291 in *Tm*GalA approximately assume the position occupied by the S102 and C101-C132 cystine residue in the *O. sativa* enzyme. The *Tm*W191–*Os*D17 replacement is similarly non-conservative, but is compensated for, in part, by the *Tm*C428–*Os*M217 substitution. The conserved catalytic nucleophile and general acid/base residues in GH27, represented by D130 and D185 in the *O. sativa* α-galactosidase are clearly observed to have structural homologs in *Tm*GalA, namely D327 and D387, respectively.

Thermal Denaturation of WT and TmGalA Mutants. The melting behavior of the WT and mutants revealed that D327G had a nearly identical T_{melt} to the WT (89°C), whereas as D220G and D387G melted close to 83°C (Table 3.3). Despite the role of D387 in catalysis, mutation of this amino acid adversely affected the stability of the enzyme. D220 is positioned near the active-site pocket and the mutation affected the enzymatic activity of the enzyme as well as the stability of the protein, possibly due to elimination of a salt bridge with K325.

pH Dependence of Glycoside Hydrolysis. WT hydrolysis of Gal-PNP followed a standard bell-shaped, pH dependence, as expected for glycosidases, presumed to result from the ionization of the two active-site carboxylic acids (5) (Figure 3.8 a,b). D327G was similar to the WT, with a slight increase in pH optimum for k_{cat} (Figure 3.8 a). D387G, however, showed increasing activity with increasing pH (Figure 3.8 a,b). This has also been shown for the β-xylosidase from *Bacillus stearothermophilus* (5) and the α-arabinosidase from *Geobacills stearothermophilus* (55) acid-base mutants. WT *Tm*GalA

showed a typical bell-shaped pH curve, as expected and seen for other wild-type glycosidases. In the case of the D387G, at high pH values there is no reduction in activity likely because the acid/base residue is not present. D220G exhibited a standard bell-shaped pH profile with a pH optimum about 5.0 (results not shown), but with reduced activity relative to the WT protein.

Catalytic Properties of D327G and D387G Mutants. A series of aryl α-galactosides were used to examine the relationship between the kinetic parameters of WT and D327G. For the WT, Brønsted plots and coefficients (see Figure 3.9) indicated that there was a linear dependence on aryl α-D-galactopyranosides with respect to leaving group ($\beta_{lg} = -0.05$ for both k_{cat} and k_{cat}/K_M). D327G, on the other hand, had a linear dependence (Brønsted coefficient $\beta_{lg} = -0.3$ for both k_{cat} and k_{cat}/K_M) only at higher pK_a. The catalytic rate of the WT relative to D327G was 200-800-fold higher, while the K_M values of D327G ranged from 0.17 lower to nearly the same. This resulted in a lower (100-1700 fold) catalytic efficiency for the mutant relative to the WT. The reaction rate for WT and D387G was compared for galactoside substrates with different leaving groups, Gal-PNP (poor leaving group) and Gal-DNP (good leaving group). The WT had approximately a 1500-fold higher k_{cat} on the Gal-PNP substrate than did D387G. However, with Gal-DNP as the substrate, the k_{cat} ratio was comparable for D387G and WT (Table 3.4).

External Nucleophile Rescue. External nucleophile rescue was attempted using azide for D327G and both azide and formate for D387G. As expected, a noticeable restoration of hydrolytic activity for the nucleophile mutant (D327G) was observed when azide was added into the reaction with Gal-PNP. The increase in rate was about 30-fold but clearly dependent on exogenous nucleophile concentration (Figure 3.11). It is seen that even small concentration of azide (40 mM) restored the activity of this mutant up to 10-fold, representing a typical behavior for nucleophlie-less mutant of retaining glycosidases. The presence of azide and formate had minimal effect on hydrolysis rate or catalytic efficiency with Gal-PNP for either the WT or D387G (Table 3.4). There was also no increase in the catalytic rate or efficiency for the WT in the presence of either

azide or formate with Gal-DNP. There was, however, a 6-fold increase in the reaction rate with Gal-DNP in the presence of both azide and formate compared to citrate buffer.

The influence of NaN_3 on the hydrolysis products of Gal-PNP with D387G was examined by NMR. During hydrolysis in the presence of 100 mM NaN_3, an additional signal with a chemical shift at $\delta = 5.554$ ppm appeared as well as the expected signals of α- and β-anomers of D-galactose. After 60 min, the concentration of a new product was estimated to be 0.57 mM (57% of all products obtained). The assignment of chemical shifts for this compound showed it to be N_3αGal*p*: δ ^{1}H: 5.554 (1H, d, $J_{1,2}$ 4.63 Hz, H-1), 4.048 (1H, ddd, $J_{5,6a}$ 7.25 Hz, $J_{5,6b}$ 4.94 Hz, $H_{5,4}$ 1.31 Hz, H-5), 3.989 (1H, dd, $J_{4,3}$ 3.31, H-4), 3.933 (1H, dd, $J_{2,3}$ 10.2 Hz, H-2), 3.771 (1H, dd, $J_{6a,6b}$ 11.8 Hz, H-6a), 3.742 (1H, dd, H-3), 3.7418 (1H, dd, H-6b).

DISCUSSION

Several common strategies were used to confirm that GH36 *Tm*GalA, like clan-D members GH27 (7), uses a double-displacement retaining mechanism (57), which involves two consecutive steps (glycosylation and deglycosylation) proceeding via two transition states surrounding the formation of a covalently-linked glycosyl-enzyme intermediate (57, 62). As a result of nucleophilic substitution in each step, the C1-atom of the sugar ring is inverted twice, resulting in an overall net retention of the initial stereochemistry. Furthermore, primary amino acid sequence analysis of GH36 enzymes with respect to those in GH27 revealed a number of conserved acidic amino acid residues. Three mutant versions of *Tm*GalA, D220G, D327G, and D387G, were shown to have less than 1% residual activity on Gal-PNP, consistent with recent reports that residues corresponding to D327 and D387 were the nucleophile and acid-base residues on a α-galactosidase from the hyperthermophile *Sulfolobus solfataricus* (6).

Crystallographic data available for glycoside hydrolase clan-D have only been available for GH27 enzymes from *H. sapiens, G. gallus, O. sativa,* and *H. jerocina*. These data have been used to investigate the structural features of lysosomal storage disorders, such as Fabry and Schindler diseases, and show highly conserved catalytic pockets within GH27 (18). Structural alignment of *Tm*GalA (1ZY9) confirmed the

structural and catalytic similarities GH27 and GH36 (Figures 3.2 and 3.5). Superposition of *Tm*GalA catalytic pocket with *O. sativa* α-galactosidase indicated that the *Tm*GalA nucleophile to be D327 and the acid/base residue to be D387, based on alignment with *O. sativa* residues D130 and D185, respectively, as shown in Figure 3.7. This result points to the shared common ancestry of GH27 and GH 36 families (52). These residues are separated by approximately 6.3 Å, rather than the 5.5 Å generally found for retaining enzymes (53), which is consistent with an average separation for GH27 catalytic residues of 6.5 Å (18). As indicated by structural alignment, the N-terminal domain of *Tm*GalA (residues 1-179) has no homologous structure in GH27 enzymes, as exemplified by the *O. sativa* α-galactosidase (Figure 3.5). This domain, which is found in longer variants in a number of other GH36 α-galactosidases, makes an important contribution to the enzyme active site located within the $(\beta/\alpha)_8$ barrel by providing the key substrate-binding residue W65. Deletion of this domain would likely produce an incorrectly folded and non-functional enzyme variant.

To characterize the *Tm*GalA mutants in comparison with the WT, two kinetic parameters, catalytic efficiency (k_{cat}/K_M) and k_{cat} for hydrolysis of Gal-PNP, were measured at different pH values. Dependence of both parameters on pH produced bell-shaped curves that reflect ionization in the free enzyme and the free substrate (k_{cat}/K_M *vs* pH) and enzyme/substrate complex (k_{cat} *vs* pH). In the case of k_{cat}, these curves illustrate the pH-dependence of the rate-limiting step. For the WT and D327G, this rate-limited stage is de-glycosylation. Mutating D327G may result in a concomitant change in the ionization state of the acid/base residue, resulting from the shift in pH optimum and a hindrance for both stages of the reaction. The pH-dependence of catalytic efficiency (k_{cat}/K_m) changed more drastically than may be explained by the decrease in substrate affinity at low pHs due to mutation. In the case of D387G, the activity does not decrease at increasing pH. In this case, the pK_a of the remaining catalytic residue was raised and the nucleophilic residue was more effectively de-protonated at higher pH, hence the reaction rate increase (4, 42). This characteristic response for an acid-base mutant enzyme is observed for *Tm*GalA D387G.

Brønsted analysis of glycosidase-catalyzed hydrolysis of aryl glycosides has been shown to be a valuable tool for distinguishing between rate-limiting steps for a particular

substrate. In retaining glycosidases, the glycon group is cleaved during the first step of the reaction producing a covalently-bound intermediate, which reacts with an exogenous neutral nucleophile. Kinetic constants characteristic of the proposed mechanistic pathway for a retaining glycoside hydrolase (Scheme 1) are designated as follows: $k_{cat} = k_2k_3/(k_2 + k_3)$; $K_m = (k_{-1} + k_2)k_3/(k_2 + k_3)k_1$; $k_{cat}/K_m = k_1k_2/(k_{-1} + k_2)$.

$$E + S \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} ES \xrightarrow{k_2} E\text{-}S^* \xrightarrow{k_3} E + P_2$$
$$\downarrow$$
$$P_1$$

Scheme 1

The first step (glycosylation) will depend on leaving group ability of the aglycon. A strong correlation between enzyme activity (k_{cat}) and the pK_a of a departing aglycon indicates that glycosylation is a rate-limiting step for a substrate with a 'good' leaving group (pKa < 8). In this case $k_2 >> k_3$ and, therefore, $k_{cat} = k_3$. Since the parameter k_{cat}/K_m is the apparent second-order rate constant for the first irreversible step of the catalytic reaction, then the break in the dependence should arise when the substrate binding reaches a diffusion control regime (10^8 s^{-1} M^{-1}).

The steady-state parameters for k_{cat}, K_m, and k_{cat}/K_m for the WT and D327G were determined and used to generate Brønsted plots (logarithms of k_{cat} and catalytic efficiencies versus the leaving group pK_as, a shown in Figures 3.9 and 3.10). The linear relationship for the k_{cat} versus pK_a shows that there is no change in the rate-limiting step for WT-catalyzed hydrolysis of aryl α-galactopyranosides. The WT exhibited small slopes (β_{lg} nearly -0.05 at both plots, Fig. 3.9 A and B), suggesting that the de-glycosylation of the glycosyl-enzyme intermediate (k_3) is the rate-limiting step for all substrates tested. A similar effect has been observed for retaining α-glycosidases, such as α-*L*-arabinofuranosidase from *Geobacillus stearothermophilus* (*56*), α-mannosidase from *Drosophila melanogaster* (51), α-*L*-fucosidase from *T. maritima* (58), and α-1,4-glucan lyase of Gracilariopsis (39). In contrast, retaining β-glycosidases exhibit a non-linear decrease in kinetic parameters with pK_a (3, 35, 45, 50). The replacement of the nucleophile, residue D327, led to a decrease rate in the glycosylation step for less reactive substrates (pK_a > 8). A catalytic efficiency value (k_{cat}/K_M) of nearly 10^5 s^{-1} M^{-1} for the

WT, and no decrease in kinetic parameters with pK_a, suggests a rate limitation related to k_1. Since the k_{cat}/K_M indicates no diffusion-control, there may be an additional reversible step, which slows overall hydrolysis. Melting behavior of D327G and WT was nearly identical melting temperatures, indicating that D327 is not involved in stabilization of the protein, as would be expected for the nucleophilic residue. Rescue of D327G with azide produced an increased hydrolysis rate that was about 30-fold higher than buffer and clearly dependent on exogenous nucleophile concentration (Figure 3.11). It was shown that even small concentrations of azide (40 mM) restored the activity of this mutant up to 10-fold, representing a typical behavior for nucleophile mutants of retaining glycosidases.

The acid-base reside performs a dual role in catalysis, acting as a general acid proton donor during the glycosylation step, and then as the general base during de-glycosylation (5). For substrates with poor leaving groups (high pK_a), the acid-base residue affects the rate of catalysis more so than for substrates with good leaving groups. When the acid-base residue is mutated rendering the enzyme inactive, it affects both the glycosylation and de-glycosylation steps of hydrolysis. In the glycosylation step, there is generally a significant decrease the rate of hydrolysis for substrates with both good and poor aglycones. However, substrates with good leaving groups require less assistance from the acid-base catalyst and, therefore, retain a greater hydrolysis rate relative to the WT enzyme. Glycosylation represents the first non-reversible step in glycoside hydrolysis, and defines the k_{cat}. Thus, the rate-limiting step can be determined by comparing k_{cat} values for substrates with different leaving groups. Increasing the k_{cat} can sometimes change the rate-limiting step from glycosylation to de-glycosylation, which is characterized by a decrease in K_M as the glycosyl enzyme intermediate accumulates. For WT *Tm*GalA, the k_{cat} values are similar, regardless of the substrate (Table 3.4). However, D387G had a 240-fold increase in activity towards Gal-DNP compared to Gal-PNP, indicating that the rate-limiting step changes from glycosylation for Gal-PNP to de-glycosylation for Gal-DNP (4). This is confirmed by the apparent reduction of K_M on Gal-DNP, which holds true for the substrate even in the presence of external nucleophiles.

Restoration of activity using external nucleophiles (azide and formate) was attempted on both WT and D387G. The WT did not show any increase in activity in the

presence of these nucleophiles, on either Gal-PNP or Gal-DNP. No restoration of activity was detected for the D387G mutant enzyme in the presence of azide or formate when Gal-PNP was the substrate. This is consistent with the *S. solfataricus* α-galactosidase, where it was shown that there was no restoration of activity by azide (6). On the Gal-DNP substrate, the D387G mutant in the presence of azide was 7-fold more active. This increase in activity is probably the result of azide aiding the de-glycosylation step, as indicated by the increase in K_M from less than 0.012 mM to 0.025 mM. Formate had a similar effect on the increase in activity.

^{1}H NMR analysis of D387G-catalyzed hydrolysis of Gal-PNP displayed an overall net-retention of the anomeric configuration of C1-atom at the glycon moiety of the substrate. The effect of an exogenous nucleophile (azide) on the hydrolysis of Gal-PNP for D387G-mutant was analyzed by NMR and did not reveal a significant restoration of mutant hydrolytic activity. However, a new product with α-configuration, N_3-α-Gal*p*, was observed and should should have been formed only at the de-glycosylation stage of the process. Formation of N_3-α-Gal*p* in the case of D387G-mediated reaction, and the absence of such a result with the WT, is likely due to weakening of the base catalyst in the D387G mutant, suggesting that D387 is an acid/base catalytic amino acid residue. These data, combined with the pH effect and structural alignment and analysis, conclusively indicate that D387 is the acid/base residue for *Tm*GalA.

This study clearly highlights the common active site structure and catalytic mechanism shared between GH27 and GH36 enzymes in Clan GH-D. It is anticipated that this contribution to understanding protein structure-function relationships in this Clan will serve to direct future efforts focusing on catalytically engineering these enzymes for biotechnological purposes.

ACKNOWLEDGMENTS

DAC and RMK acknowledge support from the Biotechnology Program of the US National Science Foundation and Energy Biosciences Program of the US Department of Energy. AAK and co-workers acknowledge the Program for Basic Research in Molecular

and Cell Biology from the Presidium of Russian Academy of Sciences (PRAS). HB thanks the Swedish Science Council (Vetenskapsrådet) for a Special Council Research Position (Rådsforskare), and the Swedish Foundation for Strategic Research (SSF) for partial funding of this research.

REFERENCES

1. **Ademark, P., R. P. de Vries, P. Hagglund, H. Stalbrand, and J. Visser.** 2001. Cloning and characterization of *Aspergillus niger* genes encoding an alpha-galactosidase and a beta-mannosidase involved in galactomannan degradation. Eur. J. Biochem. **268:**2982-2990.
2. **Apparu, M., M. Blancmuesser, J. Defaye, and H. Driguez.** 1981. Stereoselective syntheses of *O*-nitrophenyl and *S*-nitrophenyl glycosides. 3. Syntheses in the alpha-D-galactopyranose and alpha-maltose series. Can. J. Chem. **59:**314-320.
3. **Bauer, M. W., and R. M. Kelly.** 1998. The family 1 beta-glucosidases from *Pyrococcus furiosus* and *Agrobacterium faecalis* share a common catalytic mechanism. Biochemistry **37:**17170-17178.
4. **Bravman, T., V. Belakhov, D. Solomon, G. Shoham, B. Henrissat, T. Baasov, and Y. Shoham.** 2003. Identification of the catalytic residues in family 52 glycoside hydrolase, a beta-xylosidase from *Geobacillus stearothermophilus* T-6. J. Biol. Chem. **278:**26742-26749.
5. **Bravman, T., A. Mechaly, S. Shulami, V. Belakhov, T. Baasov, G. Shoham, and Y. Shoham.** 2001. Glutamic acid 160 is the acid-base catalyst of beta-xylosidase from *Bacillus stearothermophilus* T-6: a family 39 glycoside hydrolase. FEBS Lett. **495:**115-119.
6. **Brouns, J. J. S., N. Smits, H. Wu, A. P. L. Snijders, P. C. Wright, W. M. de Vos, and J. van der Oost.** 2006. Identification of a novel alpha-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. J. Bacteriol. **188:**2392-2399.
7. **Brumer, H., P. F. G. Sims, and M. L. Sinnott.** 1999. Lignocellulose degradation by *Phanerochaete chrysosporium*: purification and characterization of the main alpha-galactosidase. Biochem. J. **339:**43-53.
8. **Brux, C., A. Ben-David, D. Shallom-Shezifi, M. Leon, K. Niefind, G. Shoham, Y. Shoham, and D. Schomburg.** 2006. The structure of an inverting GH43 beta-xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. J. Mol. Biol. **359:**97-109.
9. **Davies, G. J., K. S. Wilson, and B. Henrissat.** 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. **321:**557-559.
10. **Dean, K. J., and C. C. Sweeley.** 1979. Studies on human-liver alpha-galactosidases.1. Purification of alpha-galactosidase-A and its enzymatic properties with glycolipid and oligosaccharide substrates. J. Biol. Chem. **254:**9994.
11. **Dean, K. J., and C. C. Sweeley.** 1979. Studies on human-liver alpha-galactosidases. 2. Purification and enzymatic properties of alpha-galactosidase B (alpha-N-acetylgalactosaminidase). J. Biol. Chem. **254:**1-5.
12. **Delmotte, F. M., J. Privat, and M. L. P. Monsigny.** 1975. Glycan-protein interactions - Synthesis of 4-methylumbelliferyl-2-acetamido-2-deoxy-beta-D-glucopyranoside, 4-methylumbelliferyl-Di-N-acetyl-beta-chitobioside and 4-methyl-tri-N-acetyl-beta-chitotrioside - Interaction of these osides with lysozyme. Carbohyd. Res. **40:**353-364.

13. **Dey, P. M., E. M. Delcampillo, and R. P. Lezica.** 1983. Characterization of a glycoprotein alpha-galactosidase from lentil seeds (*Lens culinaris*). J. Biol. Chem. **258:**923-929.
14. **Dey, P. M., S. Patel, and M. D. Brownleader.** 1993. Induction of alpha-galactosidase in *Penicillium ochrochloron* by guar (*Cyamopsis tetragonobola*) gum. Biotechnol. Appl. Bioc. **17:**361-371.
15. **Eng, C. M., N. Guffon, W. R. Wilcox, D. P. Germain, P. Lee, S. Waldek, L. Caplan, G. E. Linthorst, and R. J. Desnick.** 2001. Safety and efficacy of recombinant human alpha-galactosidase a replacement therapy in Fabry's disease. New Engl. J. Med. **345:**9-16.
16. **Fujimoto, Z., S. Kaneko, M. Momma, H. Kobayashi, and H. Mizuno.** 2003. Crystal structure of rice alpha-galactosidase complexed with D-galactose. J. Biol. Chem. **278:**20313-20318.
17. **Ganter, C., A. Bock, P. Buckel, and R. Mattes.** 1988. Production of thermostable, recombinant alpha-galactosidase suitable for raffinose elimination from sugar-beet syrup. J. Biotechnol. **8:**301-310.
18. **Garman, S. C.** 2006. Structural studies on alpha-GAL and alpha-NAGAL: The atomic basis of Fabry and Schindler diseases. Biocat. Biotransfor. **24:**129-136.
19. **Garman, S. C., and D. N. Garboczi.** 2004. The molecular defect leading to Fabry disease: Structure of human alpha-galactosidase. J. Mol. Biol. **337:**319-335.
20. **Garman, S. C., L. Hannick, A. Zhu, and D. N. Garboczi.** 2002. The 1.9 angstrom structure of alpha-N-acetylgalactosaminidase: Molecular basis of glycosidase deficiency diseases. Structure **10:**425-434.
21. **Ghazi, S., J. A. Rooke, and H. Galbraith.** 2003. Improvement of the nutritive value of soybean meal by protease and alpha-galactosidase treatment in broiler cockerels and broiler chicks. Brit. Poultry Sci. **44:**410-418.
22. **Gieselmann, V.** 1995. Lysosomal storage diseases. BBA-Mol. Basis Dis. **1270:**103-136.
23. **Goldstein, J., G. Siviglia, R. Hurst, L. Lenny, and L. Reich.** 1982. Group-B erythrocytes enzymatically converted to group-O survive normally in A, B, and O individuals. Science **215:**168-170.
24. **Golubev, A. M., R. A. P. Nagem, J. R. B. Neto, K. N. Neustroev, E. V. Eneyskaya, A. A. Kulminskaya, K. A. Shabalin, A. N. Savel'ev, and I. Polikarpov.** 2004. Crystal structure of alpha-galactosidase from *Trichoderma reesei* and its complex with galactose: Implications for catalytic mechanism. J. Mol. Biol. **339:**413-422.
25. **Hart, D. O., S. M. He, C. J. Chany, S. G. Withers, P. F. G. Sims, M. L. Sinnott, and H. Brumer.** 2000. Identification of Asp-130 as the catalytic nucleophile in the main alpha-galactosidase from *Phanerochaete chrysosporium*, a family 27 glycosyl hydrolase. Biochemistry-US **39:**9826-9836.
26. **Hata, D. J., and D. S. Smith.** 2004. Blood group B degrading activity of *Ruminococcus gnavus* alpha-galactosidase. Artif. Cell. Blood Sub. **32:**263-274.
27. **Heikinheimo, P., R. Helland, H. K. S. Leiros, I. Leiros, S. Karlsen, G. Evjen, R. Ravelli, G. Schoehn, R. Ruigrok, O. K. Tollersrud, S. McSweeney, and E. Hough.** 2003. The structure of bovine lysosomal alpha-mannosidase suggests a novel mechanism for low-pH activation. J. Mol. Biol. **327:**631-644.

28. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochem. J. **280:**309-316.
29. **Henrissat, B., and A. Bairoch.** 1993. New families in the classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochem. J. **293:**781-788.
30. **Henrissat, B., and A. Bairoch.** 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. **316:**695-696.
31. **Holm, L., and J. Park.** 2000. DaliLite workbench for protein structure comparison. Bioinformatics **16:**566-567.
32. **Holm, L., and C. Sander.** 1993. Protein-structure comparison by alignment of distance matrices. J. Mol. Biol. **233:**123-138.
33. **Ibatullin, F. M., and S. I. Selivanov.** 2002. Reaction of 1,2-trans-glycosyl acetates with phosphorus pentachloride: new efficient approach to 1,2-trans-glycosyl chlorides. Tetrahedron Lett. **43:**9577-9580.
34. **Jensen, M. H., O. Mirza, C. Albenne, M. Remaud-Simeon, P. Monsan, M. Gajhede, and L. K. Skov.** 2004. Crystal structure of the covalent intermediate of amylosucrase from *Neisseria polysaccharea*. Biochemistry **43:**3104-3110.
35. **Kempton, J. B., and S. G. Withers.** 1992. Mechanism of *Agrobacterium* beta-glucosidase - Kinetic studies. Biochemistry **31:**9961-9969.
36. **Krissinel, E., and K. Henrick.** 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D **60:**2256-2268.
37. **Kruskall, M. S., J. P. AuBuchon, K. Y. Anthony, L. Herschel, C. Pickard, R. Biehl, M. Horowitz, D. J. Brambilla, and M. A. Popovsky.** 2000. Transfusion to blood group A and O patients of group B RBCs that have been enzymatically converted to group O. Transfusion **40:**1290-1298.
38. **Kuhls, K., E. Lieckfeldt, G. J. Samuels, W. Kovacs, W. Meyer, O. Petrini, W. Gams, T. Borner, and C. P. Kubicek.** 1996. Molecular evidence that the asexual industrial fungus Trichoderma reesei is a clonal derivative of the ascomycete Hypocrea jecorina. Proc Natl Acad Sci U S A **93:**7755-60.
39. **Lee, S. S., S. Yu, and S. G. Withers.** 2003. Detailed dissection of a new mechanism for glycoside cleavage: alpha-1,4-glucan lyase. Biochemistry **42:**13081-13090.
40. **Lenny, L. L., R. Hurst, J. Goldstein, and R. A. Galbraith.** 1994. Transfusions to group-O subjects of 2 units of red-cells enzymatically converted from group-B to group-O. Transfusion **34:**209-214.
41. **Lesley, S. A., P. Kuhn, A. Godzik, A. M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H. E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L. S. Brinen, M. D. Miller, T. M. McPhillips, M. A. Miller, D. Scheibe, J. M. Canaves, C. Guda, L. Jaroszewski, T. L. Selby, M. A. Elsliger, J. Wooley, S. S. Taylor, K. O. Hodgson, I. A. Wilson, P. G. Schultz, and R. C. Stevens.** 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. P. Natl. Acad. Sci. USA **99:**11664-11669.

42. **Li, Y. K., J. Chir, S. Tanaka, and F. Y. Chen.** 2002. Identification of the general acid/base catalyst of a family 3 beta-glucosidase from *Flavobacterium meningosepticum*. Biochemistry-US **41:**2751-2759.
43. **Liebl, W., B. Wagner, and J. Schellhase.** 1998. Properties of an alpha-galactosidase, and structure of its gene galA, within an alpha- and beta-galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. Syst. Appl. Microbiol. **21:**1-11.
44. **Ly, H. D., S. Howard, K. Shum, S. M. He, A. Zhu, and S. G. Withers.** 2000. The synthesis, testing and use of 5-fluoro-alpha-D-galactosyl fluoride to trap an intermediate on green coffee bean alpha-galactosidase and identify the catalytic nucleophile. Carbohyd. Res. **329:**539-547.
45. **Malet, C., and A. Planas.** 1997. Mechanism of Bacillus 1,3-1,4-beta-D-glucan 4-glucanohydrolases: Kinetics and pH studies with 4-methylumbelliferyl beta-D-glucan oligosaccharides. Biochemistry **36:**13838-13848.
46. **McCutchen, C. M., G. D. Duffaud, P. Leduc, A. R. H. Petersen, A. Tayal, S. A. Khan, and R. M. Kelly.** 1996. Characterization of extremely thermostable enzymatic breakers (alpha-1,6-galactosidase and beta-1,4-mannanase) from the hyperthermophilic bacterium *Thermotoga neapolitana* 5068 for hydrolysis of guar gum. Biotechnol. Bioeng. **52:**332-339.
47. **Meier, H., and J. S. G. Reid.** 1982. Reserve polysaccharides other than starch in higher plants, p. 418-71. *In* F. A. Loewus and W. Tanner (ed.), Encyclopedia of Plant Physiology, vol. 13A. Springer Verlag.
48. **Miller, E. S., K. N. Parker, W. Liebl, D. Lam, W. Callen, M. A. Snead, E. J. Mathur, J. M. Short, and R. M. Kelly.** 2001. alpha-D-galactosidases from *Thermotoga* species, p. 246-260, Hyperthermophilic Enzymes, Pt A, vol. 330.
49. **Montgomery, E. M., N. K. Richtmyer, and C. S. Hudson.** 1942. The preparation and rearrangement of phenylglycosides. J. Am. Chem. Soc. **64:**690-694.
50. **Namchuk, M., C. Braun, J. D. McCarter, and S. G. Withers.** 1996. Fluorinated sugars as probes of glycosidase mechanisms, p. 279-293, ACS Sym. Ser., vol. 639.
51. **Numao, S., S. M. He, G. Evjen, S. Howard, O. K. Tollersrud, and S. G. Withers.** 2000. Identification of Asp197 as the catalytic nucleophile in the family 38 alpha-mannosidase from bovine kidney lysosomes. FEBS Lett. **484:**175-178.
52. **Rigden, D. J.** 2002. Iterative database searches demonstrate that glycoside hydrolase families 27, 31, 36 and 66 share a common evolutionary origin with family 13. FEBS Lett. **523:**17-22.
53. **Rye, C. S., and S. G. Withers.** 2000. Glycosidase mechanisms. Curr. Opin. Chem. Biol. **4:**573-580.
54. **Sakuraba, H., F. Matsuzawa, S. Aikawa, H. Doi, M. Kotani, H. Nakada, T. Fukushige, and T. Kanzaki.** 2004. Structural and immunocytochemical studies on alpha-N-acetylgalactosaminidase deficiency (Schindler/Kanzaki disease). J. Hum. Genet. **49:**1-8.
55. **Shallom, D., V. Belakhov, D. Solomon, S. Gilead-Gropper, T. Baasov, G. Shoham, and Y. Shoham.** 2002. The identification of the acid-base catalyst of

alpha-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. FEBS Lett. **514:**163-167.

56. **Shallom, D., V. Belakhov, D. Solomon, G. Shoham, T. Baasov, and Y. Shoham.** 2002. Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. J. Biol. Chem. **277:**43667-43673.
57. **Sinnott, M. L.** 1990. Catalytic mechanisms of enzymatic glycosyl transfer. Chem. Rev. **90:**1171-1202.
58. **Tarling, C. A., S. M. He, G. Sulzenbacher, C. Bignon, Y. Bourne, B. Henrissat, and S. G. Withers.** 2003. Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from *Thermotoga maritima* through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis. J. Biol. Chem. **278:**47394-47399.
59. **Tayal, A., V. B. Pai, and S. A. Khan.** 1999. Rheology and microstructural changes during enzymatic degradation of a guar-borax hydrogel. Macromolecules **32:**5567-5574.
60. **Thananunkul, D., M. Tanaka, C. O. Chichester, and T. C. Lee.** 1976. Degradation of raffinose and stachyose in soybean milk by alpha-galactosidase from *Mortierella vinacea* - Entrapment of alpha-galactosidase within polyacrylamide-gel. J. Food Sci. **41:**173-175.
61. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673-4680.
62. **Vocadlo, D. J., G. J. Davies, R. Laine, and S. G. Withers.** 2001. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. Nature **412:**835-838.
63. **Yang, J. K., H. J. Yoon, H. J. Ahn, B. I. Lee, J. D. Pedelacq, E. C. Liong, J. Berendzen, M. Laivenieks, C. Vieille, G. J. Zeikus, D. J. Vocadlo, S. G. Withers, and S. W. Suh.** 2004. Crystal structure of beta-D-xylosidase from *Thermoanaerobacterium saccharolyticum*, a family 39 glycoside hydrolase. J. Mol. Biol. **335:**155-165.

Table 3.1. Primer pairs for *T. maritima* α-galactosidase site-directed mutants (complimentary primer not shown)

Mutation	Primer
D103A	GG AAC CTC CAG AGC GCC TAT TTC GTG GCT GAA GAA GG
D220A	C GAG GTC TTC CAG ATA GCC GAC GCC TAC GAA AAG GAC
D220G	C GAG GTC TTC CAG ATA GGT GAC GCC TAC GAA AAG GAC
D220A	G GTC TTC CAG ATA GAC GCC GCC TAC GAA AAG GAC ATA GG
D229A	C TAC GAA AAG GAC ATA GGT GCC TGG CTC GTG ACA AGA GG
D275A	GTA TTC AAC GAA CAT CCG GCC TGG GTA GTG AAG GAA AAC GG
D299A	G ATA TAC GCC CTC GCT CTT TCG AAA GAT GAG GTT CTG AAC TG
D327A	C TAC AGG TAC TTC AAG ATC GCC TTT CTC TTC GCG GGT GC
D327G	C TAC AGG TAC TTC AAG ATC GGC TTT CTC TTC GCG GGT GC
D387G	GG ATG AGG ATA GGA CCT GGT ACT GCG CCG TTC TG

Table 3.2. Structural similarity between *Tm*GalA (PDB code 1ZY9) and GH27 members[a]

GH27 enzyme	PDB code	Input residues (PDB numbering)	Z-score	Aligned C^{α} residues	RMSD (Å)	Sequence identity (%)
Homo sapiens α-galactosidase A	1R46 (chain A)	32-421 (complete)	19.7	240	2.9	14
		32-328 ($(\alpha/\beta)_8$ domain)	22.2	236	2.7	14
Hypocrea jecorina (née *Trichoderma reesei*) α-galactosidase	1SZN	1-417 (complete)	18.6	276	2.9	18
		1-319 ($(\alpha/\beta)_8$ domain)	22.5	235	2.6	20
Oryza sativa α-galactosidase	1UAS	1-362 (complete)	21.1	268	2.8	16
		1-278 ($(\alpha/\beta)_8$ domain)	23.4	233	2.6	18
Gallus gallus *N*-acetyl-α-galactosaminidase	1KTB	1-388 (complete)	19.6	271	2.9	15
		1-298 ($(\alpha/\beta)_8$ domain)	22.5	236	2.7	15

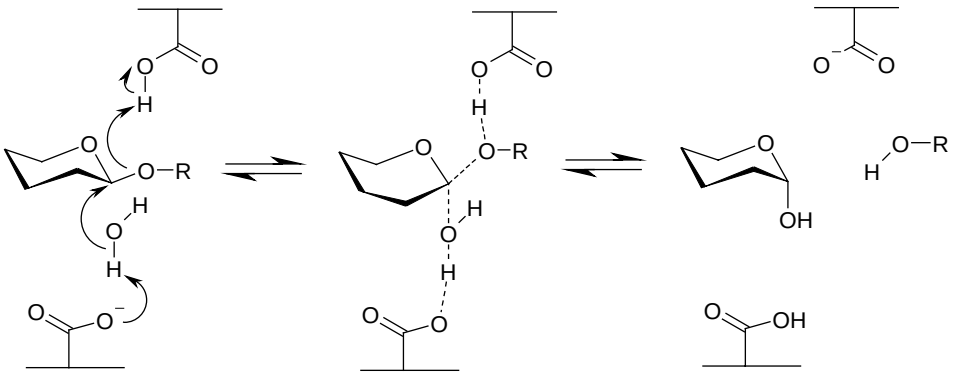
a. Pairwise structural alignments were performed using DaliLite (31) (accessed at the European Bioinformatics Institute at URL http://www.ebi.ac.uk/dali/index.html. For alignment with other complete sequences, *Tm*GalA residues 1-525 (PDB numbering) were used as input. For alignment of $(\alpha/\beta)_8$ domains, *Tm*GalA residues 180-479 were used as input.

Table 3.3: Thermal denaturation of *Tm*GalA and mutants	
Enzyme	**Melting temperature (°C)**
Wild-type	89.8
D220G	82.8
D327G	89.0
D387G	83.3

Table 3.4. Kinetics of *Tm*GalA WT and acid-base mutant in the presence of external nucleophiles

Aglycone	pKa	Enzyme	k_{cat} (s^{-1})	Ratio k_{cat} [WT/D387G]	K_M (mM)	k_{cat}/K_M ($mM^{-1}*s^{-1}$)	Ratio k_{cat}/K_M [WT/D387G]
4-nitrophenyl	7.18	**WT**	**33**	1,500	**0.052**	**620**	1,000
		D387G	0.022		0.036	0.61	
2,4-dinitrophenyl	3.96	**WT**	**23**	6.3	**0.016**	**1,400**	<4.6
		D387G	3.6		<0.012	<300	
4-nitrophenyl + azide	7.18	**WT**	**38**	1,400	**0.077**	**490**	1,400
		D387G	0.027		0.074	0.36	
2,4-dinitrophenyl + azide	3.96	**WT**	**25**	1.0	**0.023**	**1,100**	1.1
		D387G	24		0.025	960	
4-nitrophenyl + formate	7.18	**WT**	**31**	1,000	**0.044**	**700**	910
		D387G	0.031		0.040	0.77	
2,4-dinitrophenyl + formate	3.96	**WT**	**19**	0.77	**0.014**	**1,400**	1.0
		D387G	25		0.018	1,400	

a)



b)

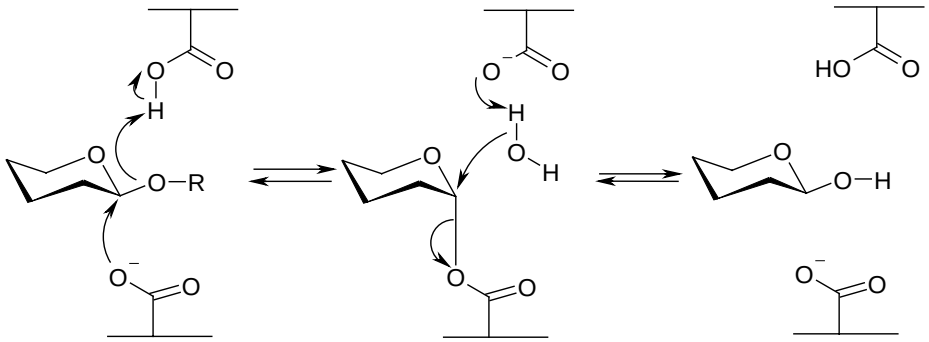


Figure 3.1: Glycosidase mechanisms. a) Inverting glycosidase mechanism in which a single displacement reaction results in the inversion of the anomeric configuration; b) Retaining glycosidase mechanism in which a double displacement mechanism generates a product with the anomeric configuration as the initial substrate.

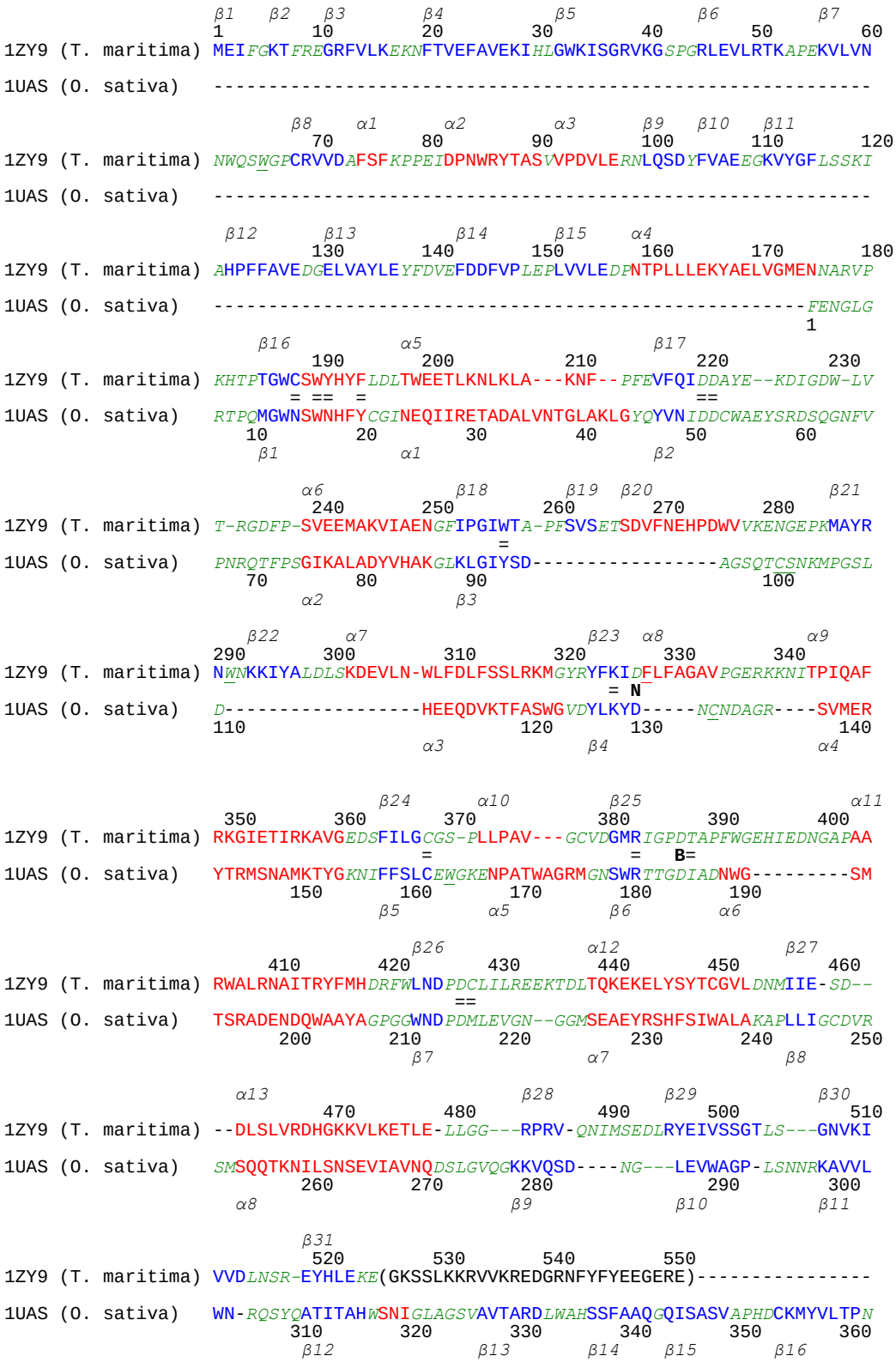

1ZY9 (T. maritima) MEIFGKTFREGRFVLKEKNFTVEFAVEKIHLGWKISGRVKGSPGRLEVLRTKAPEKVLVN
1UAS (O. sativa) ----------
1ZY9 (T. maritima) NWQSWGPCRVVDAFSFKPPEIDPNWRYTASVVPDVLERNLQSDYFVAEEGKVYGFLSSKI
1UAS (O. sativa) ----------
1ZY9 (T. maritima) AHPFFAVEDGELVAYLEYFDVEFDDFVPLEPLVVLEDPNTPLLLEKYAELVGMENNARVP
1UAS (O. sativa) ----------FENGLG
1ZY9 (T. maritima) KHTPTGWCSWYHYFLDLTWEETLKNLKLA---KNF--PFEVFQIDDAYE--KDIGDW-LV
1UAS (O. sativa) RTPQMGWNSWNHFYCGINEQIIRETADALVNTGLAKLGYQYVNIDDCWAEYSRDSQGNFV
1ZY9 (T. maritima) T-RGDFP-SVEEMAKVIAENGFIPGIWTA-PFSVSETSDVFNEHPDWVVKENGEPKMAYR
1UAS (O. sativa) PNRQTFPSGIKALADYVHAKGLKLGIYSD----------AGSQTCSNKMPGSL
1ZY9 (T. maritima) NWNKKIYALDLSKDEVLN-WLFDLFSSLRKMGYRYFKIDFLFAGAVPGERKKNITPIQAF
1UAS (O. sativa) D----------HEEQDVKTFASWGVDYLKYD-----NCNDAGR----SVMER
1ZY9 (T. maritima) RKGIETIRKAVGEDSFILGCGS-PLLPAV---GCVDGMRIGPDTAPFWGEHIEDNGAPAA
1UAS (O. sativa) YTRMSNAMKTYGKNIFFSLCEWGKENPATWAGRMGNSWRTTGDIADNWG---------SM
1ZY9 (T. maritima) RWALRNAITRYFMHDRFWLNDPDCLILREEKTDLTQKEKELYSYTCGVLDNMIIE-SD--
1UAS (O. sativa) TSRADENDQWAAYAGPGGWNDPDMLEVGN--GGMSEAEYRSHFSIWALAKAPLLIGCDVR
1ZY9 (T. maritima) --DLSLVRDHGKKVLKETLE-LLGG---RPRV-QNIMSEDLRYEIVSSGTLS---GNVKI
1UAS (O. sativa) SMSQQTKNILSNSEVIAVNQDSLGVQGKKVQSD----NG---LEVWAGP-LSNNRKAVVL
1ZY9 (T. maritima) VVDLNSR-EYHLEKE(GKSSLKKRVVKREDGRNFYFYEEGERE)----------
1UAS (O. sativa) WN-RQSYQATITAHWSNIGLAGSVAVTARDLWAHSSFAAQGQISASVAPHDCKMYVLTPN

Figure 3.2. Structure-based protein sequence alignment of *Tm*GalA (PDB 1ZY9, UniProt O33835) and *Oryza sativa* α-galactosidase (PDB 1UAS, UniProt Q9FXT4). The respective PDB file residue numbering systems are used. Secondary structural elements extracted by PyMOL from the PDB data are highlighted: sheets, blue; helices, red; loops, green italics. Secondary structure element numbering for 1UAS is from Fujimoto et al. (16). Spatially equivalent residues in the enzyme active sites are denoted with "=". The following underlined residues are also approximately equivalent in the active site: 1ZY9 W65 = 1UAS W164; 1ZY9 F328 & W291 = 1UAS C101-C132 disulfide & S102. The catalytic nucleophile residue in each enzyme is marked with "**N**", the general acid/base residue with "**B**". Residues 526-551, in parenthesis, were not observed in the crystal structure of *Tm*GalA.

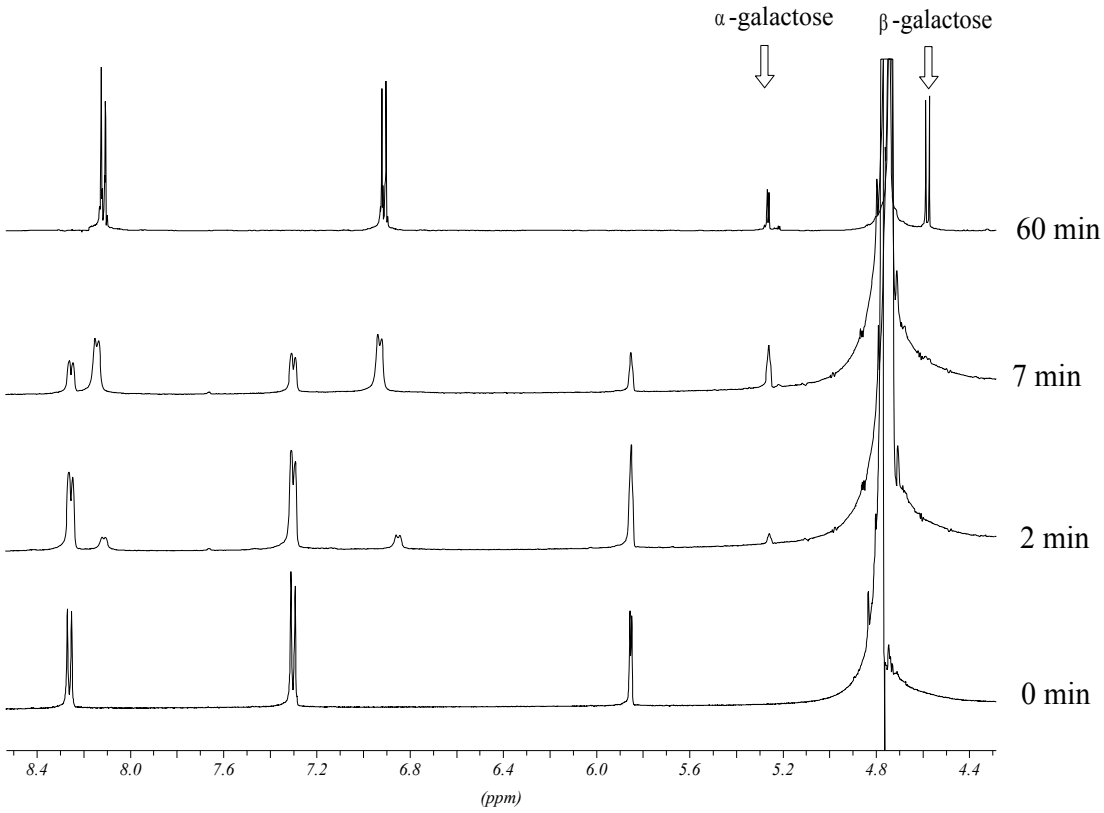


Figure 3.3. ^{1}H NMR analysis of the stereoselectivity for the WT *Tm*GalA catalyzed reaction.

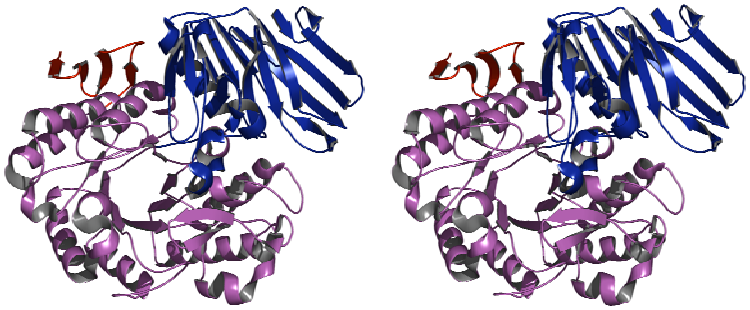


Figure 3.4. Wall eyed stereo overview of the *Tm*GalA tertiary structure and secondary structural elements. The N-terminal, $(\beta/\alpha)_8$-barrel, and C-terminal domains are colored in blue, violet, and red, respectively.

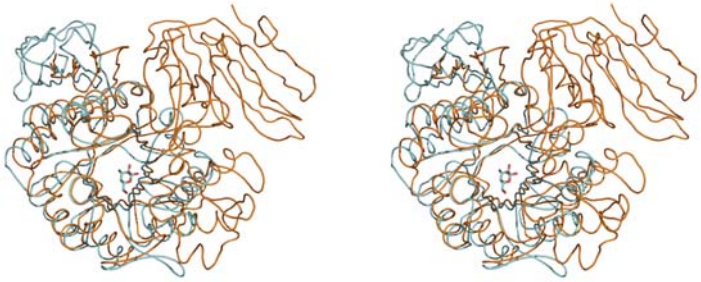


Figure 3.5. Superposition of the C^{α} traces of GH36 *Tm*GalA (PDB 1ZY9, orange) and the GH27 *Oryza sativa* α-galactosidase (PDB 1UAS, cyan). The wall-eyed stereo view is down the axis of the $(\beta/\alpha)_8$-barrel domain; the perspective is identical to that in Figure 4. Galactose, observed as an equilibrium mixture of α- and β-anomers in the *O. sativa* structure (16), highlights the position of the enzyme active sites.

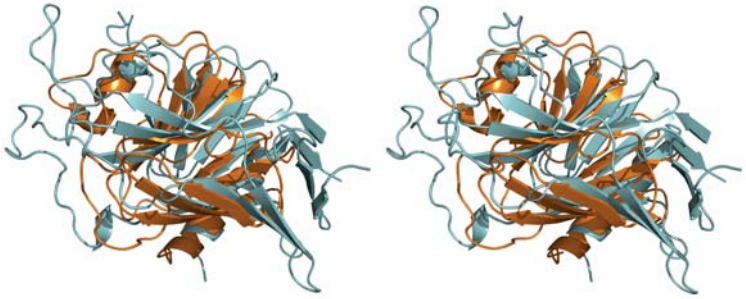


Figure 3.6. Wall-eyed stereo view of the N-terminal domain (residues 1-179) of GH36 *Tm*GalA (PDB 1ZY9, orange) superimposed on peptide D of GH38 bovine lysosomal α-mannosidase (PDB 1O7D:chain D, cyan).

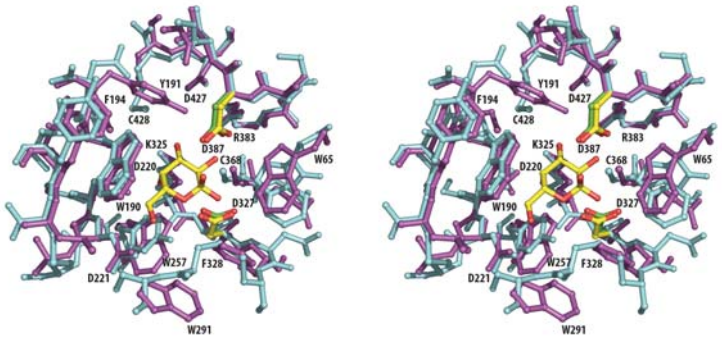


Figure 3.7. Superposition of the active sites of GH36 *Tm*GalA (PDB 1ZY9, violet) and the GH27 *Oryza sativa* α-galactosidase (PDB 1UAS, cyan). Bound galactose, as well as the catalytic nucleophile (D130) and general acid/base (D185) residues, in the *Oryza sativa* structure are shown in yellow/red. The corresponding nucleophile and acid/base residues in the *Tm*GalA structure are D327 and D387, respectively (green/red). Residue numbering is shown for *Tm*GalA, following the numbering scheme in PDB 1ZY9.

a)

b)

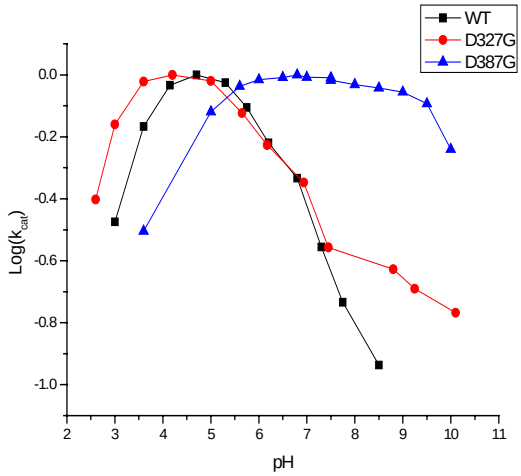
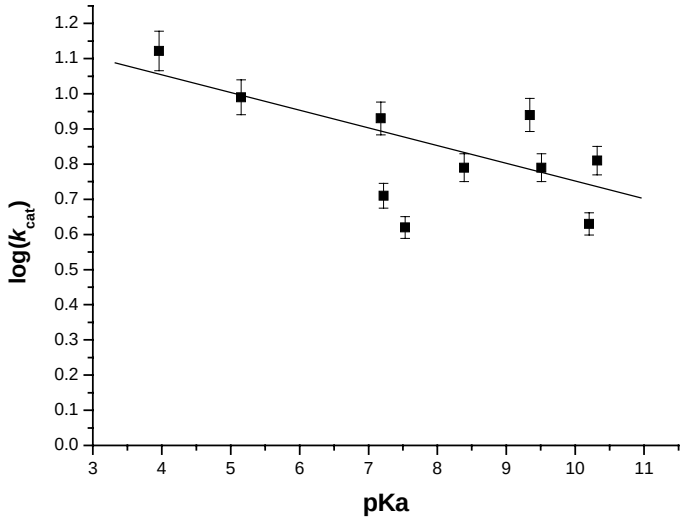


Figure 3.8: pH-dependences of the kinetic parameters. a) k_{cat}/K_M of wild-type, D327G, and D387G enzymes; b) k_{cat} of wild-type, D327G, and D387G enzymes

a)



b)

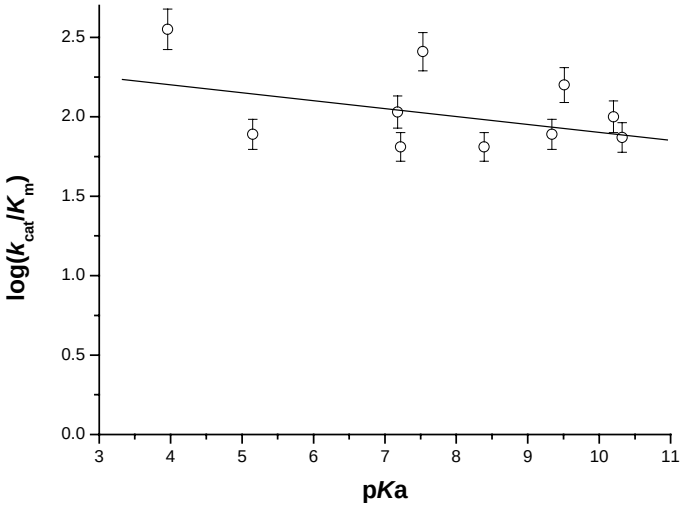
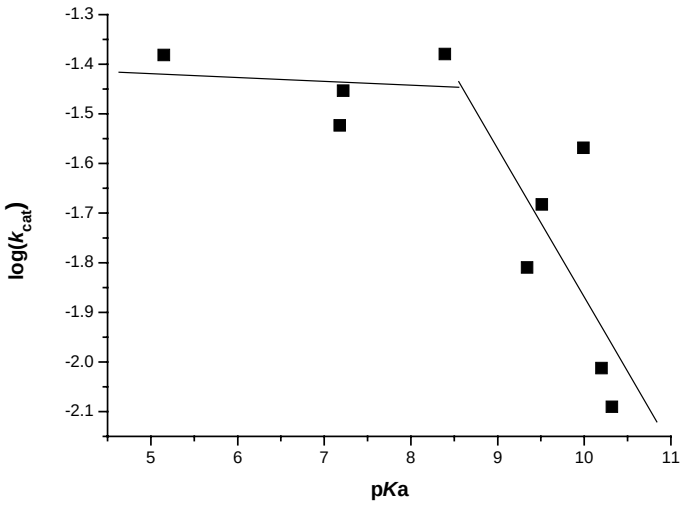


Figure 3.9: Brønsted plots of WT GalA enzyme for aryl galactosides. a) Effect of leaving group on k_{cat}, β_{lg} = -0.05; b) Effect of leaving group on k_{cat}/K_M, β_{lg} = -0.05. Leaving group ability represented as pK_a: 2,4-dinitrophenyl (4.0); 2,5-dinitrophenyl (5.2); 4-nitrophenyl (7.2); 2-nitrophenyl (7.2); 3-nitrophenyl (8.4); 4-bromophenyl (9.3); β-naphtyl (9.5); phenyl (10.0); 4-O-methylphenyl (10.2); 3,4-dimethylphenyl (10.3).

a)



b)

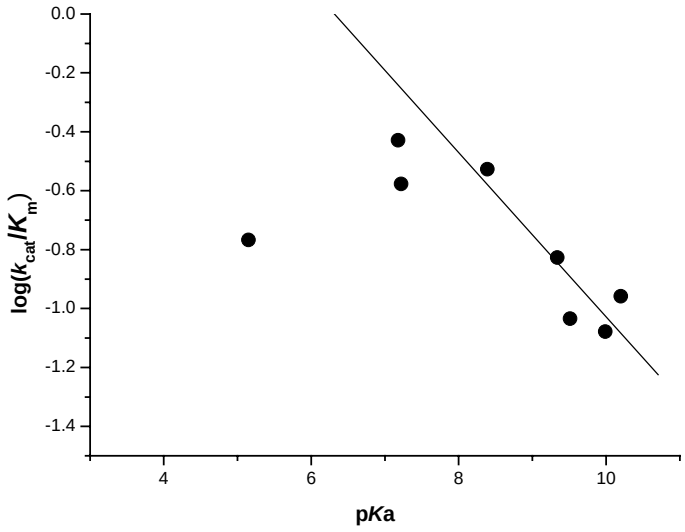


Figure 3.10: Brønsted plots of D327G enzyme for aryl galactosides. a) Effect of leaving group on k_{cat}, β_{lg} = -0.3; b) Effect of leaving group on k_{cat}/K_M, β_{lg} = -0.3. Leaving group ability represented as pKa: 2,4-dinitrophenyl (4.0); 2,5-dinitrophenyl (5.2); 4-nitrophenyl (7.2); 2-nitrophenyl (7.2); 3-nitrophenyl (8.4); 4-bromophenyl (9.3); β-naphtyl (9.5); phenyl (10.0); 4-O-methylphenyl (10.2); 3,4-dimethylphenyl (10.3).

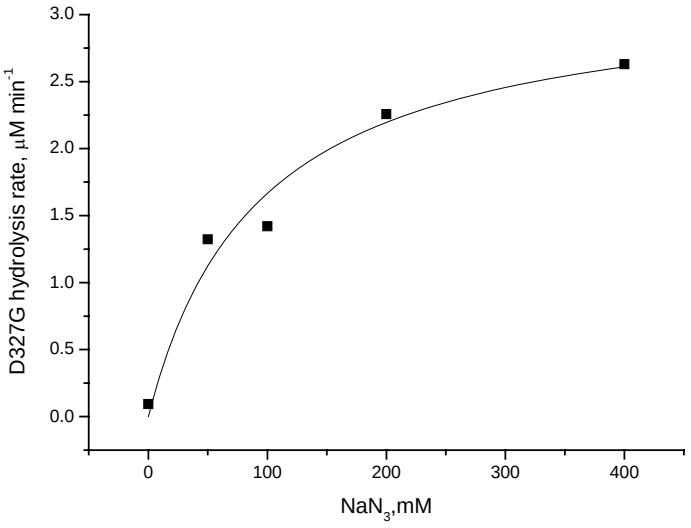


Figure 3.11. Azide resuce of the nucleophile mutant D327G.

Functional genomics-based approaches for identification and characterization of novel α-glucan acting glycosyl hydrolases in the hyperthermophilic archaeon *Pyrococcus furiosus*

Donald A. Comfort[1], Chung-Jung Chou[1], Shannon B. Conners[1], Amy L. Van Fossen[1], and Robert M. Kelly[1]*

[1]Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

Manuscript in Preparation

ABSTRACT

Bioinformatic analysis, in conjunction with whole-genome transcriptional profiling, was used to identify ORFs with potential roles in α-glucan processing in the hyperthermophilic archaeon *Pyrococcus furiosus*. In some cases, hypothetical proteins containing domains related to known glycosidases were found to hydrolyze α-glucan substrates. For example, PF0870, annotated as a hypothetical protein, shares a domain with members of glycosyl hydrolase family 57. Biochemical characterization of the recombinant version of this protein produced in *Escherichia coli* revealed it to be a novel β-amylase, capable of removing maltose from the reducing end of maltotriose. Functional genomics approaches were also found to be useful. Mixed effects model-based statistical analysis of a loop experimental design was used to track transcriptional response to growth on starch, pullulan, glycogen, trehalose, and maltose in contrast to growth on cellobiose and a mixture of cellobiose and maltose. In addition to confirming the response of several known α-glucanases, for which biochemical information had been reported previously, a number of ORFs that responded to specific carbohydrates were examined for a potential role in carbohydrate processing. In some cases, analysis was unable to link ORFs exhibiting a transcriptional response to α-glucans to biochemical properties. PF1108, annotated as a 'α-dextrin-1,6-endo-α-glucosidase', showed no activity on synthetic substrates or pullulan, starch, or glycogen. Similarly, PF1746, which has a predicted conserved domain related to glycogen debranching enzymes, showed no activity on α-glucans. In other cases, transcriptional profiling was suggestive of enzyme biochemical characteristics. A previously characterized exo-α-glucanase (Costantino *et al.*, J. Bacteriol., 172:3654-3660, 1990) that had been purified from *P. furiosus* cell extracts, corresponded to PF0132, an ORF annotated as a hypothetical protein. Transcriptional analysis indicated that this gene was highly responsive to growth on pullulan, an α-1,6-linked polysaccharide. Indeed, biochemical characterization of the native protein confirmed that this enzyme was capable of hydrolyzing panose (Glc-α-1,6-Glc-α-1,4-Glc) with comparable catalytic efficiency to maltose, and suggesting that this relates to the physiological role for this enzyme. PF0132 represents a novel glycosyl

hydrolase family, which defines a new family of these carbohydrate-active enzymes. PF0133 located just upstream of PF0132 had similar transcriptional response to the substrates studied and was investigated to see if there was catalytic activity or synergistic effects on maltooligosaccharides or pullulan in the presence of PF0132, however, none were detected. In addition to defining the functional roles of several uncharacterized ORFs in the *P. furiosus* genome, this study also reinforces the utility of functional genomics approaches, used in conjunction with biochemical analysis, to annotate microbial genomes.

INTRODUCTION

A key challenge in the post-genomics era is the functional assignment of protein-encoding open reading frames (ORFs) for which few or no insights are available through comparative amino acid sequence analysis (hypotheticals, conserved hypotheticals, or putative proteins for which annotation is based on vague relationships to domains within other proteins). These unknown ORFs typically comprise 30-40% of sequenced microbial genomes (7). In archaea, that number may be even higher; hypothetical proteins reportedly comprise 870 out of the 2065 ORFs in *Pyrococcus furiosus*, 1511 of the 2960 ORFs in *Sulfolobus solfataricus* P2, and 1057 of the 1784 ORFs in *Methanococcus jannaschii* DSM2661 (54). Determining the role of the gene products of these unknown predicted ORFs is essential for a complete understanding of the metabolism and physiology of these microorganisms (20). Establishing the biological function of predicted proteins in genome sequences is difficult, and ultimately requires biochemical and genetic characterization (19). Functional genomics approaches, in which transcriptional response patterns are attributed to physiological roles, can be used to glean useful information of potential value for genome annotation. If supporting biochemical data can then be brought to bear, not only will a confident functional assignment for a given ORF in a particular organism be possible, it will also shed light on the roles of related but unknown proteins encoded in other genomes.

Whole-genome transcriptional profiling has been used to investigate carbohydrate utilization patterns in mesophilic (5) and hyperthermophilic microorganisms (8, 9, 41, 50, 51) to define carbohydrate utilization operons and regulons. Indeed, biochemical confirmation (50) of predicted substrates for carbohydrate-binding proteins in *Thermotoga maritima* (9) has reinforced the value of functional genomics approaches for deciphering genome annotation issues. However, progress on assigning biochemical and physiological characteristics to unknown ORFs has been limited.

P. furiosus is a heterotrophic, hyperthermophilic archaeon that grows optimally at 98-100°C using peptides and carbohydrates as carbon/energy sources (17). Functional genomics have been used to investigate heat shock (63), cold shock (69), differences between peptide and carbohydrate utilization (61, 62), and growth on various carbohydrates (41) in this microorganism. Growth on carbohydrates implies that *P.*

furiosus produces a number of glycosyl hydrolases, a fact that is consistent with analysis of its genome and with previous reports on the biochemical features of specific enzymes (16, 58). The *P. furiosus* genome encodes glycosyl hydrolases capable of metabolizing cellobiose (PF0073) (68), laminarin (PF0076) (16, 22), maltodextrins and starch (PF0132, PF0272, PF0477) (14, 33, 39-41), chitobiose and chitin (PF1233, PF1234) (21, 65), pullulan (PF1935) (15, 41), cyclodextrins (PF0478, PF1939) (73), β-glucans (PF0442, PF0854) (6, 35), α-galactosides (PF0444) (67), and β-mannans (PF1208) (47). In general, α-glucan polysaccharides, such as starch and pullulan, are hydrolyzed into smaller maltooligosaccharides by endo-acting pullulanases, α-amylases, de-branching enzymes, and glucoamylases. These maltooligosaccharides are further hydrolyzed into glucose for metabolic processes, such as glycolysis. *P. furiosus* produces an amylopullulanase (PF1935*) and an α-amylase (PF0477) for extracellular hydrolysis of the α-glucans. Lee *et al.* (41) recently proposed a utilization pathway for starch hydrolysis in *P. furiosus* in which both maltodextrins and maltose are transported into the cell and metabolized. Transport of maltose occurs by the Mal-I ABC transporter (PF1739-1741, PF1744) and transport of maltodextrins by the Mal-II ABC transporter (PF1933, PF1936-1938). Intracellular processing of maltose is by the α-amylase/glucanotransferase, PF0272, with the products of glucose and maltodextrins (41). Processing of the maltodextrins occurs by an intracellular α-glucan phosphorylase which generates glucose-1-phosphate and maltodextrin$_{n-1}$ from maltodextrins (75).

Here, whole-genome cDNA microarrays were utilized to investigate the transcriptional response of *P. furiosus* to a variety of α-glucans to provide a broader perspective on the carbohydrate utilization patterns in this hyperthermophile. At the same time, bioinformatic tools were used to determine ORFs encoding predicted proteins that potentially play a role in α-glucan hydrolysis. Biochemical characterization of selected glycosidases from transcriptional and bioinformatics analysis revealed that *P. furiosus* produces a number of novel glycosyl hydrolases, some of which define new families of this enzyme class. The extent to which these and yet unknown glycosyl hydrolases encoded in the *P. furiosus* genome participate in α-glucan processing remains to be seen. But, these results suggests that a broad range of biocatalysts will be needed to convert

complex biomass substrates to simple monosaccharides in enzyme-based alternative energy and chemicals, if microbe-based systems are any indication.

MATERIALS AND METHODS

Growth of Pyrococcus furiosus on α-glucans – *P. furiosus* was grown anaerobically at 80ºC in SSM media consisting of 40 g/l sea salt, 3.1 g/l PIPES [piperazine-*N*,*N'*-bis(2-ethanesulfonic acid)], 1.0 g/l tryptone, 1.0 mL 1000 x *Vent Sulfothermophiles* medium trace metals (1.50 g nitrilotriacetic acid, 0.50 g $FeCl_2 \cdot 6H_2O$, 0.30 g $Na_2WO_4 \cdot 2H_2O$, 0.40 g $MnCl_2 \cdot 4H_2O$, 0.20 g $NiCl_2 \cdot 6H_2O$, 0.1 g $ZnSO_4 \cdot 7H_2O$, 0.1 g $CoSO_4 \cdot 7H_2O$, 0.01 g $CuSO_4 \cdot 5H_2O$, and 0.01 g $Na_2MoSO_4 \cdot 5H_2O$ per 100 ml), resazurin, brought to pH 6.8 by addition of NaOH, and supplemented with 3.3 g/l of the either glucose, cellobiose, β-cyclodextrin, glycogen, maltose, pullulan, raffinose, sucrose, starch, trehalose, or maltose/cellobiose, as previously described (63). Cellobiose, β-cyclodextrin, maltose, raffinose, starch, sucrose, and trehalose were obtained from Sigma (St. Louis, MO); glucose and glycogen were obtained from Fisher Scientific (Hampton, NH), and pullulan was obtained from Pfanstiehl Laboratories (Waukegan, IL). Cultures (60 mL) were grown for at least six passes with a 1% inoculum, before cell counts or RNA extraction were performed. Cell growth was monitored by taking one mL samples periodically and counting epiflourescent cells stained with acridine orange (27).

Only those cultures that grew to densities greater than the control culture of SSM without supplemented carbohydrates, as shown in Figure 4.1, were selected for RNA extraction and transcriptional analysis; these included cellobiose, glycogen, maltose, pullulan, starch, trehalose, or maltose/cellobiose. Three hundred sixty mL cultures were grown with a 1.0% inoculum in 500 mL Pyrex bottles. The cells were harvested in early exponential phase (~2 x 10^7 cells/ml), immediately spun at 7,500 x g for 22 minutes at 4°C. The cells were re-suspended in sugar-free SSM medium, centrifuged briefly at 14,000 x g, re-suspended in 85 µL ice cold TE buffer, and lysed with 625 µL of extraction buffer (6.2 M guanidine thiocynate; 38 mM sodium citrate, pH 7.0; 0.75% Sacrosyl). The genomic DNA was sheared by passage through a 20-gauge needle and then 62.5 µL of 2 M sodium acetate (pH 5.2) was added. Then 750 µL of acidic

phenol/chloroform (5:1; Ambion, Austin, TX) was added, tubes were vortexed and chilled on ice for 5 minutes, centrifuged at 4°C for 10 minutes at 14,000 x g, the aqueous phase transferred, and the RNA ethanol-precipitated overnight at –80°C (63). The RNA was pelleted at 20,000 x g for 20 minutes at 4°C, washed with 70% ethanol and centrifuged at 14,000 x g for 20 minutes, and the RNA was re-suspended in 10 mM Tris-HCl, pH 8.5 and further purified by the RNAqueous™ RNA isolation kit (Ambion) as per the instructions. The RNA was quantitated on a Beckman Coulter UD-640 spectrophotometer (Fullerton, CA) by OD_{260}/OD_{280} and the purity checked on an agarose RNA gel.

Microarray Design – Microarrays were run as described in detail elsewhere (41). Briefly, DNA primers were designed for all 2065 *P. furiosus* ORFs as listed on The Institute for Genomic Research website (54). Probes were generated by PCR amplification in a PTC-100 thermocycler (MJ Research, Waltham, MA) using *Taq* polymerase. The PCR products were purified using Qiagen PCR purification kits (Qiagen, Valencia, CA) and re-suspended in 50% DMSO. They were randomized and printed onto ULTRAGAP aminosilane-coated microscope slides (Corning, Corning, NY) using a Genetix Q-Array-Mini Arrayer (Hampshire, United Kingdom), and the DNA cross-linked to the slides by UV and incubated in an oven for 2 hours at 75°C.

cDNA generation and hybridization - cDNA transcripts were generated from total RNA by Superscript III Reverse Transciptase (Invitrogen, Carlsbad, CA) and random primers (Invitrogen), during which 5-[3-aminoallyl]-2'-deoxyuridine-5'-triphosphate (aa-dUTP) (Sigma) was incorporated as described previously (63). The cDNA product was purified by QIAquick PCR purification kit (Qiagen). After purification Cyanine-3 (Cy-3) and Cyanine-5 (Cy5) NHS-esters (Amersham Biosciences, Inc., Piscataway, NJ) were incorporated followed by purification with the QIAquick PCR purification kit to remove unincorporated dyes (63). Slides were scanned using the Scanarray 4000 (Perkin Elmer, Fremont CA) and signal intensities determined using Scanarray software (Perkin Elmer).

Mixed Model ANOVA analyses – Mixed model analysis of variance (ANOVA) model was applied as previously described (41, 70). Briefly, a loop design was used with reciprocal labeling to estimate dye effect. The global variance due of dye, treatment, array, block, spot effects and random error was estimated from the ANOVA model. Gene specific effects were then partitioned using a gene-specific ANOVA model.

Growth of P. furiosus for α-glucosidase – *P. furiosus* was grown overnight to stationary phase in SSM supplemented with 3.3 g/l pullulan at 80°C. This was used to inoculate a 10 L culture (0.3% inoculum) of SSM supplemented with pullulan in a New Brunswick 14 L fermenter operated at 80°C. The fermenter was run anaerobically overnight and the cells harvested at a final cell density of 4.3 x 10^8 cells/ml. The cells were harvested, concentrated by cross-flow filtration using a 0.22 μm filter (Millipore, Bedford, MA), and centrifuged to separate the supernatant. The cells were pelleted (10.0 g wet cells) by centrifugation at 7,500 x g for 20 minutes, resuspended in ~40 mL of 20 mM Tris-HCl (pH 8.0), treated with lysozyme for 30 minutes at 30°C, sonicated (10 minutes of pulsing, 10 s on, 10 s delay), followed by three passes through the French Press (Sim-Aminco) at 12,000 lbs. The lysed protein was separated from the cell debris by centrifugation (20,000 x g, 30 min) and the protein concentration determined by the Biorad Protein Assay (Hercules, CA) using Bovine Serum Albumin as standard.

Purification of α-glucosidase – The crude extract containing the α-glucosidase protein (PfαGlc) was applied to the DEAE column equilibrated with 50 mM Tris, pH 8.0. The protein was eluted by a linear gradient of 0-1 M NaCl in 50 mM Tris-HCl, pH 8.0, and activity was detected starting about 0.5 M NaCl. Fractions were screened using *p*-nitrophenyl-α-D-glucopyranoside (*p*NPαGlc, Sigma) in 50 mM sodium acetate buffer, pH 5.5 and those fractions with activity were pooled. Pooled fractions were run on a hydroxyapatite column and eluted with a linear gradient of increasing phosphate anions from 50 to 500 mM, as sodium phosphate, pH 7.2. Activity was screened, and PfαGlc started to elute at about 300 mM phosphate. Active fractions were pooled, concentrated using in an Amicon concentrator with a 10 kDa ultrafiltration membrane (Millipore) to about 5 mL. This was applied in two batches to a 300 mL Sephacryl S-200 HR sizing

column (GE Healthcare). PfαGlc eluted about 155-170 mL in 50 mM Tris-HCl, pH 7.2 + 150 mM NaCl. Active fractions were pooled and concentrated, as previously described. After concentration and immediately before application to the Biorad HIC butyl cartridge the protein was combined with 3.0 M ammonium sulfate in equal amounts to generate a 1.5 M ammonium sulfate environment. The protein was immediately applied to the HIC-butyl cartridge and a linear gradient of sulfate was run from 50 mM Tris-SO_4 with 1.5 M ammonium sulfate to 50 mM Tris-SO_4. Active protein eluted about 0.7 M $AmSO_4$.

Characterization of the native α-glucosidase from P. furiosus – Kinetics assays were performed using purified α-glucosidase from *P. furiosus*. Substrate specificity of PfαGlc was tested on the natural substrates of maltose, maltotriose, isomaltose, isomaltotrise, turanose, trehalose, and panose and run on thin-layer chromatography silica gel plates (Kiesel gel F_{254}; Merck, Rathway, NJ) using the mobile phase 1-butanol/ethanol/water (5:3:2) and resolved by dipping plate in 10% H_2SO_4 in ethanol and heating on a hot plate (72). Hydrolysis of the synthetic substrate *p*NPαGlc was characterized at 90°C in 50 mM sodium phosphate buffer, pH 5.6 run at 405 nm on Perkin-Elmer Lambda Bio 20 spectrophotometer (Perkin-Elmer Corp., Norwalk, CT) equipped with two PTP-1 Digital Temperature Controllers (Perkin-Elmer). Kinetic parameters for glucose released were determined using Amplex® Red Glucose/Glucose Oxidase Assay Kit (Invitrogen, Carlsbad, CA) as per instructions, with absorbance at 560 nm. Characterization of natural substrates was done in 100 mM sodium phosphate, pH 5.6 with 20 mM substrate and 162 ng enzyme, 50 μL final volume. Kinetic parameters were determined from Lineweaver-Burk plots.

Cloning, expression, and purification of PF0870 – The ORF PF0870 was amplified from genomic *P. furiosus* DNA using the forward primer 5'-CACACA<u>CCATGG</u>TTATGAAATTCACATATCAC-3' (*Nco*I site underlined) and the reverse primer 5'-ATATAT<u>CTCGAG</u>TTTCCCTTTATGCTCGCACCAC-3' (*Xho*I site underlined). The gene product and pET24d plasmid (Novagen, San Diego, CA) were digested with *Nco*I and *Xho*I, purified using a QIAquick PCR purification kit, and ligated together to generate the subclone. This resulted in the elimination of the stop codon of

PF0870 and the in-frame placement of the His_6-Tag of pET24d. The protein was expressed in *Escherichia coli* Rosetta(DE3) cells (Novagen) by growth overnight with no IPTG induction. Cells were lysed by French Press in 20 mM Tris-HCl (pH 8.0). The clone was sequenced to verify correct sequence at University of Georgia Integrated Biotech Laboratories. Crude extract was heat-treated at 80°C for 30 minutes to denature heat-labile *E. coli* proteins. The heat-treated extract was then applied to a His-Trap™ HP 5 mL metal affinity column (Amersham Bioscience, Piscataway, NJ) charged with Ni^{2+} and equilibrated with 20 mM sodium phosphate (pH 7.4) and 500 mM NaCl. The protein was eluted using imidazole with a step to 10 mM imidazole for five column volumes followed by a gradient of 10 to 500 mM imidazole over ten column volumes. Active fractions were assayed using 1 mM *p*NP-α-D-maltopyranoside (*p*NPαMalt, Sigma) in 50 mM sodium phosphate buffer, pH 7.0. Active fractions were pooled and concentrated using a 10 kDa Centricon filter (Millipore).

Real-time PCR of PF0870 – RNA from *P. furiosus* grown on the sugars cellobiose, glycogen, maltose, pullulan, starch, trehalose, and maltose/cellobiose was used to run real-time PCR reactions. cDNA was generated as described previously, with the exception that aa-dUTP was not incorporated into the cDNA product. The primers 5'-TATCTAGTCTTTCTTTCCAGC-3' and 5'-CCCATTCTTTTTCAATCCTTC-3' were used in reaction concentrations of 250 nM. Amplification was followed by fluorescence of SYBR-Green I dye in the iQ SYBR Green Supermix (Biorad) with a Biorad MyIQ Realtime PCR Detection System attached to a Biorad iCycler for 60 cycles with annealing and extension temperatures of 56°C and 72°C, respectively, for 30 s each and a reaction volume of 20 μL. A standard curve was generated using cDNA from *P. furiosus* grown on glycogen in four-fold dilutions from 38.4 ng to 0.15 ng. Other conditions were run with 2.4 ng per reaction of template cDNA. The PCR efficiency and cycle thresholds were calculated using the iCycler iQ software (Bio-Rad), and the fold changes were calculated by the method of Pffafl (55). A melt curve analysis was completed and confirmed that homogeneous PCR products were generated with no nonspecific amplification.

Characterization of PF0870 – Transglycosylation reactions were run in 300 μL reactions containing 100 mM sodium phosphate buffer, pH 7.0 and 20 μg of recombinant PF0870-His. Reactions were 20 mM maltose, 20 mM maltotriose, 10 mM maltotetraose, 10 mM maltopentaose, 10 mM maltohexaose, and 10 mM maltoheptaose for 30, 60, and 120 minutes, as described in Lee *et al.* (41). Additional reactions were run each containing 1% starch with 20 mM of each of the disaccharides maltose, glucose, trehalose, cellobiose, and sucrose under the same conditions previously described. pH optima was determined using the buffers sodium acetate (pH 4.2 and 5.0), sodium phosphate (pH 5.8, 6.6, 7.4, and 7.9), and Clark and Lubs (pH 8.0, 8.8, and 9.6). Thin-layer chromatography was used to resolve transglycosylation reactions as described previously. The synthetic substrate *p*NPαMalt was used to determine kinetic parameters in 100 mM sodium phosphate buffer, pH 7.0 using 5 μg of recombinant PF0870-His.

Other PF ORFs investigated – Attempts to clone selected PF ORFs was done using the primers and vectors in Table 4.1. Expression of each protein was attempted in *E. coli* Rosetta cells, grown overnight on LB medium with 0.25 mM IPTG. Cells were lysed by sonication for 20 minues (10 s pulsed, 10 s off). PF0133, PF1393, PF1746, and PF1747 were assayed in 50 mM NaP, pH 5.6 or 7.2 with 1% starch, pullulan, amylase, or glycogen for 1 hour to screen for activity. Amylose was solubilized to 2.5% in DMSO. TLC was used to monitor the hydrolysis products. Hydrolysis of *p*NP-α-maltopentaoside (*p*NPαMalt$_5$), *p*NPαMalt, *p*NP-α-mannopyranoside (*p*NPαMan), *p*NP-β-glucopyranoside (*p*NPβGlc), *p*NP-β-xylopyranoside (*p*NPβXyl), *p*NPαGlc, and *p*NP-α-galactopyranoside (*p*NPαGal) was tested for the recombinant enzymes PF1108, PF1393, PF1746, and PF1747 at 90°C for 10 minutes in 50 mM sodium phosphate buffer, pH 5.6 and 7.2. Hydrolysis of sucrose, trehalose, and turanose was attempted at pH 5.6 and 7.2 in 50 mM sodium phosphate buffer with 10 mM substrate for 45 minutes at 90°C. Raasch *et al.* (57) determined that the α-glucosidase from *T. maritima* was NAD^+, Mn^{2+}, and thiol dependent by adding cellular extract to assays, so this was attempted for recombinant PF1108, PF1393, PF1746, and PF1747 on glycogen, starch, amylase, and pullulan followed by TLC analysis. A 60 mL culture of *P. furiosus* was pelleted, resuspended in 20 mM Tris-HCl, pH 8.0, and sonicated for 10 minutes (10 s pulse, 10 s off) to lyse the

cells. Cell debris was then removed by 0.22 μM filter and protein removed by passage through a 1 kDa micron filter (Millipore). Starch zymograms were used to determine if hydrolysis clearing zones would be present for PF1108, PF1393, PF1746, and PF1747 as described with renaturation by 20 mM Tris-HCl, pH 8.0 followed by incubation at 70°C for 1 hour in 50 mM MOPS, pH 7.2. PF1460 was cloned, but not expressed.

PF1108-His was purified using the His-Trap^{TM} column as described for PF0870, and screened for activity on *p*NP-ester substrates and α-glucan polysaccharides in 100 mM sodium acetate buffer, pH 5.7 and 100 mM sodium phosphate buffer, pH 7.9 at 80°C. Ester substrates screened were 1 mM concentrations of *p*NP-α-mannopyranoside, *p*NP-acetate, *p*NP-propionate, *p*NP-caprate, and *p*NP-palmitate and 0.03% (w/v) *p*NP-butyrate were screened. Activity was screened against 1% β-cyclodextrin, starch, pullulan, and glycogen at the same conditions using the DNS reducing sugar assay (48). Hydrolysis of starch, glycogen, and pullulan was examined at 50 mM NaP, pH 5.6 and 7.2 using TLC as previously described.

PF0478 was expressed in *E. coli* Rosetta cells, heat-treated, and samples given to Drs. R. Prud'homme and S. Khan. PF1939 was expressed in BL21, heat-treated, activity determined against β-cyclodextrin as described by Yang (73), and samples given to Drs. R. Prud'homme and S. Khan.

RESULTS AND DISCUSSION

Bioinformatic analysis of P. furiosus genome with respect to carbohydrate-active enzymes – Glycosyl hydrolases are usually readily identifiable based upon primary amino acid sequences, even if functional assignment is not possible, as evident by the CAZY database (26). In an effort to identify and characterize new glycosyl hydrolases within the *P. furiosus* genome, a multi-pronged approach was undertaken involving a bioinformatic survey of the genome, functional genomics analysis using cDNA microarrays, and biochemical characterization of recombinant and native versions of proteins encoded in targeted ORFs.

Bioinformatic analysis of the *P. furiosus* genome identified a number of ORFs that could be involved in processing α-glucan substrates. ORFs were selected based upon

BLAST analysis (2), domain homology (46), and domain families (60, 64). These ORFs were cloned and expressed in *E. coli* or obtained from the Southeast Collaboratory for Structural Genomics Center (SCSGC) at the University of Georgia, as detailed in Table 4.1. Proteins were expressed from sequenced plasmids or, in the case of the clones from SCSGC, expressed proteins were verified by MS. They were investigated to determine activity against the α-glucan substrates maltose, starch, glycogen, pullulan, and synthetic substrate analogs. The cyclodextrin glucanotransferase (PF0478) and cyclodextrinase (PF1939) activities were verified against previously published reports (23, 73).

PF1108 is annotated as a 'putative α-dextrin-endo-1,6-α-glucosidase' and its strongest domain homology is to the α/β superfamily hydrolase. The α/β superfamily hydrolase is a large grouping of enzyme with greatly varied activities including esterases (44), hydrolases (34, 45), oxygenases (18), and acylases (74). PF1108 also has amino acid sequence homology to esterase family D, and is classified into the CAZY esterase family 1 (CE1). The α-dextrin hydrolase annotation comes from a 60-70 amino acid region near the N-terminus, which is related to a glycogen de-branching domain. SMART analysis (60) also identified a weak hit in the same N-terminal region related to a carbohydrate binding module. This enzyme has a homolog in *Thermococcus kodakarensis* KOD1, TK0522 (80%/87%, similarity/identity), which is annotated as a putative esterase. These two enzymes comprise the only two archaeal proteins in CE1. All other BLAST results were for bacterial enzymes with up to (41%/57%, similarity/identity). PF1108 is located on the same strand and preceding PF1109 and PF1110. PF1109 is a large protein consisting of 969 amino acids with a conserved domain from about 200-500 which is a member of COG1572. It is most closely related to *P. abyssi* PAB1790 and *T. maritima* TM1841 and a 207 amino acid region from 227-433 is related to the *Deinococcus geothermalis* putatitve type III α-amylase, catalytic region. PF1110 is a shorter protein of 172 amino acids, which is also related to PAB1790 and TM1841, as well as PF1109. In fact, a domain related to PF1110 appears repeated in PF1109 from amino acids 66-233 and 269-432, with 60% and 57% similarity, respectively. Another explanation is that PF1110 is actually PF1109 either as an extension of the protein or by a sequencing error, which have been previously noted in other regions of the *P. furiosus* genome (41). Error or not, the presence of a second

possible α-glucan processing gene adjacent to PF1108, increases the confidence of α-glucan processing.

Conserved domain analysis of PF1460 identified two overlapping domains, each indicative of α-glucan activity - pfam06202 (Amylo-α-1,6-glucosidase) and COG3408 (glycogen de-branching enzyme). The enzyme is related to another *P. furiosus* 'hypothetical protein', PF1746, which is centrally located in the Mal-I operon. These enzymes are related to a number of other archaeal enzymes, including *P. horokoshii* PH1222 (41/57% similarity/identity), *P. abyssi* PAB0689 (44/60, similarity/identity) and *Methanosarcina acetivorans* C2A MA1640 (36/52, similarity/identity), as well as some bacterial enzymes including *Bacillus halodurans* BH3691 (37/54, similarity/identity) and *Carboxydothermus hydrogenoformans* Z-2901 CHY_0669 (40/56, similarity/identity). Glcyogen de-branching enzymes are located in GHF13, and alignment of PF1746 with the amylosucrase from *Neisseria polysaccharea* 85322, a GHF13 enzyme with known catalytic residues and crystal structure, did not identify conserved catalytic residues of PF1746 (49).

PF1747 is located adjacent to PF1746 in the same operon and is annotated as a 'hypothetical protein'. Domain analysis indicates that there is a region similar to COG2152, which is a predicted glycosylase (66), and domain architecture is similar to glycosyl hydrolase family 32, which contains invertases and fructose related hydrolases. Representative proteins of the COG2152 are all of hyperthermophilic origin (both archaeal and bacterial), and related proteins are represented by fungi, archaea, and bacteria, as identified by BLAST, but no functional assignment could be determined by BLAST. Distantly related protein activities include sucrose hydrolase (sucrase), sucrose-6-phosphate hydrolase, frustosidase, and endo-α-L-arabinosidase. Alignment of PF1747 with the GHF32 β-fructosidase from *T. maritima* did not result in conservation of the known catalytic residues Asp17 and Glu190 (1).

PF1393 is annotated as a 'hypothetical protein', however, it is located in GHF57, which contains proteins of α-amylase, 4-α-glucanotransferase, α-galactosidase, and amylopullulanase activities. The assignment to GHF57 comes from a 400 amino acid sequence at the N-terminus. This was confirmed by all sequence analyses. Analysis of fifty-nine GHF57 ORFs identified five conserved regions in the hydrolase family and

seven subfamilies of enzymes, which may reflect differences in enzyme specificities (76). The sequenced *Pyrococcus* species and *T. kodakarensis* each have homologous genes: PAB1857, PH1386, and TK1436, which are the only archaeal proteins with greater than 20% identity. These four enzymes make up one of seven sub-families of GHF57 (76). PF1393 also has a helix-hairpin-helix (HhH) domain at the C-terminus, which is similar to the ERCC4-type nuclease, which has been associated with nucleotide excision repair defects (53). The HhH domain has been associated with non-sequence-specific binding of DNA (13, 28). The role of this domain in an α-glucan activity glycosidase is unclear. The domain is not conserved in the homologs – PF1393 and PAB1857 have a single HhH domain, TK1436 has a repeat HhH-HhH, and PH1386 has no HhH domain. Given the absence of this HhH domain in related proteins, this region is probably not involved in catalysis, but instead might play a role in enzyme stability or substrate binding.

Bioinformatic analysis of the *P. furiosus* genome identified the ORF PF0870 which contains a domain located near the N-terminus spanning 290 of the 597 amino acids, which has similarity to members of GHF57 (76). There three other proteins closely related to PF0870 in published microbial genomes proteins, all from other Thermococcales – PH1023 of *P. horikoshii*, PAB0644 of *P. abyssi*, and TK1743 of *Thermococcus kodakarensis* (76). Each of these homologs have at least 60% identity and 75% similaritiy at the amino acid level. Comparison of PF0870 with GHF57 primary protein sequences was used to determine the relationship between these enzymes. The conserved nucleophilic residue of GHF57 of *Thermococcus litoralis* 4-α-glucanotransferase is Glu123 (31) and the acid-base residue is Asp214 (32). Sequence alignments of GHF57 enzymes have identified proposed nucleophilic (E153) and acid-base D253 residues for PF0870 (76). The acid-base region is conserved with the other three homologs, with a consensus sequence of 251-SSDLESLVANP that is strictly conserved in the *Pyrococcus* species and with three point mutations for TK1743 as shown in Table 4.2.

Transcriptional Analysis – *P. furiosus* was grown on a number of carbohydrate sources to investigate the transcriptional response to different α-glucan linkages. Those substrates upon which growth to a density greater than 10^8 cells/mL occurred (trehalose,

starch, glycogen, maltose, pullulan, cellobiose, and maltose/cellobiose), as shown in Figure 4.1, were selected for transcriptional analysis. Despite possessing enzymes capable of hydrolyzing cyclodextrins (CDase PF1939 and CGTase PF0478) (23, 73), *P. furiosus* did not grow significantly above the control when exposed to this substrate. *P. furiosus* also possesses enzymes capable of processing sucrose (4) and raffinose (67), but was not able to efficiently utilize these substrates for growth.

During growth on the polysaccharides starch, glycogen, and pullulan, 118 genes were up-regulated relative to growth on maltose, and many of the same genes were induced for all conditions, as shown in the Venn Diagram in Figure 4.2. Of these 118 genes, 49 genes were up for all polysaccharide conditions; 38 of which were from the locus PF0682-PF0783. The genes in this locus encode a number of small proteins/peptides (< 80 amino acids), as well as a number of dehydrogenases and reductases. Many of the proteins encoded by *P. furiosus* in the region of PF0682-0736 are unique to *P. furiosus* and not found even in close relatives such as *P. horikoshii*, *P. abyssi*, or *T. kodakarensis*. Of the 49 genes up-regulated, *T. kodakarensis* contained the most homologues with 32. All genes, which changed more than two-fold relative to maltose, are listed in Appendix 4.A.

Genes identified by bioinformatics analysis were investigated by transcriptional analysis as well. There was no differential expression for PF0478, PF1393, PF1460 for any conditions. These genes may be constitutively expressed or the conditions for genes expression have not been identified. The cyclodextrin glucanotransferase (PF0478) generates circularized α-glucan molecules, which are thought to be a competitive advantage by sequestering α-glucan compounds as cyclic molecules that fewer other organisms can take up and hydrolyze (23). Average LSM expression level of PF0478 is about 4-fold below the average protein expression (12^{th} percentile), and this gene may always be transcribed at low levels. PF1393, is highly transcribed (96^{th} percentile), but not differentially expressed. PF1746 was also highly transcribed (90^{th} percentile), but differentially expressed up on all the disaccharides versus the polysaccharides, especially starch and glycogen. PF1460, which is related to glycogen de-branching enzymes, had low to moderate expression for all conditions (30^{th} percentile) and was not differentially expressed under any condition. Each of these enzymes is related to glycogen de-

branching enzymes and could be involved in intracellular processing of these substrates in a specific role not explored. PF1393 has low activity against starch, converting it into large maltodextrins, supporting this role. PF1746 is found in the Mal-I operon which is responsible for transport of trehalose and maltose and probably plays an undetermined role in modifying these compounds.

PF1108 was differentially expressed up on trehalose versus each of the polysaccharides, however, the average expression level (9^{th} percentile) was so low that the expression was not significant. Additionally, no activity against trehalose was detected by PF1108. PF1109 and PF1110 are located just downstream of PF1108, however, the transcriptional response of these genes is drastically different from PF1108. Both PF1109 (99^{th} percentile) and PF1110 (97^{th} percentile) were highly expressed, and PF1109 had the 10^{th} highest average expression level of all genes. PF1109 was up-regualted on trehalose relative to cellobiose by almost 2-fold, whereas PF1110 was up-regulated two-fold on each of the polysaccharides relative to maltose. This disparity in expression levels indicates that PF1108-PF1110 does not comprise an operon.

PF1747 (84^{th} percentile) was highly expressed up on the disaccharides versus the polysaccharides, especially starch and glycogen, in the same pattern as PF1746. This enzyme is annotated as a 'hypothetical protein' and BLAST analyses groups it with uncharacterized 'glycosylase' enzymes. Hydrolysis of substrates (maltose and trehalose) recognized by the Mal-I operon was not detected, nor was hydrolysis of turanose or sucrose.

Transcriptional analysis of PF0870 relative to other genes was not possible because the gene was not spotted on microarray slides due to a printing error. Instead, real-time PCR was used to quantitate the transcriptional response of the gene under the different growth conditions, as shown in Table 4.3. Expression was highest on starch and cellobiose (5.3- and 4.4-fold higher than pullulan, respectively) and lowest on pullulan.

Response to Disaccharides – Growth of *P. furiosus* on the oligosaccharides maltose, cellobiose, trehalose, and mixture of cellobiose/maltose resulted in 95 total genes up-regulated relative to maltose for the three comparisons, as shown in Fig 4.3. When the cells were grown exclusively on cellobiose or in a cellobiose/maltose mixture,

the genes to process cellobiose were highly differentially expressed. Transcriptional expression data of selected genes and their $\log_2$-fold changes relative to maltose are shown in Table 4.4. These genes are involved in the transport and hydrolysis of cellobiose and are expressed at similar levels, regardless of the presence of maltose. When maltose was present either by itself or with cellobiose, there were few genes up-regulated two-fold or more relative to cellobiose (see Table 4.5). Transcriptional response indicates that there is no catabolite repression response for growth of *P. furiosus* on this mixture of substrates. In fact, *P. furiosus* seems to exhibit a scavenger response, in which it is utilizing both maltose and cellobiose simultaneously.

Trehalose pathway – *P. furiosus* possesses the Mal-I operon, which is specific for transport of maltose and trehalose (29, 71). However, enzymes for intracellular trehalose processing appear to be lacking from the *P. furiosus* genome. The only identifiable trehalose-related protein annotated in the genome is encoded by the trehalose synthase gene (PF1742), which is identical to the novel glycosyl transferase trehalose synthase of *Thermococcus litoralis* (56). This enzyme reversibly synthesizes trehalose and ADP from glucose and ADP-glucose. The enzyme is also related to the *P. horikoshii* trehalose synthase (PH1035), which synthesizes trehalose, liberating UDP from UDP-glucose and glucose (59). The PH1035 enzyme was also able to hydrolyze trehalose to glucose and dephosphorylate glucose-1-phophate. A survey of trehalose active genes from ExPASy (3) identified the following enzymes used by other species to process this disaccharide: trehalase (EC 3.2.1.28), α-α-phosphotrehalase (EC 3.2.1.93), trehalose-phosphate synthase (EC 2.4.1.15/2.4.1.36), trehalose phosphorylase (EC 2.4.1.231), maltooligosyl trehalose trehalohydrolase (EC 3.2.1.141), trehalose phosphatase (EC 3.2.1.12), and trehalose phosphorylase (EC 2.4.1.64). Bioinformatic analysis of the *P. furiosus* genome, however, did not identify the presence of homologs to any of these known trehalose-active enzymes. α-α-Phosphotrehalase and maltooligosyl trehalose trehalohydrolase are members of the glycosyl hydrolase family 13, which contains the extracellular α-amylase PF0477, the cyclodextrin glucanotransferase PF0478, and the cyclodextrinase PF1939 – none of which have been shown to have α-α-phosphotrehalase activity. Trehalase is a member of GHF 37 and 65, however, no archaeal proteins are present in family 37 and

the only archaeal protein present in family 65 is the *P. horikoshii* (PH0746), which has no homolog in *P. furiosus*.

Based upon growth of *P. furiosus* on trehalose, *P. furiosus* is capable of utilizing this carbohydrate as an energy source. No trehalose hydrolyzing enzymes were identified in the genome, and intracellular trehalose processing by *P. furiosus* probably occurs by the trehalose synthase protein (PF1742). This enzyme could process trehalose with ADP by the reverse reaction, liberating ADP-glucose and glucose as described for *T. litoralis* (56). While it appears that this enzyme does not hydrolyze trehalose to glucose or have glucose dephosphorylation activity, the related trehalose synthase from *P. horikoshii* has been shown to process trehalose and glucose-1-phosphate to glucose. Transcriptional profiles of *P. furiosus* grown of trehalose does not reveal any genes expressed at higher levels on trehalose relative to maltose, that appear to be involved in trehalose utilization (See Appendix 4.A, Table 4.A.14).

Biochemical analysis of predicted α-glucan active ORFs – Activity of the ORFs identified by bioinformatic analysis were investigated biochemically for functional assignment and characterization. These ORFs were screened on the polysaccharides amylose, glycogen, pullulan, and starch as well as disaccharides and synthetic substrates. Activity was assayed in the presence of cellular extract processed through a 1 kDa MWCO membrane, to determine if any cofactors were required. Each enzyme was expressed in the soluble fraction as determined by SDS-PAGE, preceded by heat-treatment of the soluble fraction to eliminate heat-labile *E. coli* proteins.

PF1108 was screened against 1% β-cyclodextrin, starch, pullulan, and glycogen, using the DNS reducing sugar assay (48) and TLC (plates not shown); no activity was detected relative to controls for any condition tested. Activity was also assayed in the presence of cellular extract, but no activity was detected activity was detected. Activity was screened on a number of synthetic substrates (*p*NPαGal, *p*NPαGlc, *p*NPαMan, *p*NPαGlc$_5$, *p*NPαMalt, *p*NPβGlc, and *p*NPβXyl), but no activity was detected at pH 5.6 or 7.2. To determine if there was any esterase activity, this enzyme was screened on the synthetic substrates *p*NP-α-mannopyranoside, *p*NP-acetate, *p*NP-propionate, *p*NP-

caprate, *p*NP-palmitate and *p*NP-butyrate, but no hydrolysis relative to controls was observed.

Despite successful cloning of the PF1460 and numerous expression trials, PF1460 was not able to be expressed in soluble form, and thus was not investigated. No hydrolysis was detected for PF1746 on 1% glycogen, 1% starch, 1% pullulan, or 1% amylose, with or without cell extract, as determined by TLC. Nor was activity detected on the synthetic substrates *p*NPαGlc$_5$, *p*NPαMalt, *p*NPαMan, *p*NPβGlc, *p*NPβXyl, *p*NPαGlc, and *p*NPαGal. Based upon these findings, neither PF1108, nor PF1746 appear to be glycogen de-branching enzyme and their function remains to be determined.

Bioinformatic analysis of PF1747 identified some similarity to the β-fructosidase superfamily, but alignment with the GHF32 β-fuctosidase from *T. maritima* did not confirm conservation of active site residues. Still, PF1747 activity was tested on α-glucan polysaccharides and the disaccharides maltose, sucrose, trehalose, or turanose, but no activity was detected by TLC. This is clear confirmation that enzyme is not a β-fructosidase (invertase). In addition, no activity was detected on the synthetic substrates related to other specificities.

PF1393 is a GHF57 glycosyl hydrolase, but the substrate specificity has not been characterized to date. TLC indicated that there was some hydrolysis of starch at pH 7.2, however, no activity was detected on pullulan, glycogen, or amylase at this condition. TLC analysis indicates that there may be some hydrolysis of starch with the products being large α-glucan molecules as they only migrate just past the starch standard. No activity was detected on the synthetic substrates *p*NPαGlc5, *p*NPαMalt, *p*NPαMan, *p*NPβGlc, *p*NPβXyl, *p*NPαGlc, and *p*NPαGal at pH 5.6 or 7.2.

A combination of PF1108, PF1393, PF1746, and PF1747 heat-treated crude extracts were incubated with starch, pullulan, glycogen, and amylose at 90°C and pH 6.6. As determined by TLC, there was no hydrolysis of amylose, pullulan, or starch, however, there were hydrolysis products generated from glycogen. While none of the enzyme individually hydrolyzed glycogen, the combination of enzymes was able to hydrolyze glycogen as shown in Figure 4.4. Assays commonly look at only one enzyme and condition, in the cell these enzymes are present and interact with the other proteins. It is

likely that there is a synergistic effect both *in vivo* and *in vitro* that aids hydrolysis of these substrates using combinations of enzymes, such as those examined here.

Enzymatic activity of the recombinant PF0870 was screened using the synthetic substrates characteristic of GHF57 enzymes and other common glycosidases; highest activity was detected on *p*NP-α-maltopyranoside, as shown in Table 4.6. The enzyme temperature optimum was determined to be > 110°C and pH optimum is about pH 7.0, as shown in Figure 4.5 and Figure 4.6. Based upon the activity toward *p*NPαMalt, hydrolysis of maltooligosaccharides, starch, pullulan, and glycogen were tested. On *p*NPαMalt at 90°C in sodium phosphate buffer, pH 7.0, the Michaelis-Menten kinetic parameters of 962 s^{-1} k_{cat} and 82.7 μM K_M were calculate by regression to fit the Michaelis-Menten equation. The enzyme does not appear to have any transglycosylation acitivity, does not hydrolyze starch, glycogen, pullulan or large maltooligosaccharides, and only seems to hydrolyze maltotriose into glucose and maltose, as determined by the TLC (see Figure 4.7). The presence of cations or EDTA appears to have minimal effect on catalytic activity of the protein, as shown in Table 4.7.

The enzyme activity most closely resembles that of a β-amylase, however, there are some notable differences. β-Amylases are found in plants, fungi, and some bacteria and successively cleave maltose from the non-reducing end of starch, amylase, and maltodextrins. Bacterial β-amylases tend to have activity against polysaccharides such as starch, whereas the plant β-amylases tend to only have activity against maltodextrins (12). To date, the most thermostable β-amylase characterized was from *Clostridium thermosulfurogenes*, which is optimally active at 75°C (30). PF0870 now represents the most thermally stable β-amylase characterized to date. This enzyme has high-specificity for maltotriose and could be used to selectively reduce the maltotriose content in corn syrup sweeteners. The role of these protein *in vivo* is still not clear, however, the enzyme may be the final processing enzyme in the intracellular hydrolysis of maltodextrins before transfer of a glucose unit from maltose to maltodextrin by PF0272 (41).

Pullulan up-take and utilization – Based upon functional genomics and biochemical data, Lee *et al.* (41) proposed a utilization pathway for starch in *P. furiosus* in which starch is hydrolyzed to maltodextins and maltose by PF1935* and transported

into the cell for hydrolysis to glucose or glucophosphates by α-glucanotransferase (PF0272) and α-glucan phosphorylase (PF1535). The extracellular hydrolysis products are transported across the membrane by the Mal-I operon (PF1739-1750) and associated ABC-transporter (PF1739-41, 1744) or the Mal-II operon (PF1933-1939). The Mal-I binding protein (PF1739) does not have affinity towards maltotriose, but instead binds maltose and trehalose (37, 43), whereas the Mal-II binding protein (PF1938) had affinity for maltotriose and larger maltodextrins. Expression of these operons is controlled by the TrmB protein, which is thought to regulate both operons based upon the sugar binding specificity (38, 42, 43).

Using a combination of functional genomics, bioinformatics analysis, and biochemical characterization, a utilization pathway for pullulan by the hyperthermophilic archaeon *P. furious* can be proposed, as shown in Figure 4.8. Pullulan is hydrolyzed extracellularly by the amylopullulanase (PF1935*) (15, 41). These hydrolysis products are transported into the cell through either the Mal-I or Mal-II operon, but probably more dependent upon the Mal-I operon. Following transport, the resulting oligosaccharides (panose, isomaltose, isopanose, maltotriose, etc.) are then hydrolyzed by the intracellular α-glucosidase (PF0132). This pathway is proposed based upon the functional expression data for PF0132, PF0272, PF0312, PF0477, PF1535 (Table 4.8), PF1739-1751 (Figure 4.9), and PF1933-1939 (Figure 4.10) and past (10) and recent characterization and elucidation of the function of PF0132. The $\log_2$-fold change of the transcriptional response of PF1739-1751 relative to maltose is shown in Figure 4.8. Here, the genes expressed similarly on starch and glycogen, whereas pullulan induces expression similar to maltose, which is at a higher expression level than starch and glycogen as judged by LSM estimates. The Mal-II operon in PF1933-1939 is strongly expressed under all conditions and transcriptionally does not discriminate between substrates, as shown in Figure 4.10 and Figure 4.11.

The LSM values of the individual genes proposed to be involved in starch processing by Lee *et al.* (41) are shown in the Table 4.8. The proposed pullulan processing pathway is shown in Figure 4.8, with changes from the starch pathway boxed. For both starch and pullulan, initial hydrolysis occurs extracellularly by the amylopullulanase PF1935*, and the oligosaccharides are then transported by either Mal-I

or Mal-II ABC-transporters, depending on the hydrolysis product and its affinity to the binding proteins. For starch, hydrolysis to glucose occurs through a transferase reaction catalyzed by PF0272, which transfers a glucose unit to a maltodextrin generating glucose and maltodextrin$_{n+1}$ (41). PF0272 is expressed almost 4-fold less on pullulan than on maltose (LF 1.9) indicating this pathway is probably not the dominant pathway. Conversely, PF0132 is expressed most highly on pullulan (LF 1.8 v. maltose or starch) indicating that this pathway may be more important in processing the pullulan oligosaccharides. Based upon these results, it appears that pullulan hydrolysis products are preferentially processed intracellularly by the α-glucosidase (PF0132) instead of the α-glucanotransferase (PF0272) as is the case for starch and maltose.

Based upon these transcriptional leads, the activity of PF0132 was investigated on substrates containing α-1,6-glucosidic linkages. Given the protein size and transcriptional response, we hypothesized that PF0132 coded for the *P. furiosus* α-glucosidase, previously studied in its native form by Costantino *et al.* (10), and for which no gene sequence had been reported. The native form of PF0132 was purified 92-fold to near homogeneity (see Table 4.9), as described in the methods section and shown in Figure 4.12. Using purified, catalytically active, native PfαGlc, we definitively assigned the α-glucosidase to ORF PF0132 as determined by mass spectroscopy with 40.2% coverage. Biochemical analysis of PfαGlc shows that it is capable of hydrolyzing the natural substrates panose and isomaltose, oligosaccharide hydrolysis products of pullulan, as determined by TLC (Figure 4.13). Biochemical analysis of PfαGlc shows that there is a similar catalytic efficiency for panose as for maltose (see Table 4.10), further supporting this proposed pathway. This enzyme and homologs in *P. horikoshii*, *P. abyssi*, *T. kodakarensis*, and *Aeropyrum pernix* represents a new glycosyl hydrolase family (24-26), the 107th to date. This enzyme is extremely thermally stable with a melting temperature >110°C, and melts at 100°C in the presence of 1 M guanidine hydrochloride (36). It has broad substrate specificity makes it an ideal enzyme for the final processing step for generating glucose from α-glucans for processes such as high-fructose corn syrup and fermentation to ethanol (11).

PF0132 is most highly expressed on pullulan, as shown in Figure 4.14, and the genes encoded by PF0133 and PF0134 have similar expression patterns to the α-

glucosidase, PF0132. PF0133 codes for a 21.3 kDa protein with no known function; it is located upstream of PF0132, separated by 6 base pairs. Homologs are present in the Thermococcales, methanogenic archaea, carboxydotropic bacteria, and proteobacteria, indicating that the enzyme is probably not used in carbohydrate processing unless it is for turnover of storage polymers such as glycogen. PF0132 and PF0133 are located adjacent to each other in the *P. abyssi* and *P. horikoshii* genomes, but distantly located in *T. kodakarensis* KOD1. Enzymatic studies revealed no activity by recombinant PF0133 on pullulan, starch, or glycogen. PF0134 is on the opposite strand, and convergent to PF0132-0133; it terminates by overlapping PF0133 for 6 base pairs. It codes for a 22.7 kDa protein, which has weak homology in the C-terminus to a dolichol kinase domain, SEC59, and enzyme that transfers a phosphate group from cytidyl phosphate to a dolichol moiety. Given this homology, it is not surprising that the enzyme also has homology to bacterial phosphatidate cytidylyltransferase enzymes. In eukaryotes, dolichols have been associated with glycosylation pathways required for secretion of proteins (52). The role of this protein in *P. furiosus* has yet to be determined.

CONCLUSIONS

Using functional genomics, transcriptionally responsive genes were investigated for their role in processing α-glucans. This led to the functional assignment of PF0132 as the *P. furiosus* α-glucosidase. In addition, transcriptional response indicated that the gene was present during pullulan hydrolysis and subsequent biochemical characterization confirmed the enzyme activity against α-1,6-linked glucans such as panose and isomaltose. BLAST, conserved domains, and domain family bioinformatic analyses were utilized in screening the *P. furiosus* genome to identify additional putative α-glucan active enzymes. Six ORFs were identified, of which five were annotated as 'hypothetical proteins'. Five of these genes were cloned, expressed, and investigated biochemically. PF1108, PF1746, and PF1747 showed no activity towards any natural or synthetic α-glucan substrates. PF1393 appears to generate large maltodextrins from starch as determined by TLC of hydrolysis products. Another ORF, PF0870, was found to be active against maltotriose, hydrolyzing it into maltose and glucose and the synthetic

substrate *p*NPαMalt. Combining these tools available with biochemical characterization, we have been able to define characterizations for two of the 870 hypothetical proteins in the *P. furiosus* genome. The methodology combining functional genomics, bioinformatics, and biochemical characterization is applicable for any organism for which a gene chip is available and can play an important role in understanding the remaining unannotated ORFs of genomes.

ACKNOWLEDGMENTS

Funding was supplied by the National Science Foundation and the Department of Energy.

REFERENCES

1. **Alberto, F., C. Bignon, G. Sulzenbacher, B. Henrissat, and M. Czjzek.** 2004. The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. J. Biol. Chem. **279:**18903-18910.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic Local Alignment Search Tool. J. Mol. Biol. **215:**403-410.
3. **Appel, R. D., A. Bairoch, and D. F. Hochstrasser.** 1994. A new generation of information retrieval tools for biologists - The example of the ExPASy www server. Trends Biochem. Sci. **19:**258-260.
4. **Badr, H. R., K. A. Sims, and M. W. W. Adams.** 1994. Purification and characterization of sucrose alpha-glucohydrolase (invertase) from the hyperthermophilic archaeon *Pyrococcus furiosus*. Syst. Appl. Microbiol. **17:**1-6.
5. **Barrangou, R., M. A. Azcarate-Peril, T. Duong, S. B. Conners, R. M. Kelly, and T. R. Klaenhammer.** 2006. Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays. P. Natl. Acad. Sci. USA **103:**3816-3821.
6. **Bauer, M. W., L. E. Driskill, W. Callen, M. A. Snead, E. J. Mathur, and R. M. Kelly.** 1999. An endoglucanase, eglA, from the hyperthermophilic archaeon *Pyrococcus furiosus* hydrolyzes beta-1,4 bonds in mixed-linkage (1 -> 3),(1 -> 4)-beta-D-glucans and cellulose. J. Bacteriol. **181:**284-290.
7. **Bork, P.** 2000. Powers and pitfalls in sequence analysis: The 70% hurdle. Genome Res. **10:**398-400.
8. **Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger, and R. M. Kelly.** 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J. Biol. Chem. **278:**7540-7552.
9. **Conners, S. B., C. I. Montero, D. A. Comfort, K. R. Shockley, M. R. Johnson, S. R. Chhabra, and R. M. Kelly.** 2005. An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. J. Bacteriol. **187:**7267-7282.
10. **Costantino, H. R., S. H. Brown, and R. M. Kelly.** 1990. Purification and characterization of an alpha-glucosidase from a hyperthermophilic archaebacterium, *Pyrococcus furiosus*, exhibiting a temperature optimum of 105 degrees C to 115 degrees C. J. Bacteriol. **172:**3654-3660.
11. **Crabb, W. D., and J. K. Shetty.** 1999. Commodity scale production of sugars from starches. Curr. Opin. Microbiol. **2:**252-256.
12. **Dicko, M. H., M. J. F. Searle-van Leeuwen, G. Beldman, O. G. Ouedraogo, R. Hilhorst, and A. S. Traore.** 1999. Purification and characterization of beta-amylase from *Curculigo pilosa*. Appl. Microbiol. Biot. **52:**802-805.
13. **Doherty, A. J., L. C. Serpell, and C. P. Ponting.** 1996. The helix-hairpin-helix DNA-binding motif: A structural basis for non-sequence-specific recognition of DNA. Nucleic Acids Res. **24:**2488-2497.
14. **Dong, G. Q., C. Vieille, A. Savchenko, and J. G. Zeikus.** 1997. Cloning, sequencing, and expression of the gene encoding extracellular alpha-amylase

from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. Appl. Environ. Microb. **63:**3569-3576.

15. **Dong, G. Q., C. Vieille, and J. G. Zeikus.** 1997. Cloning, sequencing, and expression of the gene encoding amylopullulanase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. Appl. Environ. Microb. **63:**3577-3584.
16. **Driskill, L. E., M. W. Bauer, and R. M. Kelly.** 1999. Synergistic interactions among beta-laminarinase, beta-1,4-glucanase, and beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* during hydrolysis of beta-1,4-, beta-1,3-, and mixed-linked polysaccharides. Biotechnol. Bioeng. **66:**51-60.
17. **Fiala, G., and K. O. Stetter.** 1986. *Pyrococcus furiosus* sp. nov represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 degrees C. Arch. Microbiol. **145:**56-61.
18. **Fischer, F., S. Kunne, and S. Fetzner.** 1999. Bacterial 2,4-dioxygenases: New members of the alpha/beta hydrolase-fold superfamily of enzymes functionally related to serine hydrolases. J. Bacteriol. **181:**5725-5733.
19. **Galperin, M. Y.** 2001. Conserved 'hypothetical' proteins: new hints and new puzzles. Compar. Funct. Genom. **2:**14-18.
20. **Galperin, M. Y., and E. V. Koonin.** 2004. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res. **32:**5452-5463.
21. **Gao, J., M. W. Bauer, K. R. Shockley, M. A. Pysz, and R. M. Kelly.** 2003. Growth of hyperthermophilic archaeon *Pyrococcus futiosus* on chitin involves two family 18 chitinases. Appl. Environ. Microb. **69:**3119-3128.
22. **Gueguen, Y., W. G. B. Voorhorst, J. van der Oost, and W. M. deVos.** 1997. Molecular and biochemical characterization of an endo-beta-1,3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Biol. Chem. **272:**31258-31264.
23. **Hashimoto, Y., T. Yamamoto, S. Fujiwara, M. Takagi, and T. Imanaka.** 2001. Extracellular synthesis, specific recognition, and intracellular degradation of cyclomaltodextrins by the hyperthermophilic archaeon *Thermococcus* sp strain B1001. J. Bacteriol. **183:**5050-5057.
24. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochem. J. **280:**309-316.
25. **Henrissat, B., and A. Bairoch.** 1993. New families in the classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochem. J. **293:**781-788.
26. **Henrissat, B., and A. Bairoch.** 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. **316:**695-696.
27. **Hobbie, J. E., R. J. Daley, and S. Jasper.** 1977. Use of nuclepore filters for counting bacteria by fluorescence microscopy. Appl. Environ. Microb. **33:**1225-1228.
28. **Hollis, T., Y. Ichikawa, and T. Ellenberger.** 2000. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. EMBO J. **19:**758-766.
29. **Horlacher, R., K. B. Xavier, H. Santos, J. DiRuggiero, M. Kossmann, and W. Boos.** 1998. Archaeal binding protein-dependent ABC transporter: Molecular and

biochemical analysis of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. J. Bacteriol. **180:**680-689.

30. **Hyun, H. H., and J. G. Zeikus.** 1985. General biochemical characterization of thermostable extracellular beta-amylase from *Clostridium thermosulfurogenes*. Appl. Environ. Microb. **49:**1162-1167.
31. **Imamura, H., S. Fushinobu, B. S. Jeon, T. Wakagi, and H. Matsuzawa.** 2001. Identification of the catalytic residue of *Thermococcus litoralis* 4-alpha-glucanotransferase through mechanism-based labeling. Biochemistry-US **40:**12400-12406.
32. **Imamura, H., S. Fushinobu, M. Yamamoto, T. Kumasaka, B. S. Jeon, T. Wakagi, and H. Matsuzawa.** 2003. Crystal structures of 4-alpha-glucanotransferase from *Thermococcus litoralis* and its complex with an inhibitor. J. Biol. Chem. **278:**19378-19386.
33. **Jorgensen, S., C. E. Vorgias, and G. Antranikian.** 1997. Cloning, sequencing, characterization, and expression of an extracellular alpha-amylase from the hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis*. J. Biol. Chem. **272:**16335-16342.
34. **Juers, D. H., R. E. Huber, and B. W. Matthews.** 1999. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between beta-galactosidase and other glycohydrolases. Protein Sci. **8:**122-136.
35. **Kaper, T., C. H. Verhees, J. H. G. Lebbink, J. F. T. van Lieshout, L. D. Kluskens, D. E. Ward, S. W. M. Kengen, M. M. Beerthuyzen, W. M. de Vos, and J. van der Oost.** 2001. Characterization of beta-glycosylhydrolases from *Pyrococcus furiosus*, p. 329-346, Hyperthermophilic Enzymes, Pt A, vol. 330.
36. **Kelly, R. M., S. H. Brown, I. I. Blumentals, and M. W. W. Adams.** 1992. Characterization of enzymes from high-temperature bacteria, p. 23-41. *In* M. W. W. Adams and R. M. Kelly (ed.), Biocatalysis at extreme temperatures: Enzyme systems near and above 100 C, vol. 498. American Chemical Society, Washington DC.
37. **Koning, S. M., W. N. Konings, and A. J. M. Driessen.** 2001. Biochemical evidence for the presence of two α-glucoside ABC-transport systems in the hyperthermophilic archaeon *Pyrococcus furiosus*. Archaea **1:**19-25.
38. **Krug, M., S.-J. Lee, K. Diederichs, W. Boos, and W. Welte.** 2006. Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. J. Biol. Chem. **281:**10976-10982.
39. **Laderman, K. A., K. Asada, T. Uemori, H. Mukai, Y. Taguchi, I. Kato, and C. B. Anfinsen.** 1993. Alpha-amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus* - Cloning and sequencing of the gene and expression in *Escherichia coli*. J. Biol. Chem. **268:**24402-24407.
40. **Laderman, K. A., B. R. Davis, H. C. Krutzsch, M. S. Lewis, Y. V. Griko, P. L. Privalov, and C. B. Anfinsen.** 1993. The purification and characterization of an extremely thermostable alpha-amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. J. Biol. Chem. **268:**24394-24401.
41. **Lee, H.-S., K. R. Shockley, G. J. Schut, S. B. Conners, C. I. Montero, M. R. Johnson, C.-J. Chou, S. L. Bridger, N. Wigner, S. D. Brehm, F. E. Jenney Jr., D. A. Comfort, R. M. Kelly, and M. W. W. Adams.** 2006. Transcriptional and

biochemical analysis of starch metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **188:**2115-2125.

42. **Lee, S. J., A. Engelmann, R. Horlacher, Q. H. Qu, G. Vierke, C. Hebbeln, M. Thomm, and W. Boos.** 2003. TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. J. Biol. Chem. **278:**983-990.
43. **Lee, S. J., C. Moulakakis, S. M. Koning, W. Hausner, M. Thomm, and W. Boos.** 2005. TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers. Mol. Microbiol. **57:**1797-1807.
44. **Liu, P., Y. F. Wang, H. E. Ewis, A. T. Abdelal, C. D. Lu, R. W. Harrison, and I. T. Weber.** 2004. Covalent reaction intermediate revealed in crystal structure of the *Geobacillus stearothermophilus* carboxylesterase Est30. J. Mol. Biol. **342:**551-561.
45. **MacGregor, E. A., H. M. Jespersen, and B. Svensson.** 1996. A circularly permuted alpha-amylase-type alpha/beta-barrel structure in glucan-synthesizing glucosyltransferases. FEBS Lett. **378:**263-266.
46. **Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri, C. DeWweese-Scott, L. Y. Geer, M. Gwadz, S. Q. He, D. I. Hurwitz, J. D. Jackson, Z. X. Ke, C. J. Lanczycki, C. A. Liebert, C. L. Liu, F. Lu, G. H. Marchler, M. Mullokandov, B. A. Shoemaker, V. Simonyan, J. S. Song, P. A. Thiessen, R. A. Yamashita, J. J. Yin, D. C. Zhang, and S. H. Bryant.** 2005. CDD: a conserved domain database for protein classification. Nucleic Acids Res. **33:**D192-D196.
47. **Matsui, I., K. Ishikawa, H. Ishida, and Y. Kosugi.** August 6, 1999 1999. Methods for making and using a thermophilic enzyme as a β-glycosidase. United States Patent.
48. **Miller, G. L.** 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. **31:**426-428.
49. **Mirza, O., L. K. Skov, M. Remaud-Simeon, G. P. de Montalk, C. Albenne, P. Monsan, and M. Gajhede.** 2001. Crystal structures of amylosucrase from *Neisseria polysaccharea* in complex with D-glucose and the active site mutant Glu328Gln in complex with the natural substrate sucrose. Biochemistry-US **40:**9032-9039.
50. **Nanavati, D. A., K. Thirangoon, and K. A. Noll.** 2006. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. Appl. Environ. Microb. **72:**1336-1345.
51. **Nguyen, T. N., A. D. Ejaz, M. A. Brancieri, A. M. Mikula, K. E. Nelson, S. R. Gill, and K. M. Noll.** 2004. Whole-genome expression profiling of *Thermotoga maritima* in response to growth on sugars in a chemostat. J. Bacteriol. **186:**4824-4828.
52. **Orlean, P.** 1992. Enzymes that recognize dolichols participate in 3 glycosylation pathways and are required for protein secretion. Biochem. Cell Biol. **70:**438-447.
53. **Park, C. H., T. Bessho, T. Matsunaga, and A. Sancar.** 1995. Purification and characterization of the Xpf-ERCC1 complex of human DNA-Repair excision nuclease. J. Biol. Chem. **270:**22657-22660.

54. **Peterson, J. D., L. A. Umayam, T. Dickinson, E. K. Hickey, and O. White.** 2001. The Comprehensive Microbial Resource. Nucleic Acids Res. **29:**123-125.
55. **Pfaffl, M. W.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29**.
56. **Qu, Q. H., S. J. Lee, and W. Boos.** 2004. TreT, a novel trehalose glycosyltransferring synthase of the hyperthermophilic archaeon *Thermococcus litoralis*. J. Biol. Chem. **279:**47890-47897.
57. **Raasch, C., W. Streit, J. Schanzer, M. Bibel, U. Gosslar, and W. Liebl.** 2000. *Thermotoga maritima* AglA, an extremely thermostable NAD(+)-, Mn2+-, and thiol-dependent alpha-glucosidase. Extremophiles **4:**189-200.
58. **Robb, F. T., D. L. Maeder, J. R. Brown, J. DiRuggiero, M. D. Stump, R. K. Yeh, R. B. Weiss, and D. M. Dunn.** 2001. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: Implications for physiology and enzymology, p. 134-157, Hyperthermophilic Enzymes, Pt A, vol. 330.
59. **Ryu, S. I., C. S. Park, J. Cha, E. J. Woo, and S. B. Lee.** 2005. A novel trehalose-synthesizing glycosyltransferase from *Pyrococcus horikoshii*: Molecular cloning and characterization. Biochem. Bioph. Res. Co. **329:**429-436.
60. **Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting.** 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. P. Natl. Acad. Sci. USA **95:**5857-5864.
61. **Schut, G. J., S. D. Brehm, S. Datta, and M. W. W. Adams.** 2003. Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. J. Bacteriol. **185:**3935-3947.
62. **Schut, G. J., J. Z. Zhou, and M. W. W. Adams.** 2001. DNA microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus*: Evidence for a new type of sulfur-reducing enzyme complex. J. Bacteriol. **183:**7027-7036.
63. **Shockley, K. R., D. E. Ward, S. R. Chhabra, S. B. Conners, C. I. Montero, and R. M. Kelly.** 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microb. **69:**2365-2371.
64. **Sonnhammer, E. L. L., S. R. Eddy, and R. Durbin.** 1997. Pfam: A comprehensive database of protein domain families based on seed alignments. Proteins **28:**405-420.
65. **Tanaka, T., T. Fukui, H. Atomi, and T. Imanaka.** 2003. Characterization of an exo-beta-D-glucosaminidase involved in a novel chitinolytic pathway from the hyperthermophilic archaeon *Thermococcus kodakamensis* KOD1. J. Bacteriol. **185:**5175-5181.
66. **Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. **29:**22-28.
67. **Van Lieshout, J. F. T., C. H. Verhees, T. J. G. Ettema, S. Van der Sar, H. Imamura, H. Matsuzawa, J. Van der Oost, and W. M. De Vos.** 2003. Identification and molecular characterization of a novel type of alpha-galactosidase from *Pyrococcus furiosus*. Biocatal. Biotransfor. **21:**243-252.

68. **Voorhorst, W. G. B., I. L. Rik, E. J. Luesink, and W. M. Devos.** 1995. Characterization of the celb gene coding for beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. J. Bacteriol. **177:**7105-7111.
69. **Weinberg, M. V., G. J. Schut, S. Brehm, S. Datta, and M. W. W. Adams.** 2005. Cold shock of a hyperthermophilic archaeon: *Pyrococcus furiosus* exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoproteins. J. Bacteriol. **187:**336-348.
70. **Wolfinger, R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules.** 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Comput. Biol. **8:**625-637.
71. **Xavier, K. B., L. O. Martins, R. Peist, M. Kossmann, W. Boos, and H. Santos.** 1996. High-affinity maltose/trehalose transport system in the hyperthermophilic archaeon *Thermococcus litoralis*. J. Bacteriol. **178:**4773-4777.
72. **Xavier, K. B., R. Peist, M. Kossmann, W. Boos, and H. Santos.** 1999. Maltose metabolism in the hyperthermophilic archaeon *Thermococcus litoralis*: Purification and characterization of key enzymes. J. Bacteriol. **181:**3358-3367.
73. **Yang, S. J., H. S. Lee, C. S. Park, Y. R. Kim, T. W. Moon, and K. H. Park.** 2004. Enzymatic analysis of an amylolytic enzyme from the hyperthermophilic archaeon *Pyrococcus furiosus* reveals its novel catalytic properties as both an alpha-amylase and a cyclodextrin-hydrolyzing enzyme. Appl. Environ. Microb. **70:**5988-5995.
74. **Yau, M. H., J. Wang, P. W. K. Tsang, and W. P. Fong.** 2006. J1 acylase, a glutaryl-7-aminocephalosporanic acid acylase from *Bacillus laterosporus* J1, is a member of the alpha/beta-hydrolase fold superfamily. FEBS Lett. **580:**1465-1471.
75. **Yu, F., Y. Jen, E. Takeuchi, M. Inouye, H. Nakayama, M. Tagaya, and T. Fukui.** 1988. Alpha-glucan phosphorylase from *Escherichia coli* - Cloning of the gene, and purification and characterization of the protein. J. Biol. Chem. **263:**13706-13711.
76. **Zona, R., F. Chang-Pi-Hin, M. J. O'Donohue, and S. Janecek.** 2004. Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. Eur. J. Biochem. **271:**2863-2872.

Table 4.1 – Primers used in amplification of *P. furiosus* and status of cloning

ORF	Fwd primer (Restriction Enzyme) Reverse Primer (Restriction Enzyme)	Vector	Status	Tag
PF0132 α-glucosidase	ATATATACTAGTATGAATTACAGATATCCCCC (*Spe*I) TATATA CTCGAG TCA CTA CAA GCA ATT CCA TC (*Xho*I)	pET42a	No Clone	GST
PF0132 α-glucosidase		modified pET24d	UGA	His_6
PF0133 unknown	GAGAGACCATGGAATTGCTTGTAGTGAAGG (*Nco*I) ATA CTT CTC GAG CTG AAC CCA CCT TAC CTT GAG (*Xho*I)	pET24d	No Clone	His_6
PF0133 unknown		modified pET24d	UGA	His_6
PF0478 CGTase	GAGAGACCATGGTGGACAGATTTTACGATAG (*Nco*I) GAGAGACTCGAGGTAGATAAAAGAGTAGATAGTAGC (*Xho*I)	pET24d	Cloned	No
PF0478 CGTase	GAGAGACCATGGTGGACAGATTTTACGATAG (*Nco*I) GGATGGGTATACCTCGAGGCACCAC (*Xho*I)	pET24d	Cloned	His_6
PF0870 β-amylase	CACACACCATGGTTATGAAATTCACATATCAC ATATATCTCGAGTTTCCCTTTATGCTCGCACCAC	pET24d	Active	His_6
PF1108 Esterase	TATTCTCGGTGATCATATGAAGAG (*Nde*I) TAATATGCGGCCGCTTTGATCCATTCCTCTAAG (*Not*I)	pET21b	No Clone	No
PF1108 Esterase	GTGATTCCATGGAGAGAATGATCATGTATCG (*Nco*I) TAAAATCTCGAGTTCAAAGAGCCACAGAACGG (*Xho*I)	pET24d	Expressed No Activity	His_6
PF1393 α-amylase	GCGCGCGGATCCATGAGAGGATACTTAACTTTTG (*BamH*I) ATATATGCGGCCGCTTGTTCTAACACCTCACATAG (*Not*I)	pET24d	No Clone	His_6
PF1393 α-amylase		modified pET24d	UGA	His_6
PF1460 glycogen debranching	TATATACCATGGGAATAATATTTGGAGGCTC (*Nco*I) TATATAGCGGCCGCAATTAAAAATCCTCTCGACTC (*Not*I)	pET24d	Cloned	No
PF1746 glycogen debranching	CTCTCTCCATGGCAGTTATTCTCTCTTATAAC (*Nco*I) CTCTCTCTCGAGCATTACTATCCTGTCCTTCTC (*Xho*I)	pET24d	Cloned	No
PF1746 glycogen debranching	CTCTCTCCATGGCAGTTATTCTCTCTTATAAC (*Nco*I) ATCTTTCTCGAGATCTCGCTCTTCAAATCT (*Xho*I)	pET24d	Expressed No activity	His_6
PF1747 β-fructosidase	ATATATCATATGAATCTCAAGAGAAGGAGG (*Nde*I) TATATACTCGAGAAATAAAAGAACATGAAGAGG (*Xho*I)	pET21b	No clone	No
PF1747 β-fructosidase	ATATATCATATGAATCTCAAGAGAAGGAGG (*Nco*I) TATATGCTCGAGTAAGTCTGTTGTCCAGTAGGC (*Xho*I)	pET24d	No clone	His_6
PF1747 β-fructosidase		modified pET24d	UGA	His_6
PF1939 CDase	CACTCTCATATGTATAAGCTCGTCAGTTTC (*Nde*I) CTCTCTGCGGCCGCTATATATATCTTGTGCTAATGTAC (*Not*I)	pET21b	Expressed	No

Table 4.2: Sequence alignment of catalytic regions of family 57 glycosidases

Organism	ORF	Accession No.	Nucleophile	Acid-Base
Thermococcus litoralis		BAA22063.1	**118-GVWLTERVW**	**212-HDDGEKFGVW**
Thermococcus hydrothermalis AL662		AAD28552	**313-GGWAAESAL**	**419-TLDGENPVEN**
Pyrococcus furiosus DSM 3638	PF0870	AAL80994	**148-GYWLPENVI**	**251-SSDLESLVAN**
Pyrococcus horikoshii OT3	PH1023	NP_142934	**148-PFWLPENLI**	**251-SSDLESLVAN**
Pyrococcus abyssi GE5	PAB0644	NP_126637	**148-PFWLPENVI**	**251-SSDLESLVAN**
Thermococcus kodakaraensis KOD1	TK1743	YP_184156	**145-GYWLPEAVI**	**251-ASDLESLLGN**
Pyrococcus furiosus DSM 3638	PF1393	AAL81517.1	**180-GMWLPECAY**	**263-PYDTELFGHW**
Thermococcus kodakaraensis KOD1	TK1436	BAD85625.1	**178-GIWLPECAY**	**261-PYDTELFGHW**
Moorella thermoacetica ATCC 39073	Moth_1810	ABC20110.1	**149-GMWLPETAV**	**263-ATDGETYGHH**
			: *::* .	*:* .. .

Table 4.3 – Real-time PCR fold changes for PF0870	
Growth carbohydrate	Fold-change relative to pullulan
Cellobiose	4.4
Glycogen	2.3
Maltose/Cellobiose	3.1
Maltose	3.2
Starch	5.3
Trehalose	3.9
Pullulan	1.0

Table 4.4 – Select ORFs responding to cellobiose and a mixture of cellobiose/maltose relative maltose log_2-fold changes

ORF	Function	CEL-MAL	M/C-MAL
PF0073	β-glucosidase	4.4	3.8
PF0074	short chain alcohol dehydrogenase	3.1	3.0
PF0075	alcohol dehydrogenase	3.5	3.2
PF0360	Chitin ABC ATP-binding protein, chtD	4.0	3.5
PF0361	Chitin ABC ATP-binding protein, chtF	3.6	3.2
PF1208	β-mannosidase	2.9	2.5
PF1696	put. β-ABC sugar transporter	3.9	3.7
PF1697	put. ribose/galactose ABC transporter	4.1	3.8

Table 4.5 – Expression of genes on maltose and maltose/cellobiose relative to cellobiose, log_2 basis

ORF	Function	M/C-CEL	MAL-CEL
PF0272	α-amylase	1.9	2.5
PF0421	hypothetical protein	1.1	1.4
PF0430	phosphoribosylglycinamide formyltransferase 2	1.1	1.9
PF1344	put. maleate cis-trans isomerase	1.0	1.2
PF1935	amylopullulanase	1.3	1.3
PF1936	MalG-like maltose ABC permease	1.4	1.0
PF1938	MalE-like maltose-binding protein	1.5	1.5

Table 4.6 – Relative activity of PF0870 on synthetic substrates

Substrate	% Activity
*p*NP-α-maltopyranoside	100
*p*NP-α-galactoside	2.9
*p*NP-β-glucopyranoside	0.30
*p*NP-α-maltopentaoside	0.13
*p*NP-α-glucopyranoside	0.09
*p*NP-α-mannopyranoside	0.03
*p*NP-β-xylanopyranoside	0.01

Table 4.7 – Effect of cations on *P. furiosus* β-amylase	
Cation salt	Relative activity (%)
MOPS, pH 7.0	100
$BaCl_2$	102
$CaCl_2$	97
$MnCl_2$	94
$FeCl_2$	92
$CoCl_2$	89
EDTA	87
$MgCl_2$	79
$NiCl_2$	79
$ZnSO_4$	0

Table 4.8 – Log_2-Fold change of genes proposed to be involved in pullulan processing				
ORF	LSM MAL	LSM GLY	LSM PUL	LSM STA
PF0132	2.8	2.5	4.6	2.8
PF0272	0.2	-1.1	-1.8	-1.0
PF0312	2.2	1.8	1.7	2.2
PF0477	-1.7	-1.4	-0.2	-0.8
PF1535	-0.2	-0.6	-0.6	-0.4

Table 4.9 – Purification table for native *P. furiosus* α-glucosidase						
Step	Volume (ml)	Total Activity (U)	Total amt of protein (mg)	Sp Act (U/mg)	Relative Purification	Yield (%)
Cell extract	30.5	153	1310	0.1	1	100
DEAE	187	115	100	1.1	9.8	75
HAP	4.0	90	15	5.8	50.0	59
S-200	9.0	61	5.0	12.2	100	40
HIC-Butyl	4.7	41	3.8	10.8	92	27

Table 4.10 – Kinetic parameters of *P. furiosus* α-glucosidase			
Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}*mM^{-1}$)
pNPαGlc	21.9	0.051	429
Maltose	1600	7.1	225
Panose	278	3.2	85.6
Isomaltose	58.8	5.0	11.8

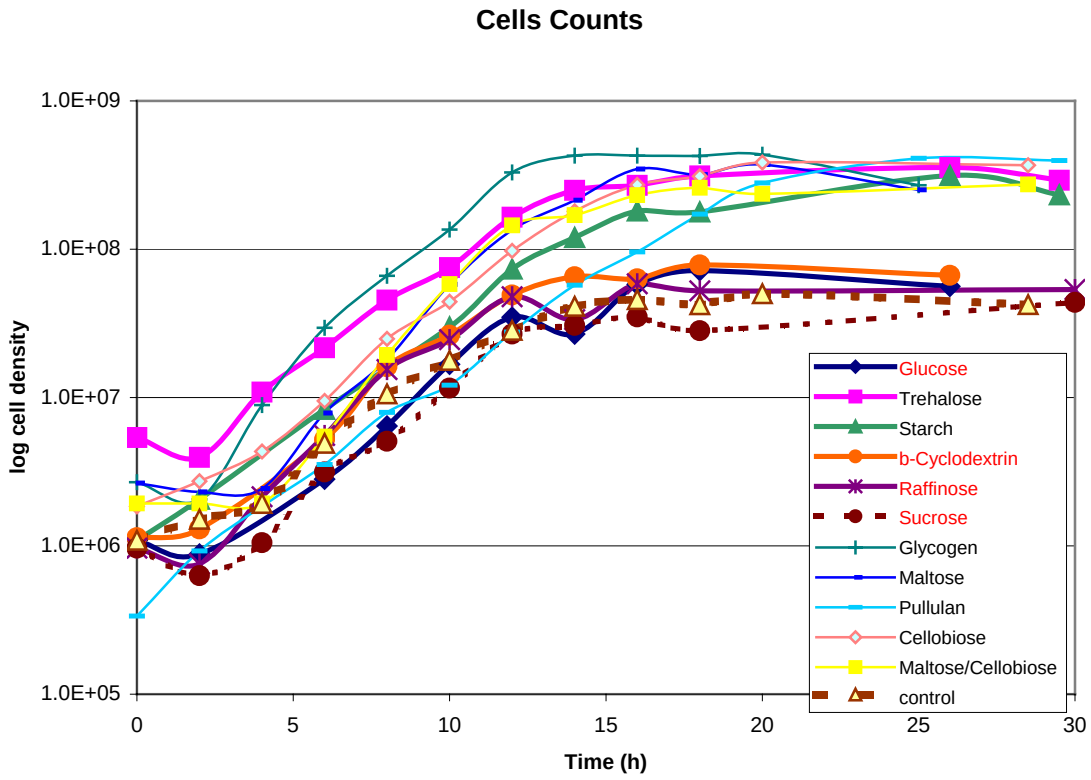


Figure 4.1: Growth *P. furiosus* on various carbohydrate carbon/energy sources.

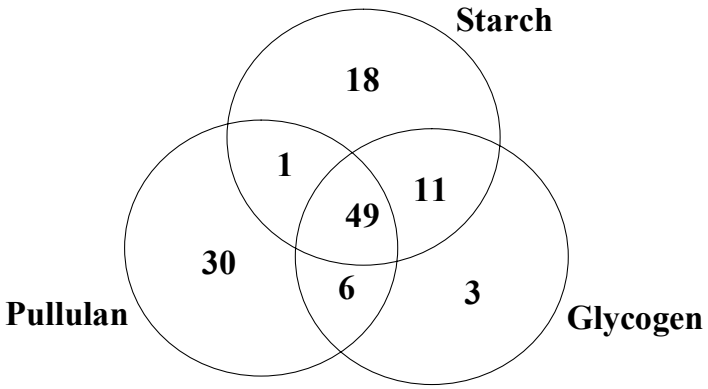


Figure 4.2 – Venn diagram of the transcriptional response of *P. furiosus* growth on polysaccharides relative to maltose. Those genes up-regulated two-fold or more relative to maltose are represented.

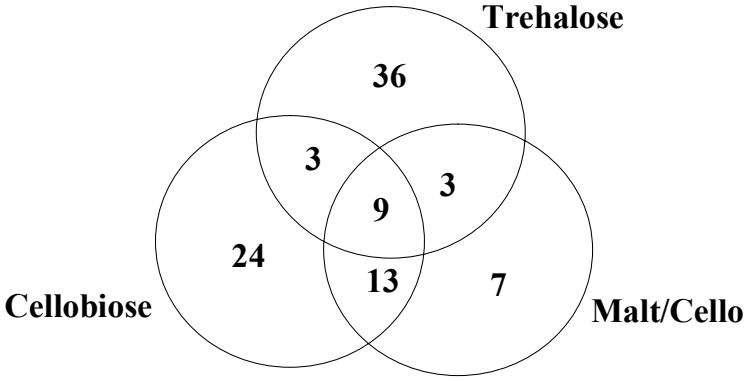


Figure 4.3 – Venn diagram of the transcriptional response of *P. furiosus* growth on disaccharides relative to maltose. Those genes up-regulated two-fold or more relative to maltose are represented.

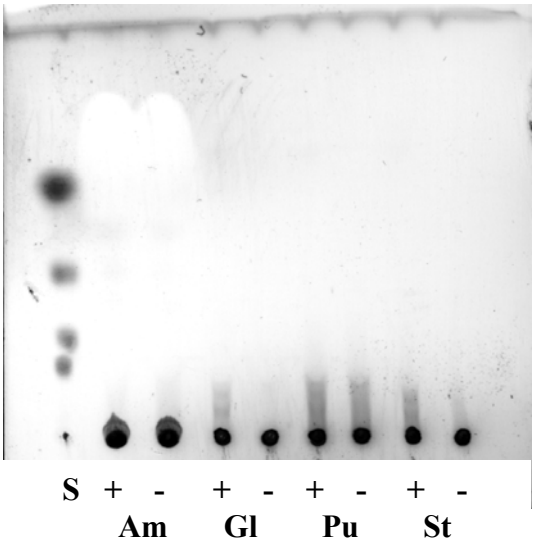


Figure 4.4 – TLC of polysaccharides incubated at 90°C for 60 minutes in the presence of PF1108, PF1393, PF1746, and PF1747. Reaction lanes (+) and control lanes (-) are as indicated. Substrates tested were: amylose (Am), glycogen (Gl), pullulan, (Pu), and starch (St). Stardards are glucose, maltotriose, maltopentaose, and maltoheptaose.

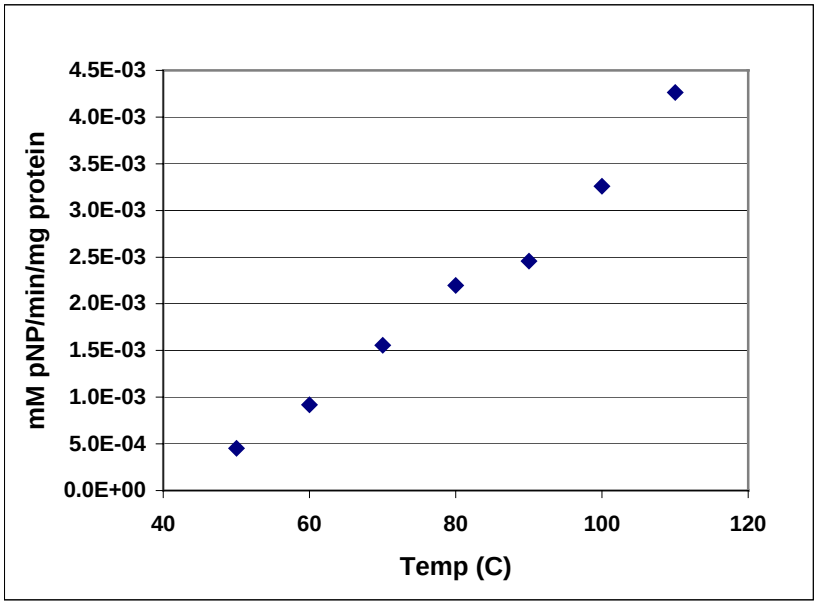


Figure 4.5 –Temperature optimum analysis of PF0870.

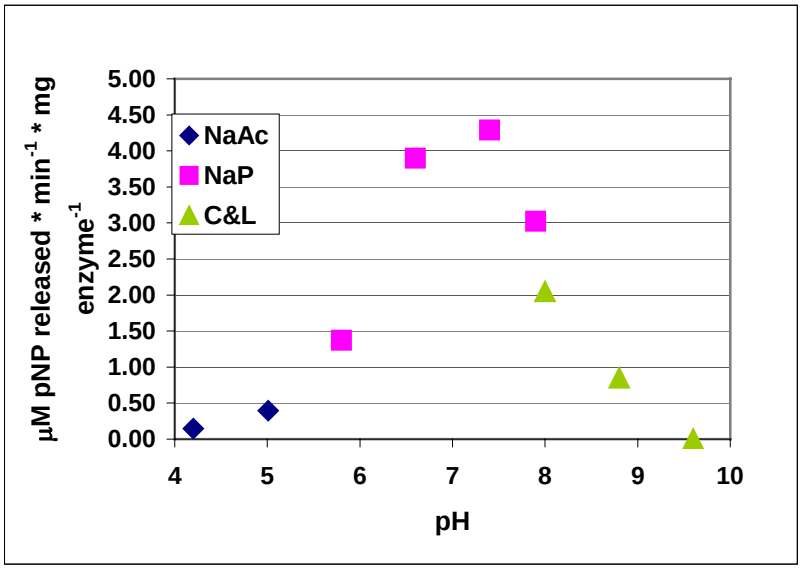


Figure 4.6 – pH optimum of PF0870. NaAc – sodium acetate buffer, NaP – sodium phosphate buffer, C&L – Clark and Lubs buffer.

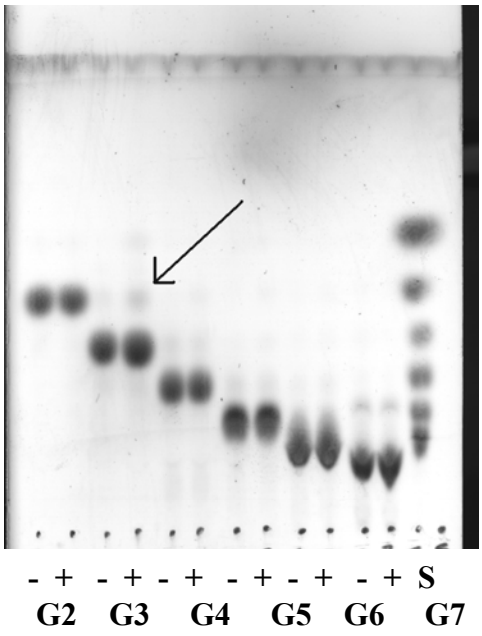


Figure 4.7 – Thin-layer chromatography of PF0870 reactions of maltooligosaccharides. – indicates a control, + the reaction lane. Number of glucose residues are indicated. Standards are in the final lane, glucose to maltoheptaose.

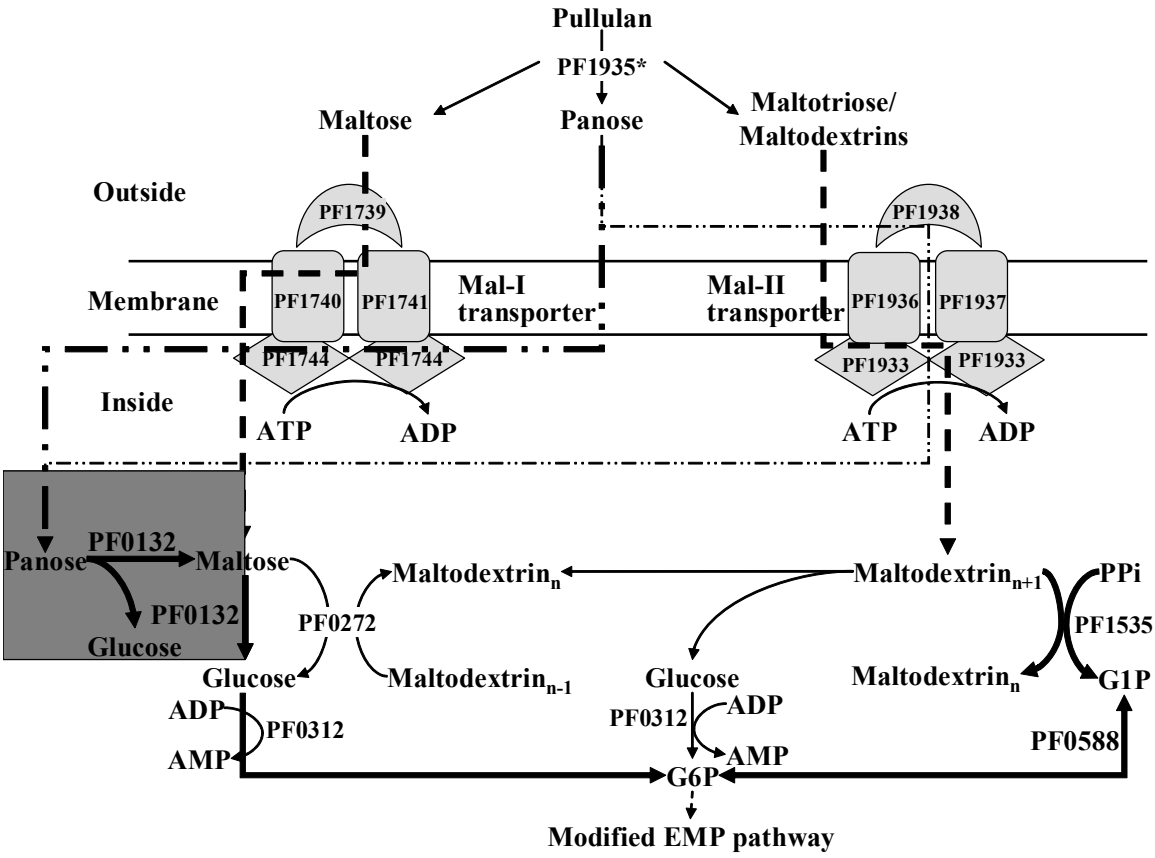


Figure 4.8 – Proposed pullulan utilization pathway. Pullulan degradation olilgosaccharides are transported across the membrane and hydrolyzed into glucose by the α-glucosidase, PF0132. Heavy lines are preferred pathway, narrow lines are alternative pathways. Changes to the starch pathway proposed by Lee *et al.* (41) are boxed in grey.

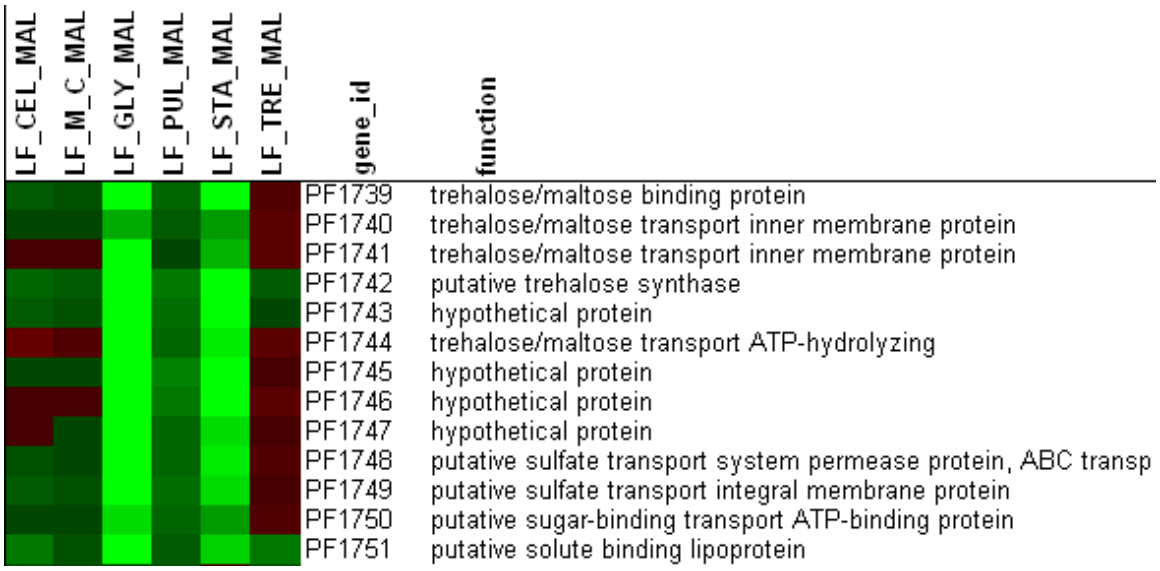


Figure 4.9 – Transcriptional response of *P. furiosus* Mal-I operon relative to maltose for growth on α-glucan substrates. Green indicates down-regulation relative to maltose, red indicates up-regulation relative to maltose.

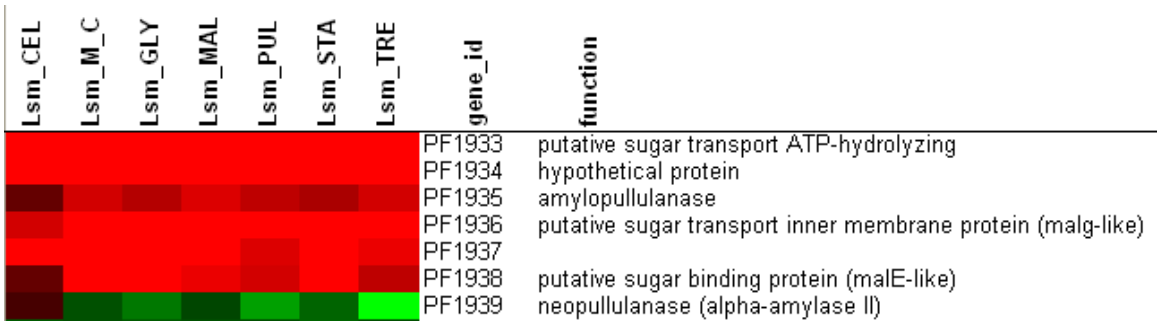


Figure 4.10 – Least squared mean estimate transcriptional response of the *P. furiosus* Mal-II operon. Red conditions are expressed more highly than average expression level, green less than average expression level.

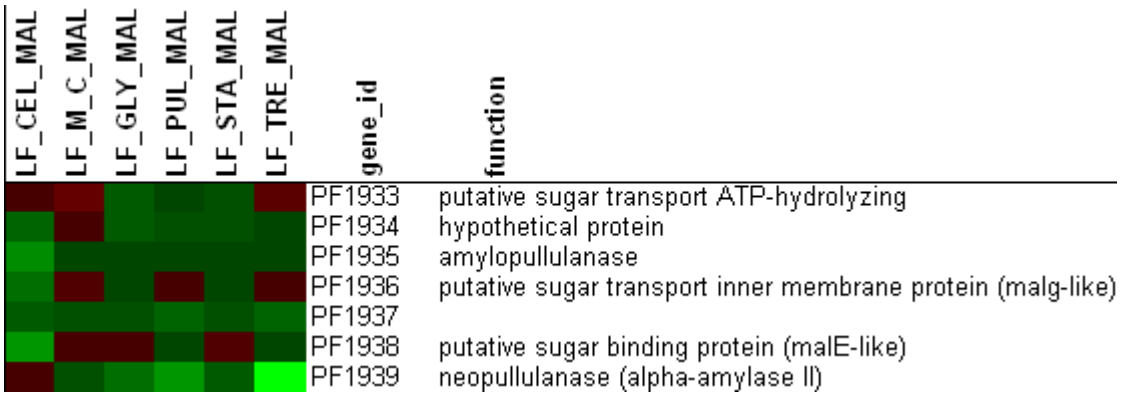


Figure 4.11 – Transcriptional response of the Mal-II operon relative to maltose. Green indicates down-regulation relative to maltose, red indicates up-regulation relative to maltose.

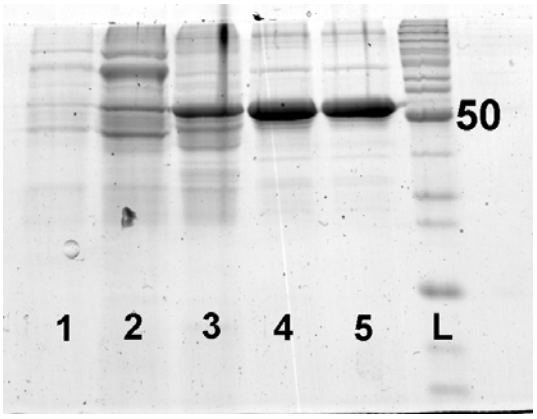


Figure 4.12 – The SDS-PAGE showing the purification of α-glucosidase. Lane 1 – Crude extract; Lane 2 – after DEAE column; Lane 3 – after HAP column; Lane 4 – after S-200 column; Lane 5 – after HIC-butyl column; L – Invitrogen Benchmark Ladder, 50 kDa band is noted.

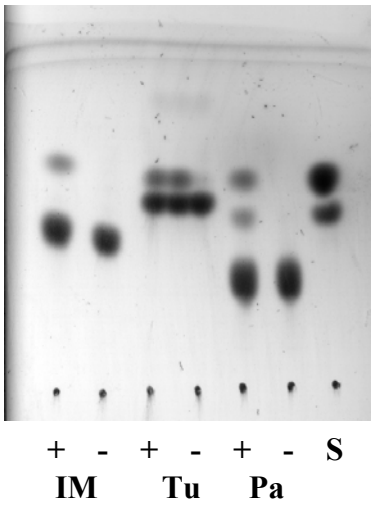


Figure 4.13 – Thin-layer chromatography showing hydrolysis of isomaltose, turanose, and panose by PfαGlc. Standards (S) were glucose and maltose. Reactions (+) and controls (-) were run on isomaltose (IM), turanose (Tu), and Panose (Pa).

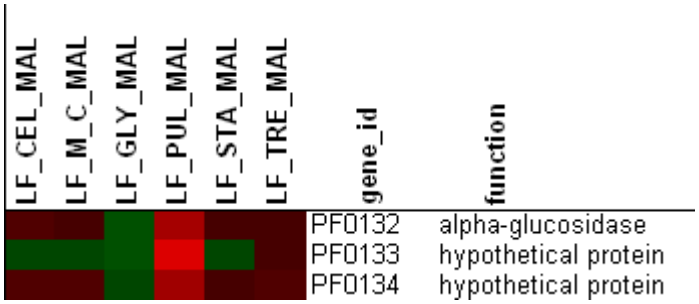


Figure 4.14 – Transcriptional response for the α-glucosidase (PF0132) gene locus relative to maltose.

Appendix 4.A

Transcriptionally Responsive ORFs of *P. furiosus*
and closest homolog in related organisms

Transcriptionally responsive genes for *Pyrococcus furiosus* when compared relative to maltose. Related proteins found in select other hyperthermophiles are also shown. Best hits had to be $<10^{-7}$ to be included. Those between 10^{-5} and 10^{-7} are designated by '?'.

Table 4.A.1 – Genes expressed more than 2-fold higher on pullulan than maltose

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0132	alpha-glucosidase		1.76		PH0222	PAB0137	TK2148	*	SSO037
PF0133	hyp protein, 191 aa		2.11		PH0223	PAB0138	TK2237	*	*
PF0134	hyp protein, 202 aa possible dolichol kinase		1.67		*	PAB1577 ?	TK2238	*	SSO0024
PF0219	putative 6-pyruvoyl tetrahydrobiopterin synthase		1.05		PH1939	PAB1123	TK0563	TM0038	SSO2412
PF0340	putative HTH transcription regulator		1.25		PH0544	PAB1376	TK0036	*	SSO1765 ?
PF0361	oligopeptide ABC transporter (ATP-binding protein) Chitin ABC ATP-binding protein, ChtF		1.16		PH0507	PAB1347	TK1756	TM1219	SSO2615
PF0450	glutamine synthetase i		1.12		PH0359	PAB1292	TK1796	TM0943	SSO0366
PF0477	alpha-amylase		1.55		*	*	TK1884	TM0364 TM1840	*
PF0686	hyp protein, 139 aa		1.04		*	*	*	*	*
PF0692	prismane protein homolog, aka hybrid cluster prot		1.29		*	PAB1707	*	TM1172	*
PF0699	hyp protein, 139 aa		1.15		*	*	*	*	*
PF0723	hyp protein, Iron permease FTR1		1.09		*	*	*	TM0417	*
PF0746	hyp protein, Predicted divalent heavy-metal cations transporter		1.06		PH0632	PAB1443	TK0652	TM1738	*
PF0769	put. glycosyl transferase, UDP-sugar		1.12		PH0401	PAB0772	TK1717	TM0756	SSO0820 ?
PF0770	UDP-glucose pyrophosphorylase		1.02		PH0413	PAB0771	TK1188	TM0862	SSO0813

Table 4.A.1 – Genes expressed more than 2-fold higher on pullulan than maltose (continued)

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0782	hyp protein, 109 aa; polysaccharide biosynthesis related protein		1.12		PH0421	PAB0783	TK1819	*	*
PF0874	membrane dipeptidase		1.13		PH1026	PAB2388	TK2040	*	SSO1864
PF1091	hyp protein, 142 aa		1.23		*	*	*	*	*
PF1279	LSU ribosomal protein L10		1.05						
PF1365	hyp protein		1.27		PH0402	PAB1764	TK0963	TM0613	SSO1834
PF1390	hyp protein, 277 aa		1.31		PH1101	*	TK0811	*	*
PF1450	hyp protein, 94 aa		1.59		PHS039	PAB0804	TK1223	TM1208	*
PF1557	hyp protein, possibly universal stress protein		1.06		PH1540	*	TK2034	*	SSO2778
PF1616	myo-inositol-1-phosphate synthase		1.05		PH1605	PAB1989	TK2278	TM1419	SSO0886
PF1622	n-type ATP pyrophosphatase superfamily		1.21		PH1608	PAB1992	TK0960	TM0197 ?	SSO0333 ?
PF1677	hyp protein, 171 aa (biotin synthase?, Pab)		1.08		PH0159	PAB2248	TK0748	TM0799	*
PF1689	transketolase C-terminal section		1.15		*	PAB0296	TK0269	TM1762	SSO0299
PF1696	putative ABC sugar transporter		1.47		PH1713	PAB0303	TK0658	TM0103	SSO2030
PF1697	putative ribose/galactose ABC transporter		1.23		PH1712	PAB0304	TK0659	TM0104	*
PF2047	l-asparaginase		1.53		PH0066	PAB2294	TK1656	*	SSO0938

Table 4.A.2 – Genes expressed more than 2-fold higher on glycogen than maltose

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0613	hyp protein, fructose-1,6-bisphosphatase	1.05			PH0759	PAB1515	TK2164	N/A	SSO0286
PF0727	hyp protein, metal domain	1.07			*	*	TK0837	TM0320	SSO2651
PF1964	hyp protein, Zn ribbon binding	1.05			*	*	TK0111	*	SSO5343

Table 4.A.3 – Genes expressed more than 2-fold higher on starch than maltose

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0191	typtone/S0 ABC permease; ABC-type dipeptide permease			1.62	*	*	TK0111	*	SSO5343
PF0192	typtone/S0 ABC permease; ABC-type dipeptide permease			1.33	PH1960	PAB1195	TK1802	TM0502 TM0533 TM0059	SSO1275 SSO1283 SSO3047
PF0193	typtone/S0 ABC ATP-binding protein			1.44	PH1959	PAB1196	TK1801	TM0498 TM0501	SSO1276 SSO2616 SSO3055
PF0194	typtone/S0 ABC ATP-binding protein			1.31	PH1958	PAB1197	TK1800	TM1151 TM0057 TM1750	SSO2615 SSO1277 SSO3055
PF0298	hyp protein, 156 aa			1.01	PH0861	PAB0822	TK0226	*	*
PF0324	hyp protein, 201 aa			1.10	PH0565 PH0564	PAB0981 PAB0982	TK1705 TK1704	*	*
PF0481	translation initiation factor eIF-2, subunit beta			1.37	PH0605	PAB0959	TK1621	*	SSO2381
PF0557	3-oxoacyl-[acyl-carrier protein] reductase			1.52	PH1901	PAB1085	TK1677	TM0019	SSO3004
PF0715	NAD(P)H oxidase			1.13	*	PAB0597	TK0284	TM0383 TM0386	SSO2560
PF0816	hyp protein, possible ion channel			1.04	PH0336	PAB1281	TK1063	TM1563	SSO2786 ?
PF1264	translation initiation factor eIF-5a			1.11	PH1381	PAB1854	TK0878	TM1763	SSO0970
PF1542	small nuclear ribonucleoprotein, putative			1.07	PHS042	PAB8160	TK0976	*	SSO6454
PF1608	hyp protein, 134 aa			1.03	PH1599	PAB0387	TK1030	*	*
PF1719	intracellular protease			1.01	PH1704	PAB0311	TK1284	TM1245	SSO1067
PF1920	triosephosphate isomerase			1.49	PH1884	PAB1208	TK2129	*	SSO2592
PF1926	recombinase, radA			1.00	PH0263	PAB0164	TK1899	TM1859	SSO0250

Table 4.A.3 – Genes expressed more than 2-fold higher on starch than maltose (continued)									
		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF1953	inosine-5'-monophosphate dehydrogenase related protein II			1.00	PH0267	PAB2191	TK0558	TM1140	SSO2740
PF2053	transcriptional regulatory protein, asnC family			1.05	PH0061	PAB2299	TK1259	*	SSO2347 SSO0157

Table 4.A.4 – Gene expressed more than 2-fold higher on pullulan and starch than maltose									
		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF1967	hyp protein, ABC type sugar transport system, UgpB		1.12	1.10	PF0025	PAB233	*	TM1235	*

Table 4.A.5 – Genes expressed more than 2-fold higher on glycogen and starch than maltose									
		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0694	flavoprotein	1.28		1.45	PH1085	PAB0596	TK0814	TM0755	SSO2515 ?
PF0719	hyp protein, 79 aa	1.02		1.30	*	*	*	*	*
PF0983	pcna sliding clamp (proliferating-cell nuclear antigen)	1.04		1.21	*	*	*	*	*
PF1133	hyp protein, 68 aa	1.01		1.17	*	*	*	*	*
PF1199	hyp protein, Ferritin-like domain	1.31		1.17	PH0568	PAB1750	TK0826	TM1526	*
PF1302	hyp protein, 147 aa; acetolactate synthase	1.06		1.21	PH1221	PAB0691	TK0626	*	*
PF1492	glycine cleavage system h protein	1.24		1.35	PH1317	PAB0559	TK0150	TM0212	SSO0920
PF1652	hyp protein, 187 aa	1.06		1.13	PH1643	PAB2009	*	*	*
PF1722	archaeal histone a2	2.45		2.00					
PF1831	archaeal histone a1	1.68		1.25					
PF1909	ferredoxin, 67 aa	1.07		1.49	*	PAB7439	TK1694	TM0927	SSO2574 ?

Table 4.A.6 – Genes expressed more than 2-fold higher on glycogen, pullulan, and starch than maltose

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0217	LSU ribosomal protein L44E	1.32	1.02	1.11	PH1940	PAB1221	TK1098	*	SSO7111
PF0246	hyp protein	1.13	1.06	1.35	PH1920	PAB1233	TK1786	*	*
PF0496	Related to TrmB	1.54	1.20	1.98	PH0799	PAB0838	TK0471	*	SSO2073 ?
PF0682	hyp protein, 60 aa	1.62	1.38	1.40	*	*	*	*	*
PF0683	hyp protein, 56 aa	1.55	1.15	1.28	*	*	*	*	*
PF0684	hyp protein, 138 aa	1.72	1.28	1.42	PH0714	*	*	*	*
PF0691	hyp protein, 128 aa; pred. transcriptional regulator	1.99	1.94	1.96	*	PAB0668	*	*	SSO0468 ?
PF0693	hyp protein, pred. Disulfide bond regulator, 158 aa	1.17	1.14	1.09	*	*	*	*	SSO11939
PF0695	hyp protein, 475 aa	1.54	1.21	1.71	*	PAB1709	*	*	SSO1869
PF0703	hyp protein, 271 aa	1.69	1.42	2.07	*	*	*	*	*
PF0704	hyp protein, 117 aa	1.32	1.46	1.17	*	*	*	*	*
PF0705	cytochrome c-type biogenesis protein	1.13	1.11	1.15	*	*	*	TM1141	*
PF0707	hyp protein, maybe 4-Oxalocrotonate Tautomerase	1.98	2.15	2.11	*	*	*	*	*
PF0708	hyp protein, Na+ efflux pump	1.55	1.57	2.04	PH0577 PH0575	PAB1438 PAB1439	TK1288 TK1289	TM1701	*
PF0709	hyp protein, helix turn helix	1.57	1.55	1.71	PH0614 ?	PAB0953 ?	TK0142 ?	TM0494 ?	*
PF0710	hyp protein, 80 aa	1.95	1.94	2.35	*	*	*	*	*
PF0716	3-hydroxyisobutyrate dehydrogenase	3.09	2.43	3.41	*	*	TK0272	*	SSO1560
PF0718	hyp protein, 247 aa	2.87	2.54	3.35	*	*	*	*	*
PF0721	hyp protein, NADPH-dependent FMN reductase	2.89	2.66	3.02	*	*	*	*	*
PF0722	alkyl hydroperoxide reductase subunit c	3.30	2.92	3.80	PH1217	PAB1673	TK0831	TM0807	SSO2121

Table 4.A.6 – Genes expressed more than 2-fold higher on glycogen, pullulan, and starch than maltose (continued)

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0724	hyp protein, Thiol-disulfide isomerase and thioredoxins	1.61	1.43	1.59	PH1130	PAB1651	TK1208	*	*
PF0725	hyp protein, Predicted CoA-binding protein	1.94	1.48	2.13	PH1109	PAB1624	TK0005	TM1435	SSO1111
PF0726	hyp protein, 124 aa	1.80	1.49	1.72	*	*	*	*	*
PF0735	hyp protein, 238 aa	1.14	1.12	1.20	*	*	*	*	*
PF0736	hyp protein, 186 aa	2.22	1.92	2.51	*	*	*	*	*
PF0739	transcriptional regulatory protein, asnC family	1.28	1.31	1.15	PH1592 PH1692	PAB0392 PAB0322	TK0834	*	SSO2652
PF0742	putative ferritin homolog	3.71	3.08	4.00	*	*	TK1999	TM1128	*
PF0747	putative proline dipeptidase aminopeptidase	3.24	2.82	3.11	PH1902	PAB2178	TK0967	TM0042	SSO0363
PF0750	hyp protein, Nitroreductase	2.14	2.03	2.12	*	PAB1763	TK0839	TM0383 ?	*
PF0751	flavoprotein	3.82	3.32	3.40	PH1085	PAB0596	TK0814	TM0755	*
PF0752	thioredoxin peroxidase	2.01	1.43	1.78	PH1217	PAB1673	TK0815	TM0780	SSO2255
PF0753	2-keto acid:ferredoxin oxidoreductase subunit beta	2.38	1.94	2.10	PH1661	PAB0347	TK0816	TM1165	SSO2816
PF0754	2-keto acid:ferredoxin oxidoreductase subunit alpha	2.37	1.76	2.04	PH1662	PAB0346	TK0817	TM1164	SSO2815
PF0755	non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase	1.98	1.24	1.77	*	*	TK0705	TM0293 ?	SSO3194
PF0765	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	2.32	1.97	2.02	PH0429 PH0618	PAB0776	TK1231	TM0583	SSO0810
PF0766	putative dehydrogenase	2.02	1.90	1.50	PH1881	PAB0775	TK0651	TM0414 TM0585	SSO3015
PF0767	Predicted pyridoxal phosphate-dependent enzyme	2.12	1.76	1.77	PH0771	PAB0774	TK2268	TM0668	SSO0897 ?

Table 4.A.6 – Genes expressed more than 2-fold higher on glycogen, pullulan, and starch than maltose (continued)

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0768	acetyl / acyl transferase related protein	1.81	1.72	1.26	PH1591 ?	PAB0773	TK0955 ?	TM0759	SSO0372
PF0776	hyp protein, PIN domain 108 aa	2.02	1.86	1.86	PH0389	*	TK0374	*	SSO1922
PF0777	hyp protein, polysaccharide biosynthesis related protein	1.86	1.62	1.70	*	PAB2422	TK1819	*	*
PF0783	hyp protein, membrane protein RfbX	2.07	2.26	2.30	*	PAB1409	TK1819	*	SSO0837
PF1036	hyp protein 120 aa	1.21	1.01	1.06	PH1127	PAB0732	TK0901	*	*
PF1287	putative ABC transporter (ATP-binding)	1.42	2.47	1.29	PH1384	PAB1855	TK0731	TM1368	SSO0925
PF1368	SSU ribosomal protein S28E	1.16	1.19	1.12	PHS040	PAB7165	TK1310	*	SSO5479
PF1499	SSU ribosomal protein S19E	1.24	1.14	1.00					
PF1572	hyp protein, amino acid metabolism regulator	1.44	1.37	1.82	PH1554	PAB2435	TK1227	*	*
PF1786	hyp protein, 110 aa	1.65	1.29	1.41	PH1738	PAB0204	TK0070	*	*
PF1881	hyp protein, DNA/RNA-binding protein Alba	1.35	1.17	1.54	PHS053	PAB7094	TK0560	*	SSO0962
PF1930	hyp protein, S-adenosylmethionine decarboxylase	1.01	1.02	1.67	PH1992	PAB1162	TK1592	TM0655	SSO0585

Table 4.A.7 – Genes expressed more than 2-fold higher on glycogen and pullulan than maltose

		Log_2 Fold Change							
gene_id	funct	GLY-MAL	PUL-MAL	STA-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0488	SSU ribosomal protein S6E	1.37	1.17		PH0615	PAB1427	TK1951	*	SSO0411
PF0681		1.43	1.15		*	*	*	*	*
PF0738	ubiquinone/menaquinone biosynthesis methyltransferase	1.00	1.04		PH1081	PAB0601	TK0729	TM0318	*
PF0760	hyp protein, transposase DNA binding	1.39	1.76		PH0585	PAB2077	TK0495 TK0850	TM0776	SSO1475
PF0771	UDP-glucose 6-dehydrogenase	1.33	1.23		PH0429	PAB0770	TK1231	TM0583	SSO0810
PF1110	hyp protein	1.02	1.31		*	PAB1790	*	TM1841	*

Table 4.A.8 – Genes expressed more than 2-fold higher on cellobiose than maltose									
		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0324	hyp protein, 201 aa (PF0325)	1.15			PH0565 PH0564	PAB0981 PAB0982	TK1705 TK1704	*	*
PF0488	SSU ribosomal protein S6E	1.02							
PF0496	hyp protein, TrmB Like	1.13			PH0799	PAB0839	TK0471	*	SSO2073 ?
PF0557	3-oxoacyl-[acyl-carrier protein] reductase	1.47			PH1901	PAB1085	TK1677	TM0019	SSO3004
PF0776	hyp protein, Predicted nucleic acid-binding protein	1.16			PH0389	*	TK0374	*	SSO1922 SSO1493
PF0983	pcna sliding clamp (proliferating-cell nuclear antigen)	1.01			PH0665	PAB1465	TK0535	*	SSO0405
PF1102	bifunctional short chain isoprenyl diphosphate synthase	1.01			PH1072	PAB2389	TK1468	TM0131 TM1535	SSO0061
PF1245	hypothetical d-nopaline dehydrogenase	1.01			PH1363	PAB1842	TK0116	*	SSO0186
PF1368	SSU ribosomal protein S28E	1.28							
PF1644	SSU ribosomal protein S9P	1.00							
PF1678	2-isopropylmalate synthase	1.40			PH1727	PAB0286	TK0283	TM0553 TM0552	SSO0977 SSO2407
PF1679	putative 3-isopropylmalate dehydratase large subunit	1.59			PH1726	PAB0287 PAB0891	TK0282	TM0554 TM0291	SSO2471 SSO1095
PF1681	hyp protein, 55 aa	1.39			PH1721	PAB0289	TK0279	*	SSO5317 ?
PF1684	acetylglutamate kinase	1.49			PH1718	PAB0292	TK0276	TM1784	SSO0156
PF1685	acetylornithine aminotransferase	1.10			PH1716	PAB2440	TK0275	TM1785	SSO0160 SSO3211 SSO2727
PF1688	transketolase N-terminal section	1.04			*	PAB0295	TK0270	TM1762	SSO0297
PF1701	chorismate mutase	1.01			*	*	TK0261	TM0155 ?	SSO0302 ?

Table 4.A.8 – Genes expressed more than 2-fold higher on cellobiose than maltose (continued)									
		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF1707	phosphoribosyl anthranilate isomerase	1.08			PH1156 ?	PAB2047	TK0256	TM0139	SSO0892
PF1708	anthranilate synthase component II	1.41			PH1356	PAB2046	TK0255	TM0141	SSO0894
PF1709	anthranilate synthase component I	1.13			*	PAB2045	TK0254	TM0142	SS0893
PF1710	anthranilate phosphoribosyltransferase	1.41			PH1598	PAB2044	TK0253	TM0141	SSO0890
PF1711	indoleglycerol phosphate synthase	1.64			*	PAB2043	TK0252	TM0140	SSO0895
PF1761	hypothetical protein	1.06			PH1674 PH1673	PAB0336 PAB0337	*	*	*
PF1881	hypothetical protein	1.01			PHS053	PAB7094	TK0560	*	SSO0962

Table 4.A.9 – Genes expressed more than 2-fold higher on cellobiose, maltose/cellobiose, and trehalose than maltose									
		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0691	hyp protein, Predicted transcriptional regulator	1.23	1.24	1.29	*	PAB0668	*	*	SSO0468 ?
PF0716	3-hydroxyisobutyrate dehydrogenase	1.20	1.46	1.48	*	*	TK0272	*	SSO1560
PF0718	hyp protein, 247 aa	1.64	2.03	1.59	*	*	*	*	*
PF0721	hyp protein, possible NADPH-dependent FMN reductase	1.36	1.20	1.41	*	*	*	*	*
PF0736	hyp protein, 186 aa	1.41	1.65	2.04	*	*	*	*	*
PF0742	putative ferritin homolog	1.47	1.51	1.66	*	*	TK1999	TM1128	*
PF0747	putative proline dipeptidase	1.07	1.40	1.61	PH1902	PAB2178	TK0967	TM0042	SSO0363
PF0751	flavoprotein	2.24	2.04	1.08	PH1085	PAB0596	TK0814	TM0755	*
PF0783	hyp protein, possible RfbX - Membrane protein	1.44	1.43	1.90	*	PAB1409	TK1819	*	SSO0837

Table 4.A.10 – Genes expressed more than 2-fold higher on cellobiose and trehalose than maltose									
		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0774	hyp protein, 78 aa	2.64		2.34	*	*	*	*	*
PF1365	hyp protein, 146 aa	1.05		1.49	PH0402	PAB1764	TK0963	TM0613	SSO1834
PF1683	n-acetyl-gamma-glutamyl-phosphate reductase	1.59		1.00	PH1720	PAB0291	TK0277	TM1782	SSO0155

Table 4.A.11 – Genes expressed more than 2-fold higher on maltose/cellobiose than maltose									
		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0477	alpha-amylase		1.33		*	*	TK1884	TM0364 TM1840	*
PF0485	cell division inhibitor minD homolog		1.09		PH0612	PAB0954	TK0952	TM1870	SSO0460 ?
PF0703	hyp protein		1.08		*	*	*	*	*
PF0739	transcriptional regulatory protein, asnC family		1.04		PH1592 PH1692	PAB0392 PAB0322	TK0834	*	SSO2652
PF0753	2-keto acid:ferredoxin oxidoreductase subunit beta		1.20		PH1661	PAB0347	TK0816	TM1165	SSO2816
PF0754	2-keto acid:ferredoxin oxidoreductase subunit alpha		1.36		PH1662	PAB0346	TK0817	TM1164	SSO2815
PF0755	non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase		1.24		*	*	TK0705	TM0293 ?	SSO3194

Table 4.A.12 – Genes expressed more than 2-fold higher on maltose/cellobiose and trehalose than maltose									
		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0707	hyp protein, put. 4-Oxalocrotonate_Tautomerase		1.03	1.24	*	*	*	*	*
PF0722	alkyl hydroperoxide reductase subunit c		1.84	1.22	PH1217	PAB1673	TK0831	TM0807	SSO2121
PF0725	hyp protein, Predicted CoA-binding protein		1.14	1.02	PH1109	PAB1624	TK0005	TM1435	SSO1111

Table 4.A.13 – Genes expressed more than 2-fold higher on cellobiose and maltose/cellobiose than maltose

		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0073	beta-glucosidase	4.38	3.84		PH0501	PAB1740	TK1761	*	SSO3019
PF0074	alcohol dehydrogenase, short chain	3.05	3.00		PH1901	PAB1085	TK1677	TM0441 TM0325	SSO3004
PF0075	alcohol dehydrogenase	3.47	3.25		PH0743	PAB1511	TK1569	TM0111	SSO2490
PF0360	Chitin ABC ATP-binding protein, ChtD	4.02	3.48		PH0505	PAB1346	TK1757	TM1220	SSO1276
PF0361	Chitin ABC ATP-binding protein, ChtF	3.61	3.25		PH0507	PAB1347	TK1756	TM1219	SSO2615
PF0710	hyp protein, 70 aa	1.59	1.42		*	*	*	*	*
PF1207	hyp protein, 76 aa	2.66	2.28		*	*	TK1762	*	*
PF1208	beta-mannosidase	2.94	2.55		PH0501	PAB1740	TK1761	*	SSO3019
PF1674	hypothetical protein	3.38	3.13		PH1732	PAB0279	TK0241	*	SSO2453
PF1676	biotin operon repressor/biotin--[acetyl CoA carboxylase] ligase	3.95	3.37		PH0147	PAB0104	TK0747	TM0224	SSO0662
PF1689	transketolase C-terminal section	3.69	3.44		*	PAB0296	TK0269	TM1762	SSO0299
PF1696	putative β-ABC sugar transporter	3.91	3.71		PH1713	PAB0303	TK0658	TM0103	SSO2030
PF1697	putative ribose/galactose ABC transporter	4.11	3.85		PH1712	PAB0304	TK0659	TM0104	*

Table 4.A.14 – Genes expressed more than 2-fold higher on trehalose than maltose									
		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0121	putative aspartate aminotransferase			1.11	PH0207	PAB2227	TK0186	TM1131	SSO0104
PF0340	putative HTH transcription regulator			1.53	PH0544	PAB1376	TK0036	*	SSO1765 ?
PF0394	UDP-n-acetylglucosamine-dolichyl-phosphate n-acetylglucosamine			1.11	PH0348	PAB1288	TK2254	TM0235	SSO0060
PF0450	glutamine synthetase i			1.64	PH0359	PAB1292	TK1796	TM0943	SSO0366
PF0456	carboxypeptidase 1			1.12	PH0465	PAB1320	TK1840	*	SSO3105
PF0468	hyp protein, Predicted metal-dependent protease			1.33	PH0451	PAB1045	TK1262	*	*
PF0486	hyp protein, Predicted membrane protein			1.01	PH0613	PAB1426	TK1267	*	*
PF0534	indolepyruvate ferredoxin oxidoreductase subunit b			1.26	PH0764	PAB0857	TK0135	TM1760.1	SSO2069
PF0612	hyp protein, 267 aa			1.06	PH0757	PAB0865	TK2203	*	SSO1962
PF0613	hyp protein, 375 aa			1.22	PH0759	PAB1515	TK2164	*	SSO0286
PF0692	prismane protein homolog			1.49	*	PAB1707	*	TM1172	*
PF0699	hyp protein, 139 aa			1.09	*	*	*	*	*
PF0704	hyp protein, 117 aa			1.11	*	*	*	*	*
PF0760	hyp protein, put transposase			1.37	PH0585	PAB2077	TK0495 TK0850	TM0776	SSO1475
PF0782	hyp protein, polysaccharide biosynthesis related protein			1.09	PH0421	PAB0783	TK1819	*	*
PF0848	hyp protein			1.63	PH1167	PAB1373	TK0050	*	*
PF0874	membrane dipeptidase			2.17	PH1026	PAB2388	TK2040	*	SSO1864
PF0891	sulfhydrogenase beta subunit			1.31	PH1290	PAB1784	TK2072	TM1758 ?	*

Table 4.A.14 – Genes expressed more than 2-fold higher on trehalose than maltose (continued)

		Log_2 Fold Change							
gene_id	funct	CEL-MAL	M/C-MAL	TRE-MAL	*Pyrococcus horikoshii*	*Pyrococcus abyssi*	*Thermococcus kodakarensis*	*Thermotoga maritima*	*Sulfolobus solfataricus*
PF0892	sulfhydrogenase gamma subunit			1.26	PH1291	PAB1785	TK2071	TM1639 TM0334	*
PF0893	sulfhydrogenase delta subunit			1.66	PH1292	PAB1786	TK2070	*	*
PF0894	sulfhydrogenase alpha subunit			1.21	PH1294	PAB1787	TK2069	*	*
PF0910	iron (III) ABC transporter, permease protein			1.03	PH1236	PAB0677	TK2019	TM0190	SSO0486
PF1091	hyp protein, 142 aa			1.62	*	*	*	*	*
PF1287	putative ABC transporter (ATP-binding protein)			2.42	PH1384	PAB1855	TK0731	TM1368	SSO0925
PF1390	hyp protein, 277 aa			1.82	PH1101	*	TK0811	*	*
PF1421	hypothetical 4-aminobutyrate aminotransferase			1.35	PH1423	PAB0501	TK2101	TM1785	SSO3211 SSO2727
PF1450	hyp protein, 94 aa Pred sub of Multisubunit Na+/H+ antiporter			1.46	PHS039	PAB0804	TK1223	TM1208	*
PF1478	hyp protein, 76 aa			1.31	*	PAB0276	*	TM1012	*
PF1557	hyp protein, Universal stress protein family			2.02	PH1540	*	TK2034	*	SSO2778
PF1622	n-type ATP pyrophosphatase superfamily			1.48	PH1608	PAB1992	TK0960	TM0197 ?	SSO0333 ?
PF1632	hyp. protein			1.53	PH1619	PAB2000	TK1232	*	*
PF1655	similar to ABC transporter (ATP-binding protein)			1.26	PH1735	PAB2148	TK0676 TK1452	TM0288	SSO3012
PF1669	hyp protein, 576 aa			1.01	*	*	TK1690	*	*
PF1677	hyp protein, BioY?			1.24	PH0159	PAB2248	TK0748	TM0799	*
PF1957	agmatinase			1.30	PH0083	PAB0050	TK0882	*	SSO0445
PF2047	l-asparaginase			1.89	PH0066	PAB2294	TK1656	*	SSO0938

Chapter 5: Hyperthermophilic Glycosynthases

Donald A. Comfort, Swapnil R. Chhabra, and Robert M. Kelly

Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

ABSTRACT

Traditional synthetic organic chemistry methods have difficulty synthesizing certain glycosidic linkages such as the β-mannosidic bond, due to ethalpic and entropic barriers. Many organisms possess glycoside hydrolase enzymes which naturally degrade these linkages. Nucleophile mutant glycoside hydrolase enzymes have been shown to synthesize glycosidic bonds, when used in combination with activated fluoro sugars, in glycosynthetic reactions. The nucleophile mutants of β-mannosidase (Man2) from *Thermotoga neapolitana*, β-endoglucanase (Tm1751, Cel5A) from *Thermotoga maritima*, and the α-galactosidase (Tm1192, GalA) from *T. maritima* were studied as potential glycosynthases. Reactions were run at elevated temperatures, at which these enzymes are stable, using α-mannooyranosyl fluoride, α-cellobiosyl fluoride, and β-galactopyranosyl fluoride as donor sugars for Man2, Cel5A, and GalA, respectively. The reactions were monitored for fluoride ion release and final products analyzed by TLC for oligosaccharide products. No glycosynthase products were detected for any reactions.

Glycosyl hydrolases are enzymes that selectively hydrolyze carbohydrates through either a single- or double-displacement mechanism resulting in inversion or retention of anomericity (23). These reactions are equilibrium based and strongly favor hydrolysis, however, synthetic products have been shown to occur under strict conditions for select glycosidases. For instance, the α-glucosidase from *Aspergillus nidulans* hydrolyzes maltose to glucose and then transglycosylates the glucose to isomaltose (11). The β-glucosidase from *Agrobacterium tumefaciens* was shown to transfer a butyl group to pNP-β-glucopyranoside after initial hydrolysis of the 4-nitrophenyl group, when 100 mM butanol was present (26). In order for these wild-type enzyme to synthesize oligosaccharide products, high concentrations of reactants are generally required, but low yields still occur because the synthesized oligosaccharide is a hydrolysis substrate for the enzyme.

One means of overcoming the hydrolysis problem for glycoside hydrolase transglycosylation products is to mutate the nucleophilic residue of a retaining glycoside enzyme so that it is inactive. This nucleophile mutant when combined with activated fluoro-sugars is capable of synthesizing oligosaccharides as demonstrated by the Withers Group at the University of British Columbia (9, 13). These mutant enzyme/fluoro-sugar combinations are referred to as glycosynthases (22). During retaining hydrolysis reactions a covalently bound glycosyl-enzyme intermediate is formed; the fluorosugar mimics this glycosyl-enzyme intermediate of hydrolysis, allowing for the transglycosylation reaction to initiate. Enzymatic glycosynthase reactions are advantageous over organic synthesis by providing a single step means of producing stereo- and regio-specific oligosaccharides, without the need for bulky protecting groups used in organic synthesis.

A number of glycosynthases have been generated to date (Table 5.1), mostly from β-specific glycosidases. The donor fluoro-sugar for glycosynthases are of the opposite anomeric conformation to the synthesis product because the reaction is single-displacement, as shown in Figure 5.1. Glycosynthases have been generated from both exo- and endo-acting glycoside hydrolases (4, 7, 13, 14, 18). The first glycosynthases mutated the nucleophile to alanine (13), however, through strategic selection of alternate

residues and directed evolution of glycosynthase enzymes, serine and glycine have been found to generate higher yields and rates (15).

Herein we discuss the attempts to generate hyperthermophilic glycosynthases from the *Thermotoga neapolitana* β-mannosidase (Man2) and *T. maritima* α-galactosidase (GalA) and β-endoglucanase (Cel5A).

Materials and Methods

Reagents: Fluoro-sugars were synthesized by Natural Products and Glycotechnology (NPG) Institute (Durham, NC). Sugars synthesized were α-mannopyranosyl fluoride (αManF), a mixed anomer of β-galactopyranosyl fluoride (GalF), and cellobiosyl-α-fluoride (αCel_2F). Media components and buffer salts were purchases from Fisher Scientific or Sigma.

Cloning, Expression, and Mutagenesis: The DNA sequence of the *T. neapolitana* β-mannosidase (man2) has been previously reported (20, 21). Using this sequence, the respective forward and reverse primers were designed to extract the gene, 5'-GAGGTTCATGAGACGTATCGATCT-3' and 5'-TATCTAGAAAGATGACTCGAGCCTTG-3'. These primers include a *Bsp*HI site in the forward primer and a *Xho*I site in the reverse primer. The subcloned gene was verified by sequencing and subsequent alignment of the sequencing results against the published gene sequence. Protein was expressed in *E. coli* Codon Plus RIL cells (Stratagene). Protein was purified by heat-treatment of crude extract and activity verified by hydrolysis of 4-nitrophenyl-β-mannopyranoside (pNPβMan) in 100 mM sodium phosphate, pH 7.0.

Sequence alignment of *Tn*Man2 with the *Cellulomonas fimi* β-mannosidase from GH2, identified E505 as the nucleophile. E505 was mutated to alanine and serine using QuikChange XL Site-Directed Mutagenesis Kit (Stratagene) using the primer 5'-CGGAAAGTTCATAAGCGCGTTCGGTTTTCAGGGAGCCC-3' and 5'-C GGA AAG TTC ATA AGC TCG TTC GGT TTT CAG GGA GCC C-3' and each compliment, respectively. An alanine acid-base mutant was generated at E419 using the primer 5'-

GTT TTG TGG TGT GGT AAC AAC GCG AAC AAC TGG GGA TTT GAC-3', and its compliment. Mutants were expressed and purified as per the wild-type mutant.

T. maritima GalA was mutated as described in Chapter 3 (Comfort *et al.*), using the primers and each respective compliment D220A, 5'-C GAG GTC TTC CAG ATA GCC GAC GCC TAC GAA AAG GAC-3' and D327A, 5'-C TAC AGG TAC TTC AAG ATC GCC TTT CTC TTC GCG GGT GC-3'.

T. maritima Cel5A serine and alanine mutants were generated by Swapnil Chhabra by alignment with the cellulase from *Schizophyllum commune* (2).

Glycosynthetic reactions: Glycosynthetic reactions using the alanine and serine nucleophile mutants of *Tm*Cel5A (Tm1751) were run in 67 mM phosphate buffer, pH 7.0 with 33 mM αCel_2F and 33 mM pNP-cellobiose using 0.33 mg of mutant Cel5A, final volume 1.5 mL. Reactions were carried out at 65 and 80°C over the course of two days. Reactions were monitored by release of fluoride ions using a ThermoOrion Fluoride ion probe (BN9609) connected to a 720A-plus ThermoOrion meter at times 2 hr, 20 hr, and 44 hr. During synthesis reactions, the fluoride ion is released, so the extent of the reaction can be measured by the F^- concentration. TLC analysis was also used to determine a shift in oligosaccharide size.

Synthetic reactions using *Tn*Man2 were run with the D505A and D505S mutants. Final reaction conditions were 65 mM Na-Phosphate buffer (pH 7.6), 62 mM αManF, 29 mM pNP-cellobiose, and approximately 0.4 mg of respective protein in a 1.5 mL volume at 80°C for 4 days. A second reaction condition was also tested using 29 mM pNP-β-D-mannopyranoside as acceptor in a total volume of 1.0 mL with 62 mM αManF, and 65 mM sodium phosphate buffer (pH 8.0), at 70°C. The reaction was monitored for F^- release.

Synthetic reactions using D220G and D327G of *Tm*GalA were carried out in 100 mM sodium acetate, pH 5.5 at 65°C, with 60 mM GalF as donor and 20 mM pNP-α-galactopyranoside as acceptor 40 hours. An additional reaction was run using 50 mM GalF and 10 mM mannotetraose in sodium acetate buffer, pH 4.5, for 48 hours. Reactions were monitored periodically for release of fluoride ions and final products checked by TLC.

Thin-layer chromatography: TLC conditions were 1-butanol:ethanol:water (3:1:1) on silica plates.

Results and Discussion

T. neapolitana β-mannosidase was successfully cloned into pET24d and expressed when transformed into *E. coli* RIL Codon Plus cells. The resulting nucleophile mutants, E505A and E505S, each had less than 5% of the wild-type activity when the heat-treated crude extract was assayed on pNPβMan. Fidelity of mutations were confirmed by full plasmid sequencing.

Glycosynthetic reactions using mutant *Tn*Man2 were run as described in the material and methods section, linking αManF and pNPβMan as shown in Figure 5.2. β-Mannosidic bonds are difficult to synthesize using traditional synthetic organic chemistry, thus the desire to find enzymatic means of synthesizing these linkages. Fluoride ion standards were used to estimate the concentration of F^- between 0.1 and 50 mM. During both reactions, the concentration of F^- was approximately 0.1 mM for the entire course of the reaction, indicating that the αManF is stable at these conditions. It also indicates that the synthesis reaction was unsuccessful. No products could be detected by TLC for either reaction.

The catalytic residues of the *T. maritima* Cel5A endoglucanase were identified by sequence alignment with the cellulase from *Schizophyllum commune*. *Tm*Cel5A has been shown to be active on a wide range of larger oligosaccharides including both mannans and β-glucans (1). Alanine and serine nucleophile mutants of *Tm*Cel5A were generated and used in glycosynthetic reactions. There reaction attempted to synthesize pNP-β-cellotetroside from αCel_2F with pNP-β-cellobioside. Reactions at 65 and 80°C did not produce any higher order oligosaccharide product as determined by TLC, Figure 5.3, or release of fluoride ions. White flocculent appeared after 24 h at 80°C, possibly indicating that the enzyme was being denatured. The mutant enzymes appeared to retain some catalytic activity as the pNP group was hydrolyzed over the course of the reaction at 80°C as determined by an increase absorbance at 405 nm, relative to the control.

The catalytic residues for *T. maritima* α-galactosidase have recently been identified as D327 and D387 for the nucleophile and acid-base residue, respectively (Comfort *et al.*, submitted). Glycosynthetic reactions were run using mutants D220A and D327A in an attempt to identify the nucleophile mutant by detecting oligosaccharide products. The donor was a mixed anomer of β-galactopyranoside fluoride and reactions were conducted using two different acceptors, pNP-α-galactopyranoside and mannotetraose. The mannotetraose reaction was designed to mimic the natural debranching of galactose from galactomannan (12, 21), as shown in Figure 5.4. Fluoride ion release was not detected for any reaction using mutant *Tm*GalA and TLC did not show any oligosaccharide product.

REFERENCES

1. **Chhabra, S. R., K. R. Shockley, D. E. Ward, and R. M. Kelly.** 2002. Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium *Thermotoga maritima* grown on glucan- and mannan-based polysaccharides. Appl. Environ. Microb. **68:**545-554.
2. **Clarke, A. J., J. Drummelsmith, and M. Yaguchi.** 1997. Identification of the catalytic nucleophile in the cellulase from *Schizophyllum commune* and assignment of the enzyme to Family 5, subtype 5 of the glycosidases. FEBS Lett. **414:**359-361.
3. **Drone, J., H. Y. Feng, C. Tellier, L. Hoffmann, V. Tran, C. Rabiller, and M. Dion.** 2005. *Thermus thermophilus* glycosynthases for the efficient synthesis of galactosyl and glucosyl beta-(1 -> 3)-glycosides. Eur. J. Org. Chem.**:**1977-1983.
4. **Faijes, M., J. K. Fairweather, H. Driguez, and A. Planas.** 2001. Oligosaccharide synthesis by coupled endo-glycosynthases of different specificity: A straightforward preparation of two mixed-linkage hexasaccharide substrates of 1,3/1,4-beta- glucanases. Chem.-Eur. J. **7:**4651-4655.
5. **Faijes, M., T. Ima, V. Bulone, and A. Planas.** 2004. In vitro synthesis of a crystalline (1 -> 3,1 -> 4)-beta-D-glucan by a mutated (1 -> 3,1 -> 4)-beta-D-glucanase from *Bacillus*. Biochem. J. **380:**635-641.
6. **Fairweather, J. K., M. Hrmova, S. J. Rutten, G. B. Fincher, and H. Driguez.** 2003. Synthesis of complex oligosaccharides by using a mutated (1,3)-beta-D-glucan endohydrolase from barley. Chem-Eur. J. **9:**2603-2610.
7. **Fort, S., V. Boyer, L. Greffe, G. Davies, O. Moroz, L. Christiansen, M. Schulein, S. Cottaz, and H. Driguez.** 2000. Highly efficient synthesis of beta(1 -> 4)-oligo- and - polysaccharides using a mutant cellulase. J. Am. Chem. Soc. **122:**5429-5437.
8. **Jahn, M., D. Stoll, R. A. J. Warren, L. Szabo, P. Singh, H. J. Gilbert, V. M. A. Ducros, G. J. Davies, and S. G. Withers.** 2003. Expansion of the glycosynthase repertoire to produce defined manno-oligosaccharides. Chem. Commun.**:**1327-1329.
9. **Jakeman, D. L., and S. G. Withers.** 2002. Glycosynthases: New tools for oligosaccharide synthesis. Trends Glycosci. Glyc. **14:**13-25.
10. **Jakeman, D. L., and S. G. Withers.** 2002. On expanding the repertoire of glycosynthases: Mutant beta-galactosidases forming beta-(1,6)-linkages. Can. J. Chem. **80:**866-870.
11. **Kato, N., S. Suyama, M. Shirokane, M. Kato, T. Kobayashi, and N. Tsukagoshi.** 2002. Novel alpha-glucosidase from *Aspergillus nidulans* with strong transglycosylation activity. Appl. Environ. Microb. **68:**1250-1256.
12. **Liebl, W., B. Wagner, and J. Schellhase.** 1998. Properties of an alpha-galactosidase, and structure of its gene galA, within an alpha- and beta-galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. Syst. Appl. Microbiol. **21:**1-11.
13. **Mackenzie, L. F., Q. P. Wang, R. A. J. Warren, and S. G. Withers.** 1998. Glycosynthases: Mutant glycosidases for oligosaccharide synthesis. J. Am. Chem. Soc. **120:**5583-5584.

14. **Malet, C., and A. Planas.** 1998. From beta-glucanase to beta-glucansynthase: glycosyl transfer to alpha-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. FEBS Lett. **440:**208-212.
15. **Mayer, C., D. L. Jakeman, M. Mah, G. Karjala, L. Gal, R. A. J. Warren, and S. G. Withers.** 2001. *Agrobacterium* beta-glucosidase: a general screen to detect enzymes for oligosaccharide synthesis. Chem. Biol. **8:**437-443.
16. **Mayer, C., D. L. Zechel, S. P. Reid, R. A. J. Warren, and S. G. Withers.** 2000. *Agrobacterium* sp beta-glucosidase is a greatly improved glycosynthase. FEBS Lett. **466:**40-44.
17. **Mullegger, J., H. M. Chen, W. Y. Chan, S. P. Reid, M. Jahn, R. A. J. Warren, H. M. Salleh, and S. G. Withers.** 2006. Thermostable glycosynthases and thioglycoligases derived from *Thermotoga maritima* beta-glucuronidase. Chembiochem **7:**1028-1030.
18. **Nashiru, O., D. L. Zechel, D. Stoll, T. Mohammadzadeh, R. A. J. Warren, and S. G. Withers.** 2001. beta-Mannosynthase: Synthesis of beta-mannosides with a mutant beta-mannosidase. Angew. Chem.-Int. Edit. **40:**417-420.
19. **Okuyama, M., H. Mori, K. Watanabe, A. Kimura, and S. Chiba.** 2002. alpha-glucosidase mutant catalyzes "alpha-glycosynthase"-type reaction. Biosci. Biotechnol. Biochem. **66:**928-933.
20. **Parker, K. N., S. Chhabra, D. Lam, M. A. Snead, E. J. Mathur, and R. M. Kelly.** 2001. beta-Mannosidase from *Thermotoga* species, p. 238-246, Hyperthermophilic Enzymes, Pt A, vol. 330.
21. **Parker, K. N., S. R. Chhabra, D. Lam, W. Callen, G. D. Duffaud, M. A. Snead, J. M. Short, E. J. Mathur, and R. M. Kelly.** 2001. Galactomannanases man2 and man5 from *Thermotoga* species: Growth physiology on galactomannans, gene sequence analysis, and biochemical properties of recombinant enzymes. Biotechnol. Bioeng. **75:**322-333.
22. **Perugino, G., A. Trincone, M. Rossi, and M. Moracci.** 2004. Oligosaccharide synthesis by glycosynthases. Trends Biotechnol. **22:**31-37.
23. **Rye, C. S., and S. G. Withers.** 2000. Glycosidase mechanisms. Curr. Opin. Chem. Biol. **4:**573-580.
24. **Sugimura, M., M. Nishimoto, and M. Kitaoka.** 2006. Characterization of glycosynthase mutants derived from glycoside hydrolase family 10 xylanases. Biosci. Biotech. Bioch. **70:**1210-1217.
25. **Trincone, A., G. Perugino, M. Rossi, and M. Moracci.** 2000. A novel thermophilic glycosynthase that effects branching glycosylation. Bioorg. Med. Chem. Lett. **10:**365-368.
26. **Watt, D. K., H. Ono, and K. Hayashi.** 1998. *Agrobacterium tumefaciens* beta-glucosidase is also an effective beta-xylosidase, and has a high transglycosylation activity in the presence of alcohols. BBA-Protein Struct. M. **1385:**78-88.

Table 5.1: Glycosynthase enzymes

Wild-type enzyme	Mutation	Glycosyl Hydrolase Family	Donor Sugar	Linkage(s) synthesized	Reference
Agrobacterium sp. β-glucosidase (Abg)	E358A	1	αGlcF / αGalF	β-(1,4)- / β-(1,3)-	(13)
Bacillus licheniformis 1,3-1,4-β-glucanase (Blg)	E134A	16	αLamF	β-(1,4)-	(14) (5)
Sulfolobus solfataricus β-glucosidase (Ss)	E387G	1	αGlcF	β-(1,3)- / β-(1,4)- / β-(1,6)	(25)
Agrobacterium sp. β-glucosidase (Abg)	E358S	1	αGlcF / αGalF	β-(1,4)- / β-(1,3)	(16)
Humicola insolens Cel7B	E197A	7	αLacF / αCelF	β-(1,4)	(7)
Agrobacterium sp. β-glucosidase (Abg)	E358G	1	αGlcF / αGalF	β-(1,4)	(15)
Cellulomonas fimi β-mannosidase (Man2)	E519S	2	αManF	β-(1,4)	(18)
Schizosaccharomyces pombe α-glucosidase	D481G	31	βGlcF	α-(1,4)	(19)
Thermotoga maritima β-glucuronidase	E476A	2	αGlcUA-F	β-(1,3)	(17)
Thermotoga maritima β-xylanaseB	E259G	10	αXyl_2F	β-(1,4)	(24)
Thermus thermophilus β-glycosidase	E338S E338G	1	αGlcF	β-(1,3)	(3)
Escherichia coli β-galactosidase	E537S	2	αGalF	β-(1,6)	(10)
Hordeum vulgare β-glucanase	E231G	17	αLamF	β-(1,3)	(6)
Cellvibrio japonicus β-mannanase	E320G	26	αMan_2F	β-(1,4)	(8)

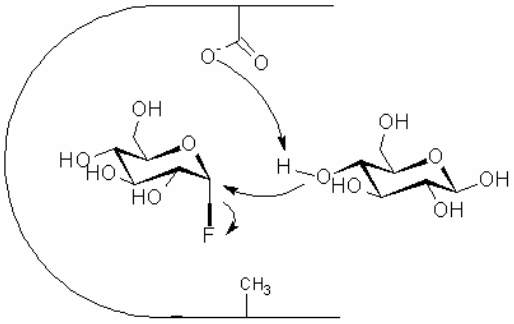


Figure 5.1 – Glycosynthase reaction mechanism uses a fluoro-sugar which mimics the glycosyl-enzyme intermediate of retaining glycosides.

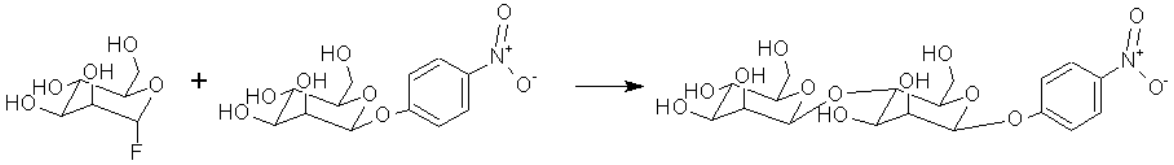


Figure 5.2: Reaction of α-mannosyl fluoride with pNP-β-D-mannopyranoside to form pNP-β-mannobioside.

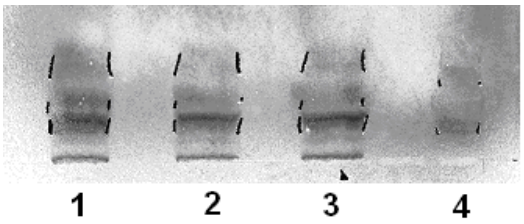


Figure 5.3 – TLC of Cel5A reaction. 1 – NucS, 2 – NucA, 3 – negative control, 4 – cellobiosyl fluoride.

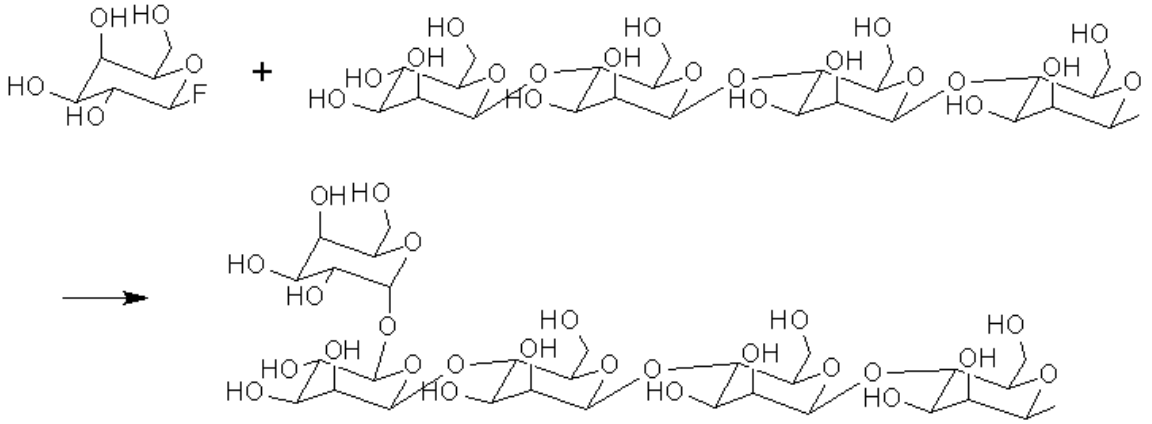


Figure 5.4: Reaction of β-galactopyranosyl fluoride with mannotetroside to form a galactomannan chain.